

**ISSUES IN
THE DEVELOPMENT OF
A WRITING SYSTEM
FOR THE
KALASHA LANGUAGE**

Gregory Cooper
PhD thesis
Department of Linguistics
Macquarie University
December 2005

CONTENTS

	Page
Abstract	iv
Certification	v
Preface	vi
Chapter 1: Introduction	1
Chapter 2: Should languages be codified in writing?	23
Chapter 3: Alternative scripts	50
Chapter 4: Phonological foundations	72
Chapter 5: Morphophonemic and syntagmatic considerations	88
Chapter 6: The first Kalasha orthography conference	115
Chapter 7: Developing a Romanized Kalasha orthography	135
Chapter 8: From letters to literature	160
Chapter 9: Conclusion	175
Epilogue	181
Appendix 1: Simplified classification of the Kalasha language	182
Appendix 2: Script catalogues	183
Appendix 3: Kalasha phoneme chart – Roman script – 1985	191
Appendix 4: Kalasha phoneme chart – Roman script – 2005	192
Appendix 5: Kalasha allophone catalogue	193
Appendix 6: Kalasha corpus sample	203
Appendix 7: Kalasha phoneme tally – Arabic-script corpus	204
Appendix 8: Kalasha phoneme tally – Roman-script corpus	206
Appendix 9: Kalasha alphabet chart – indigenous script	208
Appendix 10: Kalasha orthography questionnaire – Arabic script	209
Appendix 11: Kalasha orthography questionnaire – Roman script	210
Appendix 12: Summary of script choice issues by relevant factors	211

Appendix 13: Kalasha literacy materials – Roman script	213
Appendix 14: Genre scheme	216
Appendix 15: Kalasha literature types	221
Appendix 16: Kalasha discourse samples – Roman script	223
Appendix 17: Kalasha alphabet chart – Arabic script	225
Appendix 18: Kalasha storybook sample – Arabic script	226
Appendix 19: Kalasha pre-reader sample teachers’ instructions – Arabic script	227
Appendix 20: Kalasha alphabet chart – Roman script	228
Appendix 21: Kalasha orthographic alternatives – Roman script	229
Appendix 22: Kalasha-English glossary sample – Roman script	230
Appendix 23: Kalasha text sample – Roman script	233
Appendix 24: Kalasha-English dictionary sample – Roman & Arabic scripts	234
Appendix 25: Kalasha alphabet chart - Arabic script	237
Abbreviations and Symbols	238
Bibliography	239

ABSTRACT

Kalasha is an Indo-Aryan language spoken in the Hindu Kush mountains of Pakistan's North West Frontier Province. The Kalasha speech community has a unique culture and a rich tradition of oral literature, but there is no indigenous writing system to document them. The changes of the modern world are now threatening the viability of both the language and the culture. This thesis examines a range of issues involved in developing and implementing a Kalasha writing system, as a foundation for what could eventually become a rich heritage of indigenous literature.

The Kalasha society is a close-knit community of less than six thousand people who are very keen to preserve their language, despite heavy social influence by the dominant Kho society, numbering nearly a quarter of a million, who live throughout the whole Chitral district, including the Kalasha Valleys. According to various typologies discussed in the first chapter the Kalasha language is officially endangered. The question of whether codifying a language helps to preserve it is the subject of much debate in the literature, and this matter is discussed in the thesis. The broader question of whether a language should be codified in writing, and the value of literacy for traditional societies generally, is discussed in the context of the theory of linguistic ecology and language habitat.

Various scripts (Arabic, Roman, and others) that have been used over the centuries to represent other Indo-Iranian languages in West, Central and South Asia, are reviewed to show both the flexibility of the languages, and the political and religious pressures to which they respond. Many of these languages have been written in more than one script, some in up to four different scripts, in a relatively short period of time or over several geographical areas. This fact dispels any assumption that there might be an immutable connection between language and script.

A phonemic analysis of Kalasha is then provided, demonstrating how its unusual combinations of rhotic and nasal vowels, and retroflex and aspirated consonants, result in an inventory of over sixty phonemes. Various aspects of Kalasha morphology are outlined, with special reference to inflections and clitics. The boundaries between nouns and their inflections pose particular problems in terms of phonological analysis, and both shallow and deep levels of orthographic representation are examined and discussed as part of the strategy for dealing with allomorphs and the morphophonemic issues inherent in compounds.

Complementing this theoretical analysis, the research draws on empirical data, supplied by Kalasha writers, to examine all these issues and support the final decisions regarding the

orthographic system proposed. The data comes from natural texts, several orthography surveys, and the documented outcomes of a Kalasha orthography conference in December/January 2000/2001. This conference served as a milestone in the short history of Kalasha literacy: the choice between an Arabic or Roman script was the subject of lively and thorough discussion, with a final vote taken, and the options of shallow or deep orthographic representation were examined and addressed.

The results of this research include a proposed Roman alphabet of 22 letters, with two diacritics (tilde and apostrophe, to represent nasalized and retroflex/rhotic phonemes respectively). This will serve both the standard and peripheral dialects of the Kalasha language. The proposed orthography allows for ‘shallow’ representation, working with the phonemic rather than morphophonemic structure, although this is expected to tend toward a ‘deep’ representation over time (which the Kalasha already prefer in some cases). This issue will be revisited at the planned follow-up orthography conference.¹

By way of conclusion, the thesis presents a principled program for the documentation and publication of Kalasha literature, as resources for developing literacy. It considers the various lively oral genres and their characteristics, and how appropriately they can be rendered in writing. Publishing priorities for the newly literate community are also proposed.

¹ Dates to be determined.

I certify that
this work has not been submitted
for a higher degree
to any other university or institution.

Greg Cooper 14/12/05

PREFACE

According to the Kalasha, the thousands of foreign visitors who roam their valleys each year are either tourists, hippies or philanthropists who have come there in order to view, experience, or help them and/or their environment. Scores of journalists, filmmakers and anthropologists also go there each year, in order, ultimately, to publish, produce, or to qualify. Some of these travellers, coming from such diverse directions as Europe, North America, Japan, Australasia, and Pakistan itself, bring and leave behind aspects of their own culture. The fame of the Kalasha has spread nationally, through regular press coverage, and internationally, through the academic spotlighting of ethnographers and other researchers.¹

Paradoxically, despite all this commercial, material and even anthropological interest and patronage, the greatest threat to the Kalasha people is the loss of their very ethnicity. In order to preserve their identity their own agenda is different. Their greatest felt needs are for justice in the many dealings of the majority community which directly affect them, and for self-determination: the need for direct input into the decision-making process for planning their development. Though their culture is being documented by copious recording in all media, it is all for the consumption and benefit of outsiders. But the Kalasha people themselves (in whom their ancient culture is actually still alive), the present and future generations, will not survive culturally unless they personally can express themselves and their ethnicity, creatively and with pride.²

In this regard, history is now being written—in more ways than one. For the first time in history, they are writing down their own language, and there is now literature by Kalasha speakers in their own script. Consequently, also for the first time, the history of Kalasha society (among other topics) is also being documented by them. As they begin to read and write in their own language now they are really strengthening their roots and revitalising their culture, their identity and their sense of self-esteem, as this age of rapid change presses in on them from the outside. 'It has been proven that a community's expectation, personal dignity, and self-esteem will rise when that community's language has been written and used for education' (Pikkert 1996).

¹ See Lines (1995) for a more detailed treatment of these outside influences.

² One of the first Kalasha people to receive an education has presented a paper that discusses Kalasha perceptions and views on development issues with relation to the Kalasha community (Jan 1995).

The Kalasha people are concerned for the preservation of their inherited identity (their heritage), as well as for their currently evolving identity. They want to maintain the integrity of their indigenous character in the face of a changing world. Their own current interest in vernacular literacy and literature stems directly from their sense of need to preserve their culture against the major outside forces now bearing down on them.³ The documentation of this language will promote its preservation, but more importantly, indigenous publication will preserve and promote the actual vitality of the language and the culture.

Indigenous literature is a medium for speakers of a language to enjoy the creativity, inspiration and development of ideas that come from both reading and writing texts that have a significant level of relevance and/or interest to them. A new orthography is a uniquely created medium for that literature, which will both support and enhance an indigenous culture for generations to come.

Writing this thesis was not my motive for living among the Kalasha. But my very happy involvement with them as a community (since 1982), and my interest in linguistics, has led to this research and involvement in the first couple of decades in the history of Kalasha literature, a period which I believe will be a turning point for their future.

No important Kalasha endeavour is complete without a solemn tribute to concerned persons. So, as my personal foreword to the whole body of future Kalasha literature, I would like to acknowledge all my Kalasha friends for their extremely generous, friendly and tireless hospitality, for their very patient and willing help in many ways, and for their various valuable inputs into my linguistic research. This work is dedicated to them, and for their benefit.

G.C.⁴

³ For example, commercialism, tourism, logging, land disputes and religious pressures. See Chapter 2 for a fuller treatment of this issue.

⁴ I also very gratefully acknowledge the tireless and very valuable assistance of my supervisor, Professor Pam Peters.

INTRODUCTION

1.1 THE KALASHA

Nestled high in the Hindu Kush range of mountains of north-west Pakistan, lives a small, ancient society of people who speak a Dardic language called Kalasha. This chapter visits that remote place, introducing aspects of Kalasha society and culture (section 1.1), and their language (section 1.2). It reviews previous research on the language itself (section 1.3), and then provides a brief history of efforts at writing the language (section 1.4). The chapter concludes with an outline of the structure of the thesis (section 1.5).

1.1.1 Name

The term *Kalasha* (pronounced /ka¹ʃaʃa/), strictly applied, refers to the community and individuals who adhere to the Kalasha religion (briefly described in subsection 1.1.7), and to their language.¹ This name (and the adjective which is the same) is used by the people themselves and those closely associated with them. However, casual visitors to the area, following most of the early and even currently popular literature about the Kalasha, often refer to them as *Kalash* (pronounced /ka¹la:ʃ/ by other Pakistanis), which is not actually a Kalasha word. Other Pakistanis, who have only heard about them but have never visited, generally call them the *Kafir* (/¹kafir/) tribe because of their polytheistic religion.²

Most Kalasha who become Muslims retain their language to a greater or lesser degree, but some of them start to speak more in the neighbouring language of Khowar. Therefore, it must be pointed out here that the focus of this thesis is not only the language of the Kalasha community, but also of Kalasha-speaking Muslims, that is, all speakers of the Kalasha language.

¹ Phonetic and phonemic notation used in this thesis uses the International Phonetic Alphabet (IPA). Chapter 4 provides a short technical summary of my analysis of Kalasha phonology, and the system of transcription used here.

² *Kafir* is an Islamic theological term (in Arabic and Urdu) which means pagan, infidel, heathen, unbeliever or unbelieving.

1.1.2 Location and Environment

The Kalasha language is spoken or known by no more than a few thousand people,³ most inhabiting five very narrow valleys in the Hindu Kush range of mountains running across the remote Chitral District of the North West Frontier Province (NWFP) of Pakistan. Each of these valleys contains a river of melted snow and spring water, and most have tracks leading over mountain passes into Afghanistan. The area of three valleys where most speakers reside is known as *Kalashadesh* /ka'ʎaʂa'deʃ/) by the Kalasha people themselves, the *Kalash Valleys* by most other Chitralis, *Kafiristan* by most other Pakistanis, and the *Kalasha Valleys* by most foreigners.

By far the most common access to the area is from the south. There is a road over a 3,200-metre mountain pass, but this journey can only be made from about June to November, when there is no snow at the pass.⁴ A local airline service is scheduled from Peshawar (capital of the NWFP) to Chitral town (capital of Chitral district) up to three times each day, but these flights are often cancelled due to weather conditions at Lowari Pass. The road from Chitral town to the Kalasha Valleys is usually traversed by jeep, taking two to three hours.

Access to the area from the north is by road over the Shandur Pass, leading to and from the very remote reaches of Gilgit and Hunza. Access from the west could be made along the Chitral River across the border from Afghanistan; however, due to political unrest and bandits, this route is only used as a last resort by Chitralis (including the Kalasha) needing to get to or from Peshawar.

The topography of the Kalasha community, and its context, is shown in the two maps on the following pages. The first (Figure 1.1) is a map of Pakistan, showing where the Kalasha Valleys are situated in relation to the rest of Pakistan and neighbouring countries. Apart from a few special territories Pakistan is divided into four provinces: Sindh, Punjab, Balochistan and North West Frontier Province (NWFP).⁵ The provincial languages of these provinces are Sindhi, Punjabi, Balochi and Pashto respectively.

The second map (Figure 1.2) is of part of Chitral District, showing each of the valleys where the Kalasha language is spoken. The official names for these valleys are Bumboret,

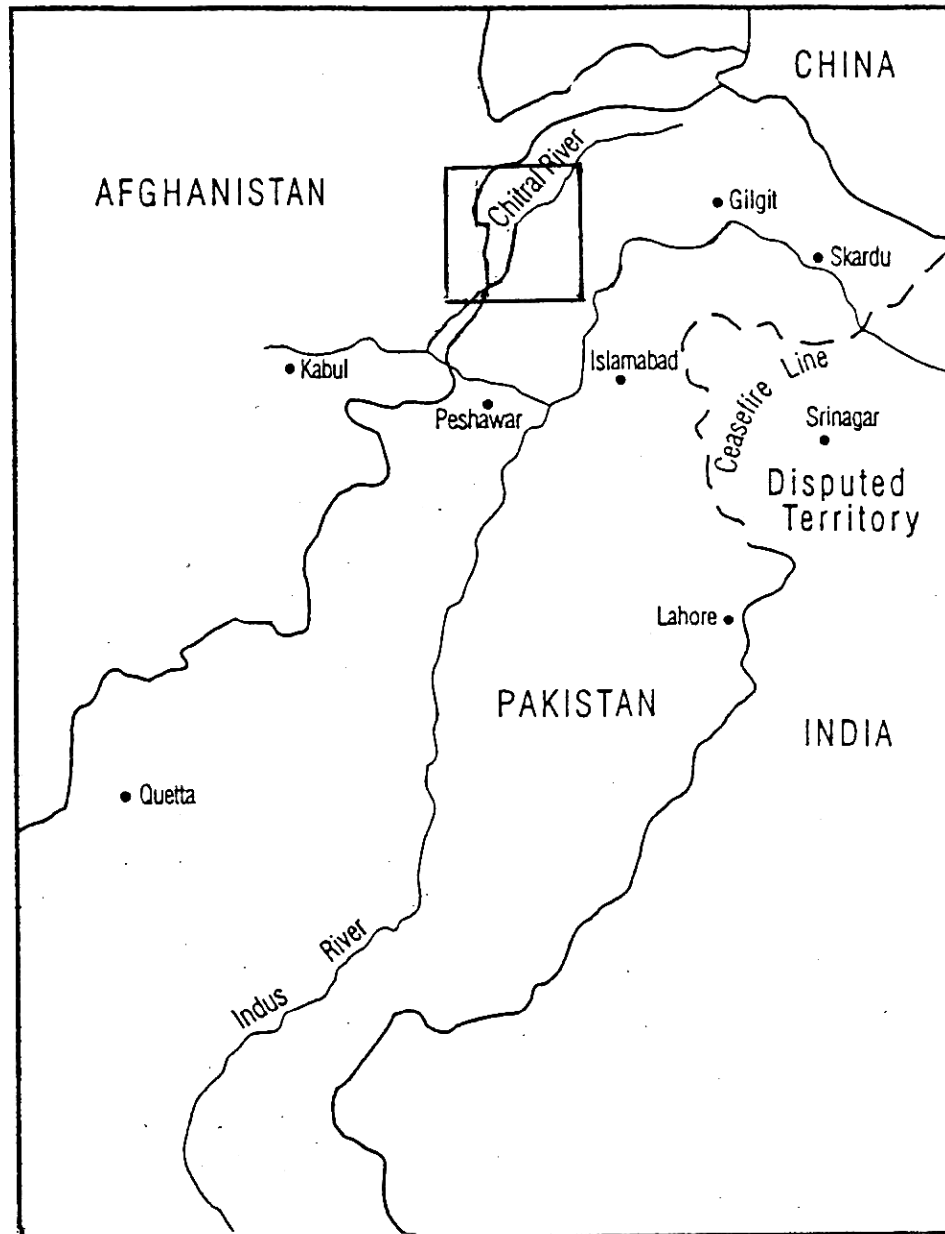
³ The population of speakers is examined more closely in the next section (1.1.3).

⁴ Lowari Pass is pronounced /'rawləj ʎap/ by the Kalasha. *Lowari* is very likely derived from *Lahori*, the Urdu adjective for the city of Lahore, where the road through this pass (eventually) leads. /ʎap/ is from the English word *top*.

⁵ These are the official Pakistan Government spellings of these provinces. Sindh is also spelt as *Sind* and Balochistan as *Baluchistan* by various other sources.

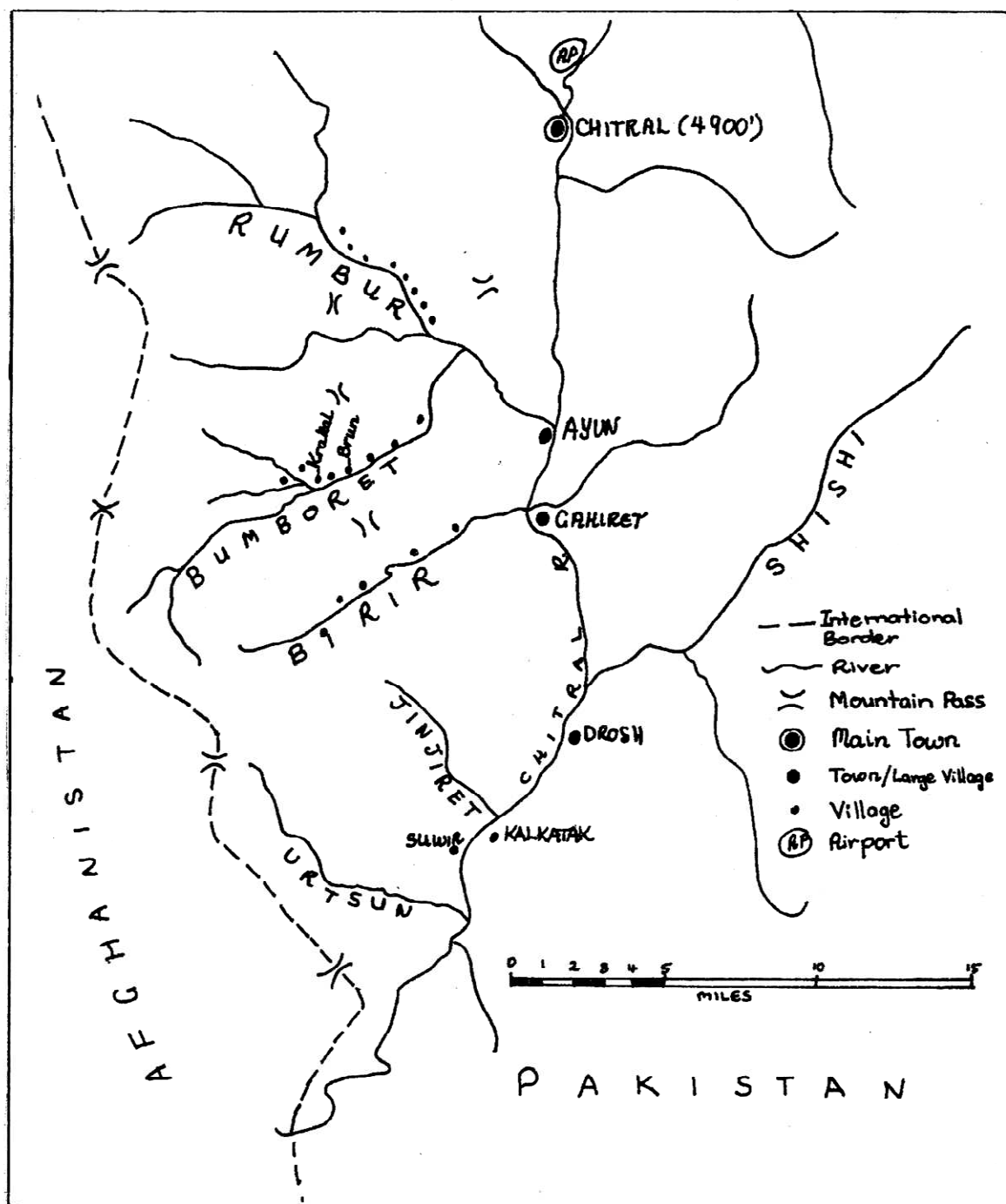
Rumbur, Birir, Urtsun and Jinjiret; however, the Kalasha community, as commonly known (i.e. adhering to the traditional religion) is situated only in the first three of these valleys.⁶ This map also shows most of the villages or hamlets. A village or hamlet may contain anywhere from about five to about 50 dwellings.

Figure 1.1: Map of Pakistan, showing inset for figure 1.2



⁶ These are the official names in both Urdu, the national language, and in Khowar, the language spoken by the majority community in Chitral District. Kalasha speakers have their own slight variations of these names, pronounced /mumo'ret/, /ruk'mu/, /bi'riu/, /uts'un/ and /d̪ɜ̃ind̪ɜ̃i'ret/. Some Muslim speakers of the Kalasha language have also been found in the villages of Suwir and Kalkatak in the main Chitral valley.

Figure 1.2: Map of the Kalasha Valleys



The villages in each valley, and the very swift, voluminous streams which they flank, are at elevations ranging from 2,000 to 2,500 metres. The terrain is very steep and landslides and river flooding are common. The climate is generally very dry, with warm summers and cold winters with snow.

The Kalasha irrigate their fields with an elaborate system of water channels and aqueducts to produce green and fertile oases in what is generally an otherwise arid landscape, though pine and cedar forests do abound on the hilltops. The Kalasha build their houses and walls with slate, carefully chipped and put in place without mortar, with beams of cedar placed horizontally every meter or so. The floors are dirt and the roofs are cedar coated with mud, with a hole or chimney in the centre for smoke. A study of Kalasha buildings has been made by Harrison (1996), a historic-buildings architect with wide experience in South Asia.

1.1.3 Population

Estimates of the Kalasha-speaking population, made over the last century, have ranged from about 1,300 to about 5,700, but an estimate of 4,500 to 5,000 Kalasha *speakers* is close to accurate, according to Decker (1992b, pp. 101–3). Being based on a linguistic survey, this figure also includes the Kalasha speakers who have converted to Islam. The actual Kalasha *community* (adherents to the traditional religion) was estimated in 1983, by Saifullah Jan, an educated Kalasha man, to be about 3,700 in number. Population surveys I conducted in 1982 and 1985 suggested a smaller estimate of the traditional Kalasha community, ranging from 1,800 to 2,300 people.

The main difficulties of taking a proper census are the remoteness of the Kalasha community, lack of resources in the local government administration, the complex mixture of ethnic groups residing in the valleys, and confusion as to the definition of Kalasha, arising from the variables mentioned above (in subsection 1.1.1).

1.1.4 Sociology

The Kalasha people live among, and freely intermingle with, several other different ethnic groups. The socially dominant group in the district are the speakers of *Khowar*.⁷ Though there

⁷ The word *Khowar* is derived from the morphemes /k^ho/, the name of the ethnic group that speaks that language, and /war/, their word for ‘language’.

is a general underlying animosity between these two groups (which occasionally explodes in a skirmish) there are also strong interrelationships between many individuals from the two groups, which transcend normal work and family divisions. So, overall, the picture is generally peaceful and cooperative, though occasionally flecked with strain.

Other minority ethnic groups living in the Kalasha Valleys include speakers of the Eastern Kativiri dialect of *Kati* (described in Decker 1992b), who, having migrated from Afghanistan in the late 19th century, have established themselves in a large village at the heads of two of the valleys (Bumboret and Rumbur). Then there are small pockets of *Gujar* people, traditionally nomadic pastoralists belonging to a very scattered ethnic group who, though spread across northern Pakistan, India and Afghanistan, still speak a common language.

The primary unit of the Kalasha community is not the nuclear family but rather the lineage, that is, extended families, with all generations living together patrilocally. The secondary level of identity is in relation to the village of residence and clan (typically one or two clans per village); thirdly (for women) to the valley of their birth; and fourthly to their religion. There is no perception of the entire Kalasha language group as a community because many speakers have converted to Islam, changing their community allegiance, though often continuing to speak the language. In the last two or three decades there have also developed political allegiances to various independent members of the Pakistani federal parliament, who represent minorities and regularly canvass their influence among the Kalasha. These political allegiances have also come to strongly affect the way certain Kalasha individuals, especially progressive young men, relate to each other.

There is no formal political system among the Kalasha people, but rather, an informal system of eldership and local councils who make religious and social decisions when necessary. Appointment to these informal bodies is by consensus of existing members, and is usually based on such values as social commitment, social influence, social achievements, wisdom and generosity.

Marriage is traditionally instituted by secret courtship and elopement, followed later by the paying of bride price and feasting. In latter times, some marriages have been arranged by the families (sometimes even prepubescently), as is the custom among the Muslims, though often it is the uncles who do the arranging, rather than the fathers.

A Danish ethnographer has described the relationships that the Kalasha have with the outside world, and how the Kalasha respond to these relationships—their strategies for survival, both culturally and individually (Sperber 1995a). She has also described Kalasha

women's dresses, accessories, body decorations and textile techniques—their intricacies, their social importance, and the changes that result due to fashion trends from within the community, as well as, again, various influences from the outside world (Sperber 1996).

1.1.5 Economy

Traditional Kalasha economy revolves primarily around their goats and sheep. On a day-to-day basis, goats provide much of the dairy component of the Kalasha diet. Less frequently, but more significantly, goats (and sometimes sheep) are objects of sacrifice for a wide variety of purposes (when their meat is also heartily consumed). Some articles of clothing, adornment and household use are by-products of goats and sheep. By far the most important and highly regarded traditional occupation is that of the indispensable male goatherd. He must take turns with his brothers at leading the mixed flocks to their summer pastures for several weeks at a time. (Cows, usually herded by young children, are also kept for their milk and meat.) Crop cultivation is the second most important work in the Kalasha Valleys, performed by both men and women. They include rotated cereal crops (wheat, corn, barley and rice), fruits, and nuts (walnuts, apricots and their kernels, mulberries, grapes, apples, pears and pomegranates).

Though traditionally a cashless society, the intrusion of commercialism means that cash (Pakistani rupees) has now become an almost indispensable ingredient in the Kalasha society. One very big issue that dominates the Kalasha community is ongoing and protracted litigation over land disputes with their Muslim neighbours, some of whom live in nearby villages and towns. Though often the winners of these lawsuits, many Kalasha families are led into impossible debts because of legal fees, which they can sometimes only escape by selling land. Either way, Kalasha-owned land is slowly disappearing.

The other more visible issue of Kalasha society is the advent of tourism, both recreational and academic. As interest in the Kalasha continues to spread, more tourists are attracted. This accounts for thousands of visitors each year, mostly from the rest of Pakistan, Europe, North America, Japan and Australasia. Though many of these are individuals, groups are also very common, and some are professionally organized tours. Journalists are usually individuals. The academic tourists also tend to be more individuals or very small groups of students or professionals in the fields of anthropology, linguistics, musicology, etc.

The influx of visitors to the Kalasha Valleys over the last two decades has created a fledgling but vigorous tourism industry, with hotels of every standard springing up in many locations, owned and operated by members of both the Muslim and Kalasha communities.

The immediate benefit is employment and cash, although, with supply still far exceeding demand, the impact on the environment has gone unchecked. This is evident in the increasing practices of timber logging and hunting, and increasing diminution of wildlife habitats, unnatural erosion and flooding, and water and soil contamination. The impact of tourism on ecology, economy and culture is equally, if not more, significant, bringing about changes in the traditional economy and systems of work, and changes in social attitudes and worldviews. All these issues are dealt with at length in a Masters thesis based on research in 1993 by a young Muslim Chitrali man who lives in Bumboret, the largest of the Kalasha valleys, (M Ayub Khan n.d.).

A number of general aid projects have been completed in the Kalasha Valleys over the last two decades, and several more are current.⁸ Many of these have been sponsored officially by the Agha Khan Foundation, the governments of Pakistan, UK, Australia, Canada, Greece, Japan, etc., and corporately and privately by several national and international foundations and individuals. However, of all these, only a Greek-sponsored school and the Kalasha teachers of the Pakistan government-sponsored Kalasha schools are actually promoting the use of the Kalasha language. Vernacular literacy is not a feature of the other projects.

Two registered indigenous cultural preservation societies, the Kalash People Welfare Society (KPWS) and the Kalash Indigenous Survival Programme (KISP), were both founded in the 1990s (the former by a committee of 31 Kalasha men, and the latter by an educated Kalasha woman), with the express purpose of funding and managing locally initiated, and carefully planned change for the benefit of Kalasha society and environment. The role of the KPWS in promoting indigenous education is mentioned in the next subsection (1.1.6).

1.1.6 Education

Since the latter half of the twentieth century, the Pakistan government has provided schools in each of the Kalasha valleys, first primary, then middle, and now a secondary school in the main valley. Education in Pakistan is not compulsory and there is no minimum age for leaving, so school is typically attended by only a minority of children. The medium of instruction of most government schools is Urdu, the national language (though for most it is

⁸ Examples include programs and schemes providing reticulated spring water, hydro-electric power, education, health and medical, tour guiding, cottage industries producing dolls, honey, etc.

not their first language). Subjects include Urdu language and literature, maths, science, English language and literature, Islamic studies and social studies.

The original schools in the Kalasha valleys were attended mainly by the Muslim children, both boys and girls. Though a few Kalasha boys and girls attended, they were usually teased and bullied by fellow students and teachers alike, for being kafirs. The exploitation of that system for the religious coercion and persecution of non-Muslim Kalasha children was a situation that the Kalasha community could not tolerate. Some of the Kalasha children could not cope with this and left, though a few persevered to various levels of education.

Because of this problem, in the early 1990s, when enough young Kalasha men were sufficiently educated to become teachers, the government agreed to establish schools exclusively for Kalasha children, and at least initially, staffed exclusively with two non-Muslim Kalasha teachers each. It was stipulated that only Kalasha teachers be employed in these schools (though that rule has been relaxed slightly since). This development has also helped to ward off religious opportunism.

Though the education is officially Urdu medium, actual instruction is also conducted in Kalasha, and some Kalasha teachers have lobbied the education department to make Kalasha literacy an official subject in the curriculum. School education is now very popular and is becoming more the norm among the Kalasha community. Student numbers have swelled exponentially, with even some Kalasha women attending. At the time of writing (2005), there are five Kalasha schools, about eleven teachers, and about six hundred students.

The constitution of the Kalash People Welfare Society (KPWS, mentioned in the previous subsection, 1.1.5) states that part of its foremost aim is ‘to preserve the Kalasha culture ... and publish the folk stories of Kalash people’. Furthermore, it aims ‘to provide education to Kalash men and women, admit them to schools and colleges and to provide them scholarships’.⁹

1.1.7 Religion

The Kalasha religion is polytheistic: one supreme creator god and several demigods, and other deified beings, some of whom are only acknowledged in one of the valleys. Other supernatural beings include demons (who not only cause gross misfortune but also instil fear and terror) and fairies (whose behaviour is malicious at worst or mischievous at best). Fairies

⁹ (Source: Kalasha People Welfare Society n.d.)

are human in form, live in remote areas of the surrounding mountains and valleys, and though their movements are clandestine, many Kalasha can testify to a strange encounter they have once had with a fairy. Many Kalasha have given serious reports of how fairies occasionally have sexual relations with Kalasha mortals.

There are many elaborate religious practices involving purification, ritual and sacrifice, which are performed at particular times; for example, various stages of the goat-herding and dairy production processes, spring, harvest and winter-solstice festivals, and the occasions of birth, death, menstruation, etc.¹⁰ In each valley there are several altars to demigods, and one or two temples containing a rudimentary altar for the goddess of family life, where some rituals are performed. Dancing and singing (usually to the accompaniment of a pair of drums) are a frequent and very popular activity practised at every festival, and also at the funerals of men. Though the seasonal festivals have religious significance they are regarded by most as merely social occasions of fun and revelry, often lasting many days.

Several studies have been made of Kafir religion in general, one of the most notable being a historical perspective in a German monograph by Jettmar (1986). Kalasha religion, in particular, has also featured in many works; for example, an examination of the practice of exorcism by shamans (Loude 1996), the status of women in the Kalasha religion (Lièvre 1996) and a thorough treatise of Kalasha religion by Jettmar (1990).

Among the Kalasha community, including their own kin, are some who have converted to Islam. It is a longstanding and ongoing trend, though it has slowed in recent years because of a government prohibition on proselytism. Some of these conversions are not primarily out of deference to the tenets of Islamic faith, but more for practical and social advantages (e.g. to avoid the very high expense of a traditional Kalasha funeral), or as an obligation (e.g. to a Muslim priest who has prayed for a sick child who then recovered). Some Kalasha convert out of a desire for a different lifestyle (e.g. Muslim women stay in the house, where work is easier than for the Kalasha women who work in the fields), or because a Muslim family has decided to marry their son to a Kalasha girl, with some payment to the Kalasha girl's family. Some thereby gain financially (e.g. by cancellation of a debt). Some are forced into conversion by trickery, (e.g. a girl's Kalasha clothes were stolen by Muslims so that she was forced to wear Muslim clothes).

As already explained, Kalasha who become Muslims sometimes retain their language for a generation or so, but some change to speaking Khowar because of social pressure. Muslim

¹⁰ A thorough treatment of the winter solstice festival has been provided by Loude and Lièvre (1984, c. 1986).

converts no longer call themselves Kalasha but Muslim. The Kalasha refer to them as *shek* (/ʃek/), which, though a term of honour in Arabic and Urdu (where it is pronounced /ʃex/), carries a mildly derogatory connotation when used by the Kalasha. There are many more of them in Birir Valley than in the other two valleys. These converts cease their participation in Kalasha customs, and the women cease to wear the traditional Kalasha dress. Female converts (usually those who have married a Muslim man) generally start switching from speaking Kalasha to Khowar, while male converts vary in the retention of Kalasha language, according to personal preference and/or social pressures bearing on them personally. All Muslim converts are eventually absorbed into the surrounding community, and while offspring born before the parents' conversion remain Kalasha, those born afterwards are automatically Muslims.

A handful of Kalasha youths became Christians in the 1960s or 70s after a Swedish missionary and a Punjabi social activist invited them to the city of Lahore to receive an education. Since then, a few others have followed the same path. One or two young Kalasha men have also become Baha'is. All these individuals have continued speaking their language, as well as their participation in traditional Kalasha customs and dress. However, a government edict has been issued to protect the community from all outside religious influence: proselytism by any outsiders (Muslims or Christians) is prohibited and punishable.

Despite this adherence by some Kalasha to different religions, the community in the Kalasha Valleys is a unique setting in that even nuclear families can coexist in remarkable harmony with two or three religions represented between them. (There is no such tolerance in the majority Muslim communities elsewhere in Pakistan.) However, the harmony is only skin deep. Outside normal family loyalties, other important matters at the community level, like land disputes and logging, contribute significantly to the general tensions felt between the two religious sodalities.

1.1.8 History

There is a popular belief among the Kalasha (and many non-Kalasha) that they are descendants of Alexander the Great. This theory is based on the fact that Alexander brought with him troops from various parts of his empire when he came to the subcontinent. Some evidence suggests that the Kalasha society originated in Syria, with a mixture of influences from other nearby cultures, including Greek. If this is true, they may have migrated from

Syria along with Alexander's General Seleucus (Trail, G 1996, p. 373). Strand (2001) also traces a possible history of the Indo-Aryans generally, including the Kalasha, from the equestrian tribesmen who originally occupied the area between the Black and Caspian Seas about four thousand years ago, and who migrated as far as South Asia.

The local history of Chitral district deserves a mention because it forms part of the relatively recent history of the Kalasha society. An Italian researcher writes, 'the former presence of the Kalasha in southern Chitral is well known to all the inhabitants of the area and is unanimously considered a historical fact' (Alberto Cacopardo 1996, p. 247). Schomberg (1938, p. 210) had earlier expressed his more daring opinion, based on historical and linguistic evidence: 'I think that unquestionably the whole of Chitral was once inhabited by one people, and those were the Kalash'.

The Kalasha were once governed and protected by a royal family who enjoyed great honour and prestige.¹¹ Eventually (some time between the 14th and 16th centuries) many of them converted to Islam. In the 19th century the British extended their empire into the Indian subcontinent, which included Chitral. However, the princely state of Chitral remained intact and retained its formal political independence, in association with both British rule, and from 1947, the government of Pakistan. It officially merged into Pakistan only in the 1960s. The Kalasha community is now officially a protected minority, but is also promoted for tourism by the Pakistan government.

A brief indigenous account of the history of the Kalasha, their traditions, and the impact of recent developments, was presented by one of the members of the community, Saifullah Jan, to the Second International Hindukush Cultural Conference in 1990 (Jan 1996).

1.2 THE KALASHA LANGUAGE

1.2.1 Main features

Kalasha (also known as Kalashamon) has been classified as belonging to the Chitral sub-branch (along with Khowar) of Dardic languages, which are spoken in the Northwestern zone

¹¹ Remnants of this princely status (in the form of wealth and social status) remain to this day among a few hundred surviving descendants of the former royal family.

of Indo-Aryan languages (Fussman 1972; Grimes 2000).¹² The Kalasha language has a large inventory of about 62 phonemes: half of the vowels are rhotic, half of them are nasal (a quarter of them are simultaneously rhotic and nasal), nearly one-quarter of the consonants are retroflex, and nearly one-third of them are aspirated. The phonology of Kalasha will be treated more fully in Chapter 4 below.

The standard order of grammatical constituents for a statement is Subject-Object-Verb. Verbs are inflected into over twenty forms, with person/number/tense being the most common. Several verbs are marked according to whether their subjects are animate or inanimate. Grammatical number is rare in nouns, only manifesting in an oblique plural inflection in a handful of words.¹³ There is no grammatical gender. There are nearly ten grammatical cases for nouns, each of which features a suffixal inflection.

1.2.2 Dialects

The Kalasha language is spoken primarily in three valleys of the Chitral district: Bumboret, Rumbur and Birir. There is some minor dialect variation between these valleys, and even between certain villages within them. There are also dialects of Kalasha spoken by a few people who inhabit the Urtsun and Jinjiret valleys, south of the three Kalasha Valleys. Those communities converted to Islam around the 1930s, and have abandoned their first language in favour of Khowar, the language of the Kho people who dominate the district.

If one dialect were to emerge as the overall standard it would be that spoken in the (central) Bumboret Valley because that is by far the most populous and developed of the Kalasha Valleys, being home to about 1,500 Kalasha speakers.

The (northern) Rumbur Valley is inhabited by about 500 speakers of Kalasha. Their dialect differs from the main one by being slower in speech, having a greater use of the glottal plosive allophone of the voiceless velar plosive /k/, and a few unique lexical items.

The (southern) Birir Valley dialect is spoken by about 1,200 people. It differs in having more lexical variation and certain unique lexical items, the use of contracted forms of a few common words, and the employment of the retroflex flap phoneme (/ɽ/), which the other dialects do not use.

¹² It should be noted that there has been disagreement over this classification of Dardic languages. Some scholars believe that although ‘Dardic’ is a convenient label for the Indo-Aryan languages of this region, it is not a distinct group in Indo-Aryan genealogical classification (Morgenstierne 1961; Strand 2001; Mock 2004).

¹³ This is exemplified in chapter 5 (section 5.1.2).

There is complete intelligibility between all three of these dialects. A more detailed description of dialect differences is given in Chapter 4 (subsection 4.3.5).

1.2.3 Bilingualism

Other languages known and used by Kalasha people, mainly by a small percentage of the men, include the national language (Urdu), the provincial language (Pashto), the district language (Khowar), and the nearest local minority language: the Eastern Kativiri dialect of the Kati language of Afghanistan, which is spoken in the highest village of two of the Kalasha Valleys, and also just across the Afghan border. English is also being learnt at a growing rate in recent years, especially among the young men.

Of these other languages, the one best known by Kalasha men and women is Khowar, the language of the local majority (the Kho) in the district of Chitral. Khowar is genetically close to Kalasha, being the only other language sharing the Chitral sub-branch of Dardic languages (see language relationship diagram in Appendix 1). However, ‘there is no report of intelligibility between these two languages’ (Decker 1992b, p. 103). Being a small minority group, and regarded as socially inferior, the Kalasha people accommodate to the Khowar language in all interactions between the communities. Kalasha men have a better knowledge of Khowar than Kalasha women do, because of the generally greater contact and trade they engage in.

Khowar is not only used by Kalasha for business but also with their Muslim relatives. Kalasha individuals who become Muslims start to favour the Khowar language in preference to their own. Within a couple of generations, the descendants of these converts grow up with no competency in the Kalasha language at all. The number of Muslim converts whose first language is now Khowar, and whose second or original language is or was Kalasha, is in the order of about a thousand.

Some variation exists in the language repertoires of the Kalasha people in the three valleys, and even within some villages. Table 1.1 below sets out the percentage of Kalasha speakers who had various levels of bilingual competence, based on surveys I conducted in 1982 and 1985 (Cooper, G 1992). *Coordinate* competence is a level at which a speaker is functionally just as competent and comfortable in L₂ as in L₁; *subordinate* competence is a level at which a speaker has some limited functional competence in L₂, less than in L₁, but more than minimal competence; and *incipient* competence is a level at which a speaker has only minimal functional competence in L₂. Bilingualism in Khowar is much greater in Birir Valley, where,

as stated in section 1.1.7, there are many more Muslim converts who have changed their language. This sector of the community is therefore set apart in separate statistics. In this table, the term Kalasha *includes* those who are Kalasha by birth, but have since converted to Islam. They are often the ones who stop transmitting the Kalasha language to their children.

Table 1.1: Bilingual competence in Kalasha community

Language	Community	Coordinate	Subordinate	Incipient
Khowar	Birir Muslims	97%	2%	1%
	Birir Kalasha	80–90%	10–20 %	5%
	All Kalasha	45%	55%	1%
Urdu	All Kalasha	1%	5%	20%

Only 1% of Kalasha females have incipient bilingual competence in Urdu, however most of them have subordinate bilingual competence in Khowar. Less than 1% of males and females have even incipient bilingual competence in English.

While the vernacular language is actually spoken by all sections of the Kalasha society in most situations, a common attitude of young Kalasha men to their own language is that it is not as ‘fast’ as Khowar for the wider social and commercial domains of interaction which some of them now enjoy. The above table shows that only about 55% of the overall Kalasha community are actually more functionally competent in Kalasha than in Khowar (including only 10–20% of Kalasha speakers in Birir Valley). However, the prevailing attitude and practice among children, women and older folk (especially the elders, who feel a responsibility to guard the traditional heritage of Kalasha culture whilst guiding their community into the future) is one of explicit preference for the use of the Kalasha language instead of Khowar, in all domains of discourse.

Krauss proposed a set of suggestions for the classification and terminology of degrees of language endangerment, whose basic terms are given in Table 1.2 overleaf.¹⁴

¹⁴ This typology was produced by Krauss following a workshop entitled ‘Language endangerment, research and documentation: setting priorities for the 21st century’, held in Bad Godesberg, Germany, 12–17 Feb 2000 (Ostler 2000). His classification and terminology were subsequently featured in Brenzinger (2001), and in revised form in UNESCO (2003).

Table 1.2 Krauss' (2001) proposed typology of language endangerment, based on intergenerational language transmission

Generation(s) of speakers	Degree of endangerment
all age groups, including children	safe
some children in all domains; all children in some domains	unsafe
parental generation and upwards	definitely endangered
grandparental generation and upwards	severely endangered
great-grandparental generation	critically endangered
none	extinct

According to the above typology, which focuses only on intergenerational language transmission, the Kalasha language would not be counted as endangered because it is spoken by all generations in the community. However, a subsequent document which deals extensively with language vitality assessment (UNESCO 2003) lists eight other major evaluative factors. One of these is the proportion of L₁ speakers, as set out in Table 1.3.¹⁵

Table 1.3: UNESCO (2003) thresholds for language endangerment, based on proportion of speakers within total population

Proportion of L₁ speakers in community	Degree of endangerment
all	safe
nearly all	unsafe
majority	Definitely endangered
minority	severely endangered
very few	Critically endangered
None	extinct

The evaluation given in this table, combined with the statistics on bilingual competence in the Kalasha Valleys, presented in Table 1.1, would suggest that the Kalasha language is definitely endangered, especially in Birir Valley. The corollary to this would be that the threat

¹⁵ The other evaluative factors in the UNESCO report are absolute number of speakers, shifts in domains of language use, responses to new domains and media, materials for language education and literacy, and external and internal attitudes to the language. By all these factors combined the endangerment status of the Kalasha language is in the balance, and could go either way.

to the Kalasha language depends on the lack of social distinction between Kalasha speakers and their non-Kalasha-speaking neighbours.

Decker (1992b) has published results of a sociolinguistic survey of the Kalasha language, with a special emphasis on dialects and bilingualism. He reports that ‘only a few Khowar-speaking families live in the Birir Valley’, but he is referring to non-Kalasha families (p. 107). He also claims that ‘respondents from all the areas said that Khowar is their best second language; however, those from Birir Valley said they could speak only a little Khowar’ (p. 110). This contradicts the data in Table 1.1; however, they are supported by additional evidence, obtained more recently (2003) by independent questioning of two Kalasha speakers, both of whom reconfirmed that ‘all Birir people know Khowar well ... [and] all Birir people know Kalasha, whether Muslim or not.’¹⁶

1.3 PREVIOUS RESEARCH ON THE KALASHA LANGUAGE

Research studies about the Kalasha language have been published by a variety of people for more than a century. Some of these have been comparative studies and others have focused purely on Kalasha.

1.3.1 Comparative areal research including the Kalasha language

Grierson (1915) published a very comprehensive series of volumes surveying the languages of India (which included what is now Pakistan). One of these volumes deals with Dardic languages (of which there are twenty-seven), and Kalasha is one of them. Grierson includes a few specimen texts and a word list. However, his data is very limited and skeletal, and his brief account of Kalasha grammar is based on Leitner (1877, see below).

Morgenstierne (1932) took a more analytical approach to the languages of the area in his short *Report on a Linguistic Mission to North-Western India*. Kalasha is one of about thirty languages he had encountered on his six-month tour in 1929, and an outline of it is sketched in a three-page account. Here he makes some useful observations on a few phonetic, lexical and morphological aspects of the language. However, not surprisingly, it is all from a comparative point of view—first comparing Kalasha with neighbouring Khowar and Kati languages, then comparing the Kalasha dialects with each other on these points.

¹⁶ Reported in an email from E Hamilton (26 January 2003), who was living in the Kalasha Valleys at that time.

Turner (1966) published *A comparative dictionary of the Indo-Aryan Languages*, in which he not only gives the ancient Vedic and Sanskrit words, but also cites the present forms of some of the original words in many Indo-Aryan languages, including Kalasha. However, the value of this work is limited by its focus on comparative lexicology.

Bashir has done extensive research in the region, which has included several field trips to the Kalasha Valleys since 1974. In a paper presented at the 12th Annual Conference on South Asia in 1983 she examined several traits of the Kalasha language in the light of Masica's (1976) 30-trait 'device for generalising typological distance', concluding that Kalasha embodies the South Asian areal feature complex, but from a 'borderland' position because of its 'preservation of Indo-Aryan traits, and its reflection of sub-areal convergences.' Bashir presents thirty-three analysed sentences to illustrate her discussion. Her subsequent PhD dissertation focused on specifically syntactic topics in Kalasha (verbs, transitive and causative relations, complements and relative clauses)—also from an areal and typological perspective (Bashir 1988b). Other more specific studies of Kalasha linguistics published by Bashir include inferentiality in Kalasha and Khowar (1988a), and involuntary experience in Kalasha (1990).

Extensive sociolinguistic surveys of the languages of northern Pakistan were conducted by the Summer Institute of Linguistics, in cooperation with the Pakistan Government Ministry of Culture, and several Pakistani institutions, between 1986 and 1991. The results of these surveys were published in a five-volume series, including one entitled *Languages of Chitral*, which has a chapter on Kalasha (Decker 1992b). After an introduction, outlining various aspects of the Kalasha social situation, this survey reports on the dialects, interaction with the neighbouring languages, second language usage patterns and proficiency, and language attitudes and vitality. Decker concludes that without any language promotion the vitality of the Kalasha language is uncertain and shaky, but that with language development it can survive.

Subsequent, more detailed research has focused on specific aspects of some of the northern Pakistani languages, resulting in a separate eight-volume series published jointly by the National Institute of Pakistan Studies (NIPS) & the Summer Institute of Linguistics (SIL) from 1997 to 1999. These include monographs on topics such as the phonetic systems, folktales, culture, morphology, grammar and lexicology of the Indus Kohistani, Kalami, Shina and Burushaski languages. Of special significance to this thesis is the seventh volume of this

series: the *Kalasha Dictionary* (Trail and Cooper 1999), discussed in more detail in subsection 1.3.2, below.

1.3.2 Kalasha-focused research

Leitner (1877) was the first to write a short description of Kalasha, among other Dardic languages, followed in 1880 by a work on a single (then) Afghan dialect of the language. Morgenstierne (1973, p. 184) later assessed its value as a pioneer work, but concluded that it was ‘not very exact, and the forms given have frequently to be interpreted – phonetically as well as morphologically – in the light of knowledge gained from other sources’.

Morgenstierne’s academic career (already noted in subsection 1.3.1, above) spanned several decades, during which he published over 200 titles. His *Notes on Kalasha* (1965) was reprinted in *The Kalasha Language* (1973), which also contained an extensive vocabulary and seventeen interlinearized texts that he collected on his 1929 field trip, and ten collected and given to him by other people. His discussion included historical phonology (comparing features of Kalasha phonology with other Dardic languages), a brief description of the Kalasha phonemic system and an analysis of its morphology. Unfortunately, nearly all of Morgenstierne’s Kalasha data was collected from one single informant in a relatively short space of time, and he acknowledges his dependence on that informant by declaring his inability to find a replacement when the informant had to leave. Morgenstierne (1973, p. 186) writes, ‘the result was that my material, as well as the grammatical sketch based upon it, remain lamentably fragmentary. The phonetical notation is in many cases vacillating and uncertain’.

Parkes has been researching Kalasha culture and language since at least 1972. His documented and published works (e.g. 1975, 1990, 1994, 1996) have focused mainly on the performance aspects and social significance of Kalasha history, oral literature, music, song-making, and ceremony—with particular attention to the orientations and roles of participants and the ‘transmission, elaboration and reception of traditional knowledge in a predominantly non-literate society’ (Parkes 1996, p. 316).

Cooper and Trail were the first to make sustained investigations of the Kalasha language, in both independent and collaborative research. Because the timing of our involvement overlapped, we agreed not to duplicate, but rather to complement each other’s work. My research, between 1982 and 2001, included investigations and analysis in areas such as sociolinguistics, phonology and morphology, and developing the linguistic, orthographic and

computational groundwork for a writing system (using both Arabic and Roman scripts). What has been documented and published as a result includes not only academic material, but also foundational literacy materials, in collaboration with Elsa Cooper and a few others. From 1982 until the present, Trail has worked extensively on Kalasha grammatical analysis, discourse analysis, lexicography (focusing especially on semantic relations in the Kalasha lexicon) and phonology. All of the abovementioned researchers speak the Kalasha language fluently, having spent considerable periods of time living with Kalasha families in their villages. The most relevant published and unpublished documentations of each are listed in the Bibliography.

A comprehensive and detailed linguistic dictionary of the Kalasha language (Trail & Cooper 1999) is written in English, but includes both Roman- and Arabic-script renderings of Kalasha words, and Urdu glosses in Urdu script. It was published as the seventh volume in the eight-volume NIPS/SIL series mentioned in subsection 1.3.1, above.

1.4 A WRITING SYSTEM FOR THE KALASHA LANGUAGE

Systematic attempts to develop a writing system for the Kalasha language began in the early 1980s after I had developed an Arabic script and an interim Roman orthography.¹⁷ The Arabic script was strongly preferred by all respondents to several orthography surveys that I conducted in different locations, in 1986, 1994 and 1996. The respondents were already literate in Urdu.¹⁸ These initiatives created a groundswell of interest in vernacular literacy over the years, which has resulted in transcriptions of texts by some dozen or so individuals, mostly adolescents. There are now scores of draft manuscripts, most of which are yet to be published because the orthography has been evolving (see samples in Appendixes 16 to 25).

With a view to publishing vernacular texts in the originally preferred Arabic script, in the mid 1980s I adapted for Kalasha a software package designed by Hoyle (1986) to print some early texts in the Naskh (flat, typewriter-) style of script. However, seeing the overwhelming preference for the Nastaliq (sloping, calligraphed) style, I later modified for Kalasha and several other northern Pakistani languages a new, commercially available Urdu word

¹⁷ The term *Arabic script* in this thesis is used in a parallel way to the common usage of the term *Roman script*. It means any script based on, and generally resembling, the script used for the Arabic language. The term encompasses as much variation and adaptation in style, alphabetic characters and conventions as the many languages it now represents also vary, as is evident in the Arabic-based scripts of Farsi, Urdu, Pashto, Sindhi, Kashmiri, to name a few. There are also various subtypes (styles, fonts) of this script.

¹⁸ Sample pages from these orthography surveys are presented in Appendixes 10 and 11.

processing and desktop publishing software package, and incorporated various adaptations of the Arabic-based Urdu script (Systems Limited & Cooper 1992).¹⁹

Two foundational Kalasha literacy books were created, designed, edited and published using the Urdu (Nastaliq Arabic-based script) software. These were a pre-reader and an alphabet book, both designed primarily for classroom use by teachers (Cooper, Hall and Cooper 1994a, 1994b), and were issued to all Kalasha schools, with some demonstration lessons to the classes.

Detailed teachers' manuals were also written to accompany each of the abovementioned books, on request of one of the teachers. Each of these was produced in both an Arabic-script edition (Cooper, E 1996a; 1996b) and an interim Roman-script edition (Cooper, E 1996c; forthcoming). These were distributed to some of the Kalasha teachers. Another interim Roman-script edition of the Kalasha alphabet book (Cooper, Hall & Cooper 2000) was produced and distributed to the teachers for each school at the orthography conference in 2001. A revision of that was then published in 2003. Several of the drafted Arabic-script manuscripts may now serve as trial early readers for those already literate in Urdu, and possibly for the Muslim Kalasha-speaking communities.

During these last two decades, the imperative was to document the language as rapidly as possible, rather than wait for the optimal research-based script and orthography to become available. This explains why several manuscripts and trial publications have already been produced. The opportunity exists now for the use and regularization of the evolving writing system through the composition, transcription and production of indigenous literature by the Kalasha community.

1.5 THE SCOPE OF THIS THESIS

The primary purpose of the research associated with this thesis was to examine and address the issues in developing a principled writing system that is most appropriate for the Kalasha language and community, and could be accepted by them as a standard, and a foundation, for the writing, editing and reading of indigenous literature, on which present and future generations of Kalasha readers and writers can build.

¹⁹ The Nastaliq script style was first developed by Persian calligraphers, and used to be the predominant style for writing Persian (giving rise to it also being referred to more specifically as *Persian script*). It is now the preferred script style for Urdu.

After an introduction to the context of Kalasha society and language, the thesis begins with the fundamental question of whether languages (like Kalasha) should be codified in writing. The arguments will be set within the context of the broader debate concerning language ecology and the effects and functions of mother-tongue literacy generally. Chapter 3 then looks at the history of scripts used for other languages in the Middle Asian region, to see whether any particular scripts have proved more durable and effective than others.

In Chapter 4, we present a phonological analysis of the Kalasha language, as the foundation for a writing system. Chapter 5 focuses on morphological and syntactic issues, and their bearing on spelling decisions and rules. The orthographic representations of variants are also discussed. Chapter 6 reports on the Kalasha orthography conference at which possible scripts and alphabets for writing Kalasha were discussed in terms of Roman and Arabic script options. The public consensus for each issue is indicated. Chapter 7 implements the conference decisions for a Roman-based Kalasha alphabet. Various concomitant rules & conventions are detailed. Chapter 8 looks at the written language from a typographical and publishing point of view, with a discussion of general strategies, methods of literacy, and types of Kalasha literature.

In conclusion, Chapter 9 looks at the limitations of this research in developing a new writing system, the means and methods used to develop it, and the new literature that will result from it. From there, some areas of possible future research are indicated.

SHOULD LANGUAGES BE CODIFIED IN WRITING?

2.1 INTRODUCTION

More than 6,800 languages are spoken in the world today (Grimes 2000), but only one-third of them have ever been written (Trosterud 1999, p. 16). Most of these (about 1,500) were written for the first time in the latter half of the twentieth century (Grimes & Gordon 1980, p. 93).

This flurry of activity grew out of the new discipline of field linguistics. It provided the tools and methods for research, analysis, and for devising writing systems for hitherto unwritten languages. While there is no lack of indigenous initiatives to codify their mother tongues, much support for these efforts comes from individuals and groups outside the speech communities. However, the significant division of interest in terms of mother-tongue literacy (which is the focus of this thesis) is not between speakers and non-speakers of the unwritten language, community members and non-members, insiders and outsiders, but rather between those who advocate and promote mother-tongue literacy, and those who do not, whichever side of the community fence they stand. The former category also includes those who advocate multiliteracies including the mother tongue. The latter category also includes those who advocate dominant-language literacy at the expense of mother-tongue literacy.

The development of a writing system for a previously unwritten language lays the groundwork for mother-tongue literacy which, in turn, provides the platform for the establishment of indigenous literature. However, the impact of literacy and literature on a previously non-literate culture is potentially enormous, because it involves substantial changes to some of the traditional sociolinguistic structures, for example, potentially creating an educated class. What happens to a language's oral literature when writing is introduced? The traditional roles of orators and storytellers will no doubt change, and as a result, the status and kudos attached to these roles will be severely challenged.

Furthermore, many societies and cultures are also influenced by the literacy and literature of other societies and cultures. The promotion of non-mother-tongue literacy can contribute to language shift and language loss. And non-indigenous literature, with its inherent philosophies, values and beliefs, often influences an indigenous society's worldview. Should

non-mother-tongue literacy be discouraged, in order to limit cross-cultural tensions in an indigenous society? Or should mother-tongue literacy and indigenous literature be developed to protect a minority society against outside influence?

It is these sorts of considerations, the potential effects of mother-tongue literacy and indigenous literature, whether internally or externally mobilized, that lead us to ask whether (and why) the Kalasha language, or indeed any language, should be codified in writing. If not, then we are faced with other questions, for example, at what level of relatedness should two different speech communities (which would naturally interact with each other) be kept separate and distinct in their oral and written literature, to avoid mutual influence? Does every instance of speech behaviour and written tradition need to be protected from every other? We begin to see that languages are not self-contained entities, but rather they are multi-level forms of human behaviour that form part of a wider network of communication.

2.2 LINGUISTIC ECOLOGY

Much has been said, written and enacted, in the latter half of the twentieth century about ecology, from biological, zoological and environmental perspectives. This concept has been extended to postulate other ecologies, for example, a sociological or anthropological ecology, which views an individual society or culture as part of a wider areal context of society/culture, and also focuses on the system that exists within a single society or culture (see, for example, Hawley 1950).

The metaphor has also been extended, by linguists, to the concept of linguistic ecology. Carl Voegelin, from his position of interest in both anthropology and linguistics, initiated the use of the term in relation to languages in the early 1960s (although actually written before 1964, this was not published until later: Voegelin, Voegelin & Schutz 1967). Voegelin and Voegelin (1964, p. 2) recognized ‘inter-language’ as well as ‘intra-language’ linguistic ecologies. They suggested a top-down, areal approach to the study of linguistic ecologies, rather than beginning with a selective, language-specific focus.

Haugen (1972, p. 325) defined language ecology as ‘the study of interactions between any given language and its environment’. He defines language, by inference, as both a psychological and a sociological code, existing only in the minds of its users, which enables them to relate to each other and to the world around them. However, he makes the point that the environment of a language is not ‘the referential world to which language provides an

index ... [rather] the true environment of a language is the society that uses it'. Thus, language ecology is about the context in which societies use their various codes of communication. In this sense it is really just a new banner for other twentieth-century fields of study, for example, 'psycholinguistics, ethnolinguistics, linguistic anthropology, sociolinguistics, and the sociology of language' (Haugen 1972, p. 327). Haugen also listed (p. 336–7) various general fields of enquiry that are relevant to forming a picture of the linguistic ecology of a (single) language: its classification in relation to other languages, the demographic profile of its users; its domains of use; the degree of bilingualism and overlap with other languages; internal variance (i.e. regional, social, 'contactual' and learner's dialects); the philological study of written texts in relation to speech; the standardization and usage of its orthography; its political significance and status in government, educational and institutional realms; attitudes to its usage (for intimate settings, for personal identification, and for status); and its relative viability in relation to other languages. In his pioneering attempt to define the concept of linguistic ecology, Haugen was focusing on how any given language fits into the wider ecology of languages.

In further developing the concept of linguistic ecology Mühlhäusler (1992, 1995, 1996) challenges the fundamental notion of 'a given language'. He follows Harris' (1979, 1980, 1990) and Grace's (1993a, 1993b) theory that languages are not distinct entities—apart from 'a few mainstream standard languages' (Mühlhäusler 1996, p. 5). The world's multitudinous languages are merely an expression of natural linguistic diversity (the loss of which is 'a historical 'accident' brought about by deliberate human agency', p. 18). He makes the additional observation that the counting and 'naming of languages in many cases is a very recent phenomenon, introduced by Europeans over the last 100 years or so' (p. 282). From this perspective it cannot be properly claimed that any one language is endangered, but rather that linguistic diversity in general is threatened. The focus is on the whole linguistic ecology.

The (synonymous) terms *linguistic ecology* and *language ecology* can be used as mass nouns to denote the general concept, or as count nouns to denote single- or multiple-language entities or networks. So, while Mühlhäusler on the one hand refers to 'the ecology of language' and 'the study of language ecology' in the general sense (p. 3), he more often uses the countable interpretation of the metaphor, for example, when he refers to the many traditional language/linguistic ecologies of Australia and the Pacific (pp. 22, 47, 54, etc.). The latter interpretation implies that it is possible to compare and contrast different language ecologies with each other. In either case, however narrowly or widely we define linguistic or

language ecology, linguistic behaviour at any level collectively contributes to linguistic diversity in wider contexts.

The strength of the linguistic ecology approach lies in its similarity to the bio-ecological model, where the whole (community) is seen to be more important than the individual members (species). All societies and all languages are part of a wider ecology. The focus of this metaphor is not so much the internal and external impacts on individual members of the ecology, that is, the languages themselves, but rather the overall organic interrelatedness of various local speech behaviours in the wider sociolinguistic context. Thus, the introduction of literacy and literature in any given speech community has a very significant effect on the linguistic ecology to which it belongs, because it can involve substantial change to a whole network of societies, cultures and languages.

The responsible approach to any major endeavour that could potentially have any effect on the environment is to first conduct an environmental impact study. There is no less reason to conduct a similar study examining any potential impact of literacy on a linguistic ecology, whether that linguistic ecology be interpreted narrowly, for just the language and culture in question, or widely, for a wider network of related cultures and languages to which it belongs.

Let us now examine the broader issues surrounding the stability of a linguistic ecology, and the factors that could either sustain or change it. These issues and factors will be explored with the support of several case studies, most from the Pacific region, as discussed by Mühlhäusler (1996) and Nettle and Romaine (2000). We will then focus more specifically on the impact of literacy in particular, and issues related to it (e.g. literature, bilingualism, education). These insights will then be applied to the Kalasha situation.

2.2.1 Stability

As is true of an environmental or biological ecology, the stability of a linguistic ecology is a major factor that contributes to its natural diversity and continuing viability. It may mean maintenance of the status quo, where nothing happens to threaten or change the ecology. Nettle and Romaine (2000) note that this sort of social stability or community solidarity was stronger ‘in epochs prior to the present one’ (p. 130) because cultural influences that might have threatened local stability from the outside were less common and less pervasive than they are now. In the current epoch they point to stabilising influences such as ‘political, geographical, and economic factors which support the maintenance of linguistic and cultural diversity’ (p. 40). For example, autonomy or non-interference from external influences

(whether political, environmental, economic, etc.) would be conducive to the stability of a local linguistic ecology. Referring to cases of minimal outside interference generally, Nettle and Romaine (p. 13) point to their own research that shows ‘quite striking correlations between areas of biodiversity and areas of highest linguistic diversity’.¹ This is especially evident in areas inhabited by indigenous peoples who, though representing only ‘around 4 percent of the world’s population, speak at least 60 percent of its languages and control or manage some of the ecosystems richest in biodiversity’.

However, the stability of a linguistic ecology does not just depend on the status quo. It can also be the result of either an equilibrium that is maintained in the face of change, or successful adjustments to changing circumstances. The stability of a linguistic ecology could be manifest, for example, as the continuing viability, or even expansion, of a language, despite gradual linguistic changes that occur over time as a result of changing circumstances or technology. The major world languages are prime examples of this overall stability, despite continuing rapid change. It is also possible for the stability of a linguistic ecology to be maintained in the face of other types of change. This is the more liberal view of external influence that Nettle and Romaine have adopted in their notion of linguistic equilibrium (pp. 89, 98), the implication being that a language ecology is not necessarily destabilized either by gradual, non-destructive changes (pp. 48–49), nor even by the death of individual languages (pp. 89–90), but more by ‘forces of homogenization’, caused by massive waves of global change (p. 98). The fact that they devote the whole of their final (eighth) chapter to strategies of sustainability *in the face of development and change* bears witness to their belief that the stability of linguistic ecologies (however that is achieved) will help to preserve linguistic diversity.

Mühlhäusler (1996, pp. 322–3) sees the stability of a linguistic ecology in terms of the ‘support system’ that will sustain it over time. His ecological theory highlights two premises: firstly, that ‘the aim of maintenance is to enable the survival of a structured diversity, rather than individual languages’, and secondly, that the emphasis should be on functional relationships between languages, rather than on the quantity of languages. He argues that linguists should be ‘shop stewards for linguistic diversity’ (Mühlhäusler (1996, p. 2). His borrowing of this industrial relations metaphor suggests that Mühlhäusler sees linguistic diversity as a status quo factor of nature that needs active protection.

¹ Linguistic diversity in this context means language diversity rather than dialect diversity.

One very important observation to keep in mind is the common assumption that the conservationist values that apply to the bio-ecological model should also apply to the model of linguistic ecology. However, this may be one area where the analogy is not perfect. It is widely accepted that, in the bio-ecological realm, the conservation and preservation of most species is a given, for the health and survival of the whole ecology.² But we cannot necessarily make the same assumption with regard to the diversity of linguistic behaviour, by which, in the context of this discussion, we really mean speech. It is neither an absolute fact, nor even a universally accepted value, that the maximum diversity of speech needs to be preserved. Haugen (1972, p. 326) pointed out that '[language] has no life of its own apart from those who use it'. Nettle and Romaine (2000, p. 5) reiterated that 'languages are not living things' and 'a language is not a self-sustaining entity. It can only exist where there is a community to speak and transmit it'. Thus we see that, in the even broader ecology of life and humanity, language is ultimately subject to the societies that use it. It is from this perspective that we now proceed in the consideration of how disturbances to, or intrusions on, stability also play a valid part in the dynamics of linguistic ecology.

2.2.2 Instability

In this section, we examine how different types of change can threaten or disturb the stability of a linguistic ecology. It is important to remember that these factors do not necessarily pose such a threat—actual outcomes are not always predictable in the real world. The power of these factors to destabilize a linguistic ecology depends on the pre-existing stability and durability of the language, the size of the society that uses it, and other more specific factors in certain situations.

Because language is a living phenomenon, and society is organic, it is sometimes difficult to separate interrelated factors of change that influence them. Some factors of change have direct impacts on a linguistic ecology, whereas others have an indirect influence, a domino effect. For the purpose of this discussion I will deal only with direct impacts on a linguistic ecology, and indirect effects will be dealt with under each one's direct cause.

As stated in the introduction to this chapter, some factors of change are obviously externally mobilized, having cross-cultural implications (e.g. the imposition of colonial rule), and others are obviously internally mobilized (e.g. natural language and dialect lability and/or

² Exceptions include those species which are both destructive and uncontrollable, such as certain pathogens and pests.

evolution). However, it is not always helpful to categorize factors of change according to whether they are external or internal. The reality is more complex than these dichotomies allow. There is a continuum of starting points, in relation to a society or culture, from which change can gain its impetus. Many factors of change start externally and then become internalized. Sometimes both external and internal forces operate concurrently. Additionally, the boundary between external and internal is arguable. National borders do not constitute useful boundaries in a discussion of linguistic ecologies, and even cultural boundaries within a nation can be blurry when seen from the point of view of a diverse whole. Haugen (1972, p. 328) agrees that the choice of region, in terms of where to draw the boundary lines for a field of focus, can be fairly arbitrary. Despite this, we will begin with some of the externally-mobilized factors of change, discussing them in terms of their imperialistic force, and spelling out the specific motivations for them. However, imperialism itself may or may not impact the language ecology of the speech community.

To illustrate some of the above disclaimers, let us take a brief look at what has happened to the Chamorro language as a result of changes to its ecology. Chamorro is spoken on more than a dozen islands, including Guam and the Northern Marianas, in a Pacific archipelago between the Philippines and Hawaii. The group of islands was colonized by the Spanish in the mid seventeenth century. Within 30 years of Spanish occupation the population of Chamorro speakers had been reduced from about 50,000 to an estimated 3,678, due to genocide, mass emigration (to the Caroline Islands), introduced diseases and natural disasters (two typhoons) (Alkire 1977, p. 19, cited by Mühlhäusler 1996, p. 106). The eighteenth and nineteenth centuries were a period of intense Hispanicization, involving new trade routes, intermarriage and social restructuring (Mühlhäusler 1996, p. 106; Day 1985, p. 172). After the Spanish-American war in 1899, the southernmost island, Guam, was taken over by USA and made into a dependency to serve as a naval and air force base. The Northern Mariana islands to the north of Guam then came under the successive controls of Germany in 1899, Japan in 1914, the United Nations in 1944, and then USA as a trust territory in 1947 and as a US Commonwealth (known as CNMI) in 1978.

As a result of all these rather drastic political changes, the Chamorros in both territories have borrowed heavily from the languages of all their colonizers, and experienced major language shifts to Spanish and then English. However, the politically forced split of the speech community in 1899, and their subsequently separate histories over the last century, has led to different outcomes in terms of linguistic ecology. The Chamorro language, ‘which was

spoken in both political spheres, came under different external influences in the two areas, was given different writing systems and exposed to different language policies' (Mühlhäusler 1996, p. 108).

On the island of Guam in the 1920s, the US government banned the Chamorro language from schools, burnt Chamorro dictionaries and enacted a law forbidding locally employed government workers from speaking any language other than English during working hours (Mühlhäusler 1996, p. 109). And yet, even after a half century of linguistic genocide, the language demonstrated a remarkable ability to withstand these extreme policies (as well as English-medium mass media), and was still surviving (Topping 1973, pp. 3–4, cited in Mühlhäusler 1996, p. 108). In 1974 Chamorro was made an official language of Guam, and was introduced into education at school, teachers college and university level. The impact of this reversal of both policy and practice is still controversial. Riley reports the subsequent further decline of Chamorro (1980, p. 332, cited in Mühlhäusler 1996, p. 112), whereas, in contrast to this, Grimes (2000) presents a more hopeful picture, reporting that Chamorro is still gaining in importance. The US Census Bureau (2002) reported about 30,000 speakers of Chamorro in Guam in 2000. This represents about 22% of the total population.

The Northern Mariana (CNMI) minority of Chamorro speakers have experienced similar rapid changes over the past century, but have escaped the force of the above-mentioned American language policies and have kept their language more intact. Chamorro is currently 'used to communicate information and other important issues within families and extended families', and as a first language, is the main characteristic of an individual's identity.³ Bender and Rehg reported that there were about 10,000 speakers, giving the language 'some hope for survival' (1991, p. 3). The US Census Bureau (2002) reported about 14,000 speakers of Chamorro in CNMI in 2000, which was 21% of the population—almost the same proportion as in Guam.

However, the most significant difference that has emerged between the Guam and CNMI linguistic ecologies is that in Guam only 18% of Chamorro speakers actually speak it more than English, whereas in CNMI about 44% of Chamorro speakers speak it more than English. This seems to suggest that the language is more stable in CNMI than in Guam: more than twice the proportion of CNMI's population speak Chamorro in preference to English, compared to Guam's population. Thus, the rate of bilingual shift is slower where outside negative interference has been less invasive.

³ (Source: Pacific Resources for Education and Learning n.d.)

Bilingual or multilingual shift can also be exemplified in another type of language ecology scenario. The emergence of pidgins and creoles has, on the one hand, sometimes played a part in the alteration or destruction of traditional forms of multilingualism. In the language ecology of Papua New Guinea, Tok Pisin has caused a language shift from several vernacular languages, for example, the Taip language (Mühlhäusler 1996, pp. 73, 127, 144; Nettle & Romaine 2000, pp. 13, 127). On the other hand, the emergence of pidgins and creoles has also facilitated intercultural communication, adaptation to change, opportunities and national solidarity. For example, Tok Pisin, being the national language of Papua New Guinea is also promoting networks of communication across the nation. Thus, overall, what is lost in language diversity is gained in social cohesion.

Linguistic ecologies are also subject to the lability of multidialectal shifts. For example, dialect levelling has occurred in Fijian, due to a wide range of external and internal factors: cross-cultural intermarriage, urbanization, literacy and education, and religious, commercial and media influences (Mühlhäusler 1996, p. 50). One effect of this dialect levelling is the fact that ‘in the 1990s, all members of the younger generation are competent in spoken standard Fijian, competence in other varieties of Fijian, being on the decline’ (p. 227). There is no denying the fact that, although the Fijian language ecology has been influenced by other language ecologies, during 200 years of contact with the outside world, standard Fijian remains one of the more viable Pacific languages (p. 230). It is debatable therefore, whether, from the perspective of the Fijian community itself, this decreased heterogeneity is a worse outcome than the increasing viability and standardization of one of them.

2.2.3 Population impacts

The most devastating group of demographic factors responsible for threatening a linguistic ecology includes those that impact on a population of the speakers, by diminishing or dispersing the actual size of the speech community. When a community has been severely affected by any event, its population can reach a critically low level, beyond which its language cannot be expected to survive. Of course, it is difficult to quantify such a threshold, because of the different combinations of factors that apply to each particular language ecology. The level of endangerment is not just a function of the absolute number of surviving mother-tongue speakers, as Crystal (2000) and Nettle and Romaine (2000) seem to suggest. As discussed in Chapter 1, one way that Krauss measures endangerment is in terms of the proportion of mother-tongue speakers in a community (though the ‘community’ in such a case

needs to be defined).⁴ A diachronic assessment of the trend of language use would give the best indication of the measure of a language's endangerment: that is, a language is endangered if the mother-tongue speaker population is decreasing over time. Some of the events that have been known to be instrumental in severely reducing or dispersing the population of certain speech communities, to the point of language endangerment, are mentioned in the following paragraphs.

Politically motivated genocide impacts on a linguistic ecology by the decimation of the whole speech community. Some examples were the Indonesian genocide of Tetun speakers in 1975 (Mühlhäusler 1996, pp. 175, 271, 337), the Russian genocide of Ubykh speakers in 1860, the El Salvadoran extermination campaign of minority Indian speech communities in the 1930s, and the Russian genocide of the Chechens in the 1990s (Nettle and Romaine 2000, p. 6).

Besides reducing the population, this sort of violation also increases the vulnerability of survivors to further abuse, which can cause their language to become less viable. Mühlhäusler (1996, p. 51) cites the case of the genocide of mainland Australian Aborigines by the new settlers. This was followed by the mass relocation of whole generations of Australian Aboriginal children from their communities between the 1880s and 1970, in order to extend their education and assimilate them into Western-style schools, single-sex boarding hostels and communities where the majority of people were white and imposed an English-only policy which punished use of the mother tongue. Mühlhäusler (1996, pp. 246–7) writes that:

for many languages, it effectively destroyed the intergenerational language transmission link by severely disrupting family structure and parent-child relationships. Children were forcibly removed from the social structure and the set of primary role-relationships which enabled the acquisition of Aboriginal language, sociocultural values, knowledge and skills. In effect, they were denied their cultural and linguistic heritage.'

Thus, these relocations contributed to the deconstruction of the population of those communities, which, in turn, caused a severe devaluation of their languages. Dyirbal was only one of the scores of Aboriginal languages that suffered in this way (Mühlhäusler 1996, pp. 61–3, 186).

The decimation of whole populations by disease, and the degradation of quality of life, has sometimes changed language ecologies. One-third of all Australian Aborigines died of smallpox in the century following European contact, wiping out many speech communities

⁴ See Table 1.3 in subsection 1.2.3 of Chapter 1.

and thus reducing the diversity of the linguistic ecology. The nineteenth-century Peruvian deportation of half of Easter Island's 3,000 Rapanui speakers, for slave labour in South America had a significant impact on the linguistic ecology of the community (Nettle & Romaine 2000, p. 198). When fifteen of the deportees returned they brought smallpox, which reduced the Easter Island speech community to about a hundred speakers. The Trumai people of Venezuela suffered a devastating influenza epidemic in 1962, which reduced their population to fewer than ten speakers (p. 10). Brazilian Indian populations have also been wiped out by disease.

The degradation of habitat can severely affect the linguistic ecology of an area by reducing the population of a speech community to dangerously low levels that threaten the viability of their language. A case in point is the 1970s dam flooding of the villages occupied by the Ugong people in Thailand. 'With the unity of the villages destroyed and their speakers scattered, the older speakers who still preserve the language have few, if any, people to speak to in Ugong' (Nettle & Romaine 2000, p. 10).

2.2.4 Language change

From their new evolutionary perspective, nineteenth-century writers spoke of new species of languages, which:

evolved in the course of their 'progress', often as a result of competition which ensured the survival of the fittest. Others looked on language change as a degeneration from the perfection of a classical paradise, which in an imperfect world could only be partially restored by eternal vigilance on the part of the guardians of good taste (Haugen 1972, p. 326).

Haugen's point is that language change is natural, and although he was applying it to the independent self-evolution of languages, the same argument equally applies to their evolution under outside influences (pp. 334–5).

Cross-cultural contact is a widespread and ongoing phenomenon that is integral in the dynamics of social and cultural change. It can involve cross-cultural influence or intervention of many kinds—social, political, religious, educational, commercial, media, etc. An example of social influence was the intermarriage by Chamorro speakers with their various colonisers (see subsection 2.2.2 above). Not surprisingly, the relevant linguistic ecologies are also impacted, as a result of which, new linguistic influences come into play, leading to the loss of

register and/or the gradual change or disappearance of various traditional domains of language.

One of the consequences of bilingualism in a language ecology is its implication for the choice of language in various domains of language use. In this context, Romaine (1995, p.30) cites Fishman et al's (1971) domains: 'family, friendship, religion, employment and education'. In addition to these social domains, Romaine (p. 31) also refers to the 'inner functions' of bilingualism (which are essentially personal domains), citing Mackey's (1968, p. 565) list of 'counting, reckoning, praying, cursing, dreaming, diary writing and note-taking', to which Romaine adds 'speech to oneself, or thinking aloud'. Romaine refers to different combinations of variables that are involved in a choice of one language or another, ('interlocutors, place and topic') and elaborates that 'in each domain there may be pressures of various kinds, for example, economic, administrative, cultural, political, religious, which influence the bilingual towards use of one language rather than the other (see Mackey 1968: 563–4)' (Romaine 1995, pp. 30–1).

Although domains of language use are often mentioned in discussions of bilingualism, the issue of the survival of certain domains is sometimes pertinent even where bilingualism is not an option, for example, where the choice becomes the use of the mother tongue or none. The loss of specialist fishing vocabulary has resulted from the commercial availability of seafood in Hawaii (Nettle & Romaine 2000, p. 56). In certain minority languages both missionary and indigenous religious groups have discouraged traditional domains of discourse and/or even introduced new religious domains. For example, Christian missionary influences have led to the loss of traditional oratory and metaphorical speech in the Ilongot language of the Philippines and to the decline of swearing in the Kilivila language of the Trobriand Islands (Mühlhäusler 1996, p. 69).

In wider contexts, language shift can occur: a particular language or dialect can be elevated over others, for the sake of expediency and/or in an attempt to promote community solidarity. Examples include the disappearance of certain dialects of the Kâte and Yabem languages of Papua New Guinea after an alternative dialect of each was chosen for literacy. The evolution and elevation of lingua francas such as Tok Pisin in Papua New Guinea, Samoan in Tuvalu (a parliamentary state consisting of nine atolls north-west of Samoa), and both Samoan and Tahitian in Tokelau (a New Zealand dependency consisting of three coral atolls north of Samoa) has drawn many speech communities together in a wider social network. The lexical and grammatical characteristics of trade languages can also significantly affect the structure

of minority languages; for example, the Tuvalu language has changed because of the influence of Samoan, and the Rapanui language of Easter Island has changed because of the influence of Tahitian (Mühlhäusler 1996, pp. 152–3, 234–5, 242).

The destruction and reorientation of communication networks (usually the direct or indirect result of some form of imperialism) can also lead to changes in language domains, language shift and/or language loss. Language policy can be very insidious, giving rise to fundamental inequalities in society, conflict and disputes involving language and the language ecology, and violation of linguistic rights. For example, when American sugar planters gradually seized control of the Hawaiian monarchy, the new provisional government that resulted eventually banned the use of the Hawaiian language in 1896 (Nettle & Romaine 2000, pp. 95–6). Poor language engineering can also impact on a language ecology, for example, inadequate or ineffective preservation strategies, discriminatory language revival policy, and switching or fluctuating policies on language use.

An example of switching in government-supported language intervention is the introduction of mother-tongue Maori literacy, which enjoyed overwhelming appeal and success in the first half of the nineteenth century. However, it was then deemed to have caused a weakening of dialectal diversity, introduced neologisms, failed to equip the community for inter-tribal diplomacy (Mühlhäusler 1996, p. 116), and even led some to develop cargo-cult beliefs (p. 217). This led to the banning of Maori, and an English-only policy was instituted in the first half of the twentieth century, resulting in ‘systematic assimilation of the Maori into mainstream society’ (p. 117). The language policy was then switched again by government-sponsored education in the 1970s, which has succeeded in stabilising Maori for bilingual use. Although the monolingual use of Maori has ceased, Mühlhäusler concedes that the debate concerning the impact of Maori literacy is continuing (p. 215).

2.2.5 Literacy

The introduction of literacy in an unwritten language opens up a new medium for it. This can have detrimental effects (e.g. reducing dialect diversity), but it can also help to preserve the language, promote internal acceptance and/or provide a pathway to modernization.

Mühlhäusler (1996) makes his views quite plain: ‘[Literacy] has led over the years to an almost total transformation of most Pacific societies and most languages spoken in the area’ (p. 212). He adds that, ‘in all instances’ (that he reviewed in his study of the Pacific), the

linguistic outcome of literacy (whether it is in national or local languages) is the acceleration of the decline of local languages (p. 215).

The introduction of literacy is seen by Mühlhäusler as a ‘changeover from orality’, characterized by increased information, information storage over time, and communication over distance (pp. 212–13). He also discusses historical effects of literacy in the Pacific, such as conceptual restructuring (in particular, views of time, pp. 236–7), social restructuring (pp. 237–8), and the loss of linguistic diversity, contributing to ‘communicative inequalities and decreasing heterogeneity’ (p. 213). It is debatable whether some of these effects of literacy actually have negative value in terms of an overall dynamic linguistic ecology.

Although the vitality of oral literature sometimes suffers as a result of the advent of literacy, it cannot be assumed that this is always the case. Oral literature encourages consistency between past and present, and yet it discourages criticism, scepticism and rational analysis; whereas written literature preserves oral tradition so that it is not silently forgotten (Jacoby 1949, p. 217, cited by Goody 1963, p. 325). There are many cases where (mother-tongue) literacy has enhanced the appreciation and use of orality. For example, after a writing system was devised for the Karaja-speaking community of Brazil, many books were developed that document their oral myths:

Karaja teachers and students use the men and women of the village as an oral library of traditional knowledge. The entire community is involved in this new kind of education just as it was in traditional culture. The development of their own literature enhances the Karaja’s sense of the value of their own culture. The oral knowledge provided by the people themselves has proved a rich resource for the schools (Nettle & Romaine 2000, p. 184).

Coulmas (1994, p. 162) claims that ‘an accepted written form greatly enhances the utility of a language, and indeed, its chances of survival’.⁵ Nettle and Romaine (2000, p. 200) agree with many others in their recommendation of mother-tongue literacy as a strategy for preserving language and diversity. Among the case studies that they cite is the Hawaiian literacy program of the 1980s and 90s, which is achieving successful results (pp. 180–3). In her discussion of writing systems, Adams (1990, p.14).makes a point that applies equally here to even the most perfect linguistic ecology: it ‘has one of two ultimate destinies: It can remain rigidly frozen, becoming imperfect by default, or it can yield to the pressures and change’.

Proponents of mother-tongue literacy do not deny that it involves radical change to a society, but their point is that mother-tongue literacy is a change that benefits and suits the

⁵ Quoted by Mühlhäusler (1996, p. 215), though he feels that this ‘widely held view ... does not appear to be universally applicable’.

speakers themselves. The example of standard Fijian cited above (subsection 2.2.2) can be seen as a positive language change because it has helped to integrate the community.

We will now look at some of the arguments in favour of mother-tongue literacy, with special reference to the Kalasha language ecology.

2.3 LINGUISTIC PRESERVATION

From the point of view of scientific interest, there is a case for trying to save the world's many endangered languages. However, in view of the rate of language death, some admit the futility of their own (and the language communities') efforts at preservation, in the face of other more powerful forces. As Dauenhauer and Dauenhauer (1998, p. 78) lament, 'the paradoxical situation is that the languages will certainly die unless we do something; but, the reality is that they may also die even if we do something. Therefore, what do we do?' Dixon (1997, p. 112) agrees that 'the loss of a language can be slowed a little; but it cannot be halted or reversed'. He points to the threshold of extinction when he states that 'no language – once it has ceased to be used in everyday life – has ever been revived' (p. 111).

Despite this bleak outlook, serious steps are being taken to turn the tide of language death by reviving them before they become extinct. One of Crystal's (2000, p. 130) six 'postulates for a theory of language revitalization' is that 'an endangered language will progress if its speakers can write their language down' (p. 138). 'If I learn to write well, my language will never disappear,' said a Machiguenga man at a writer's workshop in Peru (Cahill 2001).

So expressed, this is a tenuous argument, and may only hold true if a language is above a certain threshold of endangerment. However, where other factors and conditions allow, as we have seen in the first half of this chapter, mother-tongue literacy can help to keep some languages alive. Cahill (1999) presents case histories of several examples of endangered languages in Brazil and Papua New Guinea that have experienced well-documented revitalization, partly as a result of mother-tongue literacy programs.

The same argument from the opposite side of the coin is suggested by Crystal (2000, p. 83), namely, the devaluation and eventual loss of the minor language through lack of *literary* (i.e. written) use, which is made irrevocable by the eventual death of community members.⁶ Unwritten minor languages are at an obvious disadvantage in this respect. One reason for the

⁶ This is only one cause of what Crystal refers to as language death. The other is the lack of use of a language in its *spoken* form, for various reasons.

lack of literary use, even for *written* minor languages (and potentially leading to language death) is the promotion of non-mother-tongue literacy as a substitute for, or at the expense of, minor-language literacy. Of course, such an outcome cannot always be blamed on the use of major-language literature, and it will often occur anyway. However, an overemphasis on non-indigenous literature has been known to kill off minor languages, as Skutnabb-Kangas repeatedly maintains (e.g. 1990, 2000). Conversely, in speech communities where non-indigenous literature is already established the indigenous language has been revitalized and saved from extinction through the development and/or promotion of indigenous literature.

So far, we have been looking at linguistic preservation for the sake of the language itself, and for the sake of the culture and literature of the community that speaks it. But there are other aspects discussed by Dixon (1997). He devotes two chapters to ‘priorities’ and ‘prospects’ in relation to endangered languages. His arguments are concerned primarily with benefits to linguistics as an academic discipline, and only secondarily with benefits to the languages themselves. Dixon’s main thesis is that endangered languages should be preserved because of the unique contributions each one makes to a variety-rich corpus of human languages, which enables wider research into language typology and universals on the descriptive side of linguistics, and genetic relationships on the comparative side.

However, both the motivation and the means for such academic endeavours are questioned in a cautionary note by Kindell (2001). She seems to suggest that the purpose of language documentation is often not as much for the preservation of a language, or the support of a community, as much as it is fodder for linguists and booty for the enrichment of a linguistic treasure house:

The arguments for being involved in such documentation include the safeguarding of linguistic diversity, contributing to a knowledge base for language universals, and the western idea that knowledge in and of itself is valuable. There are some ethical questions, however. One is motivation: all too often the creation of a linguistic market. Another concerns the rights of indigenous people to their languages; many want at least collaborative research, better yet to be trained to do the linguistic research themselves; others would allow only research with direct benefit to the community.

Thus, we see perhaps the most important principle in a true consideration of linguistic ecology, that the goal of language preservation is much more likely to be achieved if the focus is not so much on the language itself, but rather on the speech community that uses it.

2.4 CULTURAL CONSERVATION and ETHNIC IDENTITY

2.4.1 Cultural conservation

Culture is very much embodied in language. Language revitalization and indigenous literature have had demonstrable effects on the conservation of cultural heritage—namely, the history, mythology, religion, traditional values and practices of a culture or subculture belonging to a specific place and time in history. The (written) literary heritage of every language in the world stems ultimately from the formulation of a basic written medium for each of those languages. The true beneficiaries of such conservation are the current and future generations of the society itself.

However, the trends of the modern era and the tide of globalization are making the world more homogeneous, and are threatening cultural diversity. The impact of globalization on local literacies is discussed by Barton (1994, pp. 3–7), in contrast to the opposing influence of multicultural diversification. Globalization tends to favour not only *lingua francas*, but also major cultures, so that besides the resulting minor language death that occurs, cultures that are embodied in minor, unwritten languages tend to be devalued in the face of cultures that are embodied in major, written languages. Furthermore, to become literate in a second language, without having access to, and the benefit of, literature in one's own language, discourages the transmission of oral traditions, and encourages the replacement of traditional values and social mores with foreign ones, and the loss of cultural pride and ethnic identity.

Because of where the nation of Pakistan is situated politically, most Pakistani subcultures and speech communities are affected by Western commercialism, and sometimes even Western culture. On the other hand, the core identity of Pakistan as an Islamic state means that the extent of Islamic values is extremely pervasive. These influences reach into even the most remote areas of Pakistan's rural sector, including the Kalasha Valleys. The Kalasha community is unique in its very long-standing resistance to the religious pressure of Islam; however there are some Kalasha individuals, especially some of those in their teens and twenties, who have chosen to subject themselves to Western and/or Islamic influences. So, even though language death is currently not a critical threat to the Kalasha community, the risk of cultural devaluation is very real. Some Kalasha individuals are focusing exclusively on second-language literacy (in Urdu and/or English) without mother-tongue literacy.

For the more traditional sector of Kalasha society, the actual documentation of a particular culture's history, collective worldview, value system and literature (in a language and format

that is accessible to the people themselves – not just to academics) reinforces a sense of their ethnic identity, self-esteem, dignity and unique status in the world at large. It in turn engenders strength, courage, cohesion and the will to survive as a minority people group.

On the other hand, multicultural diversification is a policy often held by wider communities that embrace minority indigenous cultures. The Pakistan government proudly exercises such a policy, frequently featuring the Kalasha society in its tourist and television media as an obvious example of the nation's colourful cultural diversity. While not committed to cultural conservation per se, this philosophy is at least tolerant of it, and respects the rights of ethnic minorities to maintain their culture and use indigenous languages. There is a growing worldwide concern for the linguistic human rights of ethnic communities embedded in other majority cultures.⁷

A revival of interest among the Kalasha themselves in their own culture is somewhat traceable to the spotlight of interest created by foreign anthropologists, journalists and tourists who have visited the Kalasha Valleys in the last two decades of the twentieth century. Through their enthusiasm, some of these visitors have reignited local interest in indigenous Kalasha culture, religion, history and language. Over this time, several Kalasha people have been instrumental in the documentation of their culture, with a view to conserving it. An indigenous NGO (non-government organization) of concerned Kalasha individuals, formed in the 1990s, has stated in its constitution that one of its aims is the collection and documentation of texts relating to Kalasha culture.⁸ One would hope that the opportunity for mother-tongue literacy has happened just in time to prevent the cultural devaluation of traditional Kalasha society.

2.4.2 Ethnic identity

Mother-tongue literacy also establishes and confirms the value of the indigenous appreciation and expression of ethnic identity. Coulmas (1989, p. 226) has observed that 'as the most visible items of a language, scripts and orthographies are 'emotionally loaded', indicating as they do group loyalties and identities'. But what do we mean by ethnicity and ethnic identity? Rahman (1996, p. 14) sees the major debate about the nature of ethnicity being between what he calls the *primordialists* and the *instrumentalists*:

⁷ See, for example, Martinez Cobo (1987, vol. 5, part 3, p. 11) and Skutnabb-Kangas (2000, ch. 7).

⁸ The Kalasha People Welfare Society (KPWS) is a voluntarily staffed NGO, but receives some funding from the Pakistan government.

The primordialists argue that ethnicity is felt as shared paternity, biokinsship, commonality of descent, and blood-relationship ... The instrumentalists suggest that it is a modern phenomenon, which helps social groups gain a greater share in power and wealth.

Applying Rahman's distinction to a contemporary definition of ethnic identity, from the *primordialist* position an ethnic group has shared historical and ancestral origins and involuntary group membership, whereas from the *instrumentalist* position an ethnic group is characterized by submission to distinctive value orientations and behavioural patterns, and group influence on both behaviour and identity).⁹ From the former perspective, the key concepts are *shared origins* and *involuntary*, whereas from the latter perspective, the key concepts are *submission*, *influence* and *behaviour*.

The reason that these definitions of ethnicity and ethnic groups are important to our discussion of language documentation, literacy and literature is because written media have a substantial role to play in the determination of the *instrumentalist* view of ethnicity. Kalasha ethnic identity and community solidarity are already established, due to their shared location, shared religion and shared language. Nevertheless, shared oral literature, especially on indigenous topics, will strengthen their sense of ethnicity and consolidate their social interconnection.

2.5 EDUCATIONAL OPPORTUNITIES

2.5.1 Foundational education

Mother-tongue literacy has been shown to facilitate transfer to subsequent, second-language literacy and/or literature, and also opens the door to formal education in other languages. The literature on this topic shows 'because there is strong evidence that promoting L₁ [mother-tongue] literacy skills enhances overall academic achievement, this should always be considered in planning minority education' (Hamers & Blanc 2000, p. 206). After reviewing eight case studies of literacy in primary schools in various parts of the world, Dutcher (1982, p. 40) concluded that L₁ was the better choice for initial language of instruction, especially where the following circumstances prevail: literacy skills are relatively undeveloped, parental

⁹ The parameters in this paragraph constitute the definition of an ethnic group put forward by the US National Council of Social Studies' 1976 Task Force on Ethnic Studies (Wolfram & Schilling-Estes 1998, p. 165).

support exists, and the language has a relatively lower status.¹⁰ Dutcher mentions repeatedly that the evidence is slowly building in support of the effectiveness of first-language initial education (p. 45). Having stated the ‘fact that [initial literacy] is an easier task in the language the learner speaks’, Dutcher claims that ‘evidence for this abounds, not only in everyday experience, but in the studies presented here’ (p. 46).

Dutcher particularly recommends the use of ‘the first language as the medium of instruction when children come from situations where their first language does not enjoy high status in the wider community’ (p. iii). She concludes with a hypothesis that:

a child in a subordinate group has low feelings of self-worth. These feelings influence his achievement, or lack of achievement, in a school whose classes are conducted in a language other than his mother tongue. When the initial schooling is conducted in his mother tongue, through teachers with whom he can identify, then his feelings of self-worth are enhanced, and his achievement, both in the mother tongue and in the second language, is improved (Dutcher, p. 42).

The scenario of a low-status language spoken by a socially subordinate group certainly describes the Kalasha language and society very pertinently.

It could perhaps be argued that exclusively mother-tongue literacy may reduce the potential opportunity for educational development that would otherwise only flourish in a major language. This may be true in some cases, though if a student is held back from education in a major language because his/her mother-tongue education does not advance him/her far enough to transfer, it probably points to some other problem in the society or in the logistics of the education system.

The standard examinable curriculum in Pakistan (including the Kalasha Valleys) includes Urdu language and literature, social studies, Islamic studies, mathematics, science, and sometimes English language and literature. Several of the currently eight or so Kalasha teachers have indicated their enthusiasm for, and commitment to, the cause of supplementing the compulsory state curriculum with Kalasha literacy (yet to be sanctioned by the provincial government education authority). However, Kalasha literacy has not reached the stage where it might prevent or retard the furtherance of second-language education. Nor is it likely to prevent it eventually, because of the good infrastructure for the acquisition of second-language education (in Urdu and English) at the village, district, provincial and national levels. Even though currently all educated Kalasha have become literate in a second-language

¹⁰ There was ‘a strong case’ for L₁ instruction from four of these studies, plus weaker support from one other, compared to only two cases strongly supporting L₂ instruction, and one extra weakly supporting it. The circumstances differed in each case.

first (Urdu), most are still enthusiastic about the value and benefits of mother-tongue literacy and education generally.

2.5.2 Bilingual education

Bilingualism is a skill that is usually claimed to put speakers at an advantage, not only in more than one community independently, but also between communities at one time. However, it is fair to say that this view of the value and usefulness of bilingualism has not always been held. Governments have, at various times in history, banned bilingualism; for example, the barring of the public use of the German language in USA during World War I (Kloss 1966 and Gilbert 1981, quoted in Romaine 1995, p. 247). Others have only supported transitional bilingualism in education. Romaine (2000, p. 251) explains this viewpoint:

To the extent that bilingual education is seen to legitimate an alternative point of view to the mainstream by sanctioning the use of another language and by implication the cultural values it symbolizes, it is seen [by many people] as a threat to the dominant group in society.

In contrast to this attitude, Romaine goes on to report that the US National Association for Bilingual Education argued in a testimony to Congress in 1983 that ‘there were demonstrable gains from bilingual education, as evidenced by test scores, enhanced self-esteem and community involvement’ (p. 252). This was one effort, by a special interest group, to persuade the same government that had once banned bilingualism, that in fact there was now empirical evidence for the value of including bilingual education in a school curriculum.

Padilla (1990, quoted in Romaine 1995, p. 254) argues for ‘a reconceptualization of bilingual education as a strategy for all students and not as a special instructional program for disadvantaged minorities’. This, again, is evidence of how completely views have changed with regard to bilingualism over just a few decades, with bilingual education now seen as a general strategy with value for everyone.

The traditional approach to bilingualism for ethnic minorities is through a major language (e.g. teaching English as a second language). However, Skutnabb-Kangas is of the view that transitional bilingual education programs are not good enough to keep minority languages alive, and that in fact they may lead to an acceleration of language shift (2000, chs 8–9). Her view is that mother-tongue literacy is just as viable, if not more so, as a means to bilingual competence. One is better equipped for literacy in another language if there is already a foundation of literacy in one’s own language. Dutcher (1982, p. 25) reports the conclusion

of Modiano (1973), and of many others who have reported her study of the bilingual education of Mayan children in Mexico, that:

the success of the children who learned to read in their vernacular was due to the separation of the two tasks: learning to read and learning to use a second language. The bilingually taught children learned to read in their first language and ... later they transferred their reading skills to the second language.

Hamers and Blanc analyse a number of bilingual education experiments, in which the mother tongue (L_1) is used for formal education, either simultaneously with the dominant language (L_2), or before instruction is given in the dominant language.¹¹ They conclude that:

... in all cases, the program ... motivates the child to learn through his L_1 , and develops his linguistic conceptual capacities to the extent that he will make better progress in an L_2 , than his peers schooled exclusively through L_2 . These studies provide strong support for the view that, for minority children, the acquisition of literacy skills should be disassociated from the acquisition of L_2 skills, and that formal instruction should valorize the mother tongue (Hamers & Blanc 2000, p. 345).

In similar experiments, Landry and Allard (1985, quoted in Hamers & Blanc 2000, p. 103) demonstrated that for French-speaking children in New Brunswick, the more the parents valorized the mother tongue, the better the children achieved at school. Hamers and Blanc (2000, p. 347) conclude that ‘more than the fact of teaching literacy in L_1 , it is the valorization of L_1 as a cognitive tool by the school which is responsible for the development of literacy’.

These findings provide evidence for the principle already embodied in plans for Kalasha literacy: that if the skills of reading and writing are made a special focus in the more familiar language, the way is better paved for subsequent focus on instruction and literacy in another language. Bilingual literacy among the Kalasha takes the form of Urdu and English lessons in a small, but fast-growing sector of the society, that is, school children of all ages. These have had to struggle to learn foreign languages at the same time as learning the skill of reading and writing. The teaching of literacy in Kalasha, if introduced for everyone and at primary-school level, would separate the processes of acquiring literacy skills (first) from the process of learning a foreign language (second). Literacy in Kalasha would then make most of the current generation of children trilingually literate, with very little extra effort over what is

¹¹ L_1 denotes the first language learnt in a person's life, otherwise known as mother tongue, indigenous language or vernacular. (In the context of this thesis, this is usually a minor language.) L_2 denotes any other language or languages learnt subsequent to the first. This may be another minor language or a major language, though more often it is major.

traditionally expended in teaching Urdu and English languages and literacy in an educational vacuum. Multilingual literacy then opens doors to the functional benefits of major language literacy and political and legal empowerment in the wider society.

2.6 SOCIAL ADVANTAGES

2.6.1 Social bonding

Within a community, written literature can promote connection with, understanding of, and respect for one's elders (or for youth, depending on one's point of view), thus minimising the generation gap. It can preserve and promote generational bonding within a community because, provided the older generation embraces the concept, it is able to continually validate the language of the older generation, rather than isolating them at the expense of second-language proficiency and currency. This generational bonding has a profound effect on the strength and durability of a community.

On the other hand, if an education system imposed from a major outside culture ignores vernacular languages then intergenerational cohesion is severely threatened. (This is what happened to the 'stolen generations' of Australian Aboriginal children mentioned in subsection 2.2.3, above.) Providing parallel written mother-tongue education, in some form of bilingual curriculum, can prevent the social breakdown of the community by helping to bridge the gap between the visually illiterate, but orally literate, older generation, and the orally illiterate, but visually literate, younger generation.

The Kalasha community, in particular, is socially structured by clans and lineages. Older people are respected, but even more so are the ancestors. The knowledge and praise of one's lineage promotes clan identity and pride in a way that younger generations learn to respect. Eulogies and festival speeches are an important vehicle for the expression and transmission of this clan knowledge and identity, in which Kalasha elders play a pivotal role in connecting the former and current generations.

The existence and use of Kalasha oral literature, documented in written form, could help to foster appreciation and respect for the clans, the ancestors, and even the older generation, and could help to bridge the generational gap. A middle-aged man (SK) from one of the most respected Kalasha families, once commented on the social bonding value of Kalasha literature when he said that the most senior Kalasha elder (BK) would want to take under his arm a

transcription of his own (BK's) eulogy given at a certain young man's (JK's) funeral, and to show it off with pride around the village, despite his (BK's) own inability to read it (SK, personal communication 1983).

Here we see the power of the written word in providing a bond between the generations. The middle-aged man was reminded of the value of Kalasha clan identity and Kalasha ancestry in the transcribed speech of the elder (which also confirmed his respect for the elder himself). The elder was impressed with the importance given to his words, and the praise for former generations that it engendered. In addition, the young friends of the dead man would see the importance of the eulogy given at his funeral in its being recorded in permanent form.

However, already, the younger generation of Kalasha is finding that some of the vocabulary of the older members of the community is unknown to them. The usefulness they see in their mother tongue is less as a vehicle for expressing themselves personally, and more as a means of preserving their roots. However, adult literacy is also essential to allow the older generation to participate in both the documentation process and the proliferation of indigenous literature. This will help them to maintain their traditional place in the social hierarchy as well as strengthening intergenerational ties.

Kalasha elders are not alone in lamenting the breakdown of communication between generations. It is a significant issue for the languages of many minority communities around the world, let alone the extra gulf created by the inroads of a foreign language. It already affects spoken communication, and can only be further exacerbated if there is no recourse to written communication. Conversely, written literature could encourage social bonding.

2.6.2 Social advancement

Coulmas (1996, p. 304) reminds us that 'sociologists have generally considered written language as an evolutionary advantage, since its possession increases the adaptive capacity of social systems by advancing differentiation and specialization'. This experience is borne out in many preliterate societies, where there is sometimes a measure of kudos attached to the acquisition of literacy. This kudos entails respect within the community generally and a deemed or real position of status among those who are illiterate, within the community and/or among other communities. Coulmas observes that social value is also accorded to the written medium itself. Among various cases, he cites that of India (Pattanayak 1979, p. 56, cited by Coulmas 1989, p. 198):

In India, in particular, values of permanence are attached to writing; and the written varieties of the literary languages are hence accorded higher status than vernacular speech. Moreover, written languages enjoy higher prestige than those that have not been reduced to writing. Indeed, it is often contended that a speech form only acquires the status of a proper language once it has its own script.

Not surprisingly, these attitudes are also found in many Pakistani minority-language communities. However, such awe of the literary medium, the literacy process, and those proficient in either, is more prevalent when and where education is very scarce—as in times and places characterized by little or recent contact with the outside world. As more individuals acquire literacy, the community attitude gradually changes from awe to respect for those literate individuals. Then, as the literacy rate gradually soars to majority levels, the community attitude to literates gradually changes from respect to general acceptance.

The notion that, as a new technology, literacy can shape society is, admittedly, attractively simple. Yet, at every level a balanced perspective is necessary. Coulmas (1996, p. 305) points out that:

for some time ... much hope has been placed on literacy as an agent of development, but there has been a growing recognition that in order for literacy to spread, a society must be ready for it, that is there must be a demand for literate people.

Coulmas then takes this perspective one step further by noting that ‘in recent years the focus of research has shifted from the question of how literacy affects societies to that of what societies do with literacy’.

The full potential of mother-tongue literacy has not yet been realized by the Kalasha society at large, but if it does, it will probably have the sort of initial impact previewed by the case of the elder’s awe at the written form of his own eulogy (see subsection 2.6.1, above). As the opportunity to write their own mother tongue becomes more possible, not only is the community attitude to mother-tongue literacy likely to increase in awe, respect and kudos, leading then to a wider acceptance, but also mother-tongue literacy will serve as a very convenient pedagogical bridge to majority-language literacy (and culture) over which many more Kalasha will then be able to cross, than have in the past.

2.7 THE KALASHA LANGUAGE ECOLOGY

The Kalasha society and environment has experienced a relatively high degree of outside interference, especially since the early 1980s, in terms of commerce and infrastructure. According to this factor alone, the stability of the Kalasha linguistic ecology could be under considerable threat.

The Kalasha nation was politically incorporated into the British Empire in the nineteenth century, then into the Republic of Pakistan in 1947. However, because they have always been a relatively remote minority, the impact of British rule on Kalasha society was previously very slight and indirect, hardly different from other language communities in the district. However, the more recent political control of the Kalasha by the Pakistan government has grown from very superficial to substantial just in the last couple of decades, and is now starting to reach the core of Kalasha society, with government-supported infrastructures like education, health services, the tourism industry, etc. This has brought to the fore the national language of Urdu, surpassing Khowar as the language of outside opportunity for aspiring young Kalasha. However, this change has served to awaken the Kalasha, more than before, to the value and heritage of their indigenous culture and mother tongue.

Cultural oppression does not severely impact Kalasha society in modern times. However, one exception is the racist attitude that is expressed in the current religious discrimination, ostracism and persecution of Kalasha students in Muslim schools, by both fellow students and teachers. This leads to the exclusion of many Kalasha children from participating in state education, which further disempowers them as a minority.

The Kalasha people and their environment has been severely affected by commercial interests since the early 1980s, mostly introduced by members of the more dominant Chitrali community living around them. This commercial imperialism has taken the form of land acquisition and foreign investment (gradually reducing Kalasha land rights and assets), logging (causing deforestation), and the unchecked acceleration of an infrastructure to cater for the local tourist industry. This has involved the construction of many hotels, the use of local resources and significant impacts on the environment and landscape. Such commercial endeavours imposed from the outside have brought an influx of Chitral settlers, whose language (Khowar) has taken a more prominent position as the spoken lingua franca of the area.¹²

¹² Khowar is written and read by some already-educated Khowar speakers who have a special interest in their own traditional literature.

Thus, Kalasha continues to face various challenges from other languages: from Khowar in the wider speech community (as it has for the past several centuries), from Urdu in national contexts (as it has for the past several decades), and from English in international contacts and exposure, which have increased over the past several years.

2.8 CONCLUSION

The question of whether an unwritten language should be codified into writing, has led us to consider languages in their ecological setting, where a wide range of factors, including literacy, have their impacts. These potential influences were evaluated and examined, and were found to be part of the dynamics of inevitable change. The positive effects of mother-tongue literacy, from both theoretical and practical perspectives, can be seen in many cases.

On the basis of the above discussion we may conclude that mother-tongue literacy and literature are resources that benefit the speakers of some languages by providing an indigenous medium for indigenous topics. Development of the written medium for the Kalasha language could enable Kalasha people to document their rich heritage in a permanent and widely accessible form, and to equip them with a foundation for other educational opportunities. It would also be a fresh means for the Kalasha community to exchange ideas, knowledge, creative expression, etc. The functional benefits of social mobility and opportunity are secondary, and can probably be better derived from Urdu and/or English literacy eventually.

ALTERNATIVE SCRIPTS

3.1 INTRODUCTION

The writing systems that are used to represent the different languages of the world vary almost as widely as the languages themselves. However, despite this great variety, the scripts of most languages are often familiar to readers of other languages. When comparing languages that use the same script, usually some level of phonological decipherment is possible without semantic understanding. For example, most readers of English could guess many of the sound correspondences of an Indonesian text, because both languages use a Roman script; and most readers of Urdu could phonologically decipher Arabic text because both those languages use an Arabic-based script. Conversely, sometimes a semantic understanding is possible without phonological decipherment. For example, texts written in traditional Chinese script can be read and understood by speakers of many different, mutually unintelligible languages.

There are many factors at play in the multiplicity of scripts: historical influences, geographical proximity, religious pressures, political circumstances and decrees, socio-cultural attitudes, etc. However, none of these factors intrinsically links a writing system to a language. Even linguistically based adaptations of scripts designed to improve the match between speech and symbol are often arbitrary and rarely perfect. Yet the switching between alternative scripts that continues to occur throughout history does raise the question as to whether one particular script might be linguistically better suited than others to a particular language or group of languages. The fact that certain groups of languages are written with the same script might lead us to ask whether this reflects any linguistic goodness-of-fit, or rather a set of cultural and historical circumstances that have impacted on all of those in the group. As a background to this discussion, section 3.2 (below) looks briefly at the origins and development of various scripts (including pictographic inventories, scripts, syllabaries and alphabets) and the linguistic capabilities of the two scripts that have major significance to the Kalasha situation: Arabic and Roman.

The further discussion of *alternative* scripts, that is, cases where an official script is replaced by another script, will be from two perspectives: geographical (throughout the

Middle East, West and Central Asia, and the Indian subcontinent) and historical (as it relates to this region). Though these two perspectives sometimes merge, script change in many languages can be classified into either an essentially synchronic coexistence of scripts or an essentially diachronic switching of scripts, as shown in Appendix 2.

Section 3.3 will survey the typical synchronic distribution and usage patterns of a few relevant scripts—according to language family, region and religion. However, many language communities have switched from using their original script to using a completely different script, and several have even switched consecutively to three or four different scripts.¹ These diachronic processes are discussed in section 3.4.

3.2 THE MAJOR WRITING SYSTEMS

Writing is essentially the visual representation of ideas. It began with art – initially, static pictures and icons of familiar subjects through the eyes of the artist. Pictography and iconography may have been the first stages of arranging these pictures and icons to represent ideas structurally (e.g. in a narrative form). Mnemonic devices, like knots and notches, were the first abstract means of recording a message visually, both for enumeration and as a memory aid for the recital of poems, songs and genealogies. Graphic symbols were probably first employed as identification marks, such as tattoos on cattle, or clan or tribal signs on armour. The first symbols to represent abstract ideas and semantic concepts were ideographs, essentially simplified and conventionalized pictures. In some cultures, for example, in China, this method of representing ideas has developed into a complete system of writing, giving rise to a semantically based ideographic script.

Most writing systems eventually developed from using pictographs, iconographs and/or ideographs, to the stage where they represented speech itself—employing phonological symbols. These phonological scripts are based on limited sets of symbols (graphemes) that represent the sounds (sometimes syllables) of each language. Each grapheme of a syllabary represents a separate syllable; the vowels are intrinsically included in each grapheme. Syllabaries are best suited to languages built on very simple syllable patterns. However, many phonological scripts have evolved from being syllabic to alphabetic, where each grapheme represents a phoneme, a minimally articulated significant element of speech (whether

¹ Examples of languages consecutively or concurrently employing three or more scripts are Lambadi, Saurashtra, Kurmanji, Judeo-Tat, Uyghur, Northern Uzbek, Turkish, Kazakh, and Kirghiz.

consonant or vowel). Many ideographic and syllabic scripts are still successfully used in the modern world; although it is generally accepted that an alphabetic orthography based on the phonemic principle is the most efficient system as a default for representing a language (Tauli 1977, p. 24).

The Arabic alphabet is based on the Aramaic branch of the north Semitic syllabary, but, while retaining an essentially syllabic (or semi-syllabic) means of representation, it has also adopted various alphabetic features.² For example, it uses diacritics to represent some short vowels in some word positions, as well as an elaborate combination of symbols, contractions and diacritics to represent the various word-position variants of long vowels, consonants, nasalization, aspiration, retroflexion, diphthongization, consonant clustering and doubling.

In contrast to the Roman alphabet, only some vowels in Arabic scripts have alphabetic status. In Urdu, for example, long vowels /a:, e:, i:, o:, u:/ are written with distinct alphabetic characters, while the short vowels /a, i, u/ are marked by diacritics over or under the preceding consonant (similar to Hebrew pointing). Where there is no preceding consonant (in the case of word-initial vowels) the word-initial letter for /a/ (*aleph*) is written to carry the appropriate short-vowel diacritic. Furthermore, the same Arabic-script symbol is used for each of the semivowels and the respective vowels they approximate. For example, in Urdu, one symbol is used for both /j/ and /e/, and one symbol is used for both /w/ and /o/. However, in each case, the phonemic distinction can usually be determined from the graphemic context.

Arabic scripts are written from right to left, and from the top to the bottom of the page. (The only exception to this is that numerals in Arabic scripts are written from left to right.) The script is always cursive—it cannot be written otherwise. There is no distinction of printed versus cursive styles, as in Roman script. All written language, from the quick jottings of a note, to the penning of a personal letter, to the most formal of published texts or reference materials, is based on the traditional calligraphic handwriting styles. (In fact the more sacred and formal a published text is, for example, the Koran, the more the script is made to resemble beautiful handwriting, with lots of special embellishment, and the less standard it appears.)

² The term *Arabic script* in this thesis is used in a parallel way to the common usage of the term *Roman script*. It means any script based on, and generally resembling, the script used for the Arabic language. The term encompasses much variation and adaptation in style, alphabetic characters and conventions, as is evident in the different Arabic-based scripts of Farsi (Persian), Urdu, Pashto, Kashmiri, Sindhi, Panjabi, and more than two dozen of other languages (Wikipedia 2005). There are also various subtypes (styles, fonts) of this script, as explained in footnote 1 of Chapter 6.

The letters of each word are joined according to set rules incorporating joining and non-joining forms of each letter, and these rules can never be broken. Unlike Roman script, there is no room for the whim or style of any writer to join or not to join any letters, except by these rules. All alphabetic letters can be joined from the right-hand side. However, while about three-quarters of them can also join on the left-hand side, about a quarter of them cannot. There are different forms for each alphabetic letter, depending on each letter's position in the word: word-initial, word-medial, word-final. Word boundaries are marked by word-final forms of each letter, so that spaces sometimes become redundant between words and are therefore, not inserted in some types of texts where space is at a premium (e.g. newspapers).³

The Roman alphabet has been used as a basis for the orthographies of hundreds of languages (Sztemon 2002).⁴ The Roman alphabet assigns letters to individual phonemes. It can cope with a wide range of phonemic phenomena, like diphthongs and semivowels, because all vowels and semivowels can be represented alphabetically, though often with digraphs and a small set of diacritics for specific languages. For the same reason, complex syllable structures are also very easily represented using various combinations of segments like consonant clusters. There are also separate symbols available to represent a wide variety of phonological variants and rare sounds. Furthermore, the absence or non-significance of joining rules and ligatures, allows greater variation in Roman-script symbols than is possible in Arabic script.

The International Phonetic Alphabet (IPA) is the ultimate symbolic system for writing any language. It is a special, highly modified Roman alphabet designed as a modern scientific attempt to universally and comprehensively represent every phonetically distinct element of speech, from all languages, in a single system.⁵ As such, it is theoretically able to distinguish the finest phonetic variations possible. It is a purely segmental approach to language representation and is thus most suitable to reflect every possible syllable pattern. However, the IPA is not a very suitable alphabet as a basis for a practical orthography, because not all of its distinctions are significant for any individual language. Alphabets are best matched to the relevant phonological features of each language (e.g. the aspiration, retroflexion, rhoticity and nasalization of Kalasha, as described in Chapter 4), without the need for distinctions that are not relevant in that language. One measure of the success of matching a particular

³ An orthographic space is still required in Arabic-based scripts to mark a word boundary after a non-joining letter in word-final position.

⁴ The phrase *Roman alphabet* is used here, rather than *Roman script*, to avoid any misunderstanding with another meaning of the word *script*, which entails a cursive font style. A very early attempt to prescribe scientific transcription methodologies for Roman alphabet adaptation was published by Lepsius in 1855.

⁵ A recent, thorough explanation of the IPA is set forth in the *Handbook of the International Phonetic Association* (1999).

phonological script to a language is its capacity to reflect phonemic features, both thoroughly and efficiently.

3.3 THE DISTRIBUTION OF SCRIPTS

The concept of a *typical* script arises from situations where scripts are traditionally shared by communities of people who speak languages that belong to the same language family, and/or who live in the same geographical region and/or who belong to the same religion. The scripts we are most interested in for their relevance to Kalasha are the Arabic, Roman, Cyrillic, Devanagari and Hebrew scripts (with some reference also to Greek, Armenian and a few other indigenous scripts). The following table sets out the typical correspondences between these variables.⁶

Table 3.1: Summary of the typical distributions of some relevant scripts

Script	Regions	Languages Families	Religion
Arabic	Middle East, North Africa and West Asia	Central Semitic, Turkic, western Iranian and northern and western Indo-Aryan languages (including the Dardic languages of northern Pakistan and Afghanistan)	Islam
Roman	Europe, West Asia, the Americas, sub-Saharan Africa and the Pacific	A wide variety of language families around the world, including, traditionally, the peripheral Semitic and the Turkic languages	Christianity
Cyrillic	Eastern Europe and Central Asia	Slavic and Turkic languages	Orthodox
Devanagari	India	Eastern and southern Indo-Aryan languages	Hinduism
Hebrew	Middle East	South Central Semitic languages	Judaism
Greek	Southern Europe ⁷	Greek	Christianity
Georgian	Central Asia	Georgian	Christianity

The above table indicates some measure of correlation between script type, geographical region, language affiliation and major religion.⁸ Though it clearly demonstrates typical synchronic correspondences between these variables, we will also look at some exceptions

⁶ Appendix 2 contains a more detailed list of the language families and languages of this region, and the scripts used by each, as well as a more detailed list of scripts used in this region, and the languages that use each.

⁷ The Greek alphabet was once very widespread across an empire that reached as far as Pakistan, and gave rise to the Roman and Cyrillic alphabets.

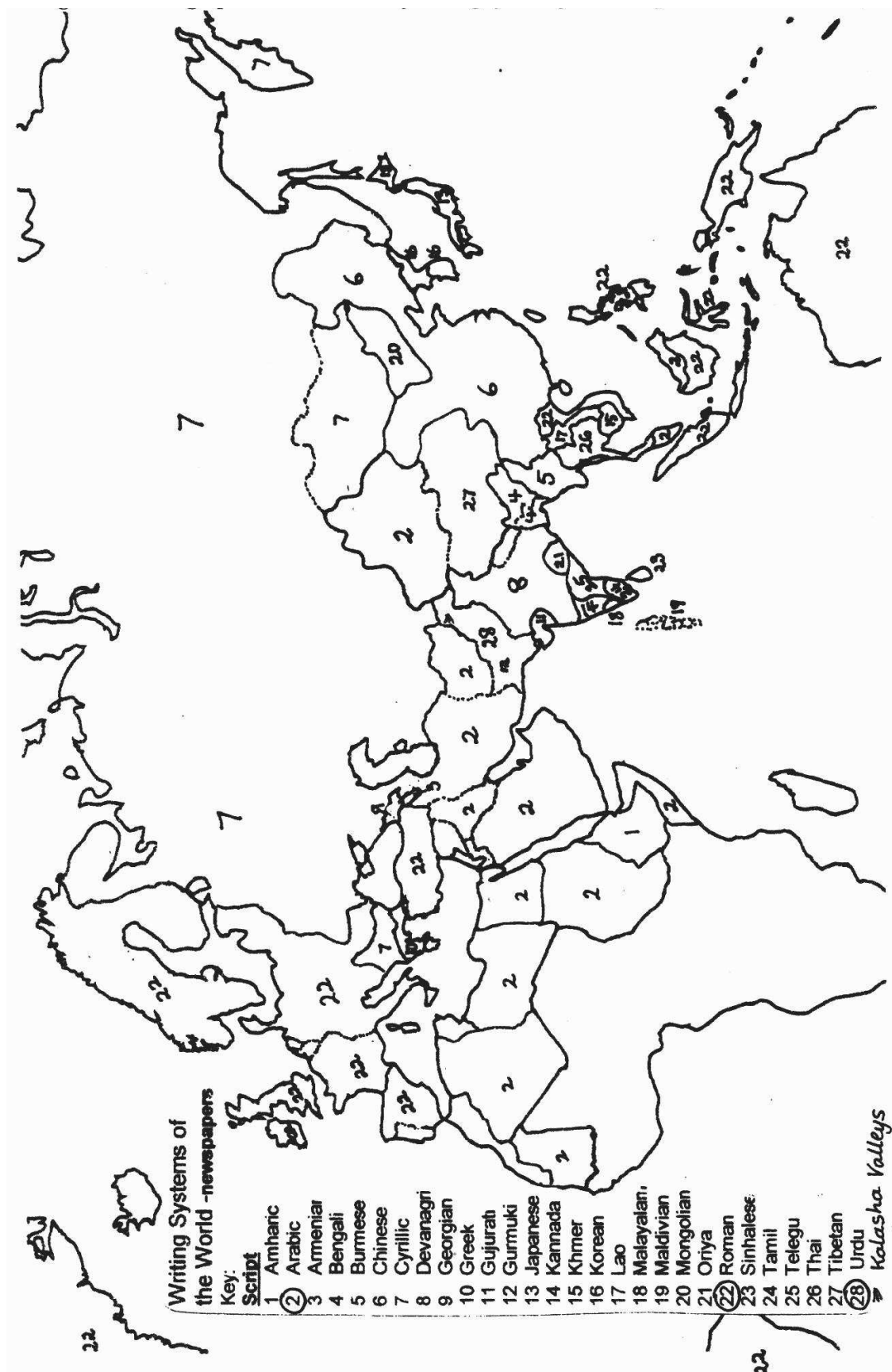
⁸ Coulmas (1989, pp. 241, 261) confirms the correlation between script and religion. Only the primary religions are listed in each case.

within the typical pattern. Regional exceptions are included in subsection 3.3.1, while religious exceptions are presented in subsection 3.3.2.

3.3.1 Regional perspectives

The map on the next page (based on a similar map in Nakanishi 1980) shows the broad distribution of scripts used in the various regions indicated in Table 3.1 above.

Figure 3.1: Geographical distribution of scripts from Europe to Asia



As the map in Figure 3.1 shows, almost all of the written languages from North Africa across to Pakistan use modified versions of the Arabic script.⁹ The Kalasha community lives within this area, and on regional considerations alone, would be expected to conform by adopting an Arabic-based script.

Regional patterns of script usage are also demonstrated with languages that are, or have been, concurrently rendered in more than one script, depending on the regions in which they are spoken. For example, several Central Asian languages spoken in geographically distinct regions employ different scripts, as the following table demonstrates.

Table 3.2: Concurrent use of scripts

	Kazakh	Kirghiz	Kurdi¹⁰	Kurmanji	Uyghur
Arabic	<i>China, Iraq</i>	<i>China</i>	Iraq & Iran	Syria, Iraq & Iran, Jordan, Bahrain, Lebanon	<i>China</i>
Roman	Turkey	Turkey	<i>Syria & Turkey</i>	Turkey, Western Europe	Turkey
Cyrillic	Kazakhstan & <i>Mongolia</i>	Kyrgyzstan	Central Asia	Armenia, Azerbaijan, Georgia, Kazakhstan, Kyrgyzstan, Turkmenistan	Kazakhstan

Here we see a subset of the wider regional pattern presented in Table 3.1: Arabic script in the Middle East, Roman script in Turkey and the West, and Cyrillic script in Central Asia.

However, we also see exceptions to the regional pattern, where scripts are being used outside their normal geographical area. Country names appearing in italic font show the exceptional use of Arabic script in China, Roman script in Syria, and Cyrillic script in Mongolia. These are just the exceptions of languages concurrently using multiple scripts in different regions. More regional exceptions are described in the following paragraph, and in subsection 3.3.2 and section 3.4 below.

In the face of the expectation of finding Arabic script in the general region of Arabic-script usage, contrary examples can be found in at least two languages spoken by neighbours of the Kalasha. These two languages are regional exceptions possibly because, along with Kalasha, they are spoken in the very peripheral north-eastern corner of this region. They are Wakhi (spoken in the Pamir mountains stretching across the upper Chitral and Hunza valleys of

⁹ In terms of the Earth's geographical coordinates, the tract of Arabic-script usage is generally between the latitudes of 15° N and 40° N, and between the longitudes of 15° W and 75° E.

¹⁰ Jahani 1989, p. 62.

northern Pakistan, the Wakhan corridor of Afghanistan, Tajikistan and north-western China); and Burushaski (spoken in the Karakoram mountains of Hunza and Gilgit valleys of northern Pakistan, and in Kashmir and northern India). Speakers of both these languages are currently experimenting with Roman script, instead of Arabic, modifying it to represent the phonological requirements of their respective languages, as explained in subsection 3.4.1, below. In addition, Brokskat (a Dardic language related to Kalasha, spoken in Jammu and Kashmir, India) has a non-Arabic and non-Devanagari script. It uses the Balti script instead, for religious reasons that are explained in subsection 3.4.2, below.

It is clear from the exceptions mentioned in this subsection that the choice of script is not predetermined, and that the link between the spoken and written forms of language is quite arbitrary.

3.3.2 Religious and political perspectives

Exceptions to the typical distribution of scripts also exist where ethnic communities spread across different countries are independently using the same script as each other, while differing from other scripts used in their immediate vicinities. This is the converse principle of the regional conformity of scripts discussed in the previous subsection (3.3.1).

The classic example of this phenomenon has been the wide use of the Hebrew script (despite more dominant local scripts) by mostly Jewish communities speaking a variety of different languages across the Middle East, North Africa and Europe, including Israel, Yemen, Iraq, Uzbekistan, Libya, Morocco, Tunisia and Germany.

Although the use of Arabic script is strongly associated with Islam, Christian communities speaking Arabic (in countries like Egypt, Lebanon, and Iraq) and Urdu (in Pakistan) use Arabic script because they are religious minorities in nations that officially use Arabic script. (By contrast, the mostly Muslim speakers of many other languages have chosen the Roman alphabet for other reasons: Turkish and Wakhi speakers in Pakistan, for phonological reasons, Indonesian because of colonization by the Dutch, and Syrian Kurdi and Chadian Arabic for other political reasons.)

The mostly Muslim speakers of various Central Asian languages (e.g. Kazakh, Kirghiz, and Osetin) use Cyrillic script instead of Arabic script, because for a long time they were political minorities in the Soviet Union, where Cyrillic script was prescribed for minority languages by the government.

These exceptions further illustrate the important point that there is no perfect script for any language, and that several alternative scripts can serve the purpose of representing any particular language. More reasons for the use of one script or another will be discussed in the next section (3.4).

3.4 THE ADOPTION AND ADAPTATION OF SCRIPTS

Even the once-perfect writing system has one of two ultimate destinies: It can remain rigidly frozen, becoming imperfect by default, or it can yield to the pressures and change. Indeed, herein lies the history of writing systems. (Adams 1990, p.14).

History is full of examples of both experimental and official adoptions and adaptations of scripts and writing systems for languages other than those for which they were initially employed. Such adaptations have featured varying degrees of dependence on extra devices like additional diacritics, digraphs, stylistic conventions, etc., and they have had varying degrees of success in terms of their teachability and durability. The ability for languages to be thus transposed into multiple scripts again demonstrates the general malleability of language with respect to orthography.

The table on the next page summarizes the direction of some script changes throughout West, Central and South Asia, at different points of time.

Table 3.2 Historical script change in West, Central and South Asia

	to Arabic	to Roman	to Cyrillic
from Arabic	n/a	Indonesian, Sri Lankan Creole Malay, Balti (India/Pakistan 1900s), Northern Uzbek, Turkish (1928), Tajik (1930), Uyghur (China 1940s), Burushaski (Pakistan 1990s), Wakhi (Pakistan 1992)	Uyghur (China 1941)
from Roman	Kazakh (China 1980), Uyghur (China 1984)	n/a	Tajik (1940), Judeo-Tat (Russia 1940), Northern Uzbek (1939)
from Cyrillic	Uyghur (China 1984)	Romanian (1800s), Osetin (1923)	n/a
from Devanagari	Kachi Koli (India/Pakistan 1940s)		
from Armenian	Kurmanji (Syria, Iraq & Iran)	Kurmanji (Turkey)	Kurmanji (Armenia, Azerbaijan, Georgia, Kazakhstan, Kyrgyzstan 1939)
from Vernacular (minor indigenous scripts)	Farsi (600s), Uyghur (600s), Kashmiri (1400s), Sindhi (India/Pakistan 1852)		Mongolian (1941)
from Georgian			Osetin (1954)

These cases of script change reveal some interesting trends. Firstly, early convergences to **Arabic** script were no doubt made for religious reasons (explained further in subsection 3.4.2, below). In later times, social considerations probably play an equal part in the choice of Arabic script, as the Muslim culture has become a major part of the social fabric in societies that have adopted Islam. Switches to the **Roman** alphabet are mostly due to social and pedagogical considerations (explained further in subsection 3.4.3, below) because the Roman script is less tied to a single religion or political entity than other scripts. Changes to the **Cyrillic** script have largely been due to the political motives of Russian leaders in the creation of the Soviet Union, where the extended Russian state exerted a powerful but relatively short-

lived influence over neighbouring countries (as exemplified several times in subsection 3.4.4, below).¹¹

Conversion to Arabic script has been happening a lot longer than conversion to Roman script. In the sample of Middle Eastern and Central Asian languages above, conversion to Arabic script has been happening ever since the 7th century and up until the mid 1980s, whereas conversion to Roman script is a relatively new and current phenomenon, beginning only in the 20th century (before WWII, and then again from the 1990s).

The continuing trends of globalization will probably have a narrowing effect on the script preference decisions of 21st century communities, where the actual range of popular and prestigious scripts is reduced, but their influence becomes more widespread. This would happen at the expense of several minor, indigenous scripts (for example, Brokskat, mentioned in subsections 3.3.1, above, and 3.4.2, below), although these vernacular scripts are not as common in the region under study as they are in East Asia and Africa. This worldwide trend of ‘script death’ would be a parallel phenomenon to the current escalating trend of language death, which affects languages of all regions.

What now follows is a discussion of the various factors that bear on the adoption and adaptation of particular scripts.

3.4.1 Phonological Representation

Probably the most fundamental hypothetical principle in adopting a script for a particular language is to choose one that is already best suited and well-fitted to the closest linguistically related language. This would ensure the maximum availability of built-in features of alphabet and script for phonological features that are common to both, as well as the closest matching fonts, requiring minimum adaptation for publishing and electronic media.

One recent focus of the script debate is the **Balochi** language (the trade language of Balochistan province of Pakistan, but also spoken in southern Afghanistan, Turkmenistan, Tajikistan, south-eastern Iran, India and Oman). The limitations of Arabic script for Balochi were seen as follows: ‘The lack of full vowel representation in the traditional Arabic script is, especially in languages with many vowel phonemes, felt as a hindrance to correct reading’

¹¹ One of many cases of a change to Cyrillic script is cited by Fishman (1977, p. xxiii): ‘Kolarz (1967) refers briefly to the reformed Arabic alphabet adopted by the Volga Tatars in the early twenties prior to the central Soviet concern for their writing system. Henze (1956) comments that “the practicality and popularity of this reformed alphabet made the Soviet introduction of the Latin alphabet among the Tatars very difficult”.’

(Jahani 1989, p. 144). On the other side of the coin, a stated reason for using a Roman script for Balochi was that it ‘represents the phonemes of Balochi better than the Arabic script’ (p. 145).

Inevitably, challenges arise when closely related languages differ in some special phonological feature that is only present in an unwritten language, and which may not be represented by an established orthographic convention in a closely related written language. This is where script adaptations become more creative, as in the need to include the representation of retroflexion, non-word-final nasalization, stress and tone in many north Pakistani languages, which are not built-in features of Arabic script.

Current moves to develop a suitable script for certain modern **South Arabian languages** with no written tradition have been reported by Hofstede (1998, p. 123):

At present, there are two systems for writing the languages. In one system, the Arabic alphabet is used. In order to represent all the phonemes with one single Arabic letter [sic], some attempts are made to create a system of modified Arabic letters. In the second system, an alphabet is used that is a mixture of IPA and a modified Latin [Roman] alphabet.¹²

The adaptation of two scripts for the **Burushaski** language, a language isolate spoken in the Hunza Valley of north-eastern Pakistan, has involved the *addition* of extra elements into both orthographies, according to Willson.¹³ The first example is the adaptation of the Arabic script by a Burushaski-speaking spiritual and cultural hero:

He took the Urdu alphabet and added different combinations of dots and the Arabic digit "4" above some consonants [to extend the range of available symbols for consonants]. He also uses the digit "2" over vowels to indicate shortness and the digit "3" over vowels to indicate rising pitch.

The second example is the adaptation of the Roman script by Willson himself, ‘using digraphs where possible (ch, gh, ng, sh, ts) and dots under English letters to show retroflex or backed versions of English sounds’.¹⁴

Besides additions to an orthography, sometimes an orthographic *deletion* or *reinterpretation* is possible when adapting a script if a phonological feature of a written language (appropriately marked) is not present in a closely related unwritten language. For example, the implication of a short vowel between contiguous consonants in Arabic script

¹² The ‘Modern South Arabian’ languages mentioned were Mehri, Socotri, Jibbāli, Baḥari, Harsūsi and Hobyōt.

¹³ S Willson 2001, pers. comm., 10 April. Willson is the author of several works on Burushaski (e.g. Willson 1996). Personal communications referred to in this chapter are responses to a small survey on script choice in north Pakistani languages, which I conducted by email among the relevant researchers. The respondents were Baart, Backstrom, Radloff and Willson.

¹⁴ S Willson 2001, pers. comm., 10 April.

does not apply to, and hence has been omitted from, the Arabic script that is used for **Khovar**, the language spoken by the closest neighbours of the Kalasha, which does not have phonemic vowel length. Contiguous consonants in Arabic-script Khovar may therefore be interpreted as a consonant cluster.¹⁵

Some languages have simultaneously adopted features from several scripts in an attempt to get the best orthographic features of each, for maximum phonological representation. An example of this approach is found in **Wakhi**, an Indo-Iranian language spoken in north-eastern Chitral District of Pakistan's NWFP, in the Hunza valley of north-eastern Pakistan, in the 'Wakhan Corridor' of Afghanistan, in Tajikistan and in China. Wakhi speakers have adopted symbols from three different scripts, in order to provide adequately for their phonemic inventory:

The Wakhi Roman script does use all 26 letters of the English alphabet, plus sixteen additional ones, most of which are simply formed by the addition of diacritics (wedges above and dots beneath) to the familiar English characters. Four Greek letters are also used (for some of the abundant fricatives), plus two Cyrillic ones ...¹⁶

The original **Balti** script was 'perhaps invented at the time of the conversion of the Baltis to Islam, about A.D. 1400' (Diringer 1968, p. 235). It was written from right to left, like Arabic, but featured mainly Roman capitals (to which unconventional phonetic values were assigned) and geometric shapes, with a generous sprinkling of diacritics to represent some consonants and all six vowels.

We have seen that phonological factors do have some relevance in the consideration of script *adoption* (choice)—and significant relevance in script *adaptation* (the actual development of an orthography), the foundation for which is the focus of the next chapter. The evidence of history shows, however, that there are usually more pressing reasons for choosing a script. Most script decisions have been made by self-determining communities based on religious, social, political, regional or pedagogical considerations. While ruling governments sometimes impose script decrees for political or religious reasons, the instigators of script change are often the majority or minority communities actually involved in speaking or planning those languages.

¹⁵ Only a small proportion of Khovar speakers can actually read or write the language. A small literary society promotes Khovar literature through the publication of both traditional texts (e.g. folklore, poetry). There is also a Khovar newspaper.

¹⁶ P Backstrom 2001, pers. comm., 6 April. Backstrom has published work about the Wakhi language and has jointly edited a volume of sociolinguistic surveys in north Pakistan (Backstrom & Radloff 1992).

3.4.2 Religious Considerations

Just as religion was a key factor in the creation and use of most writing systems in the first place (e.g. Egyptian hieroglyphics and Scandinavian runes), religious identity has continued to be the primary determinant of script choice by language communities wishing to adopt and adapt an existing script for their hitherto unspoken or unwritten language. The tendency has been for them either to prefer the script in which their religion was developed, or sometimes to deliberately avoid a script used by religious rivals.

One classic example of religious motivations impacting on the adoption and adaptation of existing scripts for a newly developed language is **Hindi-Urdu**, the language primarily derived from Hindustani, Farsi and Arabic. This language came into being when the combined armies of the region, speaking these and other minor languages, camped and fought together against the Mongol Empire that bore down on and conquered India in the 16th century.¹⁷ Eventually, in order to both identify with their own, and at the same time isolate themselves from their neighbours, the Muslims adopted and modified the Arabic script for their dialect (Urdu) and the Hindus kept the existing Devanagari script, associated with Hindustani, for their dialect of the new language (Hindi).

Similarly, the Sikh speakers of **Eastern Punjabi** (spoken in India's Punjab province) use the Gurmukhi script, which is based on Devanagari script. Again, for religious reasons, an Arabic script was adopted for **Western Punjabi**, spoken in Pakistan's mostly Muslim Punjab province (though it is rarely written).

Furthermore, where a community undergoes religious change, the tendency is to switch to a script that reflects and characterizes their new religion, for example, the switch from vernacular scripts to Arabic script for both **Farsi** and **Uyghur** speakers (most of whom converted to Islam in the 7th century). Other examples are given in Table 3.2. Referring to Uyghur, Hahn (1991, p. 91) points out that 'the pre-Islamic Uyghur script ... gradually fell into disuse as the Uyghurs developed a strong sense of attachment to the Arabic script' after the advent of Islam. He goes on to explain that, like most Islamized peoples, 'the Uyghur people have come to consider the Arabic script an essential tool for conveying not only religious matters but also virtually all native literary and artistic expressions'. Hahn feels this is the reason they tend to offer a considerable degree of resistance to the official instatement of other scripts in place of the Arabic script. He concludes that:

¹⁷ The name 'Urdu' was derived from the Hindustani (originally Turkic) word meaning 'camp', from which the English word 'horde', is also derived (through Polish), reminding us of the mass of soldiers that comprised the army of that time.

At least those Uyghurs who live in China have never truly abandoned the Arabic script, not even during those periods in which the public domain non-Arabic-based scripts were being used by official decree.

The overwhelming allegiance to Islam from the Mediterranean and the Red Sea right across to Pakistan is reflected in the fact that most communities in this region have preferred to use the Arabic script, and adapt it to represent their languages. For example, all but one of the thirty-nine Northwestern Indo-Aryan languages that have been committed to writing employ a modified Arabic script because the majority of the speakers of these languages are Muslim. The only exception in this group is the **Brokskat** language community in the Jammu & Kashmir region of India. Being Buddhists, they prefer to use a script that does not have the religious connotations of Islam. Instead, they use the Balti script, borrowed from the nearby people of Baltistan (in north-eastern Pakistan, Kashmir and India) who speak a Tibetan language. Thus, negative religious connotations have led Brokskat speakers to use the Balti script.

On the other hand, positive religious connotations have led **Balti** speakers to abandon the Balti script, and instead, to experiment with alternative scripts that characterize their own religious choices. Despite Balti being a Tibetan language, it is not written using a Tibetan script. It has undergone a series of script changes throughout history, again, mostly due to religious considerations. Early Balti literature was written in an indigenous Balti script which ‘was perhaps invented at the time of the conversion of the Baltis to Islam, about A.D. 1400’ (Diringer 1968, p. 235, quoting Grierson 1915). This invention at that time appears to have been an attempt by some, not wishing to convert, to maintain their identity.¹⁸

There have been various modest but conflicting religiously motivated attempts at establishing a new script for the Balti language in the last century, some externally initiated and some internally. In the 1900s, the British and Foreign Bible Society published three gospels and another book of the Bible in the Balti language, mostly in Arabic script. The author of a Balti grammar (Read 1934) then translated and published the gospel of John and some Psalms in Roman script. A few mullahs have translated some poems from Persian into Balti, retaining the Persian script. According to Diringer (1968, p. 235) ‘nowadays, the educated Baltis employ mainly the Persian-Arabic alphabet, which is most unsatisfactory and misleading’ (probably because of poor phonological matching).

¹⁸ It is interesting to note that an experimental Kalasha alphabet resembling this Balti/Brokskat alphabet was created by a Kalasha man, who also holds strongly to the value of the independence of indigenous religion. This is further discussed in Chapter 6 (section 6.1).

Despite this last comment, Grimes (2000) concurs 32 years later that in the Balti-speaking communities of both Pakistan and India ‘Perso-Arabic script is [still] the accepted one’. However, some of the younger generation prefer Roman script. Presumably, they are becoming less oriented to traditional values, and therefore, religious considerations are giving way to social reasons for their choice of script.

The **Purik** language community, in Indian controlled north Kashmir, provides another example of religious preference influencing script choice. It is 78% to 85% cognate with the Tibetan Balti language. “‘Purig” means “of Tibetan origin”. [These] people prefer to be culturally and linguistically identified with Tibet, although religiously with Islam’. Being ‘almost totally Shi’a Muslim’, presumably this religious affiliation has led to the ‘Persian-Arabic script of Urdu [being] used’ (Grimes 2000).

3.4.3 Social and pedagogical reasons

‘Rather than being mere instruments of a practical nature, [scripts and orthographies] are symbolic systems of great social significance which may, moreover, have profound effects on the social structure of a speech community’ (Coulmas 1989, p. 226). Social reasons for script change are almost, or equally, as influential as religious considerations (though one or both of these factors may also be determined by regional or political identities). Fishman (1977, p. xiii) mentions several historical cases in which:

newly literate communities have desired a more *distinctive* writing system, one that they could call *their own* or one that could more effectively differentiate their language from others with which they did not seek similarity but rather *dissimilarity*.

Coulmas (1989, p. 227) mentions a few more cases, but describes and illustrates both scenarios, where:

language attitudes, such as the desire to have an orthography which makes the language in question graphically similar to another or, conversely, makes the language dissimilar to another, may be irrational but they are social facts which often strongly influence the success of a proposed system.

Some contemporary examples of social considerations of script choice specific to northern Pakistan are now presented here.

Social conformity, a feeling of belonging to a larger group, was the main consideration for the adoption of an Arabic script by a delegation of speakers of **Kalami**, a Dardic language spoken in the northernmost part of Swat Valley in the NWFP of Pakistan. Baart (1997, p. 50)

has recorded the following observation concerning the very recent establishment of a writing system for that language:

In the summer of 1995, seven educated native Gawri speakers from Kalam formed a *spelling committee* and discussed a proposal for a writing system ... It was agreed by this committee that a Perso-Arabic script should be used for their language, rather than a Roman script. In general, it was felt that there should be maximum conformity of the Gawri writing system with that of Urdu.¹⁹

Urdu is the national language of Pakistan and by far the major literary language in the Kalam area. After centuries of independent rule, the Swat District was officially annexed to the rest of nation during the 20th century.

The large **Shina**-speaking society, who live on either side of the famous Karakoram Highway, are more in the mainstream of Pakistani society and have had a long history of being on the international trade route.²⁰ To them the reasons for script choice are still social, but they are more to do with perceived progress and opportunities. Radloff reflects on the practical reasons for script choice in Shina (in the Dardic family):

Shina speakers have never considered anything but an Urdu-based script. I presume it's because they know Urdu. Urdu is essential in the area for jobs, education, etc., so transition from Shina to Urdu will be helpful.²¹

Wakhi speakers in the Hunza valley of north Pakistan have given most weight to the educational factor in their script preference. Backstrom has written:

Probably the decisive factor ... to go along with the Roman-based script is the fact that the Wakhi Tajik Cultural Association decided to go with this script at one of their initial meetings in the summer of 1992. The WTCA embodies the most educated and progressive leaders of the Wakhi people. Their decision was undoubtedly influenced by the strong encouragement of their Imam to move toward English-medium education and fluency in English. The fact that their Roman script looks similar to English made it attractive as a way to hopefully simplify the leap to English.²²

Here we see the same social reason of perceived opportunities for progress as for Shina, but this time preferring Roman script. To educated Wakhi people it is a key to progress and to

¹⁹ '... and other languages in the region. As a result of this discussion, two new symbols were scrapped from the initially proposed alphabet' (Baart 1996).

²⁰ The 'Silk Road' has been a trade route through the Kohistan, Gilgit and Hunza valleys in north-east Pakistan for centuries, connecting China to the West.

²¹ C Radloff 2001, pers. comm., 2 April. Radloff is the author of at least eight works on northern Pakistani languages, including several on Shina (e.g. Radloff 1999)

²² P Backstrom 2001, pers. comm., 6 April.

upward mobility within national and global society. It is interesting to note that even the imam preferred Roman script to the usually Islamic-oriented Arabic script.

The **Burushaski** orthography is still not settled. There are currently two scripts being used for different social reasons, as Willson observes:

They wanted their language to look as much like English as possible, in order to help their children learn English ... and have found it easy to learn and use. The other orthography ... is already in use in the religious centres and by the Radio Pakistan staff.²³

The reasons for some Burushaski speakers preferring the Arabic script are cultural and religious, whereas the Roman script is seen by others as pragmatic, with a view again to future progress. Again, the use of two scripts demonstrates the point that more than one script can serve the purpose of writing a particular language, though the wider functionality is often what settles it one way or the other.²⁴

For similar pragmatic reasons to the cases described above, **Balochi** writers favoured Roman script, maintaining that it was:

the script of science and technology, which could help the Baloch develop faster in the present scientific era. It was also pointed out that the Roman script is used more throughout the world than the Arabic, and that its adoption therefore ought to make international contacts, as well as learning English, French and other international languages, easier for the Baloch (Jahani 1989, p. 145).

However, the long debate was finally won by those who favoured the Arabic script. Jahani points to:

ties to the Arabic script, which is probably the most important argument for many in this group, the fact that neighbouring languages employ this script, that many Baloch are familiar with it and that a certain body of literature has already been produced in it (Jahani, 1989, p. 149).

This brings out another principle, that precedence is also a significant factor in the establishment of a writing system, especially as it relates to pedagogical practices.

In that part of the world, ties to the Arabic script, and other means of group identification are just as much cultural considerations as they are religious ones. This identification can be

²³ S Willson 2001, pers. comm., 10 April.

²⁴ It is possible that behind the social and pedagogical reasons for the tendency of both Wakhi and Burushaski communities to choose Roman script is another motivation to underline their religious independences. Most Wakhi speakers and many Burushaski speakers belong to the Ismaili sect, which diverges greatly from orthodox Islam; and even the Shi'a Burushaski speakers see themselves as independent from the Sunni division of Islam, which dominates Pakistan. It is interesting to note that these are currently the only two languages in the region of the Kalasha to use a Roman alphabet, and that they have in common with the Kalasha their independence from orthodox Islam.

positively, as for Balochi speakers, or negatively, as for Wakhi and Burushaski speakers (as explained in footnote 22).

It is not surprising that all the same arguments for Roman script for Balochi have also been made by advocates of Roman script for Kalasha. These arguments seem tenuous to those of us in the outside world, and they may not lead to all the outcomes expected by them, but they can be very real and powerful arguments if there are few to oppose them.

A purely pedagogical argument for the teachability of a particular script could become the most powerful factor in terms of the long-term success of a script decision. It matters most to educational institutions who are involved in the development of learning and training materials and in the actual teaching of a new script, especially to children. Balochi writers who preferred Roman script to Arabic stated that ‘it is easier to learn and that its letters have fewer shapes ... English is also taught in Pakistan, which means that people are familiar with the Roman script’ (Jahani 1989, p. 145).

Baart (1996, p. 5) also reports that the **Kalami** spelling committee also had pedagogical issues in mind in their adoption of an Arabic script:

They wanted maximum ease of transfer between Kalami and the national language, so that outsiders who know Urdu will not encounter too many surprises when they see written Kalami. Also, the step from reading Kalami to reading Urdu should be made easy.

Here it is also implied that aligning the script of a minor language to that of a major, more socially powerful language, may elevate the status of the minor language by establishing its position in the wider community.

3.4.4 Political factors

Besides religious, social and pedagogical motivations, another major force in the choice of script is the decrees of governments with the political agenda of trying to unify their realms and to distinguish them from others. However, as the following examples show, it is the nature of political entities to be fickle in policy, pluralistic in practice and highly vulnerable to change.

The classic recent example of this is the Cyrillization of scripts in the USSR. One of the many affected languages was **Tajik**, which was initially written in Arabic script. In 1930, Roman script was officially accepted, and then around 1940, nearly all written minority languages in the Soviet Union switched over to Cyrillic script (Jahani 1989, p. 52).

Judeo-Tat was written in Roman script in the Russia of 1920 to 1940. From 1940, Cyrillic script was decreed there, though in Israel some Judeo-Tat literature now uses Hebrew script.

Ongoing political change has caused **Uyghur** scripts to be subject to rapid switching, especially in China. In 1937, the Arabic script version of Uyghur underwent an official reform; in 1941, the Cyrillic script was introduced; then throughout the late 1940s, Roman script was also temporarily used. In 1954, there was a reversion back to the Arabic script (with new reforms), but in the 1960s and 1970s the use of Roman script was officially decreed. In 1984, the Arabic script was again reinstated as the official script for Uyghur in China; in 1984, this was further reformed; and in 1987, a new Arabic script was introduced. Despite the imperialistic rule of the Chinese government, by continuing to issue decrees concerning language, script, and other aspects of life, there has obviously been some attitude of tolerance (or at least pluralism) toward different script types for the different ethnic minorities within its borders. This is also evident in the absence of a Chinese script for Uyghur in the government's language plans.

The abovementioned changes again demonstrate the versatility of scripts in representing language, as well as the multiplicity of (non-linguistic) factors that have a bearing on choice of script. Furthermore, in Kazakhstan, Uyghur speakers still use Cyrillic script, because of the earlier imposition of Soviet rule, and in Turkey, the Roman script is used. There the absence of powerful imperial forces in recent times has made way for the social influence of European scripts.²⁵

All of the above examples of script switching, sometimes frequent and rapid, again prove the weakness of the link between spoken and written forms of language. Appendix 2 presents a comprehensive table of reasons for script choice and change.

3.5 CONCLUSION

It should be obvious from the examples discussed in this chapter that the relationship between spoken and written forms of language is arbitrary and never fixed. The concept of an optimal script for each language is more imaginary than real. Phonological considerations, which one might assume to be paramount, have been shown to be of very little consequence in the adoption of a script. Political forces on change of script choice prove the very adaptability of

²⁵ See Shorish (1984).

language from one script to another. Religious and socioeconomic considerations are by far the strongest determinants of script adoptions.

The correlations between the use of a particular script and the factors that influence or determine the choice of that script are extremely variable, and overlap considerably, to the point of often being inseparable. Circumstances of geography, linguistics, sociology, religion and politics are inextricably intertwined with each other. Because of this, they have profound effects on each other, and combined effects on the choice and use of scripts. Script choice is by no means a circumstance set in concrete, but rather an outcome that can be determined and changed for any language.

PHONOLOGICAL FOUNDATIONS

4.1 INTRODUCTION

The first step in developing a suitable orthography for an unwritten language is to analyse its phonology, including dialectal variation. An orthography based on pronunciation minimises the distance between speech and writing. The importance of phoneme-grapheme correspondence in early literacy has been underpinned by theory, and is supported by research and tested in practice.

From a theoretical point of view, Smalley (1964b, cited in Berry 1977, p. 3) maintained that the second most important criterion of a new writing system is ‘maximum representation of speech’.¹ In reviewing Smalley’s criteria, Coulmas (1989, p. 238) concludes that ‘a new orthography should be ... founded on a phonemic analysis of the language’. Tauli (1977, p. 23) takes a teleological approach. He asserts that ‘linguistics and orthography planning cannot ignore the fact that to unsophisticated people, spelling represents pronunciation, and a letter is associated with a sound’. He supports his case by mentioning that ‘Stähle [1970, p. 71] maintains that it is essential from the social and pedagogical point of view that spoken and written language resemble each other so much as possible’. Tauli (1977, p. 24) concludes that:

the most economic system is the symbolizing of phonemes on the ground of the simple fact that it employs the fewest symbols and rules to represent speech. Such a system is certainly easier in learning to read and write. Thus the most efficient orthography is phonemic.

From the pedagogical perspective, Adams (1990, p. 82) presents research evidence (by Jeanne Chall and others) that ‘knowledge of letters and phonemic awareness were found to bear a strong and direct relationship to success and ease of reading acquisition’. And from practical experience, Pike (1947, p. 208), and many who have followed in his tradition (e.g. Shand 1979, pp. 16–21), advocated orthography design based on phonemic representation.

¹ According to Smalley, the most important principle underlying the development of a writing system is ‘maximum motivation for the learner’, based on attitudes to, and acceptance of, a writing system. However, Venezky (1970, cited in Berry 1977, p. 4) actually assigns this ‘sociolinguistic’ criterion as subordinate to Smalley’s other criteria.

However, Berry (1977, p. 9) and others have challenged the importance of this phonological criterion for the creation of a writing system, maintaining that ‘what the mature reader seeks and recognizes when he reads is not grapheme-phoneme correspondences but rather the correspondence of written symbols to the abstract lexical spelling of words’. This matter will be taken up further in the next chapter. It presents a strategic dilemma whether to provide primarily for mature readers who can cope with the irregular spelling of established orthographies, or for semi-literate people trying to cope with a newly created writing system. The latter category of learners describes the current Kalasha situation.

When a language is written for the very first time in history, a unique opportunity exists to establish an orthography that corresponds closer to the phonemic system of that language than will usually be possible thereafter; and there is no reason to avoid such an ideal where it is possible. Despite valid criticism by others of Smalley’s criterion of maximum representation of speech, Coulmas (1989, p. 229) confirms that ‘there is no reason to discard completely either the principle or employment of a phonemic-like transcription when creating an orthography.’

Chapter 3 concluded that linguistic considerations have generally played a minor role in the actual *adoption* (choice) of a script for a given language. However, phonological factors are fundamental in the *adaptation* (modification) of a script for a particular language because the distinctive sounds (phonemes) and other phonological features (e.g. tone) of each language need to be orthographically symbolized in distinctive ways.² Thus, an analysis of Kalasha phonology is an appropriate starting point in the development of a Kalasha orthography. Other issues in the determination of an appropriate writing system for Kalasha will be examined in subsequent chapters (e.g. morphophonemic and syntagmatic considerations in Chapter 5, sociological and political forces in Chapter 6, orthographic factors in Chapter 7, and typographical and technological issues in Chapter 8).

We start in this chapter, in section 1, by looking at the wider phonological context to which the Kalasha language belongs, in an areal sense. Section 2 provides an overview of the phonological features of Kalasha. Finally, in section 3, we will return to a theoretical issue: the role played by orthography in providing phonemic solutions.

² For example, the Roman script has been adapted to cater for the sounds of languages as disparate as Danish and Hungarian, and the Arabic script has been adapted to cater for the sounds of Farsi and Urdu. Some of the phonological techniques developed and prescribed for orthography planning by script adaptation can be seen in Pike (1947), Sjoberg (1968), Gudschinsky (1973), Smalley (1976), Berry (1977), Fishman (1977) Grimes & Gordon (1980), etc.

4.2 PHONOLOGICAL FEATURES OF NEARBY AND RELATED LANGUAGES

Most languages spoken in and around the North West Frontier Province (NWFP) of Pakistan have a consonantal inventory that includes plosives, affricates, fricatives, nasals, laterals and flaps, at various points of articulation. Many of these languages feature retroflex obstruents and aspirated plosives and affricates. Vowels are typically front unrounded and back rounded, with three heights, and often with nasal counterparts. Contrastive vowel length is common, although not a feature of Khowar, the closest language to Kalasha.³ Tone is not uncommon; for example, tonal processes have been analysed and presented for Shina (Schmidt & Zarin 1981) and Kalami (also known as Gawri) (Baart 1995, 1997, 1999). Burushaski, Khowar and Dameli are other languages of the area for which a distinctive low-rising tone has been observed (Baart & Sagar 2002). Diphthongs do occur, but are relatively uncommon.

A typical phoneme inventory in the region would be in the order of 30 to 60 phonemes (20 to 50 consonants and 5 to 20 vowels). Although Urdu is not indigenous to the North West Frontier Province, it is mentioned frequently throughout this thesis, as a comparison language, because it is the predominant written language of Pakistan, in which many Pakistanis (including the Kalasha) have some bilingual competence. For this reason, it has a more substantial influence throughout Pakistan in general than any other language.⁴ Urdu has been interpreted as having 63 phonemes: 40 consonants and 23 vowels (Maddieson 1984).

Aspects of phonology have been documented generally for the languages of the NWFP (Shahidullah 1964), Dardic languages (Kachru 1965; Voegelin & Voegelin 1965), South Asian languages (Masica 1992), Indo-Aryan languages (Masica 1991; Cardona 1992a), the Hindukush mountain area (Tikkanen 1995), Indo-Iranian languages (Cardona 1992b), Nuristani languages across the Afghan border (Strand 1973), North West Frontier Province (Morgenstierne 1932; Shahidullah 1964), Khowar (Skalmowski 1995; Munnings 1998a, 1998b), Kalami/Gawri (Baart 1997; Baart & Sagar 2002), Indus Kohistani (Hallberg 1992, 1999), and Shina (Schmidt & Zarin 1981; Schmidt & Kohistani 1995; Radloff 1999), all of which, along with the more major languages of the region (Farsi, Pashto, Hindko, Kashmiri and Urdu) have varying degrees of resemblance to and influence on Kalasha.

³ The loss of vowel length contrast may be a relatively recent phenomenon (see Heegård 1998).

⁴ Second to Urdu in general influence would be English. Khowar is very influential as a spoken language in the Kalasha Valleys.

4.3 KALASHA PHONOLOGY

4.3.1 Previous descriptions of Kalasha phonology

Kalasha phonology has been sketched in the past by Grierson (1915) and Morgenstierne (1932, 1965, 1973). However, each of these researchers spent only relatively short times with Kalasha speakers, because their primary work covered much vaster areas. Grierson was documenting languages all over the Indian subcontinent and Morgenstierne researched the history, archaeology, cultures and languages of the whole Hindu Kush region.

With no prior knowledge of each other, Trail and I independently began extensive and thorough research on the Kalasha language in 1982. When our paths crossed, we agreed to collaborate, and our first joint work was a 40-page phonemic summary (Trail & Cooper 1985a).⁵ Since that time, by means of further elicitation, linguistic questioning, the recording, computational analysis and transcription of texts, and the preparation of literacy and orthography materials in collaboration with indigenous speakers, a lot of extra data has been collected in the form of oral and written texts, lexical items for the dictionary, and notes.

This ongoing, systematic collection and rendering of language data has led me to a substantial amount of further examination and analysis, which, in turn, has resulted in a number of minor reinterpretations of the phonology. I have incorporated some of these reinterpretations into literacy materials, such as a Kalasha alphabet book and a pre-reader (Cooper, Hall & Cooper 1994a, 1994b, 1994c, forthcoming), as well as in my contributions to the comprehensive and continually updated lexical database that was eventually published as the first Kalasha-English-Urdu dictionary (Trail & Cooper 1999). This dictionary incorporates our interpretations of the approximately 1,000 words listed in Morgenstierne (1973) plus approximately 5,000 more entries. It is a technical work that was written for an academic readership, not really meant for the Kalasha—though they have shown considerable interest in it. The present chapter is the first formal presentation of my reinterpretations of Kalasha phonology. It takes the ‘functional view of language, in which system and structure are foundational’ (Clark & Yallop 1995, p. 389), but with added insights and influences from lexical phonology (Mohanen 1986) and experimental phonology (Ohala & Jaeger 1986). The updated summary of Kalasha phonology is represented in the revised phonemic chart that appears in Table 4.1, and in the brief *précis* that follows here.

⁵ The phonemic inventory from that manuscript appears in this thesis as Appendix 3.

4.3.2 Summary of Kalasha phonology

The Kalasha language is phonologically similar to its regional neighbours in that it has most of the phonological features listed in the previous section (4.2). Notable by their absence are labio-dental and uvular articulations, and the features of vowel length and tone. On the other hand, the Kalasha language has an unusually rich repertoire of approximately 62 phonemes: about 42 consonants and 20 vowels.⁶

Part of the explanation for this large inventory is that the Kalasha phonemic system features extra consonants and vowels in several parallel series with respect to both place and manner of articulation. There are seven places of articulation, five of them being primary. However, following Maddieson (1984, pp. 31, 76), and to some extent, the IPA:

dental and alveolar places have been collapsed together, partly because ... in a very large number of instances, it is not possible to determine if a segment is dental or alveolar in place of articulation ... and partly because a contrast between these places is unusual.

The place of Kalasha consonant articulation which is most uncommon in the global context is **retroflex** (affecting most of the obstruents, and the flap in one dialect). Although this feature is extant in the Indian subcontinent, its application to affricates and fricatives is more confined to a few of the northern languages. There are also several parallel series of consonants based on all combinations of **voicing** and **aspiration** contrasts for most manners of articulation. Three of the four approximants feature secondary manners of articulation. Two of these are lateral approximants: /l/ is palatalized and /ɭ/ is velarized.⁷ The velar central approximant is labialized to become the typical ‘labial-velar’ /w/. The status of the phoneme /ḍz/ is a matter of interpretation, as discussed in subsection 4.3.4, below.

The Kalasha vowel system features five qualities, each represented in four series/sets. The property of the primary sets is **oral**, and the properties of the secondary sets are: **nasalized**, **rhotic** and the intersection of these two sets, that is, **nasalized-rhotic**.⁸ Rhotic vowels are very unusual, occurring in less than 1% of the world’s languages (Maddieson 1984, cited in Ladefoged & Maddieson 1996, p. 313). Other languages known to have rhotic vowels are Mandarin Chinese (Maddieson 1984, pp. 252, 346), Badaga, a Dravidian language spoken by

⁶ Only about 4% of the world’s languages have more than 16 vowels (Maddieson 1984, p. 126).

⁷ Minimal pairs include /law'ek/ (‘to lie’) & /ɭaw'ek/ (‘to steal’), and /'ala/ (‘up’) & /'aɭa/ (‘that’).

⁸ Some of the taxonomic terminology used here is from Maddieson (1984), though he actually used the term ‘retroflex’, as did Trail & Cooper (1985). However, it is more appropriate to refer to this property of vowels as rhoticity. Thus, Clark & Yallop (1995) referred to such vowels as ‘rhotacized’, and Ladefoged & Maddieson (1996) now use the term ‘rhotic’.

no more than 300,000 people in about 200 villages of Tamil Nadu in south India (Lindau 1978, p. 553; Emenau 1939, cited by Ladefoged & Maddieson 1996, p. 313), and Eggon, a Niger-Congo language spoken, with about 25 dialects, by about 140,000 people in Nigeria (Lindau 1978, p. 553; Grimes 2000).

Although the *Kalasha Dictionary* (Trail & Cooper 1999, pp. xxii–xxiii) lists three diphthongs (*aw*, *ay* and *ey*, which could be interpreted as [au], [ai] and [ei] respectively), these sound sequences are more likely to be sequences of juxtaposed vocalic segments (i.e. vowels + central approximants [w] ~ [j]), which is by far the more common explanation for this phenomenon than their being phonologically unitary diphthongs (Maddieson 1984, p. 133). This reflects his general solution to the problem of ‘choosing a unit or sequence interpretation’ of diphthongs etc., by giving ‘some prejudice in favour of treating complex phonetic events as sequences (i.e. as combinations of more elementary units’ (p. 6). Clark and Yallop (1995, p. 74) confirm that ‘in many languages such sequences are simply alternative ways of transcribing diphthongs, that is, [ej] = [ei]’, i.e. [ei]. They concede that ‘questions about the occurrence of approximants such as [w] and [j] are particularly difficult to answer because of the scope for alternative analyses’ (p. 122). They are ‘questions of interpretation’ for which there is no single definitive answer (p. 75). Clark and Yallop demonstrate this dilemma for English (pp. 79–80.)

Stress in Kalasha is rarely significant, with only a handful of minimal pairs in the published lexicon, for example, [ʼaʼja] ‘mother’ and [aʼja] ‘here’, [ʼaʼta] ‘that’ and [aʼta] ‘there’, [ʼtaʼri] ‘star’ and [taʼri] ‘sugar’ (Trail & Cooper 1999).⁹ To put Kalasha phonology into proper perspective, it is worth noting that each one of the features of aspiration, nasalization, retroflexion, rhoticity and stress, occurs in less than 1% of natural text, giving all these features a low functional load (illustrated in subsection 4.3.3, below).

Kalasha phonology is summarised in the following table, which shows the range of Kalasha phonemic segments arranged by type and place of articulation.¹⁰ Appendix 5 lists all the associated allophones of each phoneme, along with their phonetic descriptions, contextual distributions and Kalasha word examples in both phonemic and phonetic notations, with

⁹ Note that in this thesis the stress mark (ʼ) appears **before** the stressed syllable. (The *Dictionary* uses an acute accent over stressed vowels.)

¹⁰ This phonemic analysis differs considerably from the one that my colleague and I originally documented (Trail & Cooper 1985a), where both nasalization and rhoticity (‘retroflexion’) of vowels were interpreted as suprasegmental features.

English glosses. The phonological notations and terminology used in the following table and throughout this thesis are taken from the International Phonetic Alphabet (IPA).¹¹

¹¹ The typeface used is *SILDoulosIPA* (SIL 1993).

Table 4.1

KALASHA PHONEME CHART

CONSONANTS

	<u>Bilabial</u>	<u>Dental/Alveolar</u>	<u>Postalveolar</u>	<u>Retroflex</u>	<u>Palatal</u>	<u>Velar</u>	<u>Glottal</u>
Plosives							
unasp vl	p	t		t̪		k	
asp vl	p ^h	t ^h		t̪ ^h		k ^h	
unasp vd	b	d		d̪		g	
asp vd	b ^h	d ^h		d̪ ^h		g ^h	
Affricates							
unasp vl		ts̺	tʃ̺	tʂ̺			
asp vl		ts̺ ^h	tʃ̺ ^h	tʂ̺ ^h			
unasp vd		dz̺	dʒ̺	dʂ̺			
asp vd			dʒ̺ ^h				
Fricatives							
vl		s	ʃ̺	ʂ̺			h
vd		z	ʒ̺	ʐ̺			
Nasals							
	m	n				ŋ	
Approximants							
central					j	w	
lateral		l	ɭ				
Flaps							
		ɾ		ɽ			

VOWELS

	front<----->back				
	<u>high</u>	<u>mid</u>	<u>low</u>	<u>mid</u>	<u>high</u>
oral	i	e	a	o	u
nasalized	ĩ	ẽ	ã	õ	ũ
rhotic	ị	ẹ	ạ	ọ	ụ
nasalized-rhotic	ị̃	ẹ̃	ạ̃	ọ̃	ụ̃

4.3.3 Relative frequency of Kalasha phonemes

In order to assess the relative status of Kalasha phonemes, and to assist in subsequent judgements as to their suitability for representation in a Kalasha orthography, Table 4.2 on the following page lists the relative frequencies of all Kalasha phonemes, based on a sample text of an indigenous story consisting of 1760 words.¹² The tally is the number of single occurrences of each phoneme in that text, though, in several cases, these phonemes occur once or even twice in a very limited set of words. The percentage figure represents the proportion of each phoneme's single occurrence (tally) compared to the total number of phonemes in the text.

¹² This text appears in Appendix 6.

Table 4.2
KALASHA PHONEME FREQUENCIES

CONSONANTS			VOWELS		
Phoneme	Tally	Percentage	Phoneme	Tally	Percentage
t	425	4.8	a	1805	20.5
s	358	4.1	i	690	7.8
k	334	3.8	o	316	3.6
n	290	3.2	u	267	3.0
m	254	2.9	e	189	2.1
j	229	2.6	ẽ	50	0.6
d	182	2.1	(24 double occurrences in 1 word, and 2 occurrences in 1 word)		
b	163	1.8	õ	38	0.4
ɬ	156	1.8	(all occurrences in 1 word)		
r	150	1.7	aː	22	0.2
h	142	1.6	(all occurrences in 1 word)		
p	141	1.6	ã	2	< 0.1
ʃ	133	1.5	iː	1	< 0.1
g	132	1.5	ãː	0	< 0.1
w	116	1.3	eː	0	< 0.1
(tʃ)	65	0.7	ẽː	0	< 0.1
ʒ	50	0.6	ĩ	0	< 0.1
(tʃ) ^h	40	0.5	ř	0	< 0.1
(dʒ)	39	0.4	oː	0	< 0.1
g ^h	38	0.4	õː	0	< 0.1
(all occurrences in 1 word)			ũ	0	< 0.1
ʂ	29	0.3	uː	0	< 0.1
l	27	0.3	ũː	0	< 0.1
k ^h	26	0.3			
ɬ	22	0.2			
t ^h	20	0.2			
(all occurrences in 2 words)					
z	19	0.2			
ɖ	15	0.2			
b ^h	3	< 0.1			
(all occurrences in 1 word)					
(tʂ)	2	< 0.1			
(tʂ) ^h	2	< 0.1			
ɖ ^h	1	< 0.1			
zː	1	< 0.1			
d ^h	0	< 0.1			
(dʒ)	0	< 0.1			
(tʂ)	0	< 0.1			
(dʒː)	0	< 0.1			
ɽ	0	< 0.1			
ŋ	0	< 0.1			

4.3.4 Marginal phonemes and phonological interpretation

As Table 4.2 shows, the relative frequencies of Kalasha phonemes vary enormously in actual use. Those phonemes with the lowest functional load are the modified vowels, the aspirated voiced consonants, the dental affricates and the retroflex flap.

As a case in point, there are relatively few minimal pairs that show contrast between most aspirated and unaspirated equivalents of voiced consonants. This could be evidence of the gradual disappearance of voiced obstruents, in a similar way to the loss of voiced aspirates in most other Dardic languages, as well as in Panjabi, Pahari, Hindko and Gujarati.¹³ Their low functional load is indicated by a low tally in Table 4.2. However, aspiration of some voiced obstruents does occur repeatedly in the same very limited set of very common Kalasha words, thus justifying their current retention in the phonemic inventory. For example:

/g ^h /	as in /g ^h õik/	‘to say’
/b ^h /	as in /b ^h aik/	‘to be able to’
/ḍ ^h /	as in /ḍ ^h onik/	‘to know’

The existence of minimal pairs for all the other phonemes with low functional load also justifies their retention in the *phonemic* inventory. However, the rarity of the occurrences of some of these phonemes in Kalasha texts will have implications on decisions relating to the *orthographic* inventory. These decisions will be dealt with in Chapter 7 (subsection 7.1.2).

There appears to be occasional contrast in analogous environments between the dental nasal [n] and a palatal nasal [ɲ] (c.f. [niʼʃan] ‘sign’ and [ɲiʼhik] ‘to emerge’). However, unless and until there is further conclusive data in the form of an absolute minimal pair, the palatal nasal [ɲ] will be interpreted as an allophone of the dental nasal [n].

There has been some uncertainty as to whether the sound [ŋ] has phonemic status in Kalasha, or not. It could be explained as an allophone of either /n/ (occurring before voiceless velar plosives /k/ and /g/), or as an allophone of /k/ or /g/ (occurring word-finally). However, the complex explanations required to justify such an interpretation (which seems to be motivated by phonemic economy) are outweighed by the more straightforward solution of accepting it as a separate phoneme.

¹³ The disappearance of voiced aspirated consonants has given rise to the development of tone in some of these languages (Baart 2003, pers. comm., 30 Oct.).

There is also an issue with the interpretation of nasalization before plosives. For example, [an'daj] 'here' could be phonemically interpreted as /an'daj/ or /ã'daj/.

[d̥ʒʰuŋkik] 'to burn' could be phonemically interpreted as /'d̥ʒʰuŋkik/ or /'d̥ʒʰũkik/

[aŋ'gar] 'fire' could be phonemically interpreted as /aŋ'gar/ or /ã'gar/

[ʃum'ber] 'before' could be phonemically interpreted as /ʃum'ber/ or /ʃũ'ber/

In this thesis, I have adopted the former interpretation in each case (i.e. oral vowel plus nasal consonant, rather than nasal vowel). However, both of these are valid interpretations and further research may point to one as preferable.

Phonemic aspiration following both the bilabial nasal [m] and the lateral approximant [l] has also been documented (Trail & Cooper 1999). However, due to the extreme rarity of this phenomenon (and thus the lack of evidence pointing to speakers' intuitions) this aspiration will be interpreted here as the phoneme /h/ following the nasal and the approximant, rather than aspirated nasal and approximant phonemes, for example, /mhal/ ('curse'), /mhalokat/ ('creation'), /lhoo/ ('sweetheart'), /lhojak/ ('flat and smooth').

4.3.5 Dialectal variation within Kalasha phonology

A general introduction to the dialects of Kalasha was given in Chapter 1 (subsection 1.2.2). It was stated there that the **Bumboret** Valley dialect would be regarded as the standard for demographic reasons. Yet, even villages within Bumboret Valley show minor dialect differences. For example, residents of the village of Krakal (/krakaʔ/) employ the voiced dental affricate phoneme /d̥ʒ/, which is not used in the dialects of other parts of the Kalasha Valleys. The absence of absolute minimal pairs in identical environments makes the status of this extra phoneme a matter of interpretation, which reflects the dynamic nature of phonology, particularly where dialects are concerned. However, even if interpreted as a phoneme, since it carries such a low functional load (as demonstrated in Table 4.2), even in Krakal, and since it is sometimes used in free variation with the similar phoneme /z/, it is not considered significant for the orthography, as the letter that will represent /z/ could substitute for it.

There are two obvious phonological differences to the Bumboret dialect in the **Birir** Valley dialect. One is the contraction of a few frequently used conjugations. For example, what is often regarded as archaic forms in the general Kalasha paradigm of the verb 'to say' are still spoken in the Birir Valley dialect, but have been contracted by the elision of intervocalic /tr/ in the other valleys, for example, /matrik/ becomes /maik/ ('to say'), etc. The reverse occurs

in the elision of intervocalic /r/ in Birir Valley, for example, for /parik/ ('to go' or 'let's go'), the Birir Valley dialect is the one that contracts this paradigm to [pa:k], /pak/, etc.

The other most notable difference in the Birir Valley dialect is the employment of the retroflex phoneme /ɽ/. For example, Birir Valley speakers say [azaɽi] for 'apricot', whereas speakers in other valleys say [azaɾi], where the intervocalic /r/ (which the other dialects do not use at all) is elided, and the two surrounding vowels are rhotic. This explanation is preferable to the alternative one of contiguous rhotic vowels in other dialects being replaced by an intervocalic /r/ in Birir, because the loss of a phoneme is more likely than the introduction of one.

It is anticipated that the Birir Valley dialect *should* be reflected in the orthography because it involves not substitution or elision, but rather the addition of a letter to represent the retroflex flap phoneme /ɽ/. This phoneme is not used in the main dialect, but the Birir Valley speakers use it between vowels that are rhotic in the main dialect but oral for them. The orthographic symbol representing this retroflex flap would be part of the standard Kalasha alphabet, but would only be important for Birir Valley writers, and not used by others.

One minor phonological difference in the **Rumbur** Valley dialect is a greater use of a word-final glottal stop allophone of the voiceless velar plosive /k/. This phenomenon is more marked in women.

Brun and Krakal are equally the largest and most populous villages in Bumboret Valley. The speech of Brun manifests the common phonological features of all three Kalasha Valleys, whereas the speech of Krakal has an additional idiosyncratic phoneme (as mentioned in the first paragraph of this subsection). Thus, the Brun dialect is the better candidate to represent the standard dialect of Kalasha. In addition, Brun is geographically central in the largest valley (Bumboret).

There is another dialect in the Kalasha community that is not geographically based, but based rather on exposure to other languages. The voiceless labio-dental fricative /f/ (phonemic in Urdu and English) and the voiceless velar fricative /x/ (phonemic in Urdu) are not phonemic in Kalasha; however, they are used by those who speak these other languages. These individuals use them as allophones of /p^h/ and /k^h/ respectively in foreign words, such as Urdu [afi'at] ~ [ap^hi'at] ('comfort'), [fel] ('fail') and [xajr] ~ [k^hajr] ('well-being'), Arabic [sa'fa] ~ [sa'p^ha] ('clean') and Farsi [xoʃ] ~ [k^hoʃ] ('beloved'). On the other hand, the voiced counterpart of /f/, that is, /v/, is not a phoneme of either Urdu or Kalasha, and is very

difficult for speakers of either language to differentiate from /w/. It is anticipated that none of these foreign phonemes will be required in the Kalasha orthographic inventory.

4.4 PHONEMIC ANALYSIS AND ORTHOGRAPHY

Although phonology is normally the foundation of orthography, elements of particular orthographies may impact back on phonological interpretation. Coulmas (1996, p. 304) refers to this phenomenon in terms of how ‘literacy influences linguistic awareness’, reporting demonstrations that ‘segmentation ability is greatly enhanced by, if not dependent on, written language skills’. Until recently, the profound effect of a writing system on language conceptualization, and as a factor in linguistic analysis, has not been well documented with empirical evidence, especially where the IPA has not been used.¹⁴

The IPA (or any other scientific phonetic transcription system) is a proper tool for phonetic analysis, and as such, is a good basis for any ultimate orthography, using any script. Despite its core set of Roman characters its usefulness is not biased to any particular family of languages, or to linguists from any particular orthographic background. The IPA’s value is based rather on the fact that it is the most comprehensive and universal system for describing phonetic phenomena.

However, in the translation from IPA to a particular conventional orthography there is sometimes a compromise or loss in biuniqueness. Indeed, the IPA is not even used for phonetic analysis by all field linguists, especially those with an areal bias, such as speakers of the language being analysed who are literate in another language. Sometimes an orthography from another language is used for analysis, instead of the IPA; for example, the linguistic analysis of Shina, as discussed below.

Even the most comprehensive conventional orthography, or combinations of orthographies, can fail to represent some features in another language with a slightly divergent phonology.¹⁵ Certain notational resources (e.g. symbolisation, orthographic conventions) will match one language more closely than another, or they may be applied in different ways by different linguists.

¹⁴ An example of how orthography may influence phonemic interpretation in regional dialects of a standard language is the English digraph *ng* to represent the segment [ŋ], which is usually interpreted as the velar phoneme; however, this sound is not a phoneme in the British Midlands and other dialects of northern England (Clark & Yallop 1995, p. 95, 120).

¹⁵ Combinations of orthographies are often used within a certain writing system (i.e. using the same script).

This phenomenon has been demonstrated by Schmidt and Kohistani (1995) who examine specific differences that emerged from two separate quasi-phonemic analyses of Shina (another Dardic language spoken in the Gilgit, Kohistan and Hunza valleys of north-east Pakistan). Neither analysis used IPA notation as a frame of reference. One was performed using only Roman script (Schmidt & Zarin 1981); the other was performed using primarily Arabic characters, with occasional resort to IPA to capture vowel quality, stress and tone. It was found that while the inventory of consonants and simple vowels were identical between the two analyses, ‘the inventory of diphthongal glides (which are difficult to represent in the Arabic system) was smaller in the second analysis’. They concluded that both solutions worked, and were therefore valid, and indeed that each interpretation could even be forced (though with some inconvenience) in the other script. However, the implications are not trivial, because alternative interpretations, based on the notational tools of the analysis, can alter language classifications in comparative studies. Their paper is an important contribution to the disciplines of both phonological analysis and orthography design, as the authors also ‘explore the question of whether it is possible to express phonemic data in less orthography-bound terms’.

The same situation could be exemplified in Kalasha, where either an Arabic or a Roman notational system could be used for analysing the retroflex flap /ɽ/ spoken in the Birir dialect of Kalasha. This would be more likely to appear in the phonemic inventory of Kalasha if an Arabic notational system were used to describe its phonology, because this phoneme, and a matching grapheme, already exists in Urdu (the natural model for Kalasha orthography) and could easily be carried over to Kalasha. However, using a Roman notational system to describe Kalasha would make it a little less likely for the retroflex flap to appear in the phonemic inventory, because this phoneme would require a special unconventional font.

Similarly, the voiced dental affricate phoneme /d͡z/ is less likely to appear in the Kalasha phonemic inventory if a Roman alphabet were used for analysis, because it would most likely require a digraph to represent it. However, even using the Arabic script for analysis of Kalasha is no guarantee that this phoneme would be recognized, because it does not exist in Urdu. Pashto (the language of the North West Frontier Province of Pakistan, where the Kalasha live) does have this phoneme, with grapheme to match, along with its voiceless counterpart /t͡s/, though Pashto would never serve as a model for Kalasha orthography because it is nowhere near as widely written as Urdu. The voiceless counterpart /t͡s/ would more likely stay in the Kalasha phonemic inventory because it has a higher functional load than /d͡z/ (see Table 4.2), meaning that it is more common and is more significantly distinct

from similar phonemes like /t/ and /s/. An Arabic notational system would have to use the little-known Pashto letter in this case, and a Roman notational system would have to use the digraph \widehat{ts} . Both of these solutions would be less than ideal, but the best alternatives under the respective circumstances.

Again, using a Roman notational system to analyse Kalasha would help to ensure the recognition of the phoneme /f/ as part of the Kalasha phonemic inventory, as a borrowed phoneme from neighbouring languages. The phonological distinction represented by the Roman letters *y* and *i*, and *w* and *o*, is lost in an Arabic-script, where these letters are shared between the semi-vowels /j/ and /w/ and their phonologically similar vowel counterparts /i/ and /o/ respectively. On the other hand, the more indigenous Kalasha alternative, /p^h/ for [p^h], is less likely to be used in a Roman orthography, because of the need to superscript for the aspiration. Otherwise, it would be confused with the Roman-script digraph /ph/, which is usually pronounced [f] (e.g. in *philosophy*). The phonemes /p^h/ and /f/ are more likely to survive in an Arabic-based Kalasha orthography, where symbols for aspirated plosives would be already available.

As Chapter 3 concluded, there is no perfect script for representing a language—nor, as we have seen here, even for analysing it. Orthographies are invented tools that do not intrinsically belong to any particular languages, and are imperfect means to capture the dynamics of speech, either analytically or representationally. The degree to which particular orthographies describe or represent particular phonologies will always be somewhat arbitrary and contrived. For these reasons, and because more cases could be made, the phonological analysis presented here for Kalasha has used IPA symbols, to ensure the maximum degree of phonetic distinction, and the maximum availability of symbols, as a basis for phonological interpretation.

In analysing the link between phonology and orthography, Ohala and Jaeger have applied a number of general psycholinguistic insights from their work in experimental phonology to the subject of the development of writing systems. They report that the experimental orthographic rendering of certain ambiguous forms by illiterate speakers often provides the evidence for the psychological reality of certain phonological patterns in their languages (Ohala & Jaeger 1986, pp. 164, 171).¹⁶ The inclusion of mother-tongue Kalasha speakers' intuitive insights in the process of orthography development is taken up further in the following chapter.

¹⁶ They acknowledge the procedural caveat for certain types of experiments that 'orthographic interference is an important [though not always overwhelming] source of artifacts in all phonological experiments involving literate subjects' (Ohala & Jaeger 1986, p. 198).

MORPHOPHONEMIC AND SYNTAGMATIC CONSIDERATIONS

5.1 INTRODUCTION

In the previous chapter, we took the first step towards developing a means for writing the Kalasha language, by examining and describing its phonological foundations. The next step is to develop principles for building on this foundation in order to construct orthographic representations of Kalasha morphemes and syntagms.

In theory, the orthographic representation of a word should reflect the phonemes that occur in it. But what happens when the phonemic shape of one morpheme varies when it becomes phonologically combined with another morpheme (e.g. with affixation)? This ‘morphophonemic’ phenomenon gives rise to the notion of allomorphs. How should an allomorph be represented at word level? Should it reflect its phonology (according to how it manifests phonemically in various contexts), or should it reflect its morphology (always according to its morphemic base form)? We find ourselves faced with this decision because, as Coulmas (1996, p. 380) puts it, ‘alphabetic orthographies vary with respect to phonological abstraction and the transparency of the relation between spelling and phonology’. Mithun (1992, p. 142) spells out the issue more precisely:

Planners must decide what level of abstraction is appropriate for a practical orthography; they must consider not only what is most accessible to the consciousness of native speakers—easiest to learn, easiest to write with confidence, and most efficient for reading—but also which of these considerations should take precedence. Furthermore, many languages have extremely complex but productive morphologies. Bound morphemes in these languages typically perform many of the functions served by separate words in European languages. Planners must decide whether they wish to preserve the identity of morpheme shapes in their spelling systems.

This dilemma is the subject of what has become the ‘orthographic depth hypothesis’, which postulates the possibility of both shallow and deep orthographies in the same language (see Lukatela et al. 1980; Katz & Feldman 1981; Feldman et al. 1983; Katz & Fieldman 1983; Bentin & Frost 1987; Frost, Katz & Bentin 1987; Frost & Katz 1989; Sebastián-Gallés 1991; Frost 1992, 1994).

A shallow representation portrays only the phonological level of language. It reflects the various phonemic manifestations of morphemes that occur in different phonological environments. It may or may not reflect underlying base forms. This method of representation, in its strictest sense, attaches to the phonemic representation discussed in chapter 4. As stated there, when translated into an orthography, a phonemic (i.e. shallow) representation is thought to be easier from the early literacy point of view.

A deep representation, on the other hand, reflects underlying base forms of morphemes or lexemes, regardless of how they might change phonologically in different environments. By this method, there are, however, sometimes alternatives when representing a morpheme that undergoes morphophonemic change: either pronunciation when combined with an affix or clitic determines how it should be represented in isolation, or the pronunciation in isolation determines how it should be represented when combined with an affix or clitic. Deep (morphemic) representations, when translated into orthography, may not be as well suited to the very early stages of literacy, because of possible morphophonemic confusion. However, since the *reading process* quickly progresses from phonemic to morphemic recognition anyway, this is probably not a serious disadvantage.

Baart provides a helpful perspective in reference to this debate:

The choice between a more abstract spelling ... and a shallow spelling ... is not going to be decided by the linguistic analysis. The best the linguistic analysis can do for us is, given a choice for an abstract spelling, help us to choose between alternative abstract spellings. But the choice between abstract and shallow as such is made on the basis of psychological and pedagogical considerations.¹

If the orthographic depth hypothesis is to be followed in the formulation of a new orthography, the choice between a shallow or deep representation would actually be a matter for consideration from several perspectives: diachronic, pragmatic, analytical, pedagogical, psychological, rational, empirical, practical, etc. We will view this choice between shallow and deep from a historical perspective first, as explained here by Coulmas (1989, pp. 230):

If anything can be learned from the history of extant alphabetic orthographies, it is that as they increase in age, they tend to move away from simple phonemic representation. This can be described as a historically conditioned deviation from the phonemic principle; but at the same time it also represents a process of accommodating in the written representation of a given language information from other systemic levels. In other words, thanks to the multi-level structure of

¹ J Baart 2003, pers. comm., 30 October.

natural languages, violating the prototypical alphabetical principle of using a single letter or combination of letters for one phoneme is not just destructive but also, in a sense, productive because deviations from, or shifts in, phoneme-grapheme correspondences do not happen randomly; more often than not they occur in a way that lets other structural units and patterns become more apparent ... Indeed, mature alphabetic orthographies encode morphological and lexical information in addition to phonemic information; and mature readers make use of this information more than they do of letter-sound correspondences.

These observations give rise to the question of the extent to which the alphabet maker should try to anticipate historical ‘corruptions’ of a strictly phonemic representation by incorporating lexical and morphological information from the start.

An attempt to anticipate a deep representation would seem to be a sensible approach, however there are hurdles: even after a very thorough morphological, syntagmatic and lexical analysis, as Coulmas points out, ‘the segmentation of words can pose considerable problems (cf., for example, Beck 1964; Gudschinsky 1970)’ because ‘the orthographic word is, partly at least, an artefact’. And yet, hypothetically, ‘the one who designs the orthography has to make non-arbitrary segmentations which will become orthographic words once the orthography is accepted and employed’. Thus, from a pragmatic point of view, Coulmas thinks that basing an orthography on premature interpretation may be too ambitious—firstly because the new orthography may contain arbitrary features that influence the speakers, and secondly because there is often a residue of segments and patterns that do not fit postulated models (p. 231). He concludes that ‘it seems wiser, easier and more economical at the initial stage to be content with a valid phonological transcription as the way to introduce a new orthography’.

Coulmas (1989, p. 232) notes also the pedagogical problem sometimes associated with a shallow representation, for example, the sometimes arbitrary and non-intuitive analysis of *phonemic* segmentation; and the fact that pedagogy involves both reading and writing processes, for which separate and sometimes conflicting issues apply. However, this point about the pedagogical problem of phonemic representation is not to be confused with the point made in the introduction to chapter 4, that pedagogical considerations assume a strong sense of phonemic awareness in literacy learners (which thus supports the concept of laying a phonemic foundation for the creation of a new orthography.)

Grimes and Gordon (1980, p. 93) also saw problems, from a pedagogical point of view, in a shallow representation:

There are two consequences of a phonemic orientation. One is the degree of inconsistency, greater in some languages than others, among the different representations of a single word or morpheme

... The other consequence of a phonemic orientation in writing is the proliferation of homographs: words that are lexically different, but are pronounced the same.

However, these considerations are more of a problem for a language like English than for a more inflected language like Kalasha.

Psychological considerations are certainly not the exclusive domain of lexical phonology, but are nonetheless a crucial concern of that theory. Mohanan (1986) steered a path between traditional structuralist-phonemic theory and generative theory (p. 7) by propounding the idea that the ‘mental lexicon’—being the powerhouse of speakers’ intuitions *that ultimately lie behind the phonologist’s choices* (p. 58)—is the rightful starting point for other levels of linguistic expression and representation. Therefore, a level of representation that is sensitive to the mental lexicon, that is, ‘lexical representation’ (pp. 10–12, 147), is the one that best captures morphophonemic processes; and a ‘syntactico-phonological’ representation will capture syntagmatic (‘postlexical’) processes (pp. 147, 185). Thus, from the mental lexicon can be derived a ‘lexical alphabet’ (pp. 173–5, 192–4) which, along with the intermediate levels of representation that make use of this alphabet, can form the basis of a practical orthography (pp. 195, 205).

Despite the fresh and continuing insights of lexical phonology and psycholinguistics, the question of what level of orthographic depth to aim for remains controversial. Shallow representation is strongly supported by three independent groups of researchers who contributed chapters to Frost & Katz (1992), each backing up their arguments with experimental evidence that word recognition usually and mainly relies on phonological cues provided via the alphabetic principle (Carello, Turvey & Lukatela; Perfetti, Zhang & Berent; Van Orden et al.). Katz and Frost (1992) suggest their own modification of the hypothesis by positing relative dependencies on prelexical (i.e. assembled) and lexical (i.e. stored) sources of phonological information, echoing Mohanan’s (1986) theory of lexical phonology. Furthermore, some of the above researchers, and others (e.g. Paap, Noel & Johansen; Colombo & Tabossi), subscribe to a ‘dual route theory’ within the orthographic depth hypothesis, which allows for lexical access based on dual routes of representation: phonological and lexical.

On the other hand, Besner and Smith (1992) present both rational and empirical evidence that a shallow representation cannot be relied upon as the only code for word recognition in the reading process, implying a preference for a deeper orthographic representation.

Moreover, Colombo and Tabossi (1992) demonstrate that their experimental subjects (36 Italian university students) did not normally choose to use a phonological strategy in reading.

From a practical perspective ‘the volume of material produced is also a factor in the optimal level of phonological abstraction represented,’ according to Mithun (1992, p. 143). She makes the point with reference to English, where the sheer volume of existing reading materials enables readers to have prodigious exposure to the spelling system. She points out that in such cases ‘assimilating the regular alternation rules necessary for interpreting a morphophonemic spelling is facilitated, because ample opportunity exists for readers to memorize and reinforce the spellings of words.’ However, for many minority languages, such as Kalasha, nowhere near as many opportunities for reinforcement will ever exist: ‘People will not begin the day by reading their language on cereal boxes and milk cartons.’ Mithun concludes that ‘a highly abstract system, necessitating a lengthy learning period, would have less chance of success in such communities’.

Finally, the constitution of the intended readership is also a factor in the orthographic depth debate. Where most are native speakers of the language there is a greater tolerance for a deep orthography because, from their own intuitive knowledge of the language, they can interpret aspects of phonology that might be otherwise masked by a deep orthography. A non-native readership would find a shallower orthography easier to understand and read (Mithun 1992, p. 143). In the case of Kalasha the vast majority of readers are expected to be native speakers, so a concession to shallow orthography is not important on these grounds.

Against the above theoretical background to the issues that relate morphology to orthography, section 5.2 of this chapter will examine, analyse and endeavour to explain the phonological processes behind various morphophonemic phenomena in Kalasha. This will help to provide an analytical resource for deciding on an appropriate level of representation, to understand what is happening morphologically, and to explore the orthographic implications, that is, how certain Kalasha words would be written at each level of representation. If a shallow method of representation is adopted the words would always be written phonemically, that is, to reflect the phonological process in each environment. Such representations will be transparent in each of the examples given. However, where alternative deep representations are possible, examples will be discussed in detail.

At the syntagmatic level, the orthographic representation of a syntagm should comprise the morphemes that occur in it. But what factors determine whether, when and how to join

morphemes, and whether, when and how to split them? The syntagmatic considerations embedded in this question will be dealt with in section 5.3 of this chapter.

We will examine all the above issues from two perspectives: the primary perspective will be the theoretical examination of morphophonemic and syntagmatic processes, and the secondary perspective will use empirical evidence of indigenous speakers' intuitive perceptions and renditions of morphemic and syntagmatic units. Let us therefore look at the empirical sources that will be used in this chapter, to provide either supportive or exceptional evidence for the theoretical arguments.

5.1.1 Description of empirical sources

The empirical evidence used in this chapter comes from four independent sources of data that I have collected, over a period of nearly twenty years, on the actual usage of, and preferences for, alternative representations of Kalasha morphemes and syntagms. These provide intuitive comments on alternative analyses of morphophonemic and syntagmatic processes that will be discussed in this chapter.

One source of data is a collection of 40 samples of correspondence from eight Kalasha correspondents over the eleven-year period between 1986 and 1997. The total corpus consists of about 10,000 words. The correspondents were all young educated Kalasha men who were experimenting with Arabic and/or Roman script in their personal correspondence with me.² In each case, their writing styles give the impression that there was no attempt to simplify their language to make allowances for the fact that I was not a mother-tongue speaker. Among other things, this implies that they were merely attempting to reflect orthographically what they perceived at the phonemic, morphemic and syntagmatic levels.

Following is a summary table setting out the authorship and dates of these letters:

² The term 'educated' in this thesis reflects the informal definition used in rural Pakistan, i.e., having passed exams at the end of junior high school. Instruction is usually given in Urdu medium, but English-medium instruction is also common in cities and some towns. The medium of instruction through which an individual has been educated significantly affects his/her approach to orthography in an unwritten language. The educated Kalasha are no exception when it comes to writing their own language. (Names have been substituted with their initials to preserve their anonymity.)

Table 5.1: *Authorship and tally of Kalasha correspondence received*

Correspondent	Total items	Distribution and occasions of correspondence											
		86	87	88	89	90	91	92	93	94	95	96	97
SK	1					1							
LK	12	3	4	3	1					1			
CK	2			1							1		
FK	19	1	2	3		1				1	3	5	3
IK	1		1										
SH	1			1									
MK	3		2	1									
DS	1		1										
TOTAL	40	4	10	9	1	2	0	0	0	2	4	5	3

The correspondents in the above table are listed in approximate order of descending age, (ranging from about 45 in 1997 down to about 15 in 1986). Five of them were later to attend the orthography conference, mentioned below as the third source of data. Their ages, educational backgrounds and occupations are included in Table 6.1 in the next chapter, which reports on that conference. Of the three correspondents who did not attend the conference, here is a brief profile of each, as at the time of writing their most recent letter, and the reason for their non-attendance:

- LK was a thirty-eight-year old children's hostel supervisor, who had studied at tertiary level (English medium). Some time prior to the conference he had moved from the province and was out of contact.
- SH was a 30-year-old postal delivery officer, who had studied at middle level (Urdu medium). During the time of the conference he had to attend to his work duties.
- DS was a 17-year-old secondary student (Urdu medium). Because the conference was held in the city during winter he had to stay back in his snowbound village to look after his aged parents. He subsequently became a tour guide, and trained as a teacher by correspondence.

A second source of data was a set of survey questionnaires, using both Arabic and Roman writing systems, which focused on alternative written forms of Kalasha words and phrases. The responses were elicited from a total of 25 educated people (including one of the very few educated Kalasha women at the time), representing the Kalasha communities from all three valleys, on four separate occasions between 1986 and 2001. The results were based on

preferences recorded in writing while respondents read (without reciting) between 3 and 18 pages of alternative forms of words, compounds and phrases, in their own choice of Arabic and/or Roman scripts. Most of them responded to the survey in both scripts.³ They indicated their preferences by marking (circling, underlining, etc.) their preferred choice of two or more alternative representations of certain words and phrases that had been carefully selected to demonstrate all the relevant issues.

A third source of empirical evidence was a set of decisions that was made at the orthography conference in Islamabad, attended by 20 Kalasha men (including teachers, students, elders and other interested individuals) from 30 December 2000 to 2 January 2001. The results of this conference are discussed in Chapters 6 and 7, but reference is also made to those results in this chapter, for the morphological insights that they provide.

A fourth source of data is the scores of letters I have received from 1998 to 2005, especially post-conference (i.e. from 2001 onwards), all of which reflect native writer intuitions, and some of which contain specific comments and preferences of orthographic interest.

5.1.2 Introduction to morphophonemic and syntagmatic phenomena in Kalasha

There are more than twenty forms of the Kalasha verb, reflecting not only the usual conjugations of person, number and tense, but also a very rich repertoire of nuances in aspect and mood.⁴ Each of these verb forms features a special (usually agglutinative) affix or auxiliary inflection.⁵ There are also nine grammatical cases for nouns, each of which features a suffixal inflection.⁶ This means that Kalasha verb and noun roots especially, are constantly

³ Sample pages of this orthography questionnaire appear as Appendix 10 (Arabic script) and Appendix 11 (Roman script).

⁴ Kalasha verb forms are many and various, falling into the following broad categories: **base** forms (root and infinitive), **tense** and **aspect** (present/future, present continuous, simple past, past hearsay, perfect participial, present perfect, past perfect, past continuous, imminent, inceptive), **mood** and **modality** (probable, imperative, optative, permissive, desiderative, obligatory), and **voice** (agentive, causative, purposive and passive) (Trail & Cooper 1999, pp. 480-6). Many verbs are also inflected according to six arbitrary verb classes (Cooper, G 1984a).

⁵ Examples of rarer fused inflections in Kalasha are /ita/, which is the participial form of /ik/, 'to come', and /t^{hi}/, which is the participial form of /hik/, 'to become'.

⁶ Examples of Kalasha noun and pronoun cases include nominative, genitive, dative, accusative, ablative, instrumental, locative, oblique and vocative (Trail, R 1996b, pp. 149-58). Gender is not significant in Kalasha, but animacy is: completely separate verbs are used to refer to the same action, depending on whether the object is animate or inanimate object.

being affixed with bound morphemes, in the form of an affix, a clitic, or an auxiliary constituent of a compound or contraction.

In addition to their morphophonemic influences on lexical roots, many of these bound morphemes also have allomorphs of their own, based on (a) phonetic assimilation, (b) grammatical class, or (c) semantic domain of the root to which they are attached. For example:

(a) A (verb or noun) suffix may have different pronunciations according to the final sounds of various roots (e.g. the diminutive suffix */-jak/* only follows the unaspirated voiceless dental plosive */t/* and the dental nasal */n/*, while */-ak/* follows any phoneme).

(b) The six Kalasha verb classes are characterized by variable inflectional paradigms, with shifting combinations of allomorphs. For example, the first person singular present suffix is unstressed */-im/* for Class 1 verbs, but stressed */-im/* for Class 4 verbs. It becomes */-am/* for Class 2 and 3 verbs and */-em/* for Causative 1 verbs.

(c) Even semantically related lexical items in Kalasha, as a class (say of nouns), can sometimes adopt irregular derivational paradigms peculiar to that semantic class (e.g. the special plural form */-an/* for a very limited set of respected and revered animate referents in the Kalasha worldview (e.g. */gaq̣erak'an/* ‘elders’, */paj'an/* ‘goats’ and */pari'an/* ‘fairies’).⁷

In the context of this very brief outline of Kalasha morphology, the remainder of this chapter is based on the chief question of determining whether, and how much, the representation of the Kalasha language should be influenced by various morphophonemic processes and syntagmatic issues. The morphophonemic processes are: devoicing of final obstruents (subsection 5.2.1), voicing of initial *[k]* in compounds (5.2.2), reduction of final consonant clusters (5.2.3), alternation of final approximants (5.2.4) and elision of intervocalic approximants (5.2.5). The syntagmatic issues are: morphemic compounding and separation (5.3.1), morphemic interpretation of clitics (5.3.2), and syntagmatic contraction (5.3.3).

⁷ Kalasha deities, though multiple, are not classified as a lexical set.

5.2 MORPHOPHONEMIC PROCESSES

5.2.1 Devoicing of final obstruents

Morpheme-final voiced obstruents in a few Kalasha morphemes are devoiced when those morphemes are pronounced in isolation. Despite the relative infrequency of their lexical occurrence (meaning that they appear in only a few words), this phenomenon is made more frequent due to their repeated use in certain common words in natural utterances. Some examples are:

[^h uguna] = [uk] + [una]	[masa ^h bi] = [ma ^h sap] + [i]
‘in the water’ ‘water’ ‘in’	‘religious’ ‘religion’ ‘of’
[ʃid ^h ik] = [ʃit] + [ik]	[a ^h wazas] = [a ^h was] + [as]
‘short ladder’ ‘ladder’ ‘-let’	‘sound’s’ ‘sound’ ‘of’
[tʃiʒ ^h g ^h as] = [tʃiʃ] + [k ^h as]	
‘a kind of grass’ = ‘garland’ ‘grass’ ⁸	

The analysis that is most likely to explain this phenomenon is that morpheme-final voiceless consonants are allophones of their voiced counterpart phonemes. For example:

/g/ = [g] non-word-finally, and [k] word-finally,
 /k/ = [k] anywhere,
 giving us /ug/ and /uguna/.

This explanation can easily be attributed to the complex phonetic process of the release of stops, which gives rise to at least a partial devoicing of voiced stops in many languages (Clark & Yallop 1995, p. 44–5). The fact that it is not a case of *utterance*-final devoicing is proved by examples such as [uk ^honi], ‘bring the water’ and [ʃit gri], ‘get the ladder’, where the devoicing is seen to be morpheme-final, not utterance-final.

Although the two allophones ([g] and [k]) of the voiced phoneme /g/ are in complementary distribution, there is a violation of the condition of *phonemic biuniqueness* (an ideal which requires that every sound in a particular environment be allocated to only one phoneme, that

⁸ It seems that in the last example, the kind of grass is named after the object for which it is primarily used, rather than vice versa.

is, that no two phonemes share a phonetic realization with overlapping distribution). In these distribution statements the sound [k], when occurring word-finally, is being interpreted as an allophone of both the phoneme /g/ and the phoneme /k/.

However, phonemic biuniqueness is not an absolute requirement in the same way as the complementary distribution of allophones (see Clark & Yallop 1995, pp. 112–13). Phonemes, by definition, contrast with each other in at least one truly analogous environment. Nevertheless, such a definition is still wide enough to allow that this contrast may be neutralized in another environment. Under this interpretation, it could be said that in our examples the contrast between the phonemes /g/ and /k/ is neutralized in word-final position. Though the occurrence of the sound [k] in this position is common to both phonemes, the occurrences do not overlap because the phonemes themselves are separate. So it would actually be more correct to say that the sound [k] is an allophone of either /g/ or /k/ (not both) in any given instance. Gudschinsky (1979, p. 15) concluded that morphemic representation is favourable where there is phonemic neutralization:

A useful rule of thumb would seem to be: when morphological distinctions which are important in the language are obscured by limitations on the distribution of phonemes—leading to neutralization of contrasts—the practical orthography should in general symbolize the underlying form of the morphemes rather than the phonemic form that is actually pronounced.

Thus, a deep representation of the vocabulary items quoted at the beginning of this subsection would be as follows:

[uk] /ug/ {ug} ‘water’	[^l uguna] /uguna/ {ug-una} ‘in the water’
[ma ^l sap] /masab/ {masab} ‘religion’	[masa ^l bi] /masabi/ {masab-i} ‘religious’
[ʃit] /ʃid/ {ʃid} ‘a log ladder’	[ʃid ^l ik] /ʃidik/ {ʃid-ik} ‘a short log ladder’
[a ^l was] /awaz/ {awaz} ‘sound’	[a ^l wazas] /awazas/ {awaz-as} ‘sound’s’
[tʃi] /tʃiʒ/ {tʃiʒ} ‘garland’	[tʃiʒ ^h as] /tʃiʒ ^h as/ {tʃiʒ ^h as} ‘a kind of grass’

The empirical evidence from the 1986–97 correspondence survey favours a deep (morphemic) representation of this phenomenon, with several correspondents spelling free (isolated) roots with morpheme-final voiced phonemes: [wa^lrek] as *wareg*, ‘other’, [aŋ^lgris]

as *an̩griz*, ‘Westerner’, [d̪ʒaˈhas] as *d̪ʒhaz*, ‘aeroplane’, [naˈras] as *naraz*, ‘angry’, [taˈʃ] as *taˈd̪ʒ* and [bek] as *beg* (both personal names). Although they could have been influenced by the Urdu spelling of all but the first of these words, there is contrary evidence to this claim in that all Kalasha writers spell the Kalasha word for ‘book’ [kitap] as *kitap*, not as *kitab*, which would reflect the Arabic/Farsi/Urdu spelling from where the word was taken. A diachronic examination of results from the orthography questionnaires of 1997 and 2001 also shows a majority preference for deep (morphemic) representation, writing voiced phonemes in morpheme-final position. In addition, an independent consensus among delegates attending the 2000/2001 Kalasha orthography conference also supported this method of deep (morphemic) representation. However, this specific issue has been one of the most controversial among the Kalasha since the conference, with consensus swinging back and forth many times. Anecdotal evidence received from three young literate speakers at the time of writing (2005) is indicating an intuitive preference for a swing back to a shallow (phonemic) representation of this phenomenon.

5.2.2 Voicing of initial [k] in compounds

The voiceless velar plosives ([k] and [k^h]), when occurring in the initial position of a few Kalasha words, are voiced ([g] and [g^h]) when preceded by another morpheme in a compound. Since there are so few Kalasha roots that take on morpheme-initial voicing, and even those roots only occur in a handful of particular compound constructions, the following list of examples is presented as being close to the sum of such words in the Kalasha dictionary (Trail and Cooper 1999), though some more may come to light over time:

[iˈstriʒa] + [ˈkuaːk] → [iˈstriʒaˈguaːk]

‘woman’ ‘child’ ‘girl’

[du] + [ˈkuaːk] → [duˈguaːk]

‘two’ ‘child’ ‘twins’

[krom] + [ˈkarik] + [-aw] → [kromgarˈaw]

‘work’ ‘to do’ ‘-er’ ‘worker’

$[\widehat{t}i\widehat{f}] + [k^has] \rightarrow [\widehat{t}i\widehat{z}^g^has]$
 ‘garland’ ‘grass’ ‘a kind of grass’

The following theory about this phenomenon seems to be the most likely explanation:

The restriction of this phenomenon to just a handful of cases suggests that it involves formations that have been fully lexicalized, and that the phonological process involved is no longer productive. An implication is that readers do not need to access the individual morphemes in their mental lexicon in order to decode the meaning of the word as a whole. Therefore there is less of a need to maintain a morphological spelling in these cases (J Baart, personal correspondence, 30 October 2003).

Empirical evidence that supports this theory is the practice of one Kalasha correspondent who spelt the bound root $\{-kar^1aw\}$ with the morpheme-initial voiced phoneme /g/: *-garaw* (‘doer’). The orthography questionnaires show a fourfold preference for a phonemic representation. A consensus among delegates attending the 2000/2001 Kalasha orthography conference also favoured a shallow representation (see Chapter 7, section 7.1.6).

5.2.3 Reduction of final consonant clusters

This phenomenon occurs only with nouns ending in consonant clusters whose first element is a nasal (/m/, /n/ or /ŋ/), or the voiceless dental plosive /t/, and whose second element matches the first by place of articulation. The consonant cluster is intact in a word-medial environment (when the morpheme is followed by a suffix), but when occurring in isolation the cluster is reduced by the elision of the second element, leaving only the first. For example:

$[han] + [-una] \rightarrow [^h\text{handuna}]$
 ‘temple’ ‘in’ ‘in the temple’

$[mon] + [-as] \rightarrow [^h\text{mondras}]$
 ‘word’ ‘s’ ‘word’s’

$[\widehat{t}^het] + [-aw] \rightarrow [\widehat{t}^hetraw]$
 ‘fields’ ‘from’ ‘from the fields’

$[dum] + [-el^i] \rightarrow [dum^h\text{bel}^i]$
 ‘lost’ + ‘s(h)e causes’ → ‘s(h)e loses’

[men] + [-aw] → ['mend̪ɔ̃aw]

‘cloud’ + ‘from’ → ‘from the clouds’

The best explanation seems to be that /han/, /mon/, /t̪sʰet/ and /d̪um/ and /men/ are the free allomorphs of the morphemes {hand}, {mond}, {t̪sʰetr} and {d̪umb} and {mend̪ɔ̃} respectively, and that we therefore represent them morphemically to account for these additional phonemes. So:

[han] /han/ {hand} ‘temple’ ['handani] /handani/ {hand-ani} ‘from the temple’

[mon] /mon/ {monɔ̃} ‘word’ ['mond̪ɔ̃ras] /mond̪ɔ̃ras/ {monɔ̃-as} ‘word’s’

[t̪sʰet] /t̪sʰet/ {t̪sʰetr} ‘field’ [t̪sʰetrani] /t̪sʰetrani/ {t̪sʰetr-ani} ‘from the field’

[d̪um] /d̪um/ {d̪umb} ‘lost’ [d̪umb'elʲ] /d̪umbel/ {d̪umb-el} ‘(s)he loses’

[men] /men/ {mend̪ɔ̃} ‘clouds’ ['mend̪ɔ̃aw] /mend̪ɔ̃aw/ {mend̪ɔ̃-aw} ‘from the clouds’

The reduced allomorphs reflect a process of weakening, leading to elision in isolated environments, rather than phonemic addition in the bound environments.

Empirical evidence that supports this interpretation is the practice of several Kalasha correspondents who spelt three free roots with the added phonemes: *putr* (‘son’), *pond̪ɔ̃* (‘five’) and *mend̪ɔ̃* (‘cloud’). A consensus among delegates attending the 2000/2001 Kalasha orthography conference also supported the interpretation of the orthographic addition of a single word-final phoneme (eliding the second phoneme where two occur, for example, {mond} instead of {monɔ̃} for ‘word’).

On the other hand, the orthography questionnaires actually showed a majority preference for a phonemic representation for this phenomenon. As with the issue of final devoicing, the level of representation of this phenomenon has also been subject to much controversy among new literates. Again, anecdotal evidence received from three young literate speakers at the time of writing (2005) is indicating an intuitive preference for a swing back to a shallow (phonemic) representation of this phenomenon.

5.2.4 Alternation of final approximants

Phonemic variance between /w/ and /ɬ/ in Kalasha occurs in a few morphological environments, such as at a morpheme boundary where /w/ is followed by a suffix beginning with a vowel. The following examples illustrate this morphophonemic alternation:

$$[\text{tʂ}^{\text{h}}\text{e}^{\text{h}}\text{tra}\text{ɬaj}] = [\text{tʂ}^{\text{h}}\text{e}^{\text{h}}\text{traw}] + [-\text{aj}]$$

‘in Chitral’ ‘Chitral’ ‘in’

$$[\text{bi}^{\text{h}}\text{ri}\text{ɬej}] \sim [\text{bi}^{\text{h}}\text{ri}\text{ej}] = [\text{bi}^{\text{h}}\text{riw}] + [-\text{ej}]$$

‘from Birir’ ‘Birir’ ‘from’

$$[\text{k}^{\text{h}}\text{aw}^{\text{h}}\text{saŋga}\text{ɬuna}] = [\text{k}^{\text{h}}\text{aw}^{\text{h}}\text{saŋgaw}] + [-\text{una}]$$

‘in the festival’ ‘festival’ ‘in’

$$[\text{hawa}\text{ɬaw}] = [\text{hawaw}] + [\text{haw}]$$

‘if it happened’ ‘happened’ ‘if’

This phenomenon is most likely to be the remnant of a historical reduction of the velarized lateral /ɬ/ to the labial-velar approximant /w/ preceding the high back vowel /u/, as also happened in French (Clark & Yallop 1995, p. 97). The phonemic representations of this small group of morphemes, would be:

$$/\text{tʂ}^{\text{h}}\text{etraw}/ \text{ ‘Chitral’} \quad / \text{tʂ}^{\text{h}}\text{etra}\text{ɬaj}/ \quad \{\text{tʂ}^{\text{h}}\text{etrawaj}\} \text{ ‘in Chitral’}$$

$$/\text{biriw}/ \text{ ‘Birir’} \quad / \text{biri}\text{ɬej}/ \quad \{\text{biriwe}\} \text{ ‘from Birir’}$$

$$/\text{k}^{\text{h}}\text{awsaŋgaw}/ \text{ ‘festival’} \quad / \text{k}^{\text{h}}\text{awsaŋga}\text{ɬuna}/ \quad \{\text{k}^{\text{h}}\text{awsaŋgawuna}\} \text{ ‘in the festival’}$$

$$/\text{hawaw}/ \text{ ‘happened’} \quad / \text{hawa}\text{ɬ haw}/ \quad \{\text{hawaw haw}\} \text{ ‘if it happened’}$$

Empirical support for this choice of representation was the consensus among delegates attending the 2000/2001 Kalasha orthography conference, although the correspondence data and orthography questionnaires showed no significant preference either way. A shallow representation that reflects the free variation of this phenomenon that occurs in actual speech would maintain the flexibility required to allow a shift to a standardized deep form at a later stage, if the community so desired.

5.2.5 Elision of intervocalic approximants

The following examples illustrate a type of morphophonemic elision of semivowels (/w/ and /j/), at morpheme boundaries preceding a vowel-initial suffix:

[p^haw] + [-una] → [p^hauna]

‘dirt’ ‘in’ ‘in the dirt’

[tʃusti] + [jar] → [tʃusti'ar]

‘alert’ ‘companion’ ‘Chistiar’⁹

[us'traw] + [uni] → [us'trauni]

‘bedding’ ‘instrument’ ‘bedspread’

Since the number of syllables is not reduced, this phenomenon would not be phonemically interpreted as a case of diphthongization. And since there is no duration difference when these contiguous vowels are intruded by the intervocalic approximants, this is interpreted as part of the phonotactic possibilities of the language. Clark and Yallop (1995, p. 74) mention this interpretation as a common explanation for this phenomenon. Here, the central approximants [w] and [j] become phonetically redundant when contiguous (across a morpheme boundary) to their vowel counterparts [u] and [i]. Therefore, we could postulate that:

[w] → [Ø] contiguous to [u]

[j] → [Ø] contiguous to [i]

giving us the alternative interpretations: /p^hauna/, /us'trauni/ and tʃusti'ar/. (For the same reason, even the independent morpheme [ha'wul^j], meaning ‘lightly’, could be interpreted as /ha'u1/.)

Because these are representations of the morphemes {phaw-una}, {ustraw-uni}, {custi-jar} and {hawul}, we could preserve these morphemic representations by retaining the approximants /w/ and /j/ when contiguous (across any morpheme boundary) to their vowel counterparts /u/ and /i/ respectively. However, respondents to the orthography questionnaire showed a slight preference in the phonological direction. This is probably because

⁹ *Chistiar* is a personal name.

intervocalic elision was only featured in the Arabic-script version of the questionnaires, and elision of one of these segments would be standard practice in the Arabic script, because in that script both central approximants, /w/ and /j/, are represented by the same symbols as their respective vowel counterparts, /u/ and /i/, and so doubling the symbol seems redundant. Since Roman script has different symbols for the central approximants and their respective vowel counterparts it would be clearer to preserve both, leading to the following deep representations: /phawuna/, /ustrawuni/, /custijar/ and /hawul/.

To summarize our findings on the most appropriate methods of representing various morphological phenomena in Kalasha, we found support for a choice of shallow (phonemic) representation in cases of voicing of initial /k/ in the second element of compounds (e.g. {kua·k} and {dugua·k}) and of the alternation of final approximants (e.g. {biriu} and {birilej}). In addition, we found inconclusive evidence for a choice of the level of representation in cases of devoicing of final obstruents (e.g. {uk}~{ug}), reduction of final consonant clusters (e.g. {put}~{putr}) and elision of intervocalic approximants (e.g. {p^hauna}~{p^hawuna}). To put this discussion into a proper perspective it is necessary to point out that each of the above morphophonemic phenomena only involves a very small group of words.

5.3 SYNTAGMATIC ISSUES

5.3.1 Compounding and separation

Compounding and separation are syntagmatic issues because, though represented at the orthographic level, word boundaries are interpreted on morphemic and syntagmemic levels, as markers of free (as opposed to bound) morphemes.

One of the many aspects of written English that is quite varied in practice is that of alternatively writing certain groups of words separately, hyphenated, or compounded. For example, *e-mail* and *in-flight* can be written without the hyphens, and *no-one*, *no-show* and *North-West* can be written as separate words.

Though writing conventions have hardly been established in Kalasha, certain groups of potential compounds present some challenges of splitting or joining. The following criteria,

from the set proposed by Pike (1947, p. 167) for the interpretation of utterances as compounds, give some guidelines. Potential Kalasha compounds are quoted within the list, as examples of each criterion, prefaced by English examples:

(a) Special arrangements of stress patterns. For example, English foregrounds the **qualifying** (adjectival or adverbial) elements of compounds by placing stress on those qualifying elements, as in the following examples (where the stressed syllable is preceded by a prime mark): *'blackboard*, *'email*, *'fireplace*, *'inflight* (not **black'board*, **e'mail*, **fire'place*, **in'flight*). Kalasha, on the other hand, places stress on the **qualified** element of compounds, as in:

/ra/ + /mut/ → /ra'mut/
 'cedar' 'tree' 'cedar tree'
 /ka'ɭaʃa/ + /mondɾ/ → /kaɭaʃa'mondɾ/
 'Kalasha' 'language' 'Kalasha language'

(b) Special phonological changes. For example, in English, 'four-by-two' is changed thus: /fɔ:/ + /baɪ/ + /tu/ → /fɔbi'tu/; one Kalasha compound changes phonologically by the process of morphophonemic substitution, explained in subsection 5.2.3 above:

/hawaw/ + /haw/ → /hawaɭaw/
 'happened' 'if' 'if it happened'

(c) Special [fixed] orders in which the morphemes occur. For example, in English 'black-and-white', 'four-by-two', 'Asia-Pacific' (not **white-and-black*, **two-by-four*, **Pacific-Asia*). In Kalasha:

/daʃ/ + /ʒe/ + /tre/ + /-a/ → /daʃʒe'trea/ (not **treʒe'daʃ* 'thir-teen')
 'ten' 'and' 'three' '-teen' 'thirteen' (lit., 'ten-and-three-teen')
 /peʃ/ + /gaŋga'i/ + /-ak/ → /peʃ-gaŋga'iak/ (not **gaŋga'iak'peʃ* 'goat-bell manure')
 'manure' 'goat-bell' (dim.) 'goat droppings' (lit., 'manure-goat-bells-mini')

(d) Inflection of compounds. For example, in English, 'lifesavers' (not **lives-savers*), and 'in-laws', but not **mothers-in-laws*. In Kalasha:

/ka'laʃa/ + /-on/ + /mondr/ + /as/ → /ka'laʃa'mondras/ (not /*ka'laʃon'mondras/)
 'Kalasha' 's' 'language' 's' 'Kalasha language's' (not /*'Kalasha's language's')

(e) The indivisibility of words by inflectional elements normally permissible for related sequences of free words. For example, in English, 'blackboard', but not '*blackboard'; and 'thoroughbred', but not '*thoroughbred'. In Kalasha:

{kanda} + {-an} + {d̪ʒaw} → {kanda d̪ʒaw} (not /*kanda and d̪ʒaw})
 'almond trees' 's' 'stand' 'almond tree stand' (not /*'almond trees' stand').

(f) The omission of certain words normally expected in a related sequence of free words. For example, in English, 'cutthroat' (not '*cut the throat'), 'toothpaste' (not '*teeth cleaning paste') and 'website' (not '*World Wide Website'). In Kalasha: /'keram/ as well as /keram'boʈ/ 'shuffleboard'.

What is not apparent from the above examples is that compared to English, Kalasha has relatively few words which could be interpreted as compounds. While the above examples of English compounds represent only a minute proportion of the total number of English compounds, the Kalasha examples would represent a significant proportion of all the compounds in Kalasha. Because they fit the above criteria, all of the Kalasha expressions listed above should be interpreted as compounds.

The rarity of Kalasha compounds makes them less problematic from a literacy point of view. As in many languages, including English, their status is likely to be in constant flux, and they are theoretically more difficult to write unambiguously in Arabic script. Arabic scripts have no hyphenation, nor even significant word spacing—only word-final and word-initial forms to mark morpheme boundaries (as explained further in Chapter 3, section 3.2, and Chapter 7, subsection 7.1.6). Yet, contrary to expectation, the Arabic-script questionnaire showed a preference for representing compounds by joining morphemes in this way, rather than splitting them with a word space. Other empirical evidence that supports this interpretation of compounding is the practice of a Kalasha correspondent who spelt one phonological compound twice without spacing and without final and initial forms: *kromgaraw* ('worker').

It was the Roman-script questionnaire that showed the opposite trend: a distinct preference for splitting, rather than joining. For this reason, though compounds obviously occur in Kalasha, they may or may not be represented as such, for reasons of visual simplicity. This is

discussed further in Chapter 7 (subsection 7.1.5A in relation to the possibility of hyphenation, and subsection 7.1.6 in relation to orthographic space and the setting of compounds).

5.3.2 Interpretation of clitics

The following phrases from sample texts illustrate morphemes that are both phonologically and grammatically bound:

[**'pari**w**ori**]

/**'pari**u **ori**/

's/he will go' 'let'

'Let him/her go.'

[**be**a'**d̥**ʒat]

/**be** a'**d̥**ʒat/

'without' 'necessity'

'unnecessary'

[**na**be'**hel**]

/**na** be'**hel**/

'un-' 'blessed'

'cursed'

These morphemes are not grammatically free because their syntactical positions are always fixed. They are therefore to be interpreted as prefixes or suffixes (marked here with a hyphen), as follows:

/-ori/ permissive: 'let' (always follows a verb)

/be-/ ~ /bej-/ negative: 'un-' (always precedes an adjective)

/na-/ negative: 'un-' (always precedes an adjective)

However, certain Kalasha morphemes pose difficulties of interpretation, because they are neither essentially bound forms (affixes), nor essentially free forms (independent words), but rather forms that are phonologically bound but grammatically free (clitics).

Clitics are phonologically bound because:

- (a) they never occur in isolation
- (b) they are never stressed

However, they are grammatically free because:

- (a) free morphemes can occur between clitics and the words they qualify (see set A below).
- (b) they qualify words from several grammatical classes (see sets B and C below).
- (c) they can occur almost anywhere in the sentence (see set C below).
- (d) Their semantic scope can even be wider than word level (see set D below).

Some problematic Kalasha morphemes are featured in the following sets of illustrative sentences. Until they have been interpreted through application of the above phonological and grammatical criteria for clitics, they will be written separately here (like words) for the sake of discussion, but with bold typeface to make their identification and positions clear. Following these sets of sentences, an interpretation of their respective statuses (as affixes, words or clitics) will be presented.

A. This first set of sentences illustrates purely syntactic (positional) variation of the segment /ne/, meaning ‘not’. Note that in the first pair of sentences the underlined verb with its inflected auxiliary /'praʃmi—'asam/ can be interrupted by the insertion of the morphemes /o/ and/or /ne/, and that in the second pair of sentences the verb with auxiliary /par'im daj/ can also be interrupted by the insertion of the morpheme /ne/.

/mimi **ne** praʃmi asam/ ~ /mimi praʃmi **o** **ne** asam/

‘you’ ‘not’ ‘forgetting’ ‘I’m’ ‘you’ ‘forgetting’ (definitive) ‘not’ ‘I’m’
 ‘I definitely have not forgotten you.’

/ne parim **daj**/ ~ /parim ne **daj**/

‘not’ ‘I-go’ ‘-ing’ ‘I-go’ ‘not’ ‘-ing’
 ‘I’m not going.’

/**daj**/ is a continuative particle.

B. The next sentence illustrates syntactic (positional) variation of the segments /ne/ ('not') and /o/ ('definitive'), purely in the changing focus of semantic reference (indicated by capitalization in the free translation). In this sentence /o/ qualifies the subject in the first clause, and it qualifies the negative particle in the second clause.

/se **o** krom **ne** kariu, **ne** **o** pariu/
 's/he' (definitive) 'work' 'not' 's/he-does' 'nor' (definitive) 's/he-will-go'
 'S/HE will not work, NOR EVEN will s/he go.'

/o/ is a definitive particle and /ne/ is a negative particle.

C. The next set of sentences illustrates both syntactic (positional) variation of the segments /mi/ (exclusive), /ta/ (contrastive) and /asta/ (inclusive), as well as variations in the focus of semantic reference (indicated by capitalization). To save space, the morpheme glosses have been included at the end of each set:

/tu **mi** maj ne paʃi ai/ 'Only YOU didn't see me'
 /tu maj **mi** ne paʃi ai/ 'You only didn't see ME'
 /tu maj ne **mi** paʃi ai/ 'You did NOT EVEN see me'
 /tu maj paʃi **mi** ne ai/ 'You didn't even SEE me'
 'you' 'to-me' 'not' 'seeing' 'not' 'you-were'

/mi/ is an exclusive particle.

/tu **ta** maj ne paʃi aas/ 'YOU have not seen me.'
 /tu maj **ta** ne paʃi aas/ 'You have not seen ME.'
 /tu maj ne **ta** paʃi aas/ 'You have NEITHER SEEN me.'
 /tu maj paʃi **ta** ne aas/ 'You have not SEEN me.'
 'you' 'to-me' 'not' 'seeing' 'not' 'you-are'

/ta/ is a contrastive particle.

/a **asta** taj ne dʒ^honim/ 'I ALSO don't know you.'
 /a taj **asta** ne dʒ^honim/ 'I don't know YOU either.'

/a taj dʒ^honim **asta** ne/ ‘I don't even KNOW you.’
 ‘I’ ‘to-you’ ‘not’ ‘I know’ ‘not’

/**asta**/ is an inclusive particle.

D. The semantic scope of clitics can even be wider than word-level, as the following examples show. In this first example /**mi**/ (exclusive) refers to a whole noun phrase, and in the next two examples, /**o**/ (definitive) and /**ta**/ (inclusive) both refer to whole clauses:

/tasa durej motʃ **mi**/
 ‘his/her’ ‘of-household’ ‘person’ (excl.)
 ‘only people from his/her household’

/ʃẽhẽ g^hõi **o** k^hẽ karim/
 ‘thus’ ‘saying’ (def.) ‘how’ ‘I-will-do’
 ‘So saying (speaking thus), what can I do?’

/maj paj ne aan g^hõi **ta** tu sahi a'mai/
 ‘my’ ‘goats’ ‘not’ ‘they-are’ ‘saying’ (contrastive) ‘you’ ‘true’ ‘you-said’
 ‘On the other hand, you are right when you say you don't have goats.’

All the morphemes exemplified above satisfy the phonological and grammatical criteria for clitics, indicated at the beginning of this subsection, and are therefore to be interpreted as clitics. I have duplicated them immediately below, using an equals sign (=) here to indicate their word-level positions. They are all enclitics, rather than proclitics:

/=daj/ continuative: ‘-ing’

/=mi/ exclusive: ‘only’; ‘even’

/=asta/ inclusive: ‘also’; ‘even’

/=bata/ repetitive: ‘again’

/=ki/ conjunctive: ‘that’

/=ta/ contrastive: ‘now, on the one hand’ / ‘neither’

/=o/ definitive: ‘the’; ‘this’; ‘but on the other hand’; ‘the previously mentioned’; ‘even’; ‘so’

The above conclusions on the attachment of potential enclitics are supported by data from correspondence, from seven letters written by three Kalasha men between 1989 and 1995, which yielded the results shown in the following table. Here, the lack of an *em* rule (as in the first of each pair) indicates their perception of the morpheme as a word, written separately from associated words, and an *em* rule (as in the second of each pair) indicates their perception of the morpheme as an affix, joined to associated words.

Table 5.3: *Treatment of potential clitics in Kalasha correspondence*

Morpheme	Number of associated words	Number of Occurrences	Number of Correspondents
<i>o</i>	29	49	3
= <i>o</i>	9	20	3
<i>daj</i>	24	40	3
= <i>daj</i>	0	0	0
<i>asta</i>	27	30	3
= <i>asta</i>	3	5	2
<i>ki</i>	1	12	1
= <i>ki</i>	1	2	1
<i>mi</i>	7	12	1
= <i>mi</i>	1	1	1
<i>ori</i>	3	5	1
= <i>ori</i>	4	5	2
<i>ta</i>	3	3	1
= <i>ta</i>	1	1	1
<i>bata</i>	1	1	1
= <i>bata</i>	0	0	0
<i>na</i>	0	0	0
<i>na=</i>	1	2	1
<i>be</i>	1	1	1
<i>be=</i>	0	0	0

The overall trend in the above table is to write these segments separately, rather than joined, which suggests that clitics should generally be represented **separately**. However, the table also shows that *=ori* and *na=* were slightly more often joined in writing. This suggests that they were perceived as affixes, and that this **joining** should be the normal orthographic practice for affixes. In the Arabic-script survey /*=ori*/ may have been written separately to avoid the unusual situation in Kalasha of adding on to an already inflected verb stem, since all other verb suffixes add on to the uninflected root.

In view of the fact there was so little data for *be=*, and the fact that it is so similar to *na=* I propose that it also be interpreted and written as a prefix, rather than an enclitic, as suggested in the discussion above. The fact that /*be=*/ was written separately even once in the Kalasha correspondence could be mirroring the orthographic convention for a similar morpheme in Urdu /*be=*/ (from where it is derived).

Furthermore, apart from a few exceptions, the *joining* of enclitics to independent morphemes was generally reserved in the correspondence for those words whose usage is more frequent with these enclitics than without them, for example:

bakio, {*baki=o*}, ‘furthermore’

kajo, {*kaj=o*}, ‘and when?’

fatalako, {*fatalak=o*}, ‘therefore’

kajasta, {*kaj=asta*}, ‘whenever’

falejo, {*falej=o*}, ‘and so’

tŋopo, {*tŋopa=o*}, ‘tomorrow’

djapata, {*djapa=ta*}, ‘otherwise’

koki, {*ko=ki*}, ‘because’

The only borderline case was *warego*, {*wareg=o*}, ‘and furthermore’, which also occurred a couple of times as *wareg o*, {*wareg o*}. These **joined** forms could remain variable in orthographic representation.

5.3.3 Contraction

Some examples of English contractions are *can't*, *let's*, and *nor'-wester*. The conventions for writing most English contractions (using the apostrophe) are more rigid than those for the writing of compounds generally, and the introduction of new ones is rare.

Kalasha contractions are relatively very few in number, though occurring in frequently used expressions, and only outside Birir Valley (where a possibly older dialect is preserving the longer forms), for example:

/ˈmatrik/ – /tr/ → /ˈmaik/ ‘to speak’

/ˈasik/ – /s/ → /ˈaik/ ‘to be’

/ˈabi/ – /b/ → /ˈai/ ‘we/you (pl.)’

These contractions may simply be represented as variants without any need for any orthographic elision marks (such as the apostrophe in European languages).

However, contractions like the following, found in all Kalasha dialects, are a different case:

/ˈhawaw#haw/ – /w#h/ + /t/ → /ˈhawat̪aw/ ‘if it happened’

Empirical evidence that this expression is understood as two words can be found in the practice of several Kalasha correspondents who wrote these expressions (in Arabic script) each time with word spacing, and therefore also with final and initial forms), for example, *hawaw haw*. However, the orthography questionnaire showed an equal result for *hawaw haw* and *hawat̪aw*. This case is discussed further in Chapter 7 (subsection 7.1.5A), but the morphophonemic representation of this phrase, like that of other contractions discussed above, could also remain in free variation (just as *can't* freely varies with *cannot* in English), though with no need for any elision mark.

To summarize our findings on syntagmatic phenomena in Kalasha, we found that, in order to reflect perceived linguistic units, compounds and contractions would be best represented without word or morpheme boundaries (and without an elision mark for contractions). But most Kalasha clitics should be represented separately, as free morphemes, in order to reflect their high degree of grammatical independence.

5.4 CONCLUSION

The Kalasha language presents a set of morphophonemic and syntagmatic issues potentially affecting its orthography. Each of these issues was taken on its own merits and examined accordingly, to determine, in each case, the best method of representation in preparation for the orthography. The conclusions varied according to the particular morphophonemic or syntagmatic phenomenon in question. In general, deep forms preserve underlying lexical and grammatical information better than shallow forms (which merely preserve phonological information). They are also supported by some of the theoretical and experimental findings relating to psycholinguistic perception and pedagogical processes, which were discussed in the introduction to this chapter. Any intervening shallow representations which Kalasha literates may introduce over time, especially those that reflect cases of free variation, would still maintain the flexibility required to allow a shift to the deep form at a more mature stage of the orthography, if the community wanted to standardize in that direction. This is likely to happen anyway, according to the theory of orthographic evolution, referred to in section 5.1 of this chapter (Coulmas 1989, p. 230). For example, ‘in English, the failure to adjust spelling to phonetic usage has allowed the writing system to change from a largely phonemic one to a morphophonemic one (Hill 1967)’ Coulmas (1992, p. 255).

Having said this, we must remember that attempts to rigidly prescribe any language behaviour on a speech community (including writing conventions) will not be able to stand against the force of natural evolution as many indigenous readers and writers continually use the orthography over time.

THE FIRST KALASHA ORTHOGRAPHY CONFERENCE

6.1 INDIGENOUS INTEREST IN KALASHA ORTHOGRAPHY

No one can be sure just how many Kalasha individuals, in times past, had ever imagined or tried writing down their language, without reference to any other person's attempt to do the same. Most likely, the thought of any such endeavour would only occur to one who had at least an elementary level of education.

It was as recently as the early 1970s that a handful of adventurous young Kalasha boys became pioneers among their community in the pursuit of education (two days' journey away in the exotic city of Lahore, capital of Pakistan's Punjab province). However, far from being honoured, they were looked down on as shirkers, who were avoiding their traditional roles as shepherds. Despite those early attitudes, in the relatively short space of time since then, circumstances and attitudes have changed rapidly, so that by the year 2000 there were about five hundreds Kalasha people (girls and boys equally, and even a few adult women) attending five Kalasha schools in three Kalasha Valleys. This is a significant increase over the estimated 200 pupils just 5 years earlier.¹

Urdu has always been the language of education in the vast majority of schools throughout Pakistan. Consequently, the natural choice of method for writing Kalasha was to use the familiar Arabic script, using a modified Urdu alphabet.² Over the last couple of decades, a few isolated and experimental attempts, using a modified Arabic script, had been made by certain educated Kalasha (e.g. Abdul Khaliq and Mir Rahim Khan).

Chapter 1 (section 1.4) detailed my own early experimental research and development of Kalasha orthography, in both Arabic and Roman scripts, in the 1980s and 1990s. Texts written using Arabic script were both written and read by Kalasha people. Texts written using

¹ At the time of writing (2005) there were nearly six hundred students attending Kalasha schools.

² The term *Arabic script*, as explained in footnote 2 of Chapter 3, means any writing system based on, and generally resembling, that system used to write the Arabic language. The Arabic script has been adapted in style, alphabetic characters and conventions to represent many other languages around the world, including Urdu. The most popular Urdu script style bears a closer resemblance to the traditional Farsi (Persian) style of Arabic script (*Nastaliq*) than to the purely Arabic style (*Naqsh*). *Nastaliq* is also preferred by the Kalami community, in Swat District which neighbours Chitral District where the Kalasha live (Baart 1996, p. 6). Because of the geographical proximity of Urdu usage to Kalasha, the experimental Arabic-based Kalasha script has also used the *Nastaliq* style, rather than the *Naqsh* style.

a Roman alphabet were used for most linguistic research, analysis, text processing and a lexical database. The *Kalasha Dictionary* (Trail & Cooper 1999), which is the publication of that database, uses both Arabic and Roman writing systems.³ A Roman writing system was used in all these research situations for reasons of both its familiarity (to the researchers) and its ease of use in computational software. Occasionally, texts written in the Roman writing system were used in literacy surveys that I conducted among the Kalasha community, though they were usually unpopular, due to the unfamiliarity of Kalasha individuals with the Roman writing system, and also to negative connotations from exposure to the irregular use of the Roman alphabet in the English language.

Since the early 1990s a few Kalasha students started to study in tertiary institutions (in nearby towns of Drosh and Chitral, and in the provincial capital, Peshawar) where English is the medium of education. The idea of using a Roman-based alphabet occurred to a couple of them as possibly a better medium for writing their language. Again, this was seen as a foreign concept and was not viewed seriously, though one particular man, Injinier Khan, the first Kalasha primary school teacher, continued privately experimenting with a European-based alphabet.⁴

In mid 2000, when Injinier Khan realized that the *Kalasha Dictionary* was a comprehensive and consistent version of what he had also envisioned, Kalasha orthography became a hot topic in an unprecedented way. He and Taj Khan Kalash Sharakat (a student in the 1980s and 1990s) pursued the issue vigorously, initially in competition, but latterly in cooperation, seeking to establish an orthographic foundation for Kalasha literature. It was a result of their initiative, sparked by the interest, enthusiasm and remote help of a Hungarian researcher Anna Haraszti, that several meetings of teachers and elders were convened during the (northern) summer of 2000, in the Kalasha Valleys, to discuss this very issue of the choice

³ In this thesis, the phrase *Roman script* has been avoided in some contexts, to avoid confusion with the *Roman* style of typefaces (e.g. *Times New Roman*) and/or handwritten scripts, neither of which is meant here. Instead, the phrases *Roman writing system* and *Roman alphabet* will more commonly be employed because, besides its left-to-right direction, it is the design and nature of the individual letters that is more in focus. In this sense, it is to be seen as a parallel system to the Arabic script. There are also features of Roman alphabet writing systems that do not pertain to individual alphabetic letters (e.g. diacritics, superscriptions, hyphenation and other orthographic conventions). These are features, not of any actual *Roman alphabet*, but of the *Roman writing system* as a whole. The term *script* will still sometimes be used to denote writing systems generally (including Roman ones, and especially when contrasted with Arabic scripts).

⁴ A copy of Injinier Khan's alphabet chart was finally sent to me by Anna Haraszti on 6 November 2001, and it appears in Appendix 9. It contains 50 to 52 symbols, including 5 oral vowels with their rhotic, nasalized and rhotic/nasalized counterparts, some normal and retroflex consonants, but no aspirated consonants. It appears to have had some Greek influence, but in many ways it fits the description of the original Balti alphabet, described in Chapter 3 (subsection 3.4.1). That was a 1400s invention to maintain the non-Islamic identity of the Baltis, but it has since been abandoned by them and adopted by speakers of Brokskat (a language closely related to Kalasha) for similar reasons of religious independence (see Chapter 3, subsection 3.4.2).

of script for the Kalasha language. At those meetings, a preference for a Roman alphabet was clearly indicated by verbal consensus. This was confirmed in another meeting in October 2000, at which an invitation was then made for me to come and provide technical input: to familiarize them with the implications of such a choice and to help them implement it.

6.2 THE CONFERENCE

As a result of the development of indigenous interest in Kalasha orthography, as outlined in the previous section (6.1), and to address the issues that had been raised, a special conference was held from 30 December 2000 to 2 January 2001. It was held in Islamabad, and was attended by 21 Kalasha people (most of them teachers, students and elders).⁵ Table 6.1, following, gives a brief and approximate profile of those who attended.

⁵ The fact that no women attended reflects cultural gender differences in both decision-making roles and mobility.

Table 6.1: Profile of indigenous participants in the Kalasha Orthography Conference

NAME ⁶	AGE in 2000	EDUCATIONAL LEVEL in 2000	EDUCATIONAL MEDIUM	OCCUPATION
UB	51	none	n/a	civil works contractor
AK	50	middle	Urdu	hotel owner & entrepreneur
MJ	43	middle	Urdu	civil works contractor & hotel owner
SK	42	middle	Urdu	Contractor
AU	39	tertiary	Urdu & English	Teacher
DM	39	tertiary	Urdu & English	Teacher
EK	36	tertiary	Urdu & English	Teacher
CK	35	secondary	Urdu	high school teachers' aide
TK	35	secondary	Urdu	border police constable
WK	35	tertiary	Urdu & English	Teacher
FK	31	secondary	Urdu	hotel owner & interpreter
IK	31	Middle	Urdu	tourist guide
MRK	29	Tertiary	Urdu & English	teacher
MG	27	Primary	Urdu	watchman & cook
SA	25	Tertiary	Urdu & English	teacher
SS	24	Primary	Urdu	tourist guide
IKB	20	Tertiary	Urdu & English	student
TKKS	20	Tertiary	English	student
ZK	18	secondary	Urdu	student
FA	15	Middle	Urdu	student
S ⁷	10	Primary	Urdu	student

This table demonstrates the cross-section of Kalasha society represented in the delegation, in terms of age, educational level (and medium of instruction), and occupation. Not shown in the table is an equally wide variety, within this group, in terms of social status, socio-political influence, dialect, residency (representing the three main Kalasha valleys, as well as the town and city dwellers) and level of contact with the West. The participants also represented at least nine different clans, which represents about one-third of the total number of clans in Kalasha society. Though many of the conference participants bear the surname *Khan*, which

⁶ Names have been substituted with their initials to preserve their anonymity. According to Kalasha tradition, these delegates are listed in descending age order, rather than in alphabetical order.

⁷ S is a son of one of the teachers. He was present in all sessions, attentive to all discussions, and offered his responses privately.

is a very common name in the North West Frontier Province of Pakistan, none of them is closely related to any of the others.

AK and UB are recognized elders of the Kalasha community in two of the Kalasha valleys. Despite their different educational backgrounds, their travel to, and presence at, the conference was a token of their encouragement and support for Kalasha literacy. Furthermore, their active and committed participation suggested that they believed in the potential usefulness and relevance of a formal writing system for the whole Kalasha society, not just for the younger generation or educated individuals (teachers and students).⁸

Almost all of those who attended the conference had studied through Urdu medium, which would normally be expected to lead to a preference for Arabic script. Though most Pakistanis learn English as a subject at school, relatively few become fluent, because of lack of practice. However, IKB and TKKS were educated in English-medium schools (in the cities of Lahore and Rawalpindi respectively), and are therefore proficient in English. AK and EK are also very fluent in English. EK's tertiary studies were in an English-medium college in the nearby town of Chitral, and AK used English in his employment with aid organizations, and as a hotel owner, with many foreign tourists.

The venue of the national capital, Islamabad, was chosen because the only reliable way in and out of their heavily snowbound mountain district is a day and night trip through Afghanistan, which is closed for transit of non-locals like myself. Video, audio and written records were made of the conference.

As leader of the conference I first presented several seminars on relevant topics of phonology and grammar, heavily illustrated with Kalasha examples. Despite the volume, theoretical nature and novelty of this linguistic information, intense interest and attention was maintained by all. Later unsolicited feedback proved that these sessions were the highlights of the conference, giving the Kalasha participants a rare insight into how language works, particularly their own.

Of all the alternative writing systems discussed in chapter 3, the two most likely to represent Kalasha are the Arabic and the Roman. For this reason, I tabled the equal consideration of both as the main agenda for the conference. So, rather than just concentrating on the Roman writing system (their preference as reported to me a few months earlier) I presented a thorough overview of both writing systems, outlining and illustrating the

⁸ This is by no means to claim that the whole society saw or sees value in a writing system or mother-tongue literacy. Many are interested, but many are indifferent.

implications of each, in relation to a host of different factors, with the aid of parallel examples, both printed on handouts and written up on a whiteboard.⁹ These factors, as they relate to Kalasha, are explained in detail below, forming the major part of the subject matter of this chapter.

Frequent interjections were made, and much lively discussion ensued in every session. In this respect, the conference was more like a workshop, with many questions and contributions collectively aimed at the common goal of determining a script and establishing orthographic conventions for Kalasha. Despite much lively and sometimes heated debate, the spirit of the conference was one of working together and harmony. Participants were not afraid to disagree, but when they did agree common assent was usually expressed by oral consensus.

Following the series of sessions in which the script issue had been presented and discussed at length, a vote, in the form of a secret ballot, was cast by all the participants to formally choose a script option for their language as an agreed foundation for further literacy development. Blank ballot papers and pens were distributed to all Kalasha people present, and ample time was given for voters to record their preferences. During this session due care was taken to avoid all interpersonal communication and observation that might have influenced voters one way or the other. Only occasional instructions were spoken. After a consensus was reached that vote collection should begin, a box was carried around, into which participants were asked to throw their folded ballot papers. The box was then carried to the front of the meeting and the votes were counted by two people, with one scrutineer observing. As each ballot paper was opened, a running tally was recorded on the whiteboard for all to see. Fifteen votes had been collected. All fifteen were unanimous in their choice of Roman script. Possible explanations for a less than maximum representation of all participants are that the least literate participants lacked the confidence to vote, and may have abstained for other reasons. The Kalasha vote administrator did not seem to cast a vote himself, yet considering his role in convening and coordinating the conference, and promoting Roman script, it is reasonable to assume that his vote would also have been for Roman script. Only one of the fifteen votes was for both Roman and Urdu scripts, as concurrent alternatives, so that the choice could be available to the individual, according to his or her own literacy skills. There was no vote for solely Arabic script.

The presentation and discussion of the result of that vote is featured in the conclusion to this chapter. After the choice of script was finally agreed, a raft of further subsidiary decisions

⁹ Several demonstration texts in both scripts were also presented.

had to be made during the second half of the conference, which relate to the specific implementation of that script. These include the issues relating to morphophonemic and syntagmatic considerations, as discussed in Chapter 5. The resultant decisions are documented in Chapter 7.

6.3 FACTORS OF CHOICE BETWEEN ARABIC AND ROMAN WRITING SYSTEMS

The major motivations for the adoption and adaptation of existing scripts for languages in general were presented in Chapter 3, divided into the major categories of phonological, religious, sociological, pedagogical and political. Here in this chapter, where the choice has been narrowed to only the Arabic and Roman options for Kalasha orthography, and in the Kalasha situation, we will consider the factors of choice between these two writing systems, in two broad categories: sociopolitical (including educational and pedagogical) and phonological. As stated in the previous section (6.2), all the subsumed factors discussed in the remainder of this chapter were systematically presented to the conference, preceding the formal vote for a script.

The significance of each of these factors is first briefly stated below in general terms, and then the way in which each script handles each factor is contrasted, with specific reference to the Kalasha situation. However, it must be noted that a mere count of these factors in favour of one choice or the other would be an invalid method of deciding between them, as each one is weighted differently. Moreover, due to the subjective nature of such an assessment, no attempt is made here to attach individual weights to the various factors. These particular ratings were not indicated at the orthography conference, and no votes were taken on each of these issues at the time. However, additional comments in each subsection do indicate the preference of the Kalasha literates themselves, expressed at the orthography conference, as well as in private written and spoken communication at other times. A table summarising these preferences by factor was compiled, and appears as Appendix 12.

6.3.1 Sociopolitical factors

Sociopolitical factors were discussed at length in Chapter 3, and as being equal or second only to religious considerations in importance. But positive religious considerations for script choice are of far less importance for the Kalasha because, being a preliterate society, and

having their own indigenous religion, they have no religious literature, nor do they consider any other society's religious heritage as relevant for them.¹⁰ Therefore, against the background of the evidence given in Chapter 3, the primacy of sociopolitical factors must be considered in the choice of a writing system for Kalasha. Various sociopolitical factors relating to specifically Arabic and Roman writing systems are treated here, with the respective implications of each factor for the Kalasha situation.

A. SCRIPT FAMILIARITY AND ACCEPTANCE

The introduction of a new script that is already more familiar (e.g. to students who learn it at school) is more likely to be accepted and promoted than a script that seems foreign and/or less familiar.

Arabic script is the most familiar by far, being the medium of education in all schools in the Kalasha Valleys (see Chapter 1, subsection 1.1.6), and it is the script used for most books, newspapers, magazines, signs and labels throughout Pakistan. Arabic script is viewed with more favour and acceptance in most of the wider contexts to which the Kalasha belong (e.g. broadly cultural, educational and national). It is the script of Pakistan and the script of Islam, the national religion.

The **Roman** alphabet is less familiar, but is increasingly used in both commercial advertising (marketing products and services) and official signage attached to the properties and projects of the government, as well as those of foreign aid and development providers/facilitators. The readiness of the Kalasha to adopt a Roman alphabet (previously quite novel to them), the preference they indicated for it in preliminary meetings (referred to in the introduction to this chapter), and their willingness to consider it in the conference, suggests that they do not consider script familiarity (which favours Arabic script) to be a strong factor.

A script that is readily accepted and positively regarded in the wider sociopolitical contexts beyond the Kalasha valleys (e.g. in cross-cultural, commercial and administrative settings, both near and far) would stand a good chance of helping to engender good relations with other majority language groups. Most international visitors to the Kalasha Valleys (tourists) use a Roman alphabet in their own languages. This is important for the Kalasha because it is with

¹⁰ There are, however, significant negative religious attitudes, felt by most of the Kalasha community, toward the dominance of Islam in the area. This could be enough for them to foster negative attitudes about the Arabic script.

these visitors that they generally share a greater mutual respect than with their immediate fellow-Pakistani neighbours.

In relation to this factor, most of those who attended the orthography conference stated that local sociopolitical acceptance of a script was of minimal significance to them, and in fact that its international reception was also unimportant. Because of their distinct and unique culture they see themselves as Kalasha first and Pakistani second.

B. SCRIPT AS AN ART FORM

A script that is seen to preserve a writing tradition that is meaningful for the community has greater social value than a script that ignores or devalues such writing traditions. Also, a script that is deemed to have some intrinsic value, or to be attractive, might be regarded as a preferable option to a script that has no special value attached or beauty associated with it, or one that has a negative value or a perceived ugly appearance.

Arabic script has preserved the ancient and highly developed art of calligraphy. Scribes and calligraphers were the guardians of orthography. Calligraphy had become so commonplace that until the 1980s daily newspapers in various Arabic scripts right throughout the world were written by hand by assemblies of calligraphers every night. However, in more recent times, this art has become a skill, like the routine performance of classical music, though much more utilitarian. Technological development has now resulted in the widespread use of computerized versions, their quality partly based on how closely they can copy the traditional handwritten shapes (proportions, orientation, degrees of curve, varying thickness, etc.) The truly creative side of calligraphy is now reserved for original renditions of poetic and religious texts (especially in Persian culture and the Farsi language) and for display fonts in the graphic arts. The modern use of an Arabic script, in these contexts at least, is a deliberate preservation of this art form. Arabic script, in even its most standard form, has been described as having aesthetic appeal. This is possibly due to factors such as its slope, curves, balanced position of diacritics, variations in thickness, complexity, traditionally manual formation, etc.

The calligraphic transcription of **Roman** alphabetic lettering was chiefly a functional foundation for modern Roman orthographies. Therefore, the modern use of Roman alphabets does not usually preserve the calligraphic art form of a former era. Roman calligraphy is now mainly confined to display functions and text headings, rather than for continuous text or for

any particular genre. Calligraphers and graphic artists are probably among the few who would regard standard Roman letters per se as having aesthetic appeal.

The artistic history of scripts is not seen to be a significant value for Kalasha people, who do not share the literary history of their neighbours. It is not surprising therefore that the Kalasha, while very familiar with the appearance of the Arabic script, but traditionally not practitioners of it, see no beauty in the Arabic script. Beauty is in the eye of the beholder. Though presented at the Kalasha orthography conference as an issue of wider significance, it was never taken up for discussion—a poignant reminder of its perceived non-value. In short, the conference participants voiced a consensus that the appearance of the script is not as important an issue to them as its functionality.

C. INTEGRATIVE VS. DISTINGUISHING VALUE

A new script that conforms to the existing script of a neighbouring/majority language could promote social and political harmony. Sociopolitical integration is very highly valued in some cultures and subcultures.

One of the greatest felt needs of the Kalasha people, in their position as the social, religious and economic underdogs of both the local and national communities to which they belong, is political power and leverage. The use of **Arabic** script for the Kalasha language would probably promote political benefits of mutual value both for the Kalasha as well as for the dominant community that surrounds them.

The choice and use of a **Roman** alphabet for Kalasha would not only reap no political benefit, it might even exacerbate current relations by the very nature of its non-conformity to the official (Arabic) script of the country. But this factor is not of major significance to the Kalasha because, being a very distinct minority, they have no expressed concern for political integration whatsoever, either at the conference or at any other time. Their contrary sentiments have been expressed in frequent complaints about the political majority.

On the other hand, a script that sets a culture apart could be a preferred choice if that setting apart is perceived as having some psychological benefit to the community in question. For example, the use of a script that is separate from one that is used by a neighbouring or majority language could foster a sense of freedom, self-esteem and/or coming of age. Cultural conservation and ethnicity were presented as powerful reasons for literacy in Chapter 2, (subsections 2.4.1 and 2.4.2).

The use of **Arabic** script for the Kalasha language would not make the Kalasha feel special or independent compared to neighbouring communities. However, by very virtue of the fact that it would still be a new script, specially adapted for their language, there would still be some psychological benefit.

The prospect of a **Roman** alphabet for the Kalasha language could encourage a general feeling of self-esteem (of which they are in short supply) among those of them who have had most outside contact, and a sense of identity with the outside world from where they receive most recognition and moral support.

This distinguishing value of script choice was expressed by several of the more educated conference participants as having great significance for those Kalasha (including themselves) who are most keen to preserve their culture. For others, who are more interested in only their own education and personal advancement, this factor has little relevance.

D. LEGAL ACCESS

A script that resembles the script used in official, legal and judicial proceedings could encourage easier access to these services. The reason for this is that the use of any script trains users in some of the features of other languages that use the same script. Throughout Pakistan, these proceedings are all conducted in English, so members of any minority communities who are used to a Roman alphabet would have an advantage in interpreting official, legal and judicial documents over those who are not. This demonstrates the issue of political and legal empowerment, discussed in Chapter 2 (section 2.2 and subsection 2.5.2) as a reason for literacy.

The exclusive use of an **Arabic** script for Kalasha could inhibit access to official and legal services by keeping Kalasha defendants at a perceived distance from the judicial process because legal documents are written in the English language, which uses a Roman alphabet. In Pakistan, English is the language of the public service sector and the courts. Therefore, the readers and writers of Kalasha or any language that uses a **Roman** alphabet could have an advantage over those who can only read and write an Arabic script. This is because a Roman alphabet could serve as a lead into the English language, or at least to the point of not being quite as daunting as English.

Many Kalasha individuals have expressed the importance of this issue for them in their situation, so it represents a strong weighting factor for them. In particular, those who have been involved in court cases over land disputes (e.g. FK and several others who were not at

the conference) have expressed frustration at their inability to access the judicial system and have sought help from foreigners to interpret the legalese of court documents.

E. EDUCATIONAL OPPORTUNITIES

An indigenous script that resembles the script used in educational institutions could encourage a greater access to education.

In Pakistan, Urdu is the language of most primary, middle and high schools. Therefore, readers and writers of Kalasha in an **Arabic** script could have a psychological and pedagogical advantage in Urdu-medium education over those who did not use an Arabic script for Kalasha.

In Pakistan, English is the language of some primary and secondary schools and all higher-education institutions. English-medium education is generally perceived as being better quality and having much wider scope than Urdu-medium education. Therefore, readers and writers of Kalasha using a **Roman** alphabet would have an educational advantage over those who did not use a Roman alphabet for Kalasha.

The rapidly increasing proportion of Kalasha society that is now accessing education close to home includes primarily young people and their teachers, who perceive the value and scope of a Roman writing system as a key to further education. Teachers at the conference led the consensus that the immediate advantages of the more familiar Arabic script are outweighed by the longer-term advantages of a Roman alphabet. Despite the initial difficulties, the latter was seen as a better educational investment for the longer term and for future generations.

F. CULTURAL CONNOTATIONS

A script whose connotations link it with a majority culture that uses a similar script could work either in favour or against the minority culture. The most common cultural associations of the **Arabic** script are with Pakistan, Saudi Arabia and/or the Muslim world. This is perceived as either a neutral or a negative association by the Kalasha, and a positive one by the majority community of Pakistan.

The immediate connotation of the **Roman** alphabet in Pakistan is with Western culture, because of its association with English (even though many Western visitors do not use English, and some Eastern visitors do.) This is perceived by the more conservative and less educated sector of the majority community (especially those neighbouring the Kalasha valleys) as a negative association, due to their theory that the West is (morally and politically)

bad, even though most of them are more pragmatically receptive of the West. This connection could be seen as a positive one by the Kalasha whose attitude to Western culture is more favourable.

At the conference, it was the recent school leavers (e.g. ZK, IKB and TKKS) who reacted more negatively to Arabic script, because of its cultural connotations. It was these same individuals, plus a couple of others who have had relatively more contact with Westerners than other Kalasha (e.g. FK and SS), who positively preferred the Roman alphabet because of its cultural connotations. The older participants tended toward a neutral stance on this issue, neither being threatened nor enticed by cultural connotations of existing scripts. Indeed, the majority of Kalasha participants were more interested in the pragmatic reasons for script choice, as the next subsection (6.3.1G) demonstrates.

G. ORTHOGRAPHIC ASSOCIATIONS WITH OTHER LANGUAGES

Certain scripts carry orthographic associations with other actual languages that use the same script. These similarities can relate to the actual *alphabets* that are adopted, and the formal *spelling* that is established. The implications can be positive or negative.

A script that uses a complex *alphabet* or alphabetic variants is less likely to be easily adopted than a script that has a simple but adequate alphabet with no alphabetic variants.

Alphabetic letters in the **Arabic** script have multiple graphemic variants, based on word position. Briefly, there is a word-initial, a word-medial and a word-final form (most of these with joining ligatures also attached), besides the fundamental alphabetic form (which is the word-final form with no joining ligatures). Most of the forms are heavily truncated when written word-initially and word-medially. Diacritics therefore play an important part in distinguishing similar base forms. There are also the issues of representing word-medial nasalized vowels in Kalasha, for which there is no form in most common Arabic-based scripts, and word-initial vowels, for which Arabic script requires an *aleph* (letter *a*) prefix to distinguish some of them from the central approximants (/j/ and /w /), which are represented with the same symbols. All this constitutes a steep learning curve for the beginning student, but it is par for the course when learning Urdu, so that such learning is accepted as normal.

The **Roman** alphabet also has graphemic variants. Rather than being word-initial allographs, as in Arabic script, Roman capitals constitute a set of sentence-initial allographs (which also have several other functions, as explained in Chapter 7, subsection 7.2.3A).

However, Roman script is less dependent on these variants than Arabic script is on its variants.

Many conference participants mentioned other features and examples of Roman scripts, particularly English orthography, in various formal and informal discussions throughout the duration of the conference. These topics (e.g. capitalization, digraphs and apostrophes) were brought up and discussed by those who were familiar with English because they teach it (EK, AU, DM), or speak it fluently (TKKS, IKB, ZK, AK and FK) or at least have some theoretical understanding of it (MRK, SA). Their exposure to and knowledge of the Roman alphabet predisposed all of these participants to a willingness to at least tackle the issues involved in its adoption and adaptation for the Kalasha language.

Although it was not specifically expressed at the conference as a contributing factor to script choice, correspondence over the last fifteen years from some of those most interested in literacy (e.g. FK, TKKS, CK, SK, EK, AU and ZK) also demonstrates their growing preference for the simpler Roman alphabet in its ability to cope with unique features of Kalasha phonology, without the complicating issue of word-position variants.

The way *spelling* corresponds with sound, in each of the Arabic and Roman writing systems, can create different associations with other scripts.

An **Arabic** script for Kalasha would predispose Kalasha students to the conventions of Urdu spelling. This would be easy for most, firstly because most of them are learning the Urdu language in Arabic script anyway, and secondly because Urdu spelling is quite easy and consistent compared to English spelling.

A **Roman** alphabet for Kalasha might remind Kalasha students of the many problems and irregularities of English spelling. This would be daunting for most, firstly because most Pakistanis usually learn the process of reading and writing (and spelling) in a foreign language (whether Urdu or English), and secondly because English spelling is quite complex, and not very consistent, compared to Urdu spelling.

Due to their escalating exposure to the outside (literate) world, and a growing interest in the possibility of indigenous literacy, Kalasha students and teachers are becoming increasingly ready for the challenge of Romanized alphabetic spelling of their own language. Furthermore, their growing confidence with the Roman alphabet, inspired by the perceived prestige of the English language, has made them less daunted by negative connotations of English spelling. This positive attitude to a Roman alphabet was evidenced at the conference by their ability to accept usages that are more restricted for Roman characters than the English

language allows (e.g. the letter *c* for the voiceless postalveolar affricate). These spelling issues are discussed in more detail in Chapter 7.

H. PEDAGOGICAL FACTORS

All the issues involved with the actual practice of having to teach, and learn to use, a particular script to represent a particular language are undeniably crucial factors in considering the suitability and ultimate success of such a script. One indicator of this suitability is the ease with which a proposed script is read and written by those to whom it needs to be taught. In Chapter 3 (subsection 3.4.3), this factor was illustrated by the case of writers of the Balochi language of Pakistan, who preferred a Roman writing system to Arabic because ‘it is easier to learn’ (Jahani 1989, p. 145). The same factor can be applied to the choice between an Arabic or a Roman writing system for the Kalasha language.

A script in which alphabetic letters may be taught and demonstrated using any position in the word has an advantage over scripts in which alphabetic letters may only be taught and demonstrated in certain restricted word positions.

Alphabetic letters in **Arabic** script are most clearly taught and demonstrated in word-final position where they most closely resemble the alphabetic forms of the letters. Alphabetic letters in non-word-final positions are not easily recognized. This adds orthographic limits to other constraints (e.g. lexical) when designing early literacy materials.

Roman characters can be taught and demonstrated in any word position, as there is no difference in form. This maximizes the possibilities of design in the preparation of early literacy materials.

The teachability of a script was not part of the frame of reference for many of the conference participants, although this was a very significant factor for a few of the tertiary-educated individuals (e.g. teachers EK, AU and MRK, and tertiary student TKKS), all of whom have been involved in the development of literacy materials.

6.3.2 Phonological factors

As stated in Chapter 3 (subsection 3.4.1) and Chapter 4 (section 4.1), phonological factors are a fundamental theoretical consideration in script choice. We will now look at how this principle is applied to the choice between Arabic and Roman alphabets for the Kalasha language. Among the most notable features of Kalasha phonology that have particular

implications for either or both of these two scripts, are the features of rhoticity, retroflexion, nasalization and aspiration. Although each one of these features occurs in less than 2% of words in natural texts (see Chapter 4, Table 4.2), half of the vowels are rhotic, half of the them are nasalized (a quarter of them are simultaneously rhotic and nasalized), nearly one-quarter of the consonants are retroflex, and nearly one-third of them are aspirated.

I. RETROFLEX CONSONANTS AND RHOTIC VOWELS

Having an established convention to represent retroflexion and rhoticity could be a weighting factor in choosing a script.

Arabic script (as adapted for Urdu script) does have an established convention to represent retroflexion of consonants. This is by writing a smaller version of the shape of one of the (non-retroflex) consonants (called [tʃ], a letter for /t/) as a diacritic over the top of certain other non-retroflex consonants (sometimes replacing other diacritics that would otherwise be there). In Kalasha, the same convention could be extended to the extra retroflex consonants it has, and to each of the rhotic vowels it has, with no compromise. However, this addition of extra diacritics would add considerable complexity to the overall appearance of the script.

Roman-script representation of retroflexion and rhoticity has generally been achieved with the use of a dot below the character (see Pike 1947, pp. 19, 221, though its use for retroflexion has long been a standard in the literature on South Asian languages). Other options that have been used are attaching a right-hand descending tail for retroflexion and a right-hand hook for rhoticity (as in the IPA). Several orthographic options have been proposed for Kalasha retroflexion and rhoticity, including capitalization, acute or grave accent, or an apostrophe or opening single quote mark following the character.

Since retroflexion and rhoticity are major dimensions of the Kalasha phonemic inventory, indigenous orthography planners have realized the importance of a convention to adequately represent these phonological phenomena, and the need for such a convention as an essential facility of a potential script. Some of them had previously expressed their reluctance to overload the Kalasha orthography with the conventional diacritic that represents retroflexion in Urdu script, by suggesting slightly simpler but unconventional alternatives. This point was reiterated at the orthography conference when the implications of Arabic script were discussed. Too many diacritics were seen to be inefficient because they required going back over many words to add what, in some cases, would be up to ten individual marks.

J. NASALIZED VOWELS

An established convention to represent nasalized vowels in a certain script could be a weighting factor in choosing that script. Nasalized vowels in Kalasha can occur in the following word-position distributions:

(a) word-final

(b) word-medial

and in the following immediate environments:

(c) with no contiguous vowel (which occurs most commonly)

(d) in combination with a contiguous vowel (which may itself be either nasalized or oral)

In this section, following some initial background discussion, we will consider the comparative merits of both Arabic and Roman writing systems in their ability to represent the various distributions of nasalized vowels.

Although nasalized vowels would only occur in less than 2% of words in natural text (see phoneme frequencies in Table 4.2 of Chapter 4) the Kalasha nonetheless see the need to adequately represent nasalization alphabetically. They see a precedent in Urdu, in its handling of nasalization (see subsection (a), below), and also the significant part that it plays in the phonemic inventory of Kalasha.

(a) Word-final nasalized vowels

Arabic script does have an established convention to represent word-final nasalized vowels. This is by writing a variant of the letter for /n/ (without its normal diacritic) following the character.

Roman writing systems do not have a set or special convention for representing word-final nasalized vowels. One option that has already been used is a tilde either over or following the character. There is a widespread practice throughout India and Pakistan, when using the Roman script alternative to their respective orthographies, of representing word-final nasalized vowels with the Roman lower-case or upper-case letter for the consonant /n/, i.e. *n* or *N*. A few Kalasha writers may have also occasionally used this convention in an informal way, but, in any case, the ambiguity that necessarily results from this practice has not disqualified a Roman writing system for Kalasha. In other words, nasalization has not been an issue of script choice for the Kalasha, but rather just a factor to deal with, once a script has been chosen.

(b) Word-medial nasalized vowels

Arabic script does have an established convention to represent word-medial nasalized vowels that precede consonants, but the convention is quite obscure and not widely practised. This is by writing a word-medial form of the letter for the consonant /n/, but with an extra diacritic (resembling a breve) on top of the letter. If a nasalized vowel that precedes another consonant (a situation which also occurs in Kalasha) is written in Urdu script using this special nasalized vowel variant of the letter for /n/ it is interpreted by most readers as the consonant /n/, because the special symbol is so rarely used, or even known, in Urdu.

Roman writing systems do not have a special convention for representing word-medial nasalized vowels. Options that have already been used for other languages, and suggested for Kalasha, include a tilde over or following the character, or the lower-case or upper-case letter for /n/ (*n* or *N*).

This factor has not been an important issue to the Kalasha in their actual consideration of script choice, probably because of its relatively rare occurrence (see phoneme frequencies in Table 4.2 of Chapter 4). However, a few Kalasha who have written Kalasha extensively with Arabic script (e.g. FK) realize its shortcoming in dealing with word-medial nasalized vowels. Specific options are discussed in Chapter 7 (subsection 7.2.5).

(c) Single nasalized vowels

There is no special problem associated with the representation of single nasalized vowels in either Arabic or Roman writing systems, which has not already been dealt with in the preceding subsections.

(d) Multiple nasalized vowels

Urdu in **Arabic** script does not have an established convention to represent each member of a multiple nasalized vowel set because there is no commonly used word-medial nasalization symbol. The only option for Kalasha, where multiple nasalized vowels are common, would be to mark only the last vowel, as is done in Urdu. This would work only if it was impossible to have a nasalized and an oral vowel together, which is the case in Kalasha anyway.

Roman writing systems do not have a set or special convention for representing multiple nasalized vowels. Options that have already been used or suggested include a tilde over or following the characters, or the lower-case or upper-case letter for /n/.

The Kalasha conference participants had no problem with the marking of only the second in a pair of nasalized vowels in Arabic script, nor with the tilde-marking of both vowels using Roman alphabetic letters, as suggested above. Again, specific options are discussed in Chapter 7 (subsection 7.2.5).

The whole issue of representing nasalized vowels turned out to be a more complex situation than the conference participants were anticipating. They generally either oversimplified the problem examples, or admitted the difficulty and complexity of an Arabic script representation. Bending or extending the rules of Arabic script to cope with these different phonological environments in Kalasha was something they were more reluctant to do than adopting or even creating new rules in a Roman writing system that was new to them.

K. ASPIRATED CONSONANTS

An established convention to represent aspiration in a certain script could be a weighting factor in choosing that script.

Arabic script does have an established convention to represent aspiration of consonants. This is by writing a special aspiration symbol following the character. **Roman** writing systems do not have a set convention for representing aspiration. Options that have already been used include a superscripted version of the letter for /h/ (^h) following the character, or the letter for /h/ itself following the character.

This factor actually has no theoretical consequence for the choice of scripts, as far as the Kalasha are concerned, since both Arabic and Roman writing systems have simple and adequate ways of representing aspiration. However, if the Roman solution were not superscripted, three meanings of /h/ may need to be taught: glottal friction by itself, aspiration of a consonant, and an orthographic function as the second element in voiceless postalveolar fricative and affricate digraphs, such as *sh*, *zh* or *ch* (though the last could easily be represented merely by the letter *c*).

6.4 CONCLUSION

The choice of script, determined by formal vote at the conference was for the **Roman** alphabet. This decision reflects the pedagogical, sociopolitical and orthographic factors that the Kalasha associate with the Roman alphabet, and presumably coincides with the set of issues (A to K) presented in this chapter (which had been drafted well before they were

presented and discussed at the conference). The Roman alphabet seems to be the superior choice for the current and future sociopolitical circumstances of Kalasha society—apart from its suitability from a phonological point of view.

Though the Roman alphabet was seen by most as less familiar and more of an initial challenge than the Arabic script, the steeper learning curve was not considered a serious obstacle to developing a new orthography. The teachers repeatedly asserted that a Roman alphabet would be easier to teach and write. Putting a Roman writing system into use was considered the better choice for future generations.

Most of the factors considered in this chapter would be relevant for other languages around the world where similar sociopolitical or phonological circumstances prevail, especially those where Arabic and Roman writing systems are the most likely alternatives, as exemplified in various sections of Chapter 3.

The choice of the Roman writing system then entails further issues as to how its alphabet and orthographic rules should be specifically modified to represent the Kalasha language. Those issues were also discussed at the orthography conference, and will be discussed in Chapter 7.¹¹

At the end of the conference each of the teachers was presented with literacy materials for use in their respective schools. An account of literacy planning developments that have occurred since the conference is presented in chapter 8, section 8.4.

¹¹ An exhaustive discussion of all specific orthography issues at the conference was not possible, due to time constraints and delegates' mental saturation levels after four days.

DEVELOPING A ROMANIZED KALASHA ORTHOGRAPHY

7.1 USING THE ROMAN ALPHABET FOR THE KALASHA LANGUAGE

Having reported the issues discussed at the Kalasha orthography conference, and the consensus decision for a Roman writing system, let us look further at the implications of this choice. Radloff (1996, p. 9) points the way:

Creating an alphabet is only the first step in reducing a language to writing. The most difficult step is to introduce spelling rules or standardisation so that words can be written consistently by different people and easily read by all. Such spelling rules must be developed scientifically, or reading the words will be difficult.

In this chapter, we will firstly examine specific *orthographic* issues that would arise in the use of a Roman alphabet for Kalasha (section 7.1), and then consider orthographic options that flow out of specific *linguistic* features of the language (section 7.2). Again, orthography conference discussions will be referred to where relevant, bearing in mind that for most of the participants, their greater familiarity with Arabic-script orthography is sometimes reflected in their viewpoints about orthography in general, and even about Roman orthography in particular.

Several special features of the Kalasha language require specific orthographic treatment. Some of the issues involved would be essentially the same, irrespective of whether an Arabic, Roman or other writing system were used, and some would be specific to only one writing system. In this chapter, we will deal with both the generic issues and those specific to the Roman writing system. On each issue, relevant opinions and comments by Kalasha conference participants will be featured. Some of these were expressed by informal vote, either verbal or by show of hands; others came out in the course of informal discussion.¹ The outcome of this chapter will be a preliminary code of orthographic principles, graphotactic

¹ In some cases even non-responses were significant: for example, if there were no objections to a proposal, if they experienced difficulty in understanding a concept, or where a matter was inconclusive. However, opinions and queries were expressed very freely throughout the conference.

recommendations, and spelling conventions, as agreed by oral consensus or at least majority agreement by the Kalasha.

7.1.1 Roman alphabet spelling models

The International Phonetic Alphabet (IPA) is not purely a Roman alphabet because it involves symbols from various European alphabets. It would be less than ideal as a practical orthography unless most of the IPA symbols required were already familiar to the language community, or there is some other independent reason why they would be more suitable than other alternatives. However, for the Kalasha, as for even educated speakers of many other languages, the IPA inventory includes many unfamiliar symbols.

The increasing exposure of the Kalasha community to the English language, which led, in the first place, to the serious consideration of the Roman alphabet as an orthographic model, has familiarized many Kalasha speakers with the various orthographic conventions used in English. For this reason, it would make sense for Kalasha orthography to conform to those elements of English orthography that are simpler or less confusing than the more obscure IPA or Roman orthography alternatives. A case in point would be the use of the English symbol *y*, instead of the IPA symbol *j*, which also represents that sound, but which would be confused with English *j*. Another example would be the use of the English symbols *j* and *c*, instead of the IPA symbols dʒ and tʃ , which also represent those sounds, but are unnecessarily complex for a practical orthography. This case is discussed further in subsection 7.2.2, below.

One could argue for an even more comprehensive match to an English-based orthography, perhaps embracing elements of various proposed English spelling reforms and/or ‘user-friendly’ English spelling schemes. The new conventions could conform to spelling rules with the least ambiguous and most common assignments of Roman alphabetic letters for particular English phonemes (e.g. *oo* for /u/, *ee* for /i/), and other orthographic practices, such as the doubling of consonants in stressed syllables. However, such schemes rely too heavily on the Anglo-centric model, and are not recommended for a practical Kalasha orthography.

Many simpler, English-based solutions are nevertheless still possible. For example, the English letter *u* (but only where it stands for /u/) would be better for rendering Kalasha /u/ than the English rendering of *oo*, which, being a digraph, is a more complex representation. And, as pointed out above, the English letter *c* is preferable to the IPA equivalent, but it is also preferable to the English digraph *ch*, which would be needed to distinguish the Kalasha

aspirated affricate /tʃ^h/ from the unaspirated /tʃ/. Other issues and problems with an English-based orthography will be raised in subsequent subsections (e.g. alphabetic constituency in 7.1.2, consonant digraphs in 7.1.3).

7.1.2 The constituency and order of the Kalasha orthographic inventory

Which letters should comprise the Kalasha orthographic inventory? The English alphabet consists of 26 letters (representing more than 40 phonemes), whereas the Kalasha language consists of 62 phonemes (as presented in Chapter 4, subsection 4.3.2). This would suggest that we need all the available letters of the English alphabet, and many more.

Some alphabetic letters that cannot be satisfactorily matched to Kalasha phonemes may need to be either reallocated or scrapped altogether. For example, the letter *f* could not have its traditional English meaning in Kalasha because /f/ is not a Kalasha phoneme—its closest equivalent in Kalasha is /p^h/, which would be represented in a way that is consistent with other aspirated phonemes. Similarly, the letter *v* could not have its traditional meaning in Kalasha, because Kalasha [v] is one of the allophones of /w/, so the letter *w* would suffice. The Roman letters *c* and *q* do not represent distinct phonemes in English, so they too are natural candidates for either reallocation or omission. (An option has already been proposed, in subsection 7.1.1, for the letter *c*.)

However, before resorting to the invention of new characters it makes more sense to utilize existing orthographic devices for alphabet extension, for example, digraphs and diacritics. In fact, most orthographies are designed in accordance with this principle (see Sztemon 2002). These two devices are discussed at more length both generally (in subsections 7.1.3 and 7.1.4 respectively), and specifically (at various points in section 7.2, as required for the representation of specific phonological features of Kalasha).

Alphabetic order is another issue that cannot be ignored. The *Kalasha Dictionary* (Trail & Cooper 1999), being primarily in English, uses the English alphabetical order. However, some previous unpublished versions of it used the Urdu alphabetic order, in preparation for what was then anticipated to be an Urdu-script-based orthography. The Arabic and Urdu alphabetic orders group sets of letters together that share the same base shape, which aids in the recognition and teaching of the alphabets, making elementary literacy easier than if letters

were mixed.² Interesting, and of potential relevance even for a Roman orthography, is the fact that each of these groups of letters reflects the shared manner and/or point of articulation of its member phonemes, as can be seen in the following schematic display of sets of letters, where both the order and grouping of Arabic-based alphabets has been applied to Kalasha phonemes.³

Table 7.1: Roman-script Kalasha phonemic inventory in Arabic alphabetic order

a	aː					front low vowels
b	p	t	t̪			non-velar plosives
ḏẓ	ḏẓ̣	t̪ʃ	t̪ʂ	ts	dz	affricates
d	ḍ	r	ɾ			voiced dental plosive and flap
z	ʒ	ẓ				voiced oral fricatives
s	ʃ	ʂ				voiceless oral fricatives
k	g					velar plosives
l	ɭ					approximants
m	n	ɲ				nasals
o/w	oː	u	uː			back vowels/
h						glottal fricative
i	iː	e/j	eː			front non-low vowels

Aspirated consonants are not shown above because in Arabic-based scripts they are digraphs that do not feature separately in alphabets. For this reason, neither are they included in the primary sorting order of dictionaries written in Arabic-based scripts. Instead, the order of words containing aspirated consonants is based on the representation of the aspiration symbol as the letter *h*, resulting in words beginning with these aspirated consonants being embedded within the set of entries for unaspirated consonants, for example, *pe ... pho ... po*.

² This order is also used in the worldwide standard for Arabic-based scripts in electronic media (Unicode 2003).

³ Each row of phonemes is represented by a common base shape in Arabic-based alphabets, and the order within each row reflects actual alphabetic order, though there is no significance to the columns in this schematic display.

However, this pattern of phonetic similarity within each of the above groups of letters is insufficient justification to adopt an Urdu-based alphabetic order for Romanized Kalasha, given the confusion it would cause in its being an unorthodox order for Roman letters. In addition, Arabic-based alphabets make no distinction between /o/ and /w/, nor between /e/ and /j/, so the issue of relative order between them does not ever arise. For these reasons, the English-based alphabetic order is the recommended option for Kalasha phonemes.

Obviously, the secondary sorting order that applies for aspirated consonants in Arabic-based orthographies (as exemplified above) would also apply for a Roman orthography. And likewise, other consonant digraphs (distinguished by underlining here) would also be sorted secondarily within the main order, for example, *se ... sh ... si ... te ... th ... ti ... ts*.

Similarly, alphabetic letters representing retroflex and rhotic phonemes would logically follow their respective non-retroflex/rhotic counterparts in a Roman orthography, as they do in Arabic-based orthographies. Though these secondary ordering conventions were not specifically voted on by the Kalasha, no objections were raised to the alphabetical order proposed here when it was featured, showing both primary and secondary sorting, in demonstration editions of literacy materials at the orthography conference.

7.1.3 Digraphs

A digraph is a sequence of two alphabetic symbols that represent a single phonemic segment, consonant or vowel. They have been established in the writing of various European languages (e.g. Asturian, Galician, Polish and Spanish *ch*; Albanian *dh*, *gj*, *ll*, *rr*, *sh*, *th*, *xh* and *zh*; Balkan Romany *kj*, *gj*, *lj* and *nj*; Welsh *dd*, *ff*, *ng*, *ll*, *ph*, *rh* and *th*) whenever a phonemic inventory exceeds the number of letters in an alphabet, and they will be equally useful for newly written languages like Kalasha.⁴ We will only consider English digraphs here because the Kalasha are more familiar with English than any other language with a Roman alphabet. To avoid having to invent new symbols for phonemes that Kalasha and English share in common, and for which English consistently uses digraphs, it is recommended that Kalasha also use some of these. Specific recommendations are put forward in subsections 7.2.1 (fricatives), 7.2.2 (affricates) and 7.2.6 (aspirated consonants).

⁴ Indo-European examples taken from Sztemon 2002.

Another general issue with digraphs is whether both orthographic elements should be subject to marking for special linguistic features (e.g. retroflexion or aspiration), or only one of the elements? And if only one, which one? This matter is dealt with for particular phonemes in subsections 7.2.3 (retroflex consonants and rhotic vowels) and 7.2.6 (aspirated consonants).

7.1.4 Diacritics

Diacritics are auxiliary symbols that extend the range of an orthography. Alphabetic writing systems present a rich variety of diacritics that have developed over centuries for many different languages, for example, to distinguish phonemes in many Slavic, Turkic and North Caucasian languages, and to mark tone in many tone languages, including Vietnamese and the Pinyin orthography of Chinese (Coulmas 1989, pp. 44, 246). Commonly used Roman diacritics include acute, grave and circumflex accents, cedilla, háček, tilde, umlaut, etc.

In handwritten or printed text, the use of diacritics can be problematic. If a diacritic is written outside a letter it will extend the vertical or horizontal space normally required for that letter. This can lead to orthographic sprawl, where letters and/or words may appear disjointed, and therefore hard to read, or where lines of text may take up an inordinate amount of vertical or horizontal space, restricting formatting or layout potential. On the other hand, a complicated diacritic written within a letter will reduce the visual resolution of that letter. This can lead to orthographic crowding (in either handwritten or printed text), where certain letters and/or words may appear cramped, and therefore hard to read.

Other disadvantages of diacritics are as follows: They are easily omitted due to speed or oversight. Accidental misplacement leads to possible confusion as to their intended position, in relation to nearby letters. Relatively small diacritics (e.g. dots) may not be bold enough to be noticed. Diacritics with a high functional load (e.g. those that provide contrast in a phonemically, morphemically or semantically similar environment) are especially susceptible to accidental misuse or omission if they are small.

Most European orthographies except English use diacritics liberally, and some (e.g. Hungarian) function quite adequately with a prodigious use of diacritics. However, for the abovementioned reasons, the less diacritics there are in a *newly created* orthography, the less there are for new readers to learn to differentiate, less chance of accidental omission, and less effect on handwriting speed and efficiency.

For those communities that have already been exposed to a Roman alphabet with very few or no diacritics, too many diacritics in their own new orthography can make the writing system look unwieldy, awkward and fragmented. Holm (1980, p. 86) cites an example of the Navajo encounter with diacritics in a new orthography:

Their orthographic expectations are those of English. What seems to deter many adult Navajos from learning to read Navajo are the diacritics: all those funny little marks, which are contrary to their (English-induced) orthographic expectations.

The presentation of diacritics common in several non-English Roman orthographies (e.g. acute and grave accents, háček, umlaut) was met with confused responses by the Kalasha participants at the orthography conference, probably because of an aversion to unfamiliar diacritics. Keeping the range of diacritics to a minimum would help keep the new Kalasha orthography simple and efficient, for the reasons discussed above. Technical issues to do with the use of diacritics are discussed further in Chapter 8.

7.1.5 Punctuation

Punctuation marks commonly used at sentence-level in English, such as full stops, question marks, commas, quotation marks, dashes, colons, parentheses, etc., and even word-level punctuation, such as apostrophes and hyphens, can often be carried over to perform the same or similar functions in a newly written language. Neither the need, nor solutions, for punctuation were discussed at any length by participants at the Kalasha orthography conference. The choice of specific punctuation marks to be introduced into the Kalasha orthographic repertoire is best determined by teachers, who are involved in teaching literacy. For example, full stops and question marks, which also have their counterparts in Arabic scripts, are more likely to be useful than semicolons and square brackets.

Some punctuation marks, such as apostrophes and hyphens, could have uses as phonological or syntagmatic markers in Kalasha, if such uses did not interfere with any normal punctuation functions they might have.

A. WORD-LEVEL PUNCTUATION

The use of the **apostrophe** to mark syntagmatic contraction in English (as in *can't*, *o'clock*, etc.), could very well be used in Kalasha orthography, where elision causes two words to contract into one, for example, with {hawaw haw} becoming /hawawaw/, we could write the

apostrophe in place of the elided phonemes, as *hawat'aw*.⁵ Also, the elision of the phoneme sequence /tr/ in the non-Birir dialects could be marked by writing an apostrophe in place of this sequence, for example, with /matrik/ becoming /maik/ we could write it as *ma'ik* ('to say'), etc. And when the reverse occurs, where the Birir Valley dialect is the one that contracts another paradigm, writers there could write *pa'k* for /parik/ ('to go' or 'let's go'), etc. The introduction of such an orthographic convention would be a stylistic decision based on a concern to indicate the formal origins of certain words, using the deep-level representation discussed in Chapter 5.

However, this use of the apostrophe to mark elision in Kalasha orthography would be confusing if it was also being employed to mark any other linguistic features, for example, retroflex consonants or rhotic vowels. At the conference, the decision to use the apostrophe to mark retroflexion and rhoticity, as reported in subsection 7.2.3C, preceded the discussion of elision, so by the time this topic was reached there was no interest in redeploying the apostrophe to mark elision. In fact, they preferred not to mark elision at all, leaving the apostrophe free for other roles.

The English use of the **hyphen** sometimes indicates a level of association between morphemes that is closer than separated words but not as close as a fully compounded word. (In some cases it serves as a transitional device for certain pairs of morphemes that are on their way to becoming compounded.) It could have this same function in Kalasha, and/or the added function of demarcating the affixed clitics that were discussed in Chapter 5 (subsection 5.3.2). For example, words beginning with the prefix *na-*, meaning 'not' could be written as, for example, *na-behel*, 'cursed', and words beginning with *bej-*, meaning 'without, non-, un-, -less' could be written as, for example, *bej-ad̤ʒat*, 'unnecessary'. However, hyphenation is unlikely to be popular with the Kalasha. Firstly, its usage as such would most probably always be a matter of controversy, as it is in English. Secondly, a hyphen-like dash is used to end sentences in Urdu, which is not a helpful association to have in one's mind when trying to use the same symbol for a different purpose within sentences in another language.

When presented with the option of using hyphenation for word-level punctuation the Kalasha participants at the conference rejected it. There is therefore little incentive to use it,

⁵ Through much of this chapter, examples of the proposed Kalasha orthography are represented using the IPA symbols for Kalasha phonemes, in an italic font. Proposed Roman orthography will be introduced progressively, and a final simplified Roman alphabet for Kalasha will be presented at the end of the chapter, after all the orthographic issues have been discussed and had solutions proposed for each.

unless required for the pedagogical purposes of disambiguation. The use of a soft hyphen when words are broken at the end of lines was not discussed at the conference, so this issue would need to be reviewed after the Kalasha have gained some experience in typesetting and publishing.

Word **spacing** is usually used in Roman alphabetic writing systems to mark word boundaries. Both of these are novel concepts to those who have learnt Urdu, where orthographic spacing is not significant, and very often not even used. As stated in Chapter 3 (section 3.2) and Chapter 5 (section 5.3.1), word-level spacing in Urdu, Farsi and Arabic is often marked by special word-final forms, so a space is sometimes redundant.⁶

B. SENTENCE-LEVEL PUNCTUATION

Direct speech in the Kalasha language is marked grammatically by the use of the participle /g^hōi/ immediately following the direct quotation. This may be translated into English as ‘saying’. For example, /a dura ais, g^hōi amaaw/, translated freely is “‘I was at home”, he said’, but literally is “‘I at-home was”, saying, he said’. This very common Kalasha word makes the need for quotation marks almost redundant in written text, except that it does not indicate where the quote starts. The issue of quotation marks was not raised at the orthography conference due to lack of time. Again, they are better left aside until those who teach literacy can establish the need for them. In such a case, the English conventions should be used, as the Kalasha will be familiar with them anyway.

Other common marks of English sentence-level punctuation, such as full stops, commas, question marks, dashes and colons, would all be useful in Kalasha orthography. All have direct equivalents in Arabic orthographies (though horizontally rotated for the opposite direction of text), so most would be familiar. There is no Arabic equivalent of the Roman dash, other than the Arabic dash which is a full stop.

Sentence-level spacing in English is marked by punctuation (e.g. the equivalents of the full stop, question mark), so a space is actually redundant, though often inserted, perhaps to reflect a slight pause between sentences, which does not always occur between words. Sentence spacing marks sentence boundaries, in addition to sentence-level punctuation. However, Urdu writers often insert this space *before* the sentence-final punctuation mark, and new writers of Kalasha in Arabic script tend to follow the same practice. This may reflect a perception that

⁶ An orthographic space is still required in Arabic-based scripts to mark a word boundary after a non-joining letter in word-final position.

sentence-final punctuation is not a word-level feature, and that therefore, a sentence boundary marker should stand by itself, not attached to the last word in the sentence.

7.1.6 Compounds

When two or more English morphemes are very frequently used with each other, they are sometimes orthographically compounded, with or without any joining mark (which in Roman script is the hyphen). Examples of such morpheme sets in Kalasha (in Roman orthography), with possible renditions, are:

kaṭaṣa + *mondr* → *kaṭaṣa mondr* ~ *kaṭaṣa-mondr* ~ *kaṭaṣamondr*

‘Kalasha’ + ‘language’ → ‘Kalasha language’

kanda + *d̄ṣaw* → *kanda d̄ṣaw* ~ *kanda-d̄ṣaw* ~ *kandad̄ṣaw*

‘almond’ + ‘grove’ → ‘almond grove’

mraṭṣ + *waki* + *ṣofi* → *mraṭṣ waki ṣofi* ~ *mraṭṣ-waki-ṣofi* ~ *mraṭṣwakiṣofi*

‘mulberry’ + ‘harvest’ + ‘Spring festival’ → ‘mulberry harvest festival’

Whereas writers of English tend to join morphemes, Pakistanis (including Kalasha) tend to be splitters rather than joiners. English compounds are frequently unpacked on Pakistani signs, for example, *Wel come*, and in Urdu script, certain Urdu words (and even some English words) are written with both word-final and word-initial forms, but without a word space. In the following examples, full stops are inserted in both the phonemic transliteration of the Urdu spelling, as well as in the English gloss, to indicate where the grapheme boundary (i.e. between word-final and word-initial forms) occurs, and capitals are used to represent word-initial forms in the Urdu spelling:

islam.abad

fæsəl.abad

esosi.eʃən

‘Islam.Colony’

‘Faisal.Colony’

‘Associ.Ation’

‘Islamabad’

‘Faisalabad’

‘Association’

In the first two examples, these word-final and word-initial forms (without a space) indicate a morpheme boundary, and in the third example, they indicate phonetic stress.

Writing the above sets of Kalasha morphemes separately would explicate the constituent morphemes. This is also the preferred practice of the Kalasha, as shown in Chapter 5 (subsection 5.3.1). However, the morphemes of certain common Kalasha phrases may be joined into compounds. If so, this will probably happen naturally and without a lot of deliberation.

7.1.7 Foreign words

A. COMMON WORDS

Every language faces the issue of how to represent foreign words orthographically. Usually, this happens without much planning, sometimes more closely reflecting the spelling of the source language (e.g. French *champagne* is spelt the same in English)—occasionally even to the point of importing foreign diacritics in formal presentational styles (as in English *café*, *façade*, *piñata*, *háček*, etc.). In other cases a foreign word is modified, to a greater or lesser degree, to make it conform more closely to the phonemic and/or orthographic inventory and conventions of the target language (or that of another donor language), or to the already established spelling of similar words in the target language. The spellings of the English words *campaign*, *linguistic* and *psyche* were modified when imported from other languages: *campaign* from French *campagne*, *linguistic* from French *linguistique*, and *psyche* from Greek *ψυχή* (where script transliteration was made at the expense of phonemic transliteration). The Japanese use special alphabets (Katakana and Hiragana) to cope with the problem of writing foreign words (especially Chinese) that do not conform to their own orthographic system.

The advantage of using orthographic conventions from the source language is that they preserve the foreign identity and etymology of words. This would be a ‘deep’ orthographic representation. The advantage of modifying words to conform to the receptor language is that it facilitates their pronunciation, using phonemic recognition. This would be a shallow orthographic representation. In certain situations, foreign spelling, based on the source language of foreign words, may enhance the status of the receptor language and its new orthography, especially if rules are developed for the spelling of particular features of foreign languages. However, this is more likely to happen where the borrowing language has highly developed literature, which is not yet the case with Kalasha.

Using receptor language spelling, at a shallow, phonological level, the words *hotel* and *school* imported into the Kalasha spelling scheme would be *hoʈol* and *sukul* respectively. There are two processes at work here: one is the closer conformity to the target language phonemic system (in this case Kalasha pronunciation /hoʈol/ and /sukul/), and the other is the abandoning of peculiar donor language orthography (in this case the digraphs *ch* and *oo* representing the phonemes /k/ and /u/ respectively).

B. PLACE NAMES

When it comes to foreign place names, the Kalasha are also likely to indigenise their spelling, according to reflect their own verbalisation and pronunciation. So, in Kalasha, *UK* would be written as *lanḑan* (*London* being the most commonly heard geographical name from the UK), *France* as *p^harans* (the labio-dental fricative being replaced by the aspirated bilabial plosive), *Germany* as *ḍẓarman* (from *German* the adjective), *Australia* as *iṣṭrelia* (their /str/ consonant cluster is always prefaced by the phoneme /i/), *USA* as *amrika*, etc.

Not surprisingly, some countries of the Middle East and Asia are better known in most languages throughout the region (including Kalasha) by other non-English names, for example, *Greece* is /junan/, *Egypt* is /misar/, *Syria* is /aram/, and *China* is /tʃin/. Therefore, the Kalasha are likely to represent the names of these countries to match.

C. PERSONAL NAMES

The principle of the orthographic indigenisation of personal names is generally preferred, and is already practised by many Kalasha individuals.⁷ Several decades ago a Kalasha elder named his daughter after the English word *vaseline*, and his son after the English word *regiment*. In Kalasha indigenised orthography, these names are spelt *waḥlim* and *radẓiment* respectively, to approximate their indigenous phonemic interpretations. In a neighbouring hamlet, four Kalasha brothers, in turn, were named after the English words ‘master’, ‘commander’, ‘general’ and ‘engineer’. Using the new, conference-approved Roman orthography the first three brothers would spell their names something like *maṣṭer*, *kaṃander*, and *ḍẓarnel*, for the same reason of phonemic approximation. The fourth brother, who became the first Kalasha teacher, has spelt his name as *Injinier*. This is partially an indigenisation and partially a source language swap of the English sequence *ie* for *ee* to represent the vowel /i/. Another elder from that same hamlet named his three daughters after the English words ‘London’ (meaning UK), ‘glass’ and ‘election’. Their names are now spelt *lanḑan*, *iglas*, and *lakṣan* respectively. A Kalasha woman nicknamed *BBC* (because of her reputation for passing on information) would probably be referred to as *bibisi* in writing,

⁷ It should be pointed out that the Kalasha place less significance on names than many Westerners do. In practice, personal names are rarely used by Kalasha. They usually use kinship terms for personal address, such as /aja/ ‘mum’, /dada/ ‘dad’, /baja/ ‘brother’ and /baba/ ‘sister’, and relational expressions in third-person references, such as /kharima ajas/ ‘Karim’s mother’ and /ramasena putras/ ‘Ramasen’s son’.

because the Kalasha hear the name *BBC* (on short-wave radio) far more than they ever see the initialism in print. Besides this, the indigenous spelling would remind the Kalasha of a very common and distinctively Kalasha woman's name, *Bibi*.

Although the Kalasha generally follow this principle of indigenized spelling of foreign words, they tend not to abide by the principle for already established personal names that they have imported into the language, often of Arabic origin. Most people with these foreign names (even other Pakistanis) prefer to keep the official transliterations of them, which are more familiar throughout the Muslim world, whether they are writing with an Arabic or Roman alphabet. The orthographic recognition of a personal name tends to be more important than its phonemic matching in Kalasha. For example, *Saifullah* for /sajp^hula/, *Mohammed* for /mamat/, *Jinnah* for /d͡ʒana/, *Faizi* for /p^hajzi/.

A Kalasha boy born in 1980 was named Taj (a common Middle Eastern name meaning 'crown' in several languages, including Arabic and Urdu). Of course, this is pronounced as /tat͡ʃ/ in Kalasha phonology, but the conference decision (discussed at length in Chapter 5, subsection 5.2.1) was to orthographically preserve word-final phonemic voicing, thus: *Taj*. Despite this, however, he has sometimes signed his correspondence as *Tac*, which reflects the unique Kalasha phonetic pronunciation of the word. Indeed, the spelling of personal names is very flexible in Pakistan generally, not only among the Kalasha.

7.2 ADAPTING THE ROMAN ALPHABET FOR KALASHA PHONEMES

7.2.1 Fricatives

The Roman letters *s* for /s/ and *z* for /z/ would naturally be used in the same way in Kalasha. The Roman digraphs *sh* for /ʃ/ and *zh* for /ʒ/ could also be useful in Kalasha. The fact that the letter *h* would also be used for aspiration does not pose a problem because the phonemes /ʃ/ and /ʒ/ cannot be aspirated in Kalasha. So the letter *h* in the digraphs *sh* and *zh* would function as the auxiliary element of the digraphs, following the English precedent in *sh*. Retroflex fricatives are discussed below in subsection 7.2.3, which deals with retroflex consonants generally.

7.2.2 Affricates

As mentioned above in subsection 7.1.1, above, and in line with the decision to conform largely to English orthography, the letter *j* would most naturally stand for the Kalasha affricate /d͡ʒ/, except for its word-final allophonic variant [t͡ʃ], where the letter *c* could be used. The English digraph *ch*, meaning /t͡ʃ/, would be confusing in Kalasha because /t͡ʃ/ can be aspirated with phonemic significance in Kalasha. For this reason, the Kalasha phoneme /t͡ʃ/ would be better represented by the letter *c*, which does not have a unique phonemic allocation in English, and the Kalasha phoneme /t͡ʃ^h/ by the digraph *ch*, in line with the representation of other aspirated consonants. These suggestions were accepted by all but one of the 21 orthography conference participants. One queried this novel (non-English) use of the letter *c*.

There are also two English letter sequences, *ts* and *dz* which could be employed to represent the Kalasha affricate phonemes /ts/ and /dz/ respectively. At first, these suggestions were met with mild uncertainty, a not unexpected reaction, as the Kalasha are still getting used to the idea of creating alphabetic digraphs to represent single phonemes in their language. There being no better suggestions, these notations were assented to by all the Kalasha conference delegates. However, in Chapter 4, subsection 4.3.4, the point was made that for some phonemes with lower functional loads there could be orthographic implications, such as the inclusion in or exclusion from the alphabetic inventory, and in subsection 4.3.5, it was suggested that the alveolar affricate /d͡z/ in particular, might be excluded, as it only occurs in a minor dialect—hence, perhaps, the initial uncertainty about it.

Retroflex affricates (including aspirated retroflex affricates) are discussed in the next subsection (7.2.3), which deals with retroflex consonants generally. Aspirated affricates are discussed in subsection 7.2.6 that deals with aspirated consonants generally.

7.2.3 Retroflex consonants and rhotic vowels

Retroflexion and rhoticity are pervasive aspects of the Kalasha phonemic system, as we saw in Chapter 4. However, apart from Arabic and Devanagari scripts, most writing systems have no convention for representing retroflexion or rhoticity. Most of the orthographic devices proposed or used to represent retroflexion and rhoticity using a Roman alphabet involve some kind of diacritic attached to or appearing with normal consonants. They have included (1) a right-hand hooked descender attached to the bottom of retroflex consonants, and a hook

attached to the right-hand-side of rhotic vowels (as in the IPA); (2) a dot below the character (Grierson 1915; Pike 1947; Morgenstierne 1973; Bashir 1983; Cooper, 1984b; Trail & Cooper 1985a, 1985b, 1999; Decker 1992b; and Baart 1997); (3) acute, grave or circumflex accent, macron, breve or háček, all above the character, or a cedilla attached to the bottom of the character, (all tested with the Kalasha during and after the 2000/2001 orthography conference, with little response); and (4) an apostrophe following the character (discussed at length in subsection B below). Other kinds of orthographic device for retroflexion and rhoticity are (5) underlining, (6) the addition of *r* before or after the retroflex or rhotic character, and (7) capitalisation (discussed at length, in subsection A below).⁸

The use of diacritics associated with the Roman alphabet was introduced in subsection 7.1.4, above. In particular, the hook and the dot would be difficult to write accurately and clearly by hand, and complex fonts including them are not easily accessible to the Kalasha community. The more conventional Roman diacritics (e.g. acute, grave or circumflex accents, macron, breve or háček) would be easier to write accurately and clearly by hand, and the appropriate fonts for computer use would be accessible to the Kalasha community. However, most of these diacritics are used with vowels, or consonants with no ascenders. In Kalasha, they would be required for use above the consonant letters *t* and *d*, whose ascenders would interfere, in computer fonts, with the diacritics. Furthermore, the availability of computer fonts even for these more conventional diacritics is still an issue in the Kalasha valleys, as will be discussed further in Chapter 8.

A. CAPITALS FOR RETROFLEXION AND RHOTICITY

If capitalisation is used as means of marking retroflexion and rhoticity, the retroflex consonants of Kalasha would be: *C, D, J, R, S, T* and *Z*, and the rhotic vowels of Kalasha would be written thus: *A, E, I, O* and *U*. Such a convention avoids all the problems of diacritics and special characters discussed earlier in this chapter. Roman capital letters are very widely known and recognised, and are on every keyboard. There is therefore one symbol available for each retroflex or rhotic phoneme, and so perfect consistency is maintained.

⁸ Underlining was thought by some to be too easily confused with other uses of underlining, for example, emphasis or syllabic stress. The addition of the letter *r* before or after a retroflex consonants or rhotic vowel, was used with additional diacritics by Morgenstierne (1932), and is commonly employed in some Australian Aboriginal languages. However, in Kalasha this would be confused with consonant clusters already containing the letter *r*, e.g. /pruʂt/.

Capitals were used to represent retroflexion in the unpublished Kalasha-English and English-Kalasha lexical inventories and dictionaries (Trail & Cooper 1985b, 1987, 1997), as well as in various papers by Trail (1987, 1996a, 1996b). As a result, capitalisation also came to be used in correspondence by several Kalasha over those years. This method has also been used in various published works on other Indo-Aryan languages. For example, capitals were used for retroflexion to represent Kalasha and Khowar by Bashir (1988b, 1990, 1996), as an editorial policy for the publication of various authors' articles on languages of the Hindu Kush region (Bashir & Israr-ud-din 1996), and in Khowar by Munnings (1998a, 1998b). This method was also used in the analysis of Kalami by Baart (1995), and in a Roman alphabet-based transcription scheme developed for the Sylheti language of Bangladesh in the 1990s, representing allophones in free variation: [t] and (retroflex) [ɖ] (e.g. [kita] and [kiTa], both meaning *what*) (Lie et al. 1999).

The main problem with the use of capitalisation to represent retroflexion and rhoticity is that it interferes with the other accepted uses of capitals in Roman orthographies, such as marking a new sentence, to distinguish proper nouns from common nouns, to mark the first person singular pronoun in English, to indicate grammatical emphasis, to make printed text appear visually clearer or more important, to distinguish acronyms from normal words, etc. If any of these uses were kept, one could never be sure in some environments whether a capital has a phonemic or a grammatical or other function. Any interference between concurrently operating but conflicting functions of capitalisation may not be easily tolerated—and it might create confusion. Capitalization for retroflexion and rhoticity would also occur word-medially, which could also lead to confusion.

If capitals were used for Kalasha retroflexion and rhoticity, they would entail the need not to use them for any of the ordinary alphabetic functions of capitalisation when writing in the Kalasha language. For example, the marking of sentence boundaries would have to rely only on sentence-level grammatical punctuation (e.g. full stops, question marks) and not on the additional (redundant) word-initial capitalisation of sentence-initial words. However, the convention of sentence-initial capitalisation was deemed important by the majority of those who expressed a view at the conference. To deviate from this convention was perceived by them as a backward step, so they preferred not to depart from standard English practice on this matter.

B. APOSTROPHES FOR RETROFLEXION AND RHOTICITY

The apostrophe as a diacritic option for representing retroflexion and rhoticity in Kalasha was first proposed by Hamilton in 2000, and incorporated in Cooper, Hall, Cooper, & Hamilton (2000).⁹ Using this convention, the retroflex consonants in Kalasha would be written thus: *c'*, *d'*, *j'*, *r'*, *s'*, *t'* and *z'*, and the rhotic vowels of Kalasha would be written thus: *a'*, *e'*, *i'*, *o'* and *u'*. Though the apostrophe is a word-level punctuation mark in English, it is not needed for this purpose in Kalasha. It avoids the abovementioned problems specifically associated with capitals, and unlike some other diacritics, it does not occupy extra vertical space. It is also part of the regular Roman-alphabet keyboard (also meaning that no special font is required).

Another possible advantage of the apostrophe for retroflexion is that, being a diacritic device, it reflects a familiar facility for representing retroflexion in Arabic script, which also entails the insertion of a diacritic over the appropriate letters.

One specific problem with the use of the apostrophe has been observed. After handwriting a word in cursive script there is often not enough space left between letters to insert an apostrophe, which can occur almost anywhere (sometimes in more than one position) in the word. While the English apostrophe usually precedes *s*, its usage for Kalasha retroflexion would require it to occupy very narrow spaces between all sorts of letters (in both handwritten and printed texts), some tall and some very close. There appears to be no significant difference between the use of a shift key for capitalisation before typing a letter, and the use of the apostrophe key after typing a letter—both involve two keystrokes.

The delegates at the 2000/2001 Kalasha orthography conference preferred to use the apostrophe rather than continue with the convention of capitalisation. The main reason given was the potential confusion between a phonemic function of capitals to mark retroflexion and rhoticity and other useful functions of capitals that they deemed important enough to want to preserve (especially for marking the beginning of sentences and proper names). A review and assessment of this matter was planned as an agenda item for discussion at the second Kalasha orthography conference.¹⁰

On a typographical note, especially for publishing purposes, the apostrophe symbol ('), which is usually the same as the English single closing quote mark, could easily be substituted

⁹ A precedent exists for this way of modifying a series of consonants in 'the use of the prime (') to mark a palatalized consonant in common transliteration systems of Russian' (J Baart 2003, pers. comm.).

¹⁰ At the time of writing (2005) this has not yet been convened.

by a similar symbol, for example the prime (') or the grave accent (`). Various options were presented to and discussed by the Kalasha with a view to choosing between them as a means of marking retroflexion and rhoticity. However, in the end they decided that neither of them should be prescribed, and that they could be used in free variation. They anticipated that the fonts available and to be used by them would vary, and that when handwritten, they would not necessarily be distinct.

C. RETROFLEXION AND RHOTICITY IN COMPLEX GRAPHEMES AND GRAPHEME CLUSTERS

Complex graphemes include (1) *digraphs*, (2) *aspirated consonants* (which could be represented as digraphs or with diacritics—as discussed in subsection 7.2.6 below), and (3) *nasalized vowels* (which would include diacritics—see subsection 7.2.5 below), whereas **grapheme clusters** include any submorphemic sequence of contiguous graphemes. The question arises here because the convention for marking retroflexion or rhoticity could be applied to either or both elements in digraphs, and could interfere with diacritics used to mark some other feature (e.g. nasalized vowels or aspirated consonants). Using either method of placement, how would we mark complex graphemes or grapheme clusters for retroflexion and/or rhoticity?

(1) For most *digraphs*, this is not an issue, because none of the characters chosen to represent retroflex or rhotic phonemes in Kalasha is a digraph. Though the alveolar affricates / \widehat{ts} / and / \widehat{dz} /, would be written as *ts* and *dz* orthographically, their retroflex counterparts / $\widehat{tʂ}$ / and / $\widehat{dʂ}$ / would be written simply as *c'* and *j'* respectively, after the representation of their postalveolar counterparts / $\widehat{tʃ}$ / and / $\widehat{dʒ}$ / as *c* and *j* respectively. And though the postalveolar fricatives / $\widehat{ʃ}$ / and / $\widehat{ʒ}$ /, would be written as *sh* and *zh* orthographically, their retroflex counterparts / $\widehat{ʂ}$ / and / $\widehat{ʐ}$ / could be written simply as *s'* and *z'* respectively.

(2) There are three retroflex *aspirated consonants* (i.e. the retroflex versions of / $\widehat{tʃ}^h$ /, / $\widehat{tʰ}$ / and / \widehat{d}^h /). Should the aspiration symbol also be marked for retroflexion or not? For example, using the digraph alternative for aspiration, would we write 'field' as *c'hetr*, *ch'etr* or *c'h'etr*? (Or using the diacritic alternative for aspiration, would we write it as *c^hetr*, *c^h'etr* or *c^h'etr*?) From a phonetic point of view, only the first element of aspiration digraphs should be marked for retroflexion, because it carries the primary phonemic information (being concerned with articulatory position). The second element represents aspiration (which is phonemically secondary, being concerned with egression of air). Retroflexion also precedes aspiration in the

actual production of speech. The precedent set by Urdu in Arabic script also suggests that only the first letter of a digraph (in that case, aspirated consonants) should be marked for retroflexion. When presented with this issue at the Kalasha orthography conference, all participants also preferred to mark only the first element of aspiration digraphs, illustrating perhaps their intuitive grasp of some of these arguments.

(3) The chosen method of marking *nasalized vowels* is a tilde over the top of the letter. This presents no problem with the rhoticity diacritic which follows the letter, for example, *ã'*, *ẽ'*, *ĩ'*, *õ'*, *ũ'*.

In **grapheme clusters** with two or more contiguous retroflex or rhotic segments the Kalasha prefer to mark each retroflex or rhotic phoneme separately in their correspondence (e.g. *prus't'*, meaning 'good', and *az'a'i*, meaning 'apricot').

7.2.4 Velarized lateral approximant

The lateral approximants /l/ and /ɭ/ are the only Kalasha phonemes for which a single English letter (*l* or *L*) could be used as a base form, where neither member of the pair is retroflex, rhotic, nasalized or aspirated. This means that for one of these phonemes we need to either invent a new symbol or mark or modify the letter *l* with another orthographic device that does not mean retroflex/rhotic, nasalized or aspirated. Unless we are to invent yet another new symbol or diacritic, it seems best to use the regular retroflex notation (exceptionally) for this as well, that is, to mean velarized (as opposed to palatalized) with regard to lateral approximants. The justification for this is that velarization is like retroflexion, since both involve a retraction (of the tongue) in the point of articulation (Clark & Yallop 1995, p. 65). Moreover, there is no retroflex counterpart of the lateral approximant phoneme /l/ to confuse it with. Although this is a phonetic explanation, which would not be appreciated or even understood by most Kalasha, they perhaps had an intuitive sense of the similarities. At any rate, it was confirmed by a majority vote at the orthography conference in favour of this proposal, namely *l* for /l/ and *l'* for /ɭ/.

7.2.5 Nasalized vowels

The most common symbol already used for nasalized vowels in Roman alphabetic systems is the superscripted tilde (˜), and this convention was proposed for representing nasalized

vowels in the Kalasha language. Most Kalasha participants at the orthography conference agreed that it looked foreign, though some interpreted this negatively (because it is not indigenous) and others positively (for its distinguishing value). However, in the absence of any agreed alternative at the time, the majority of conference delegates accepted this convention.

The issue of fonts for electronic media comes in here, because symbols with overhead tildes have to be specially inserted, and overtyping is not a straightforward process with standard word processors. A special symbol set may be used to produce these letters thus: *ã, ê, ï, ô, û*. However, where this is not possible for typographical reasons, a suggested alternative is to type the tilde following the vowel symbol, thus: *a~, e~, i~, o~ and u~*.

Since the Kalasha have chosen diacritics to represent both retroflex/rhotic and nasalized phonemes (apostrophe and tilde respectively), the question arises as to how to write those phonemes that possess both qualities, namely the five nasalized-rhotic vowels, in those alternative situations when the tilde also has to be written after the letter. Which diacritic should precede the other? Alternatively, should one be superimposed on the other? The latter option might seem appealing in the interests of economy of space, if it were not for the need for a special font.¹¹

From a phonetic point of view, rhoticity is a primary factor because it is concerned with the point of articulation, preceding secondary factors like nasalization, which are concerned with manner of articulation. The example of Urdu in Arabic script also suggests that the marking of rhoticity could take precedence, as Arabic-script retroflexion is marked on the letter, while nasalization follows it. Where special fonts like superimposed tildes are not available to depict nasalization, for example, in emails and chat sessions, it is preferable that the rhoticity diacritic also take precedence from a reading point of view, because the apostrophe, being a smaller, more compact diacritic than the tilde, would be more visible when closer to the letter, where it has a better chance of being noticed. Using these simple fonts it was found that the majority of the Kalasha orthography conference participants also (without any prompting) preferred to write rhoticity before nasalization, again illustrating their intuitive grasp of some of these arguments and/or their familiarity with parallel facility in Arabic-based Urdu orthography. Where these typographic fonts are not possible, the set of vowel phonemes that are both rhotic and nasalized would be written thus: *a'~, e'~, i'~, o'~, u'~*,

¹¹ See the discussion on vertical and horizontal space required by diacritics at the beginning of this chapter (subsection 7.1.4).

u'~. However, where they are available, the diacritics would be reversed in order of priority, for reason of economy of space, thus: *ā'*, *ē'*, *ī'*, *ō'*, *ū'*.

The fact that, in some correspondence, the Kalasha mark only one vowel (per word) with a nasalization symbol, for example, *ishpō'i* ('flute'), *pac'hīak* ('bird'), *ga'wa'hī'ak* ('rabbit'), confirms the phonemic differentiation of nasalized and oral vowels. In other cases, where two phonemically nasalized vowels do occur together, Kalasha respondents will write both as nasalized, for example, /ãã/ 'yes'.

The fact that the majority of Kalasha writers at the conference verbally voted not to mark vowels that are contiguous to a nasal consonant, for example, *no'a* ('under') and *moa* ('maternal uncle'), also supports the phonological interpretation that vowels in these environments are phonemically oral. This practice would be a deep representation of the phonology, deviating from Urdu, where vowels contiguous to nasal consonants are written as nasalized vowels.

7.2.6 Aspirated consonants

How can aspiration be written with a Roman alphabet? In Urdu script, aspiration is marked by writing the thirty-fourth letter of the Urdu alphabet (one of the two letters in Urdu that are used for the phoneme /h/) immediately following the letter that signifies the phoneme being aspirated. Normally this letter is called {ha-e-do-tʃʌʃmi}, literally 'h-of-double-goggle', because its non-word-initial form resembles a pair of double-lens goggles. When used as an aspiration marker it is called {ha-e-mʌxlut-ut-tə'lʌffəz}, literally 'h-of-mixture-of-pronunciation', or, more specifically, {ha-e-məl'fuzi}, literally 'h-of-aspiration'. In an early version of an Arabic-based Kalasha alphabet (Cooper, G 1986b), this same Urdu letter was adopted to represent not only Kalasha aspiration, but also the Kalasha phoneme /h/ (instead of using a separate symbol, as in Urdu).

This economy of symbolisation was trialled by Kalasha individuals for over ten years with no observed or reported confusion or ambiguity. As a result, and based on the success of that experiment, it was proposed as the most natural, efficient and convenient way of marking aspiration using a Roman alphabet, to use the (non-superscripted) Roman symbol for the phoneme /h/. Not surprisingly (given that they were used to this concept in principle), the Kalasha agreed to this proposal at their conference.

This means that the aspirated consonants in Kalasha would be represented as eight alphabetic digraphs: *bh*, *ch*, *dh*, *gh*, *jh*, *kh*, *ph*, *th*, as well as *tsh*. (The last would actually be a trigraph if *ts* was already established as a digraph for the Kalasha affricate \widehat{ts} /, as proposed in subsection 7.2.2 above.) Four of these orthographic sequences could be thought of as ambiguous by some, because of their phonemic meanings in English: *ch*, *ph* and *th* are written for the English phonemes $\widehat{tʃ}$ /, /f/ and /θ/ respectively, and *tsh* for the English phoneme sequence /tʃ/ (as in ‘potsherd’ and ‘boatshed’). However, the English interpretation of the first of these (*ch*) would not alter its Kalasha phonemic meaning, since the digraph *ch*, in the pre-vocalic environments in which it can only occur in Kalasha, is always phonetically aspirated in both English and Kalasha anyway ($[\widehat{tʃ}^h]$). The English meanings of the digraphs *ph* and *th* are generally not well understood by non-English speakers, so their phonemic associations are relatively weak. In addition, the trigraph *tsh* is very rare in both English and Kalasha, so no serious problem of ambiguity would arise.

The convention of using the symbol for /h/ to denote aspiration has also been used in the Indological font (used for other Indic languages) that employed a Roman writing system, for example, the Sylhetti orthography developed to provide a ‘close correspondence with Bengali transliteration and orthography for easier reference in a Bengali dictionary’ (Lie et al. 1999). Other languages of northern Pakistan with emerging Roman-based orthographies have also followed this same principle.

The convention of superscripting the *h* to show that it is part of a single, aspirated phoneme segment, as in Pike (1947, p. 7) and the IPA, would be less practical to use in a new orthography, because it requires extra font adjustment on computers.

7.2.7 Phonemic stress

Stress does not feature regularly in most European languages using Roman orthographies. Neither is it marked in Urdu, Farsi, Arabic, etc., because in these languages it is not phonemic. Neither is it in Kalasha, except for a few isolated examples where pairs of words are phonemically identical but differ only in stress placement, for example:

[ʼaʃa] ‘mother’	[aʼja] ‘here’
[ʼtari] ‘star’	[taʼri] ‘sugar’
[ʼpara] ‘you(pl.) (will) go’	[paʼra] ‘I/you(sg.) went’.

In view of the rarity of this phenomenon, and the fact that grammatical contexts would largely disambiguate these cases, I suggest that it need not be marked generally. Where stress needs to be marked for pedagogical purposes (e.g. for use in primers and early readers) an ad hoc method of marking can be used, such as underlining of the stressed vowel or the stressed syllable. Unlike bolding or italics, this can be performed in handwriting as well as on computer, and unlike the single quotation mark or accents, would not be confused with other orthographic markings, such as the apostrophe for retroflexion and rhoticity. In the *Kalasha Dictionary* (Trail & Cooper 1999) stress was marked by an acute accent, over the stressed vowel of every polysyllabic word, because this work was written for non-native speakers of Kalasha.

7.3 CONCLUSION

Throughout this chapter, the finer points of Kalasha orthography, based on the Roman alphabetic system, have almost all been made to conform (by the Kalasha speakers themselves) to the English alphabet, and to English spelling and writing conventions more broadly (with the addition of devices for marking retroflexion, rhoticity and nasalization). This is because of their rapidly increasing exposure to, and interest in, the English language, especially in the schools, and of the ever-increasing impact of the Internet, especially email communication, even on this remote village community.

In summary, to represent an inventory of Kalasha phonemes using the symbols, digraphs and diacritics discussed throughout this chapter, we arrive at a comprehensive *orthographic inventory*, presented here in the recommended order:

**a a' ã ã' b bh c c' ch c'h d d' dh d'h dz e e' ě ě' g gh
h i i' ĭ ĭ' j j' jh k kh l l' m n ng o o' ō ō' p ph r r'
s s' sh t t' th t'h ts tsh u u' ũ ũ' w y z z' zh**

This orthographic inventory is based on the following set of 22 base forms:

a b c d e g h i j k l m n o p r s t u w y z

The question now remains as to whether the official Kalasha alphabet should be taught as a set of over 60 graphemes (representing most or all of the phonemes) or 22 letters (representing the orthographic base forms). The former approach is a maximal one, taking a shallow, aurally oriented perspective, whereas the latter approach is minimal, with a deep, visually oriented perspective.

Alternatively, could there be some compromise between these two sets? For example, the Urdu alphabet adds to its set of orthographic base forms only graphemes that represent the *retroflex* phonemes of Urdu, because these graphemes are orthographically single elements, whereas graphemes for *aspirated consonants* and *nasalized vowels* are not included in the Urdu alphabet because they are digraphs. Early Roman-script literacy materials in Kalasha, have been influenced by previous long-lasting trials of Arabic-script orthography, by following yet another middle path between this Urdu precedent and the maximal approach mentioned above. To the 22 Kalasha base forms were added alphabetic letters for 7 retroflex consonants, 3 to 5 digraphs, 5 rhotic vowels, and the 1 velarized lateral approximant, resulting in a set of 38 to 40 letters. Only the aspirated consonants and nasalized vowels were excluded from these early versions of the Kalasha alphabet.

On theoretical grounds, it seems best that the official Kalasha alphabet be formulated as the set of 22 letters (the orthographic base forms) for the following reasons:

(1) An alphabet comprising a minimal base set breaks the pedagogical process into stages (representing basic oral, retroflex, rhotic, aspirated, and nasalized phonemes in Kalasha). The graphemes not included in the official Kalasha alphabet would be taught as part of the spelling (in the same way that the English digraphs *sh*, *ch*, *th*, etc., although representing English phonemes, are taught subsequent to the English alphabet). This approach has been proposed by one of the Kalasha teachers, Anis Umar, who wrote (in a fax dated 2 May 2002):

I think that we must create an easy way for the Kalasha teachers and students, so they can accept the new alphabet. If there are many letters they will be afraid to learn it. On the other hand we must have letters for every Kalasha sound if we want to save our language with the Kalasha accent (sic). My proposal is to divide the alphabet in[to] five lists.

Anis Umar's lists are the 22 basic letters 'for someone who wants to start learning', plus the retroflex and rhotic phonemes, the digraphs, the nasalized vowels and the aspirated phonemes. At least one other teacher has also since expressed his reservation concerning the inclusion of digraphs in the alphabet, for the same pedagogical reasons.

(2) The exclusion of digraphs from the Kalasha alphabet follows the precedents of English, Urdu and many languages, whose alphabets also contain no digraphs or diacritics, even though that means that certain phonemes are not represented in the alphabet. Essentially, the alphabet functions quite well as a set of unique symbols that represent *most* phonemes in a language. Its ultimate quiddity is not that it is a complete inventory of orthographic

equivalents to every single phoneme, but rather that it is a practical and useful set of symbols which (even in combination) depict the language efficiently.

(3) Most of the Kalasha phonemes requiring digraphs and diacritics to represent them have very low functional loads (less than 1% occurrence in natural text, according to analysis summarised in chapter 4, subsection 4.3.3), so their exclusion from the alphabet for the above two reasons, will have no serious consequences.

Despite these arguments, there has been some degree of ongoing discussion relating to the issue of reducing the alphabet, so it will no doubt warrant review at a future Kalasha orthography conference.

FROM LETTERS TO LITERATURE

In Chapter 6 a Roman alphabetic system was chosen to represent the Kalasha language, and in Chapter 7 we addressed some of the finer points of writing Kalasha with a Roman alphabet and according to Roman orthographic conventions. We also incorporated some elements of English orthography that the Kalasha have adopted, in line with the increasing use of English in their education system. What remains now is to look at the implementation of the new Kalasha orthography as a medium for Kalasha literature. The implementation includes technical and technological issues of typography (section 8.1), desktop publishing (8.2), Internet accessibility (8.3), literacy initiatives (8.4), the strategic issues of planning literature, (8.5), discourse types and genres (8.6), and a suitable range of these genres for the new body of Kalasha literature (8.7).

8.1 TYPOGRAPHICAL REQUIREMENTS

Orthography decisions have typographical implications. The availability, constituency and usage of fonts are the chief considerations in this regard. (The focus here is not on display font varieties, but rather on the need for special symbols that some fonts contain along with their standard set.) Dependence on special fonts can be avoided at the foundational stage of orthography development if the ordinary symbols of the Roman alphabet are used with minimal diacritics. However, some diacritics are unavoidable to reflect and distinguish the full range of 62 phonemes of Kalasha (including 20 vowels). The use of the apostrophe symbol to mark retroflex consonants and rhotic vowels, and the use of the tilde to mark nasal vowels, are principles coming out of Chapter 7 that will keep the proposed Kalasha typography to a limited set of symbols. The Kalasha community's adoption of many English orthography conventions in the writing of their language, and the availability of conventional English keyboard resources means that typographical solutions are ready to represent their indigenous phonology.

8.2 DESKTOP PUBLISHING

A writing system needs to have workable publishing software available. The Roman alphabet is used in the widest range of software, maximising opportunities for linguistic data analysis, text processing and development of literacy materials in minority languages.

It might be argued that most communities that are just developing a new orthography for the first time are not likely to have reached anywhere near the level of technological sophistication where computers are used for print or electronic publishing. However, a growing number of Kalasha individuals are becoming computer literate, and will want to engage in the desktop publishing of their own indigenous texts, using whatever software is available, easily accessible and user-friendly. Currently, and for the foreseeable future, this software is in English.

Rhydwen (2000, pp. 104–5) testifies to the impact of computers with English software for Aboriginal literacy workers at the School of Australian Languages:

Almost all computers use English software ... Yet Aboriginal literacy workers, who ... were [previously] discouraged from developing literacy skills ... were delighted by the advent of computers. Not only did they enable them to write at all, but they enabled them to produce low-cost literacy materials on-site in the local language for use in local schools.

A key to the Kalasha language becoming publishable in a similar way would be the acquisition of appropriate technology and skills for desktop publishing and printing. Hardware and software that suit their third-world situation need to be found (donated to the community or cheaply acquired by them), not necessarily state of the art, but affordable and requiring minimum technical assistance. Such technology might include typewriters (still very common in Pakistan), screen-printing apparatus, photocopiers, and basic computers (with appropriate software for word processing, desktop publishing and file management). These low-tech solutions and techniques could be used as a primary, interim and/or peripheral technology. The idea of establishing of a publishing house has been suggested to them, where all such technologies could be used communally, where training and experience in publishing could be shared, and where literacy and training materials, as well as various types of literature, could be produced.

8.3 INTERNET ACCESSIBILITY

Roman alphabetic orthography has far better potential than Arabic script to interface with the Internet. The alignment of the vulnerable Kalasha language with a writing system that will carry it furthest into the global network of human interaction will not only maximize communication opportunities, but also establish the language as a valid and viable medium of written discourse.

Being a third-world country the private affordability of computers in Pakistan is a luxury compared to the West. For this reason, Internet cafés are now ubiquitous in Pakistani cities and towns. Thousands of individuals, especially young men, regularly frequent these Internet cafés to browse the web, chat and engage in email correspondence. There are now Internet cafés even in the frontier town of Chitral, just two hours by jeep from the Kalasha valleys. In mid 2000 only one Kalasha person had an email account, but by the end of that year, six had acquired them. This trend illustrates the acceleration of technological change and of community attitudes and practices that result.

The new modes of human interaction which result from email technology, and their compatibility with traditional ways of expression in an oral-aural culture, is the subject of a paper on the impact of email on Iroquois speakers of Mohawk (MacDougall 2000, pp. 91–8). They reported a tension between the way they thought and how their language appeared on screen. The linear format of the written language promoted logical thinking and a heightened self-awareness. Despite the unprecedented nature of these experiences, the computer-literate Iroquois enjoy the functional advantage of being ‘heard’ over large distances rapidly. The impact of email technology on the Kalasha community may be similar to this, in that they too are traditionally an oral-aural culture.

Though it will still be a long time before many Kalasha people own a computer, these newly computer-literate Kalasha individuals are already writing email messages in their own language using variations and approximations of the orthography developed in this thesis. Their use of communal computers in the Internet cafés of Chitral town is facilitating their communication with the outside world, as well as enabling them to read with relative ease, in English, information that is posted on the Internet from the outside world.

8.4 LITERACY INITIATIVES

The new Kalasha orthography is very likely to establish an interest in and a motivation for vernacular literacy. Motivation is the single most important factor in acquiring literacy, even more important than the teaching methods, the quality of teaching, the capability of the teacher, or the adequacy of the writing system. However, these other factors do boost and foster existing motivation (Lee 1982, p. 9).

Interest in Kalasha literacy started slowly (see a description of early literacy initiatives outlined in Chapter 1, section 1.4), but accelerated in the lead-up to the orthography conference (discussed in Chapter 6). In the Kalasha Valleys there is now enough momentum and interest in education generally to ensure the successful teaching of literacy in both Urdu and English, through the schools. The educational processes of reading and writing are being successfully taught in these schools, so the principles of transfer from English to Kalasha literacy can easily be taught through carefully researched and tested literacy materials.

An application to the NWFP government requesting its sanction for the official inclusion of Kalasha literacy as a subject in the school curriculum is still awaiting approval. At the time of writing (2005), there has been no significant progress in this matter. In the meantime, some Kalasha teachers have planned to hold literacy transfer lessons, through extracurricular classes, to be the primary means of teaching Kalasha literacy.

In 2003 the schools were equipped with the *Kalasha alphabet* book that was published in 2000 using the newly developed Roman writing system, that was published in that year. Subsequent use of all the above reading materials has been very enthusiastic, but very intermittent, due to the lack of follow-on materials, which is due, in turn, to the lack of resources for their production. In 2004, two young men (one of them self-taught in Kalasha literacy and the other educated to tertiary level) began conducting an itinerant Kalasha-literacy teacher-training program on a voluntary basis, using materials that my colleagues and I have developed. However, negative attitudes subsequently developed among some of the Kalasha teachers (employed by the government) who saw themselves as bypassed, and their social status as threatened. The government teachers are willing to continue promoting Kalasha literacy, but they expect to be remunerated for their services, which at this stage would have to be outside of their official school hours.

8.5 LITERARY PLANNING

Having established that motivation is the single most important factor in the success of a literacy program (see previous section, 8.4), Lee (1982) then takes a step back to argue that the single most important factor in motivating speakers of a language to become literate in the first place, is the existence of a body of literature that they want to read. A body of literature to access is even more motivational for literacy than a new orthography. To illustrate this point negatively Lee cites the following case:

Wendell (1975) discusses the problems of trying to motivate Indians of Mexico to learn to read when there was nothing for them to read except literacy primers ... If there is not an accompanying body of literature which is considered relevant, the motivation is likely to diminish before full literacy is attained.

This has been a critical issue in the Kalasha society where, for over 15 years, only tentative literacy materials have been available, because texts that have been collected over that time for a body of indigenous literature could not be mass produced, due to the lack of a standard Arabic or Roman orthography. The risk of diminishing motivation among the Kalasha community was a real one, until a plan was developed to settle the orthography, of which this thesis is the research dimension.

A useful consideration in the planning and pursuit of this goal of literature development is the role of genres. Matthews (1995, pp. 35–46) has postulated a genre-based approach to literacy. Her basic premise is developed thus:

Spoken language is between a speaker and audience who are physically close at a particular time ... Written language, however, separates the speaker and hearer by time and location ... As ethnic groups come in contact with a dominant literate culture they are faced with the need for language forms not bound by time and space. As new genres are accepted from the literate society into the vernacular language, there is an increasing need to master the genres of the dominant culture. Increasing mobility and culture contact mean that face to face modes of language are no longer adequate.

Since genres may be oral or written, the relevance of Matthews' premise is that some of the genres adopted from a dominant culture will necessarily be written genres, which were obviously never known or used before in the preliterate ethnic group. Matthews applies the concept of genre theory to literacy, but it is equally applicable to literature development, not only as a process subsequent to a literacy program, but also as an anticipated process to

motivate a literacy program. A genre-based approach to literacy will create an expectation of a variety of literature types.

It is ultimately the newly literate community that will want to assign their own priorities to these genres or literature types, according to their own interests and cultural values. Little has been written about this in the literature, because field linguists are often more interested in analysing and/or documenting the minority language, than in literacy, and anthropologists in studying and documenting the culture. Literacy specialists usually follow a pragmatic approach that reflects and follows the newly literate community's expressed interests and preferences.

In a creative writers' training approach to literature development, Duerksen et al. (1989, p. 5) propose that 'production of easy folk stories from within the culture can be developed first' and also that 'some writing exercises should be included through which the writers explore their language by means of descriptive words, emotive words, colors, sensations, and so forth.' Then, focusing on the implementation of community development initiatives from outside the culture, they propose that those who are trained as writers will themselves read literature in other languages to find solutions to a variety of health, environmental and economic challenges. Relevant topics might include nutritional variety, agricultural techniques, animal health, reforestation, disease control, childcare, family planning, farm practices, retail management, safe water supply, sanitation, and waste management. 'In those areas where writers know there is a felt need and they have acquired significant helpful information, they will write booklets in their mother tongue on those subjects.'

Robyn Terrey, a literacy specialist who worked in the Philippines, indicated to me her ideal order of criteria for the introduction of indigenous literature: first the **predictable**, then the **relevant**, then the **new**.¹

The **predictable** category of literature includes fundamental texts that bridge the gap between oral and written genres. They can be predictable in the sense that both the form and content of newly written texts in this category are already familiar to readers, so they anticipate the text as they read, which aids in their recognition of frequently occurring forms. Examples of this approach include songs and stories that have been recited and heard many times before. Much of traditional Kalasha oral literature is embedded in these songs and recitations, and a small amount in formal banter and riddles. (Within the discourse scheme

¹ Telephone conversation in August 2002.

presented below these additional types of oral literature are not strictly genres, but rather surface structure media, serving as vehicles for various genres.)

However, literature can also be predictable if it is written in a style that leads readers to anticipate phonemes, orthographic letters, words, sentences, paragraphs or topics based on deliberate and repetitive patterns that are established by the author. Therefore, literacy primers can be designed to teach recognition of these elements by repetition and built-in redundancy. Some such primers have been drafted in Kalasha, using topics from everyday situations.²

Examples of the **relevant** category of literature that may be introduced into a literacy program were reported to me by Mary Stringer, a literacy specialist with SIL International (which has made vernacular literacy one of its primary goals).³ Stringer cited the experience of Robert Young, who worked in the Benabena language community (in Papua New Guinea) from 1956 to 1992:

Some narrative stories that some of the people wrote (edited for length) were popular. Then the Maths book was very popular because it helped the men to check the coffee prices. Robert wrote a diary of interesting things that he did with the young men and let them read it. That was popular because the young men were involved in the activities.

The pragmatic approach referred to here relies exclusively on the acceptance and popularity of various literature types by the community, thus foregrounding what is *relevant for them*. However, relevant literature may also include topics that are possibly less popular, but deemed useful (say, by an indigenous literary committee), in terms of the community's current situation or needs. Examples would include texts belonging to the Information genre, discussed further in section 8.6, below.

New literature can also be both popular and practical. Either form or content may be new, so it may feature either new topics in familiar genres, or familiar topics in new genres. The latter proved less successful in Stringer's experience of literacy development in the Waffa language of Papua New Guinea (from the 1960s to the 1980s):

We began quite early (as soon as we had readers and writers) with a weekly Newssheet ... including a description of a current, unusual happening in the village, a message of importance from another village ... a calendar for the month ... and any other topical information of interest.

² Examples of topics include conversations about getting ready to go somewhere, working together, making walnut bread, making each other's goats fight in a competition, ode to parents, festival song, personal advice, and tongue twisters.

³ Email received on 11 September 2002.

The outcome was not what the literacy agents expected. The villagers ‘did not want to read the articles, because they were about things that they already knew. They wanted to read about things they did not know!’⁴

This community apparently had a penchant for both topics and genres of novel interest. Of course, what is relevant or new for some of the community may not be so interesting for others. Potts (1987, p. 27) identifies the need for different literature types for different sectors of society:

Once an audience has been defined, members of that group can be encouraged to decide what kinds of literature they would like to read, and to formulate ways in which it can be produced.

In order to test interest in new genres, various sectors of a community can be introduced to different kinds of literature that could be available to them, if they so desired. A further discussion of traditional and new genres is taken up in the next section (8.6).

8.6 DISCOURSE TYPES AND GENRE CATEGORIES

Members of the newly literate community will want to determine the type, content and form of their own literature themselves. In Chapter 2, section 2.4.1, the importance of texts on the topic of local culture was highlighted by their special mention in the constitution of the indigenous Kalasha People Welfare Society. However, a full repertoire of potential literature types in Kalasha may be derived from the range of topics that the Kalasha find predictably familiar, relevant, novel and interesting to talk and/or sing about or recite in their oral culture.

Some of these topics are listed in the table below, classified into groups of discourse genres, which could just as easily apply to literature as they do to speech. The existence and use of various genres are windows into a culture—they realize cultural values. These genres are more fully described in Appendix 14, but they are classified here into major discourse types using part of Longacre’s (1983) framework of what he calls the four *notional types* of monologue discourse (Narrative, Procedural, Behavioural and Expository).

Reference is made to this framework here because Longacre posits it as a map of universal notional categories, a perspective from within which it should be helpful to view the range of Kalasha discourse (and literature) types.

⁴ Email, 11 September 2002.

Table 8.1 Some of Longacre's discourse types with corresponding genres and Kalasha topics

DISCOURSE TYPE	GENRE	TOPICS THAT ARE PREDICTABLE, RELEVANT AND/OR NEW FOR THE KALASHA
Narrative	Anecdote	personal affairs and anecdotes, romances, family relationships, Kalasha-Muslim relations, religious conversions, menstrual hut issues, debts, ceremonial purity, tourists
	Story	people in situations that invoke humour or wisdom
	Fairytale	encounters with fairies and evil spirits, occult phenomena
	Myth	mythology, supernatural history
	Fable	stories involving humans and animals which illustrate life situations
	Legend	ancient history, tales of former times
	History	heroes and heroic acts, merit feasts, mighty men of valour, war, kingdoms, major political allegiances, regional control, political and royal relationships and tactics
	Report	news, current affairs, social issues, land rights, logging, legal cases, weather
Procedural	Instruction	agriculture, arboriculture, construction (houses, walls, aqueducts), forms and records, goat herding, food preparation, weaving, religious practices
Behavioural	Exhortation	relationship advice, marriage guidance, behavioural expectations, ceremonial rites, festival participation, sexual taunts, prayers
	Eulogy	acts of courage, enterprise, generosity and heroism
Expository	Information	aid and development topics, apiculture, astronomy, biology, environmental issues, foreign customs and culture, health, hygiene, linguistics, literacy, medical (diagnosis, treatment options), nutrition, technology
	Description	customs and culture, anatomy, botany, cottage industries (honey, dolls, souvenirs)

Most of the genres listed in the above table are already used in traditional Kalasha oral literature, although Report, Instruction, Information and Description would be new genres, with discourse structures imposed from outside the culture, and requiring trained writers to draft them. Besides being able to write on any topic in these new genres, those trained as creative writers will also be able to write on new topics in any existing genre.

All the above genres vary in the way they would exemplify the arguments for vernacular literature mentioned in Chapter 2. Very broadly, all genres would be useful for the aim of preserving the language (subsection 2.3), and most of them would be useful for the purpose of cultural conservation (subsection 2.4.1). This is particularly important in the transmission of both language and culture to the younger generation (especially through the Narrative and Behavioural discourse types). The exceptions are that Information would best serve the goal of foundational education (subsection 2.5.1) and Anecdotes and Stories would be most useful for bilingual education (subsection 2.5.2). The new genres of Information and Description would be more beneficial for the purposes of social advancement (subsection 2.6.2). The new genre of Reports would be most useful for bilingual education (subsection 2.5.2).

In the details of the previous paragraph, we have a comprehensive answer to the question posed in Chapter 2 (*Should languages be codified in writing?*). A different part is played by each of the genres in establishing the case for writing the Kalasha language. The traditional genres occur naturally, and the new genres occur in response to a changing world. The survival of all of these genres consolidates the versatility and richness of the language, and strengthens its ability to survive.

8.6.1 The default discourse type

In his classification of discourse, Longacre (1983, pp. 3–7) maps the monologue discourse types onto a grid of two basic parameters. The first of these he calls *Contingent Temporal Succession* at the notional, deep/semantic level, or *Chronological Linkage* at the surface level. Both Narrative and Procedural discourse types are characterized by this parameter.⁵ However, Behavioural and Expository discourse types do not feature any contingent temporal succession or chronological linkage. They are characterized instead by logical (including topical) linkage.

The second of Longacre's two monologue discourse type parameters he calls *Agent Orientation* at the notional, deep/semantic level, or *Participant Reference* at the surface level. Both Narrative and Behavioural discourse types are characterized by agent orientation and participant reference, whereas Procedural and Expository discourse types do not feature this parameter.

This grid of parameters is predictably reflected in the actual set of Kalasha genres. **Narrative** discourse genres are characterized, strongly and markedly, by both contingent temporal succession (with very clear and obvious tail-head chronological linkage, based on participles) and agent orientation (using a rich set of pronouns and participant case suffixes). An example of chronological linkage is worth highlighting here because of its importance in the scheme of Kalasha genres. Tail and head linkage in this and the following example is shown with consecutive pairs of underlined words in the Kalasha transcription, as well as in the English glosses:⁶

⁵ Longacre further classifies chronological linkage into three configurations, in terms of the way that elements of consecutive paragraphs relate to each other: head-head, tail-head and summary-head.

⁶ These texts also appear in Appendix 15, with Kalasha orthographic transcriptions added.

/se motʃ to gordok k^hodʒik-a defa paraw. bo defa pa-i sapraw.
 ‘that’ ‘man’ ‘that’ ‘donkey’ ‘searching-to’ ‘far’ ‘went’ ‘very’ ‘far’ ‘Go-ing’ ‘found’
 ‘That man went far to search for the donkey. Going very far, he found it.’

sapra-i o gordok amaraw. mar-i o tasa mos aʃaw. aʃ-i abomaw./
 ‘Find-ing’ ‘then’ ‘donkey’ ‘killed’ ‘Kill-ing’ ‘then’ ‘its’ ‘meat’ ‘ate’ ‘Eat-ing’ ‘vomited’
 ‘Finding the donkey, he killed it. Having killed it, he ate its meat. After eating he vomited.’

Procedural discourse in Kalasha is characterized only by contingent temporal succession, with the same use of tail-head chronological linkage as for Narrative discourse, but with no agent orientation.⁷ The following text about making rope is an example:

/awel dʒats tʃ^hiʃin. p^hato to ʃa-haj karin. ʃa-haj ka-i to o
 ‘firstly’ ‘goat hair’ ‘card’ ‘then’ ‘it’ rolled’ ‘do’ ‘Rolled’ ‘make-ing’ ‘it’ ‘then’

sawza-i kondʒin. kond-i to du lifen. du lifa-i to o ʃen./
 ‘make-ing’ ‘spin’ ‘Spin-ing’ ‘it’ ‘two’ ‘twist’ ‘two’ ‘Twist-ing’ ‘it’ ‘then’ ‘ply’

‘Firstly, they card the goat hair. Then they make it into a roll. Having made it into a roll then they spin it. Having spun it, they twist it into two strands. Having twisted it they ply it.’

Behavioural discourse in Kalasha is characterized only by agent orientation, with highlighted and repeated participant reference, but with no chronological linkage. For example:

/ʃabaʃ o taj hatja ej maj baja mutimire nawaw/
 ‘bravo’ ‘oh’ ‘your’ ‘sake’ ‘hey’ ‘my’ ‘brother’ ‘Mutimire’ ‘grandson’
 ‘Bravo to you, oh my brother, grandson of Mutimire’

⁷ This is similar to the absence of agent orientation to accompany imperative verb forms of English procedural discourse, such as instructional text.

Expository discourse in Kalasha is characterized by neither of Longacre's parameters. For example:

/biriu-ay motʃ bo ʃiʃojak. istriʒa asta bo ʃiʃojak. biri-ay ek asta
 'Birir-in' 'people' 'very' 'beautiful' 'women' 'also' 'very' 'beautiful' 'Birir-in' 'one' 'other'
 bo maʃur iʃnehari shiau. se o ɖa. saw-in pi ʃia iʃnehari bo pruʃt.
 'very' 'famous' 'thing' 'is' 'that' 'one' 'wine' 'all-'s' 'from' 'this' 'thing' 'very' 'good'

'The people in Birir are very beautiful. The women are also very beautiful. In Birir, there is one other very famous thing. That is wine. Of all things this is very good.'

Although expository discourse in Kalasha features neither chronological linkage nor participant reference, the above example does however demonstrate the liberal use of textual cohesion devices that can make this discourse type very lively. The expressions /asta/ ('also' and 'other'), /se o/ ('that'), and /ʃia/ ('this') provide the *topical* linkage that was mentioned at the beginning of this subsection as a characteristic of expository discourse. Furthermore, the word /biriu/ ('Birir') somewhat replicates participant reference by its repetitive foregrounding at the head of two of the sentences.

The fact that the Narrative discourse type is the only one that is positive for both of Longacre's basic discourse parameters might suggest that Narrative is a default discourse type. Narrative, especially ad hoc oral narrative monologue, is the only discourse type where both of these parameters come into play in a natural way.

The importance of contingent temporal succession has been demonstrated by Linde (1981, pp. 84–114) as one of the primary principles of discourse coherence. Its surface structure equivalent, chronological linkage, is constructive in the true sense of the word because it puts together unstructured discourse elements into a structured framework. Chronological linkage is not always forward, linear or paramount in a given discourse, but it provides a default system for the expression of oral literature.⁸

⁸ Chronological linkage can also go back and forth, follow a circular or spiralling path, and/or be framed by a logical/topical linkage system at a higher level.

The relevance of chronological linkage to the development of literature in a newly written language is that it provides the structure and momentum that is necessary to sustain fluency for new topics and/or genres in the new written medium of language. It could equally be said of the participant reference parameter, which intersects with chronological linkage in the Narrative discourse type, that it is necessary to sustain maximum human interest for new topics and/or genres in the new written medium of language. These characteristics of Narrative would suggest the need to foreground it as a discourse type in any planning of a literature repertoire for a newly written language.

8.7 A KALASHA LITERATURE REPERTOIRE

What follows is a proposed order in which various genres of indigenous literature could be introduced into the Kalasha community in particular, in a way that would maximize the existing momentum of their enthusiasm and interests, and the potential for development of literacy skills.

Traditional genres:

1. Story, Fable, Anecdote
2. History, Fairytale, Myth, Legend
3. Exhortation, Eulogy

New genres:

4. Report, Description, Instruction
5. Information

It is to be expected that the traditional genres would be introduced first, as these are more predictable and familiar. Most of these genres belong to the Narrative discourse type. They are very predictable in length and consist of an inherent structure of elements (e.g. clear chronological linkage and regular participant reference). The new genres are unpredictable in length and consist of imposed discourse structures (e.g. limited chronological linkage and little or no participant reference), making them more of a challenge to read and write in Kalasha. However, experience in reading will be quickly gained with practice, and as mentioned in the previous sections (8.5 and 8.6), writers could be trained in the adoption and application of new genres.

The subject matter for category 1 genres (from the above list) is common property among the Kalasha, and can therefore be documented by any educated individual. Categories 2 and 3

genres embody oral literature that is vested in certain sectors and individuals within the society, for example, elders and orators, who are the traditional custodians of the assets of Kalasha cultural lore. Probably the most significant of all Kalasha oral texts are performed in song at festivals. They include eulogies and are extremely solemn and respectful in tone. In stark contrast to these are other songs, also in the same two categories, but performed on different occasions, which are extremely flippant and coarse. (If Kalasha poetry existed it would also fit into these two categories.) Even one of the younger members of the Kalasha community has affirmed the value of maintaining the Kalasha oral culture in written form. His expression of support for this endeavour forms the epilogue to this thesis. The transmission of texts in categories 2 and 3 genres has traditionally been oral, so a new written medium for the Kalasha language opens up opportunities, on the one hand, for increased current exposure and preservation for posterity via a new mode of transmission, but on the other hand, potential fragmentation of the traditional mode and practice of oral transmission. Yet, only one of the custodians of Kalasha oral literature appears anxious about such a change, because he makes a living from selling his oral knowledge to foreign anthropologists.

In order to maximize the positive opportunities of a new written medium, the Kalasha People Welfare Society, an indigenous NGO (non-government organization), has a stated goal to document as much of the traditional Kalasha heritage as possible (belonging to genre category 2 above). To this end, and to address any negative consequences of a new written medium, Kalasha teachers have taken the initiative of introducing an extra-curricular school subject, comprising topics of Kalasha heritage and culture, involving several Kalasha elders who have accepted the responsibility to impart their knowledge verbally to the students. This they are glad to do, as it is an immediate and measurable objective in the perpetuation of their heritage, independent of any written medium (though they have also welcomed the prospect of an alternative medium for their knowledge).

With respect to the new genres, the documentation of topics in category 4 (from the above list) would be more the domain of occupational specialists (e.g. goat herders, farmers and weavers), whereas topics in category 5 genres would more likely be introduced by trained writers, adapting texts and knowledge from outside the culture.

Several Kalasha texts from all the traditional genres have already been committed to manuscript, using either an Arabic script or an interim Roman orthography (both pioneered in

the 1980s).⁹ Texts in either writing system can be converted and orthographically regularized by computer to the new, standard Roman-based Kalasha writing system as developed in this thesis. It will be a relatively simple matter to prepare these already drafted texts for publication, alongside the introduction of newly drafted texts.

However, the potential for the successful development of Kalasha literature lies not in the linguist's skill in designing or modifying an indigenous orthography, nor in the pens of Kalasha writers, crafting texts that are meaningful and relevant, but rather in the community of Kalasha readers and writers. They will invest meaning into indigenous literature by recognising it as a new medium, with a variety of (old and new) genres, through which can be woven all the colours and textures of the human intellect.

⁹ The Kalasha alphabet that I developed in Arabic script appears as Appendix 16. The interim Kalasha alphabet that I developed in Roman script is the same as presented in Chapter 7, except that capitalization was the device I used to mark retroflexion and rhoticity then, instead of the apostrophe. A sample of this interim Roman alphabet appears in Appendix 19. Examples of Kalasha literature that has already been drafted in Roman script appear in Appendixes 21 to 23.

CONCLUSION

9.1 CONSTRAINTS ON THIS STUDY

This thesis has explored a spectrum of issues relevant to the development and implementation of a writing system for the Kalasha language. The quest for the most appropriate orthography that would best suit the language and the community has led us to a range of perspectives, including regional, sociological, political, religious, linguistic, technical and literary, from which relevant factors could be examined. However, from many perspectives we find that not all the characteristics of a theoretically ideal orthography can feature in a real writing system. Various constraints and limitations can arise in the process of researching and designing an integrated and consistent system to represent a living language.

Ideological opposition to the concept of writing a language, because of the fear that it will precipitate the breakdown of indigenous culture, was discussed in Chapter 2 (section 2.1). Such opposition can hinder the course of orthography research and development, especially in a small community, where consensus is important and/or where literacy is not highly valued. An effective and durable writing system for any language needs to be owned by the whole community that speaks that language.

The convening of the orthography conference in 2000/2001 was a test of the Kalasha community's interest in, and commitment to, indigenous literacy. The fact that it brought together several rival personalities in what turned out to be amiable cooperation toward a common cause was one measure of its success. All participants had already embraced the concept of writing Kalasha, and the process of implementing it was met with much anticipation and interest. Every known implication and option relating to proposals for a Kalasha writing system was presented to the conference and discussed vigorously and thoroughly by all participants, with no substantial disagreements. However, as indicated in chapter 6 (section 6.2), no one could claim that every individual sees value in a writing system or mother-tongue literacy. Many are interested, but some are not interested.

Orthographic conformity may be an issue in the establishment of a new writing system, where there is pressure to adopt some orthographic precedents, from other languages in the region, which may not entirely suit the target language. These pressures may be linguistic, arising from shared phonological features; or from religious, sociological or political considerations. Conceivably, some regional precedents may be overlooked in the development of a new orthography, but those most relevant to the Kalasha language were addressed in Chapter 3. In any case, there is plenty of scope for orthographic innovation and experimentation in the Kalasha situation because of their distinct identity and independent stance in the region.

Dialect variation can imbue a standard orthography with fuzzy edges, because no one system can suit all natural varieties of a language, whether they result from phonological, lexical or grammatical differences. The total picture of dialectal variation can be limited by insufficient research. Several dialect studies have been conducted of the Kalasha language, and their findings, summarised in Chapter 4 (subsection 4.3.5), have been taken into consideration to develop a practically polylectal orthography. Of course, ongoing dialect variation could eventually reach the stage where it is no longer captured in the system presented here. However, a continuing policy of orthographic flexibility could allow the adaptation of the system to represent evolving speech behaviours.

Morphophonemic complexities need to be dealt with when designing an orthography, to cater for minor but systematic variations in natural discourse, and sometimes interim or arbitrary decisions need to be made. These issues, brought up in Chapter 5, raised the question of using deep or shallow orthographic strategies for the Kalasha language, and required some compromises in the matching of symbols for sounds. Surveys eliciting preferences for various options in the representation of morphological and syntagmatic phenomena were used for psycholinguistic reference in discussing these dilemmas. Data from unsolicited texts was also used wherever possible, to enlarge the sample, though the problems of demographic skewing (e.g. with variables of gender and education) are inherent in it.

Mutability over time is a factor that must be recognised and conceded in relation to any orthography. Languages constantly undergo natural and gradual metamorphoses, changing to suit the speakers—each community in each generation. As a result, orthographies may also

evolve over the course of time. The way a language has been codified at any one point of time may not necessarily represent the language quite as aptly in subsequent generations and eras. In addition to natural language change, orthographies themselves can also be changed by deliberate design. The Kalasha orthography presented in this thesis could change, or be changed, to reflect and follow the inevitable fluidity of linguistic practice, the growing awareness of orthographic issues, and shifting orthographic preferences, not to mention the influence of new literary genres, and the availability of new technology and media.

9.2 FUTURE RESEARCH IN KALASHA LANGUAGE AND ORTHOGRAPHY

The changing nature of languages and orthographies opens up the question of future research in the Kalasha language and orthography. Further linguistic research may have implications for Kalasha orthography, as outlined below, refining the analysis of some details of the Kalasha language—or simply underlining aspects of it that are already adequately represented.

Comparative orthographic research should continue to monitor parallel initiatives that evolve in neighbouring minor languages, enabling a cross-fertilisation of ideas, principles and practices. For those languages that have recently been codified (e.g. Khowar, Wakhi, Kalami, Burushaski, Shina, Hindko and Balochi) particular attention should be given to the use of their new orthographies, and to the development of their new literatures. But the unwritten languages of the region (e.g. Phalura, Dameli, Gawar-Bati, Kati, Kamviri, Yidgha, Gujar, all spoken in Chitral District) should also be monitored, because interest in indigenous orthographies could develop there in the future.¹

Phonological research could look more closely at the frequencies of marginal phonemes of the Kalasha language, for example, to more accurately determine the phonemic status of the voiced alveolar affricate [d͡z], the retroflex flap [ɽ], and the velar nasal [ŋ], and the phonemic interpretation of pre-plosive nasalization, the palatal nasal [ɲ] and the rare sequences [mh] and [lh]. Such research could suggest adjustments to the orthography as it continues to develop and change over time. The phonetic qualities of other Kalasha allophones (e.g. the

¹ A description of the orthographies of some of the written languages mentioned here, and a discussion of the reasons for the adoption of their respective scripts, was presented in Chapter 3.

unreleased plosives and affricates and the bilabial and labiodental fricatives) should also be examined in more detail, with particular attention to place of articulation, though this is unlikely to have orthographic implications.

Lexicological research could be expanded to capture more Kalasha archaic forms, to enhance understanding of phonological processes embedded in the language. Likewise of interest are the traditional and genuine Kalasha alternatives to foreign words that have crept in (mostly from Khowar and Urdu) over recent decades, due to increasing language contact situations. Further research could also be conducted in the specialised terminology of certain semantic domains, such as history and mythology, male/female vocabulary differences, the vocabulary of social relationships, abstract philosophical and religious concepts; and cultural practices such as food preparation, weaving, women's fashion and health issues, childbirth and the care of infants. Another area of lexicological enquiry awaiting investigation is the semantic relationships among words that share certain phonemic and semantic features with each other (e.g. /d͡ʒaw/ 'stand of trees' and /d͡ʒ^haw/ 'fence made of branches').

Dialect surveys need to be extended to investigate demographic and sociological factors of the varieties of Kalasha spoken especially in the Urtsun, Jinjiret and Shishi Valleys. There, whole Kalasha communities became Muslim some time prior to the middle of the twentieth century, and use of the Kalasha language fell into abeyance in favour of Khowar, the language of Chitral District. This is particularly relevant for the Urtsun Valley where a revival of the Kalasha language has been recently reported.²

Morphological research would be enhanced by the collection of more Kalasha texts (both spoken and written), and the creation and analyses of larger corpora, in order to further explore and analyse the significance of variable morphological and syntagmatic phenomena, such as voicing/devoicing, compounding and clitics. This could support refinements of some aspects of the morphological analysis and orthography provided here.

² The Khowar-speaking Utsun community have realised that when they reverted to speaking Kalasha (their original language before conversion to Islam), their opponents in a land dispute case could no longer understand them, and they won their case. This has led to a resurgence in the use of the Kalasha language. (Information supplied from Imran Khan, cited by R Trail in an email dated 4 December 2002.)

Orthographic research would be needed to periodically review the alphabetic system presented in this thesis, in order to ensure that it continues to reflect, as closely as possible, the intuitions and wishes of the community, even as they change. The administration of such research must respect and promote the sense of ownership of the orthography by the community. An orthography committee could be created to monitor the actual usage of the Kalasha orthography, and the practices and preferences that evolve, with particular regard to the finer points of spelling and punctuation. Future research could also include an assessment of the possibility of a parallel Arabic script to represent the Kalasha language for particular sectors of the society (e.g. the Muslim communities of Kalasha speakers).

As pointed out in chapter 6, it would be useful to elicit further feedback from all the participants of the 2000/2001 orthography conference, to gauge how they view the conference in retrospect: its usefulness, relevance, significance, practical implications, etc. It would also be revealing to find out the perceptions and attitudes of others in the community who did not or could not attend, both those who were/are interested and those who were/are not. A second Kalasha orthography conference is planned for the future, where all issues concerning Kalasha orthography would be reviewed and assessed. At the time of writing (2005) the follow-up conference has not yet eventuated, due to lack of funding.

Literacy surveys could be extended to (a) measure literacy performance and activity in the Kalasha, Urdu and English languages, and (b) gauge changing community attitudes to literacy generally. These may be better administered by non-Kalasha researchers who can do it more objectively. Such surveys would need to be elicited primarily from the schools, but also somewhat from the wider community. Community views of the researcher, and the researcher's role in the whole literacy initiative, would also warrant investigation. Does the community focus more on literacy itself, and its development in the context of their own culture, or on its foreign-sourced facilitation?

Discourse studies could be conducted to further analyse the natural discourse structures associated with each of the Kalasha genres. The results of such research could be used to understand, document and preserve Kalasha oral literature, especially in the context of literacy developments, and to train creative writers in the composition of Kalasha literature using both traditional and new genres.

Literary research should be conducted by an educationally oriented committee of Kalasha individuals to explore the types of literature that will suit and interest various sectors of the Kalasha readership. In this regard, one perspective should focus on the older generation, exploring how best to maintain oral culture and traditional knowledge and practices, and to promote and facilitate community cohesion and solidarity. Another focus should be on the younger generation, seeking how best to equip them, through indigenous literature, to preserve Kalasha history, heritage, language and culture; and to use their own literature as a contemporary medium, both for cultural and intellectual enrichment, and for creative and thoughtful communication.

EPILOGUE

by Taj Khan Kalash Sharakat¹

I think that it's a very crucial step we are taking to create a language in written form, which is very impressive and unique as well in its oral existence. Kalashamon has been very exceptional among accompanying Dardic languages, especially because of keeping ancestors' religion. Somehow the religious philosophy and social setup and other important sectors have vital impact on the language, especially when it is not written. The philosophy and real sense of expression is the essence of oral tradition. It comes from inside of the people and each word spoken is distinct as it explains the situation or perspective [of each] one in word. In oral languages people have more profound and immense ways of expression and a distinct sense of feeling, which may alter or sound somewhat different in written expression. While listening to a Kalasha man or woman singing an ancient folk song, encircled by group of admirers, could be good example. People pay attention particularly on lyrics and expression and it moves them deeply. This unifying impact of oral language, symbolic expression, philosophy of the language, traditional ideology and all other elements like rituals, legends, songs and stories, which are drawn from them, in my opinion need to be captured ...

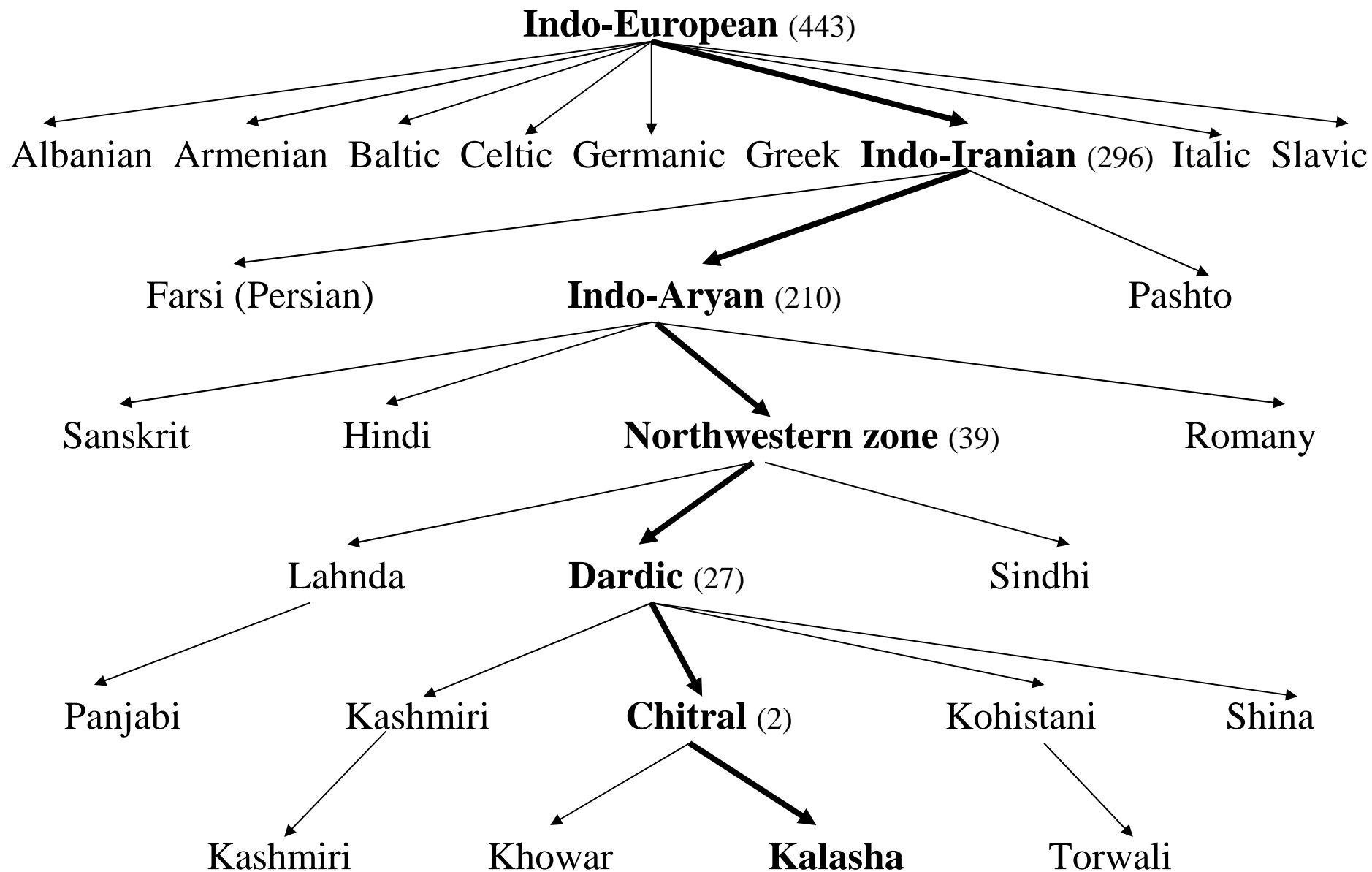
Giving the honor and opportunity to the women, elders, shepherds, and the people who have never been to school, to write freely is not only [a] thought but an intuition of mine. This will be the CLASSICAL KALASHAMON or ONJESTA MOND [pure language] ... on which ... Kalasha literature will follow.

It is extremely difficult, almost impossible to form ONJESTA MOND but this [is] what I would like to do ... Giving the opportunity to the uneducated Kalasha intellectuals to become aware of their indigenous knowledge, who are unconscious of what they possess orally, and bringing forth that into literature, in my opinion, is the land mark where we make ourselves literally distinct. We are not just learning the way of writing a language, but to enrich and strengthen the Kalasha way of life.

10 October 2002

¹ Taj was a 22-year-old student at the time of writing this passage.

Appendix 1: Simplified classification of the Kalasha language



Appendix 2: Script catalogues

This appendix contains annotated lists of some languages of relevance to Kalasha, with special reference to script alternatives, choice, change and/or variance. The languages are classified in the following four ways

A. Language families: This is the major catalogue, where relevant languages are grouped by families, and the script situation is explained for each language. This section contains information on the types of script that are, or have been, used for each language, various reasons for choice between them and some interpretative comments.

B. Script types: Arabic, Roman, Cyrillic, Devanagari, Hebrew, Armenian, Greek, Indigenous, None.

C. Settings of script alternatives: Diachronic (historical), Synchronic (geographical).

D. Reasons for script alternatives: Political, Religious, Sociological, Cultural, Miscellaneous.

The nature of these groupings means that repeated mentions of most languages will be found within many sections and subsections. Furthermore, classification is often difficult because reality is nowhere near as clear cut as these categories imply.

Within each section and subsection the actual languages are listed according to their language families. They have been ordered firstly by linguistic proximity to the Kalasha language, secondly by geographical proximity to the Kalasha Valleys, and thirdly, alphabetically by language family name. We start from Kalasha itself and progressively move out into wider circles. For example, other Dardic languages are listed by increasing geographical distance from where Kalasha is spoken. Other Indo-European languages are listed by progressively decreasing linguistic relationship to Kalasha.

All other languages are listed alphabetically by their language family names. There will always be disputes over language classification. The scheme offered here is taken from Grimes (2000), as is much of the information about scripts that is not otherwise referenced. The previous (13th) edition of that work was quoted as ‘the most authoritative source on the languages of the world’ (Ostler & Rudes 2000).

A. LANGUAGE FAMILIES

INDO-EUROPEAN | INDO-IRANIAN | INDO-ARYAN | NORTHWESTERN ZONE | DARDIC | CHITRAL

Kalasha (Bumboret, Rumbur and Birir valleys of south-west Chitral District of NWFP of Pakistan)

Preferred from early 1980s: Arabic script

Preferred from mid 2000: Roman script

Khowar (whole of Chitral District of NWFP of Pakistan)

Arabic script used mainly by a small group of educated enthusiasts who have formed the *Anjuman-e-Taraqqi Khowar* (Association for the Development of Khowar), which publishes a magazine, some folk tales, historical chronicles, etc.

INDO-EUROPEAN | INDO-IRANIAN | INDO-ARYAN | NORTHWESTERN ZONE | DARDIC | KOHISTANI

Kalami (northernmost part of Swat District of NWFP of Pakistan)

Arabic script

INDO-EUROPEAN | INDO-IRANIAN | INDO-ARYAN | NORTHWESTERN ZONE | DARDIC | SHINA

Shina (Gilgit and Hunza valley in north-east Pakistan)

Arabic script

Brokskat (Jammu & Kashmir in India)

Balti script

INDO-EUROPEAN | INDO-IRANIAN | INDO-ARYAN | NORTHWESTERN ZONE | DARDIC | KASHMIRI

Kashmiri (Jammu & Kashmir in India and Azad Kashmir in Pakistan)

Sarada script (related to Devanagari) is the ancient script of Kashmir.

Arabic script. Literature can be traced to the 1400s.

INDO-EUROPEAN | INDO-IRANIAN | INDO-ARYAN | NORTHWESTERN ZONE | LAHNDI

Panjabi, Western (Pakistan, India, etc.)

Arabic script used, but not often written in Pakistan.

Cf. Eastern Panjabi of India, which uses Devanagari script (Indo-European | Indo-Iranian | Indo-Aryan | Central Zone | Panjabi).

Jakati (Ukraine, Afghanistan, Moldova)

Probably unwritten

INDO-EUROPEAN | INDO-IRANIAN | INDO-ARYAN | NURISTANI

Kati (Afghanistan, Kalasha Valleys and other valleys in south-west Chitral district of NWFP of Pakistan)

Unwritten

INDO-EUROPEAN | INDO-IRANIAN | INDO-ARYAN | NORTHERN ZONE | WESTERN PAHARI

Dogri-Kangri (Jammu & Kashmir in India)

Jammu (Dogri): Arabic script

Himachal Pradesh (Kangri): Devanagari script

INDO-EUROPEAN | INDO-IRANIAN | INDO-ARYAN | NORTHWESTERN ZONE SINDHI

Sindhi (Sindh district in Pakistan)

Pre 1852: Khudawadi script

Post 1852: Arabic script

INDO-EUROPEAN | INDO-IRANIAN | INDO-ARYAN | CENTRAL ZONE | RAJASTHANI | UNCLASSIFIED

Gujuri (India, Pakistan, Afghanistan)

Devanagari script

Arabic script

Lambadi (India)

Each of the three dialects uses a different script:

Maharashtra: Devanagari script,

Karnataka: Kannada script,

Andhra Pradesh: Telugu script.

INDO-EUROPEAN | INDO-IRANIAN | INDO-ARYAN | CENTRAL ZONE | WESTERN HINDI | HINDUSTANI

Hindi-Urdu (India, Pakistan)

India: Devanagari script

Pakistan: Arabic script

INDO-EUROPEAN | INDO-IRANIAN | INDO-ARYAN | CENTRAL ZONE | PANJABI

Panjabi, Eastern (India, Kenya, Singapore, etc.)

Gurmukhi script, a variant of Devanagari;

Bhatyiana [dialect] uses Devanagari script.

Cf. Western Panjabi of Pakistan, which uses Arabic script (Indo-European | Indo-Iranian | Indo-Aryan | Northwestern Zone | Lahnda).

INDO-EUROPEAN | INDO-IRANIAN | INDO-ARYAN | CENTRAL ZONE | GUJARATI

Koli, Kachi (lower Sindh Province in south-east Pakistan)

Arabic script.

Some older people use Gujarati script, related to Devanagari.

Saurashtra (India)

Has had its own script for centuries.

A modern version developed in the late 1800s.

Since the end of the 19th century, books have been printed using Telugu, Tamil, Devanagari, and Saurashtra scripts.

Currently an adapted Tamil script is most commonly employed, using superscript numbers and a colon to show sounds not used in Tamil.

INDO-EUROPEAN | INDO-IRANIAN | INDO-ARYAN | EASTERN ZONE | BENGALI-ASSAMESE

Sylheti (Bangladesh)

Lie et al. (1999, p. 73) developed a transcription scheme based on the Roman alphabet. They make the point that ‘Roman script has been used and advocated in systems of transliteration of Indic languages for some time ... For example, **Konkani** ... [spoken] between Bombay and Goa, and **Kashmiri**.’

INDO-EUROPEAN | INDO-IRANIAN | IRANIAN | EASTERN | NORTHEASTERN

Osetin (Georgia, Azerbaijan, Germany, Hungary, Kazakhstan, Russia, Tajikistan, Turkey, Turkmenistan, Ukraine, Uzbekistan)

1844 to 1923: Cyrillic script

1923 to 1939: Roman script

1938 to 1954: Georgian script

1954 to Cyrillic script

INDO-EUROPEAN | INDO-IRANIAN | IRANIAN | EASTERN | SOUTHEASTERN | PAMIR

Wakhi (north-eastern Chitral District of NWFP and Hunza valley of northern Pakistan, Afghanistan, Tajikistan, China)

Roman script

INDO-EUROPEAN | INDO-IRANIAN | IRANIAN | WESTERN | NORTHWESTERN | BALOCHI

Balochi (south-western Pakistan, south-eastern Iran, southern Afghanistan, Turkmenistan, Tajikistan, India, Oman)

Roman script

Arabic script

INDO-EUROPEAN | INDO-IRANIAN | IRANIAN | WESTERN | NORTHWESTERN | KURDISH

Kurdi

Iraq & Iran: Arabic script

Syria & Turkey: Roman script

Central Asia: Cyrillic script

Kurmanji

Previously: Armenian script.

Turkey: Roman script

Armenia, Azerbaijan, Georgia, Kazakhstan, Kyrgyzstan: Cyrillic script.

Syria, Iraq, and Iran: Arabic script

INDO-EUROPEAN | INDO-IRANIAN | IRANIAN | WESTERN | SOUTHWESTERN | PERSIAN**Bukharic** (Israel, Uzbekistan)

Hebrew script

Farsi (Iran, Afghanistan, etc.)

Pre-7th Century: Pre-Islamic script (Grimes & Gordon 1980)

Post-7th Century: Arabic script

Tajiki (Tajikistan, etc.)

Pre-1930: Arabic script

1930-1940: Roman script

1940 onwards: Cyrillic script

INDO-EUROPEAN | INDO-IRANIAN | IRANIAN | WESTERN | SOUTHWESTERN | TAT**Judeo-Tat** (Russia)

Russia 1920 to 1940: Roman script

Russia 1940 onwards: Cyrillic script

Israel: Recently private literature uses Hebrew script

Azerbaijan: ?

AFRO-ASIATIC | SEMITIC | CENTRAL | SOUTH | ARABIC**Arabic, Chadian**

Roman script

Arabic, Moroccan (Morocco)

Arabic script

Hebrew script

Arabic, Judeo- (Israel)

Hebrew script

Arabic, Judeo-Tunisian (Tunisia)

Hebrew script.

Arabic, Standard

Saudi Arabia etc.: Arabic script

Israel: Hebrew script

AFRO-ASIATIC | SEMITIC | SOUTH | SOUTH ARABIAN**Mehri** (Yemen, Oman, Kuwait)**Mehri: Bathari** (Oman)**Soqotri** (Yemen)**Jibbali** (Oman)**Harsusi** (Oman)**Hobyót** (Oman, Yemen)

ALTAIC | TURKIC | EASTERN**Uyghur**

The Arabic script underwent reforms in its rendition of the Uyghur language in China in 1937 and 1954.

Cyrillic script has been used in China since 1941.

Roman script was also used temporarily prior to the 1950s, and officially in the 1960s and 1970s.

The Arabic script was reinstated as the official script for Uyghur in China in 1984, followed by further reforms in 1985.

A new Arabic script was introduced in 1987.

Kazakhstan: Cyrillic script

Turkey: Roman script

(See also Shorish 1984).

Uzbek, Northern (Uzbekistan)

Formerly: Arabic script

Then: Roman script

Currently: Cyrillic script.

Uzbek, Southern (Afghanistan)

Arabic script, orthography not yet standardized.

ALTAIC | TURKIC | SOUTHERN | AZERBAIJANI**Azerbaijani, North (Azerbaijan)**

Official: Roman script

Popular: Cyrillic script

Azerbaijani, South

Iran & Iraq: Arabic script

Afghanistan: Most do not know the Cyrillic script

ALTAIC | TURKIC | SOUTHERN | TURKISH**Gagauz (Moldova)**

Similar to Turkish

Cyrillic script introduced in 1957

Turkish (Turkey) (See Heyd 1954 & Lewis 1981)

Pre 1928: Arabic script

Armenian script

Greek script

Post 1928: Roman script.

ALTAIC | TURKIC | SOUTHERN | TURKMENIAN**Turkmen**

Turkmenistan: Cyrillic script

Afghanistan: Arabic script.

(Some better-educated persons can read Cyrillic script)

Iran: Arabic script.

ALTAIC | TURKIC | WESTERN | ARALO-CASPIAN**Kazakh**

Kazakhstan & Mongolia: Cyrillic script

Turkey: Roman script

China: Had an official Roman alphabet; since 1980 uses a modified Arabic script.

Iran: Any who are literate would use Arabic script.

Kirghiz

Kyrgyzstan: Cyrillic script

Afghanistan: Afghan Kirghiz do not read Cyrillic script

China: Arabic script

Turkey: Roman script

ALTAIC | TURKIC | WESTERN | PONTO-CASPIAN

Karaim (Lithuania, Israel, Ukraine)

Cyrillic script

Hebrew script.

Kumyk (Russia, Turkey)

Cyrillic script

AUSTRONESIAN | MALAYO-POLYNESIAN | WESTERN MALAYO-POLYNESIAN | SUNDIC | MALAYIC | MALAYAN | LOCAL MALAY**Indonesian**

Arabic script

Roman script

Malay (Malaysia, Brunei, Indonesia, Myanmar, Singapore, Thailand, etc.)

Arabic (Jawi) script

Roman (Rumi) script

CREOLE | MALAY BASED

Sri Lankan Creole Malay

Previously: Arabic (Jawi) script

Now: Roman script

NORTH CAUCASIAN | NORTHWEST | ABKHAZ-ABAZIN**Abaza**

Russia: Cyrillic script

Turkey: Roman script

NORTH CAUCASIAN | NORTHWEST | CIRCASSIAN**Adyghe**

Russia: Cyrillic script

Turkey: Roman script

SINO-TIBETAN | TIBETO-BURMAN | HIMALAYISH | TIBETO-KANAURI | TIBETIC | TIBETAN | WESTERN**Balti** (North-eastern Pakistan, India)

Vernacular script, now obsolete for Balti (though still used for the Brokskat, a Dardic language of Indian Jammu and Kashmir).

Arabic script used by older generation and Roman characters by the younger.

Purik (India)

People prefer to be culturally and linguistically identified with Tibet, although religiously with Islam.

Perso-Arabic script of Urdu used unless they have been to the university.

LANGUAGE ISOLATE**Burushaski**

Arabic script

Roman script

B. SCRIPT TYPES**1. Arabic/Urdu**

Kalasha	Kachi Koli	Hobyót
Khowar	Kurmanji	Uyghur
Kalami	Farsi	Southern Uzbek
Shina	Moroccan Arabic	Northern Uzbek
Kashmiri	Standard Arabic	South Azerbaijani
Western Panjabi	Mehri	Turkish
Dogri-Kangri	Soqotri	Turkmen
Gujuri	Jibbali	Kazakh
Urdu	Harsusi	Kirghiz

2. Roman/Latin

Kalasha	Mehri	North Azerbaijani
Sylheti	Soqotri	Gagauz
Wakhi	Jibbali	Turkish
Kurmanji	Harsusi	Kazakh
Judeo-Tat	Hobyót	Kirghiz
Somali	Uyghur	
Chadian Arabic	Northern Uzbek	

3. Cyrillic

Kurmanji	Northern Uzbek	Kazakh
Tajiki	North Azerbaijani	Kirghiz
Judeo-Tat	Gagauz	Karaim
Uyghur	Turkmen	Kumyk

4. Devanagari

Dogri-Kangri	Lambadi	Eastern Panjabi
Gujuri	Hindi	

5. Hebrew

Bukharic	Judeo Arabic	Karaim
Judeo-Tat	Judeo-Tunisian Arabic	
Moroccan Arabic	Standard Arabic	

6. Armenian

Kurmanji	Turkish	
----------	---------	--

7. Greek

Turkish		
---------	--	--

8. Indigenous

Brokskat	Saurashtra	Somali
Lambadi	Farsi	Uyghur

9. No script

Jakati	Kati	
--------	------	--

C. SETTINGS OF SCRIPT ALTERNATIVES**1. Diachronic (Historical)**

Kalasha	Kurmanji	Gagauz
Kachi Koli	Judeo-Tat	Turkish
Saurashtra	Uyghur	Kazakh
Sylheti	Northern Uzbek	

2. Synchronic (Geographical)

Khowar	Judeo-Tat	South Azerbaijani
Kalami	Chadian Arabic	Turkmen
Shina	Moroccan Arabic	Kazakh
Kashmiri	Judeo Arabic	Kirghiz
Western Panjabi	Standard Arabic	Karaim
Lambadi	Uyghur	Kumyk
Kurmanji	Southern Uzbek	
Tajiki	North Azerbaijani	

D. REASONS FOR SCRIPT ALTERNATIVES**1. Religious**

Hindi-Urdu	Standard Arabic	Northern Uzbek
Farsi	Uyghur	Judeo-Tunisian Arabic
Judeo Arabic	Southern Uzbek	

2. Political

Hindi-Urdu	Northern Uzbek	Kazakh
Tajiki	Gagauz	Kirghiz
Uyghur	Turkish	Kumyk

3. Sociological**(a) Modelled on English**

Kalasha	Wakhi	North Azerbaijani
Sylheti	Northern Uzbek	

(b) Modelled on Arabic/Urdu

Kalami	Kashmiri	Kazakh
Shina	Western Panjabi	
Khowar	Northern Uzbek	

4. Cultural

Brokskat

5. Miscellaneous (multiple or uncertain reasons)

Dogri-Kangri	Karaim	Mehri
Hindi-Urdu	Tajik	Soqotri
Kachi Koli	Judeo-Tat	Uyghur
Saurashtra	Chadian Arabic	South Azerbaijani
Bukharic	Moroccan Arabic	Turkmen

Appendix 3: Kalasha phoneme chart – Roman script – 1985

CONSONANTS

	<u>Bilabial</u>	<u>Dental</u>	<u>Alveolar</u>	<u>Alveopalatal</u>	<u>Alveopalatal retroflexed</u>	<u>Palatal</u>	<u>Velar</u>	<u>Glottal</u>
Stops								
unasp vl	p	t			T		k	
asp vl	ph	th			Th		kh	
unasp vd	b	d			D		g	
asp vd	bh	dh			Dh		gh	
Affricates								
unasp vl		ts	c		C			
asp vl		tsh	ch		Ch			
unasp vd		dz	j		J			
asp vd		dzh	jh					
Sibilants								
vl		s	sh		S			h
vd		z	zh		Z			
Nasals								
	m	n		ñ				
Approximants								
		l	ĩ					
Flaps								
			r					
Semivowels								
w						y		

(Dental and alveolar affricates and sibilants are grooved.)

VOWELS

	<u>front</u> unrounded	<u>central</u> unrounded	<u>back</u> rounded
<u>high</u>	i		u
<u>mid</u>	e		o
<u>low</u>		a	

DIACRITICS

~ nasalisation. This occurs on all vowels, both normal and retroflexed.

(caps) retroflexion. This occurs on all vowels, both oral and nasalised.

‘ stress. This occurs unpredictably in the word, and at times causes meaning difference as in /a’ya/ ‘my mother’, and /aya/ ‘here’.

Appendix 4: Kalasha phoneme chart – Roman script – 2005

CONSONANTS

	<u>Bilabial</u>	<u>Alveolar</u>	<u>Postalveolar</u>	<u>Retroflex</u>	<u>Palatal</u>	<u>Velar</u>	<u>Glottal</u>
Plosives							
unasp vl	p	t		t'		k	
asp vl	ph	th		t'h		kh	
unasp vd	b	d		d'		g	
asp vd	bh	dh		d'h		gh	
Affricates							
unasp vl		ts	c	c'			
asp vl		tsh	ch	c'h			
unasp vd		dz	j	j'			
asp vd			jh				
Fricatives							
vl		s	sh	s'			h
vd		z	zh	z'			
Nasals							
	m	n					
Approximants							
central					y	w	
lateral		l	l'				
Flaps							
		r		r'			

VOWELS

	-----front-----			-----back-----	
	high	mid	low	mid	high
normal	i	e	a	o	u
retroflex	i'	e'	a'	o'	u'
nasalised	ĩ	ẽ	ã	õ	ũ
nasalised & retroflex	ĩ'	ẽ'	ã'	õ'	ũ'

Appendix 5: Kalasha allophone catalogue: phonemic inventory with allophones

<u>Phonemes</u>	<u>Allophones</u>	<u>Description, Distribution and Examples</u>
/p/	[p̚]	unreleased unaspirated voiceless bilabial stop word-final /d̪ɪp/ [d̪ɪp̚] ‘tongue’
	[p]	unaspirated voiceless bilabial plosive anywhere /paʃ/ [paʃ] ‘fleece’
/p ^h /	[ϕ]	voiceless bilabial fricative word-medial /ti'phak/ [ti'ϕak] ‘agreement’
	[pp]	aspirated voiceless bilabial trill rare: word-initial /'phutʃiŋg/ ~ /'phutʃiŋg/ ['pputʃiŋ] ‘goat excrement’
	[p ^h]	aspirated voiceless bilabial plosive non-word-final /p ^h aʃ/ [p ^h aʃ] ‘greatly’
/b/	[b]	unaspirated voiced bilabial plosive non-word-final /bas/ [bas] ‘day’
/b ^h /	[b ^h]	aspirated voiced bilabial plosive word-initial /b ^h as/ [b ^h as] ‘heat’

<u>Phoneme</u>	<u>Allophones</u>	<u>Description, Distribution and Example</u>
/t/	[t̚]	unreleased unaspirated voiceless dental stop word-final ʔts ^h at/ [ts ^h at̚] 'pool'
	[t]	unaspirated voiceless dental plosive anywhere /tan/ [tan] 'one's own'
/t ^h /	[t ^h]	aspirated voiceless dental plosive non-word-final /t ^h an/ [t ^h an] 'agreement'
/d/	[d]	unaspirated voiced dental plosive non-word-final /daj/ [daj] '-ing'
	[d ^h]	aspirated voiced dental plosive non-word-final /d ^h aj/ [d ^h aj] 'sheaf'
/ɖ/	[ɖ̚]	unreleased unaspirated voiceless retroflex stop word-final /mut/ [muɖ̚] 'tree'
	[ɖ]	unaspirated voiceless retroflex plosive anywhere /paɖi/ [paɖi] 'cummerbund'
/ɖ ^h /	[ɖ ^h]	aspirated voiceless retroflex plosive non-word-final /a ^h ɖi/ [a ^h ɖi] 'bone'

<u>Phoneme</u>	<u>Allophones</u>	<u>Description, Distribution and Example</u>
/ɖ/	[ɖ]	unaspirated voiced retroflex plosive non-word-final /ɖak 'karik/ [ɖak 'karik] 'to melt butter'
/ɖ ^h /	[ɖ ^h]	aspirated voiced retroflex plosive non-word-final /ɖ ^h ak/ [ɖ ^h ak] 'waist'
/k/	[ʔ]	glottal stop word-final (Rumbur dialect) /eg/ [ekʔ] 'one'
	[k]	unaspirated voiceless velar plosive anywhere /kam/ [kam] 'clan'
/k ^h /	[x]	aspirated voiceless velar fricative sometimes: non-word-final in loan words /khajr/ [xajr] 'well'
	[k ^h]	aspirated voiceless velar plosive non-word-final /k ^h am/ [k ^h am] 'level'
/g/	[g]	unaspirated voiced velar plosive non-word-final /gum/ [gum] 'wheat'
/g ^h /	[gh]	aspirated voiced velar plosive non-word-final /g ^h um/ [g ^h um] 'noise'

<u>Phoneme</u>	<u>Allophones</u>	<u>Description, Distribution and Example</u>
/ts̃/	[ts̃]	unaspirated voiceless dental affricate anywhere /ts̃ak/ [ts̃ak] ‘lukewarm’
/ts̃ ^h /	[ts̃ ^h]	aspirated voiceless dental affricate non-word-final /ts̃ ^h ak/ [ts̃ ^h ak] ‘pus’
/dz̃/	[dz̃]	voiced dental affricate non-word-final in Krakal/Brun dialect /dz̃a'raj/ [dz̃a'raj] ‘rennet’
/tʃ̃/	[tʃ̃]	unreleased unaspirated voiceless postalveolar affricate word-medial before stop or affricate /ʼatʃ̃tuo/ [ʼatʃ̃tuo] ‘after’
	[tʃ̃]	unaspirated voiceless postalveolar affricate anywhere /tʃ̃at/ [tʃ̃at] ‘false argument’
/tʃ̃ ^h /	[tʃ̃ ^h]	aspirated voiceless postalveolar affricate non-word-final /tʃ̃ ^h at/ [tʃ̃ ^h at] ‘pool’
/dʒ̃/	[dʒ̃]	unreleased voiced postalveolar affricate word-final /mendʒ̃/~mẽdʒ̃/ [mentʃ̃]~[men] ‘cloud’
	[dʒ̃]	voiced postalveolar affricate non-word-final /dʒ̃a/ [dʒ̃a] ‘wife’

<u>Phoneme</u>	<u>Allophones</u>	<u>Description, Distribution and Example</u>
/ḍ̥ʰ/	[ḍ̥ʰ]	aspirated voiced postalveolar affricate word-initial /ḍ̥ʰa/ [ḍ̥ʰa] ‘filtration’
/ṣ̥/	[ṣ̥ʰ]	unreleased unaspirated voiceless retroflex affricate word-final /mraṣ̥/ [mraṣ̥ʰ] ‘mulberry’
	[ṣ̥]	unaspirated voiceless retroflex affricate anywhere /buṣ̥uṣ̥ik/ [buṣ̥uṣ̥ik] ‘to smile’
/ṣ̥ʰ/	[ṣ̥ʰ]	aspirated voiceless retroflex affricate non-word-final /ṣ̥uṣ̥ʰik/ [ṣ̥uṣ̥ʰik] ‘dried fruit’
/ḍ̥ʒ/	[ḍ̥ʒ]	voiced retroflex affricate non-word-final /ḍ̥ʒanḍ̥ʒer/~ḍ̥ʒāḍ̥ʒer/ [ḍ̥ʒanḍ̥ʒer] ‘chain’
/s/	[s]	voiceless dental fricative anywhere /sak/ [sak] ‘extremely’
/z/	[z]	voiced dental fricative non-word-final /zak/ [zak] ‘lard’
/ʃ/	[ʃ]	voiceless postalveolar fricative anywhere /ʃak/ [ʃak] ‘vegetables’

<u>Phoneme</u>	<u>Allophones</u>	<u>Description, Distribution and Example</u>
/ʒ/	[ʒ]	voiced postalveolar fricative anywhere [ʒa] ‘until’
/ʂ/	[ʂ]	voiceless retroflex fricative anywhere /ʂak'aːk/ [ʂak'aːk] ‘suddenly’
/ʐ/	[ʐ]	voiced retroflex fricative non-word-final /ʐak/ [ʐak] ‘thick’
/h/	[h]	voiceless glottal fricative non-word-final /ha/ [ha] ‘breath’
/m/	[m]	bilabial nasal anywhere /mo/ [mo] ‘don’t!’
/n/	[ɲ]	voiced alveolar nasal rare occurrences /ni'hik/ [ɲi'hik] to emerge’
	[n]	voiced dental nasal anywhere /no/ [no] ‘nine’
/ŋ/	[ŋ]	voiced velar nasal non-word-initial /traŋ/ [traŋ] ‘narrow’

<u>Phoneme</u>	<u>Allophones</u>	<u>Description, Distribution and Example</u>
/j/	[j]	voiced central palatal vocoid approximant non-word-final /‘aja/ [‘aja] ‘mother’
/w/	[v]	voiced labiodental fricative before front vowels /ni'weʃik/ [ni'veʃik] ‘to write’
	[w]	voiced central labial-velar vocoid approximant anywhere [‘awa/ [‘awa] ‘grandmother’
/l/	[lʲ]	voiced palatalized dental lateral approximant anywhere /law'ek/ [lʲaw'ek] ‘to lie’
/ɭ/	[ɭ]	voiced velarized dental lateral approximant anywhere /ɭaw'ek/ [ɭaw'ek] ‘to steal’
/r/	[ɾ̥]	voiceless dental flap utterance-final /t͡ʂʰiɾ̥/ [t͡ʂʰiɾ̥] ‘milk’
	[r]	voiced dental flap elsewhere /ra/ [ra] ‘cedar’
/ɽ/	[ɽ]	voiced retroflex flap intervocalic (Birir dialect only) /az̠'aɽi/ [az̠'aɽi] ‘apricot’

<u>Phoneme</u>	<u>Allophones</u>	<u>Description, Distribution and Example</u>
/i/	[i]	high front unrounded vowel anywhere /i̯ts/ [i̯ts] ‘bear’
/e/	[e]	mid front unrounded vowel anywhere /et̩s/ [et̩s] ‘eye’
/a/	[a]	low front unrounded vowel anywhere /at̩s/ [at̩s] ‘today’
/o/	[o]	mid back rounded vowel anywhere /ot̩s/ [ot̩s] ‘herbal’
/u/	[u]	high back rounded vowel anywhere /uts/ [uts] ‘spring’
/ị/	[ị]	rhotic high front unrounded vowel non-word-initial /pị/ [pị] ‘press!’
/ẹ/	[ẹ]	rhotic mid front unrounded vowel non-word-initial /hẹ/ [hẹ] ‘theft’
/ạ/	[ạ]	rhotic low-mid front unrounded vowel anywhere /t̩ạ/ [t̩ạ] ‘clap’

<u>Phoneme</u>	<u>Allophones</u>	<u>Description, Distribution and Example</u>
/ɔ̃/	[ɔ̃]	<p>rhotic mid central rounded vowel</p> <p>non-word-initial</p> <p>/ʈʂɔ̃i/ [ʈʂɔ̃i] ‘parasite’</p>
/ũ/	[ũ]	<p>rhotic high back rounded vowel</p> <p>anywhere</p> <p>/ʈʂũi/ [ʈʂũi] ‘plait’</p>
/ĩ/	[ĩ]	<p>nasal high front unrounded vowel</p> <p>non-word-initial</p> <p>/paʰĩntʂak/~paʰĩtʂak/ [paʰĩntʂak] ‘thin’</p>
/ẽ/	[ẽ]	<p>nasal mid front unrounded vowel</p> <p>anywhere</p> <p>/hẽʃ/ [hẽʃ] ‘like this’</p>
/ã/	[ã]	<p>nasal low front unrounded vowel</p> <p>anywhere</p> <p>/hãʃ/ [hãʃ] ‘horse’</p>
/õ/	[õ]	<p>nasal mid back rounded vowel</p> <p>non-word-initial</p> <p>/ʈʂʰõʷik/ [ʈʂʰõʷik] ‘to summon’</p>
/ũ/	[ũ]	<p>nasal high back rounded vowel</p> <p>anywhere</p> <p>/brũa/ [brũa] ‘Brun (village)’</p>
/ĩ̃/	[ĩ̃]	<p>rhotic nasal high front unrounded vowel</p> <p>no occurrences found yet</p>
/ẽ̃/	[ẽ̃]	<p>rhotic nasal mid front unrounded vowel</p> <p>anywhere</p> <p>/pẽ̃/ [pẽ̃] ‘palm’</p>

<u>Phoneme</u>	<u>Allophones</u>	<u>Description, Distribution and Example</u>
/ã̃/	[ã̃]	<p>rhotic nasal low-mid front unrounded vowel anywhere</p> <p>/ḍḗḗã̃/ [ḍḗḗã̃] ‘walnut kernel’</p>
/õ̃/	[õ̃]	<p>rhotic nasal mid central rounded vowel anywhere</p> <p>/põ̃/ [põ̃] ‘leaf’</p>
/ũ̃/	[ũ̃]	<p>rhotic nasal high back rounded vowel non-word-initial</p> <p>/pũ̃/ [pũ̃] ‘grape harvest’</p>

Appendix 6: Kalasha corpus sample

This is a sample of one of the eight texts in Roman script that was used to construct a phoneme tally. The final tally appears in Appendix 7.

laDer khan'a ochu'nik: kay'mina kay'mina eg ba'ca as'ta. bo dawLatman' as'ta. ba'caas som bo naw'kar as'ta. bo phawj' ta'sa som as'ta. ma'gam ba'caas tre chu as'ta. putr'o ne as'ta. te tre chu'Lasi bo shisho'yak ba'ca bo khoshan' as'ta. copami'na uST'iLa. uST'i dram'i'a pai jaga'La, ki ta'sa nogor' bo shisho'yak. se ma'iLa ki, o may nogor' kimon' shisho'yak. may naw'kar khē pruST kai krom kar'in day ghō'i ma'i tan hardi' khoshan' kar'iman as'ta. eg ad'uo ba'ca shēhē' ka'da ki, tan tre chu'Lasi dawat'una khoj'iLa dawat' zhu ba'ca ga'La. chu'Lasi kay ki'a mon ne di'ta. piST'aw ad'uo ge'ri dawat' di'ta. Saw tre chu'Lasi dawat'una i'ta a'an. tara' i'ta nis'i a'an. nis'i ahu' mahu' zhu'La. ahu' zhu'i to, ba'ca chu'Lasi kay sha'ma sual' ka'da ki, may ga'Da chu, tay ku'ra pay'da kai a'au? ga'Da chu'Las bo soc kai to ma'i a'au ki, may da'da may pay'da kai a'au ghō'i ma'iLa pha'to ba'ca tu bo pruST may chu. tu bo pruST ghō'i ta'sa Sa'baS di'ta. pha'to machuma'ra chu'Las kay sual' ka'da may chu, tay ku'ra pay'da kai a'au? se ba'ta soc'as moc'ay pai a'au. bo soc kai to, se ba'ta ma'iLa ki, da'da tu may pay'da kai as'as ghō'iLa akhe'ri sual' chu'tyak chu'Las kay ka'da. se'o bo shisho'yak as'ta. ba'ca chu'tyak chu'Las kay ma'iLa ki, tay'o ku'ra pay'da kai as'au? chu'tyak chu'Las bo dunu'na, bo dunu'na to ma'iLa ki, may'o khoday' pay'da kai as'au. pha'to ba'ca bo kahar'i hu'La. kahar'i thi, ma'iLa ki, ge'ri tay kay sual' kar'im day, ki tay ku'ra pay'da kai a'sau ghō'i ma'iLa. pha'to chu'tak chu'Las ma'iLa ki, da'da ho'ma saw khoday' pay'da kai as'au ghō'i ma'iLa pha'to ba'ca kahar'i thi, hu'kum.

Here is part of the same text as interlinearized by R Trail:

```
\tx laDer khan'a          ochu'nik.
\mr laDer khan-a          ochunik
\mg Lader Khan-GEN.3S story
\ft Lader Khan's story.

\tx kay'mina kay'mina ek ba'ca as'ta.
\mr kay -mina kay -mina ek ba'ca as -ta
\mg when-TIME when-TIME one king be.ANIM-PAST.HEARSAY.3S
\ft Once upon a time there was a king.

\tx bo dawLatman' as'ta.
\mr bo dawLatman as -ta
\mg very rich be.ANIM-PAST.HEARSAY.3S
\ft He was very rich.

\tx ba'caas som bo naw'kar as'ta.
\mr baca-as som bo naw'kar as -ta
\mg king-OBLQ.3S possession many servant be.ANIM-PAST.HEARSAY.3P
\ft The king had many servants.

\tx bo phawj' ta'sa som as'ta.
\mr bo phawj t -asa som as -ta
\mg many soldier RMOT-3S.OBLQ possession be.ANIM-PAST.HEARSAY.3S
\ft He had a great army.

\tx ma'gam ba'caas tre chu as'ta.
\mr ma'gam baca-as tre chu as -ta
\mg but king-GEN.3S three daughter be.ANIM-PAST.HEARSAY.3S
\ft But he only had three daughters.

\tx putr'o ne as'ta.
\mr putr-o ne as -ta
\mg son -CNTR not be.ANIM-PAST.HEARSAY.3S
\ft He had no son.

\tx te tre chu'Lasi bo shisho'yak as -ta
\mr te tre chu -Lasi bo shisho'yak as'ta.
\mg those three daughter-3P.NOM very beautiful be.ANIM-PAST.HEARSAY.3P
\ft Those three daughters of his were very beautiful.
```

Appendix 7: Kalasha phoneme tally – Arabic script corpus (60 texts)

FREQUENCY COUNT OF ALL KALASHA PHONEMES
FROM 60 SELECTED TEXTS IN KALASHA SCRIPT
corpusa.txt (279K)

Segment	# of Words	# of Occurrences	% Reduplication	% of All Phonemes
a	18266	27677	52%	> 24%
i	7565	8561	13%	> 7% (actually more)
y	6981	7252	4%	> 6% (actually less)
o	5963	6459	8%	> 5%
s	5268	5540	5%	> 4%
n	4985	5324	7%	> 4%
h	4611	5118	11%	> 4%
u	4577	5077	11%	> 4%
r	4930	5050	2%	> 4%
t	4854	4936	2%	> 4%
k	4350	4537	4%	> 4%
m	3666	3879	6%	> 3%
d	2701	2852	6%	> 2%
b	2016	2190	9%	> 1%
p	1949	1970	1%	> 1%
g	1656	1675	1%	> 1%
L	1502	1526	2%	> 1%
e	1330	1425	7%	> 1% (actually more)
z	1222	1226	0%	> 1%
l	1114	1137	2%	> 1%
s	933	1038	11%	< 1%
c	888	889	0%	
w ^h [v]	749 810	833 894	11% 10%	
j	749	753	1%	
T	643	655	2%	
kh	612	612	0%	
th	455	455	0%	
ph	447	451	1%	
ch	355	361	2%	
gh	346	352	2%	
Q	352	352	0%	
D	340	350	3%	
A	325	343	6%	
e~	281	285	1%	
C	169	176	4%	
Ch	144	145	1%	
sh	129	131	2%	
I	108	108	0%	
i~	101	101	0%	
E	88	88	0%	
jh	65	66	2%	
v	61	61	0%	
U	54	56	4%	
? R	54	54	0%	
zh	47	47	0%	
bh	46	46	0%	
? N	22	29	32%	
Z	26	26	0%	
ts	20	20	0%	

J	9	12	33%
Th	10	10	0%
o~	6 ?	6 ?	0%
? nh	3 ?	3 ?	0%
o~	3	3	0%
E~	1	1	0%
? dz → /z/ ?	1	1	0%
? mh	1	1	0%
a~	0	0	(should be more?)
A~	0	0	(should be more?)
I~	0	0	(should be more?)
u~	0	0	(should be more?)
U~	0	0	(should be more?)
dh	0	0	
dzh	0	0	
Dh	0	0	
tsh	0	0	
Jh	0	0	
lh	0	0	
Lh	0	0	
	<u>27595</u>	<u>112348</u>	(av. 4.1 phonemes per word)

Appendix 8: Kalasha phoneme tally – Roman script corpus (8 texts)

FREQUENCY COUNT OF ALL KALASHA PHONEMES
FROM 8 SELECTED INDIGENOUS TEXTS
corpusr8.txt (12K)

Segment	Number of Words containing	Total occurrences	% of All Phonemes
a	1321	2002	> 24%
i	584	631	> 7% (actually more)
u	366	458	> 5%
k	367	385	> 4%
n	320	358	> 4%
r	341	344	> 4%
y	304	317	> 3% (actually less)
e	305	312	> 3%
t	293	300	> 3%
m	284	290	> 3%
o	264	270	> 3%
w	217	249	> 3%
s	221	229	> 2%
d	204	213	> 2%
b	185	189	> 2%
l	168	176	> 2%
h	165	167	> 2%
g	166	166	> 2%
p	155	155	> 1%
sh	147	149	> 1%
l	144	147	> 1%
j	93	93	> 1%
z	84	84	> 1%
S	63	71	< 1%
T	56	58	
D	46	52	
zh	51	51	
c	45	45	
th	41	41	
A	31	38	
ph	34	34	
kh	33	33	
o~	26	27	
gh	25	25	
e~	16	23	
ch	21	21	
O	21	21	
O~	11	11	
bh	11	11	
Ch	10	10	
U	7	7	
a~	6	6	
u~	5	5	
ts	5	5	
C	5	5	
wh	2	2	
Dh	2	2	
E~	1	1	

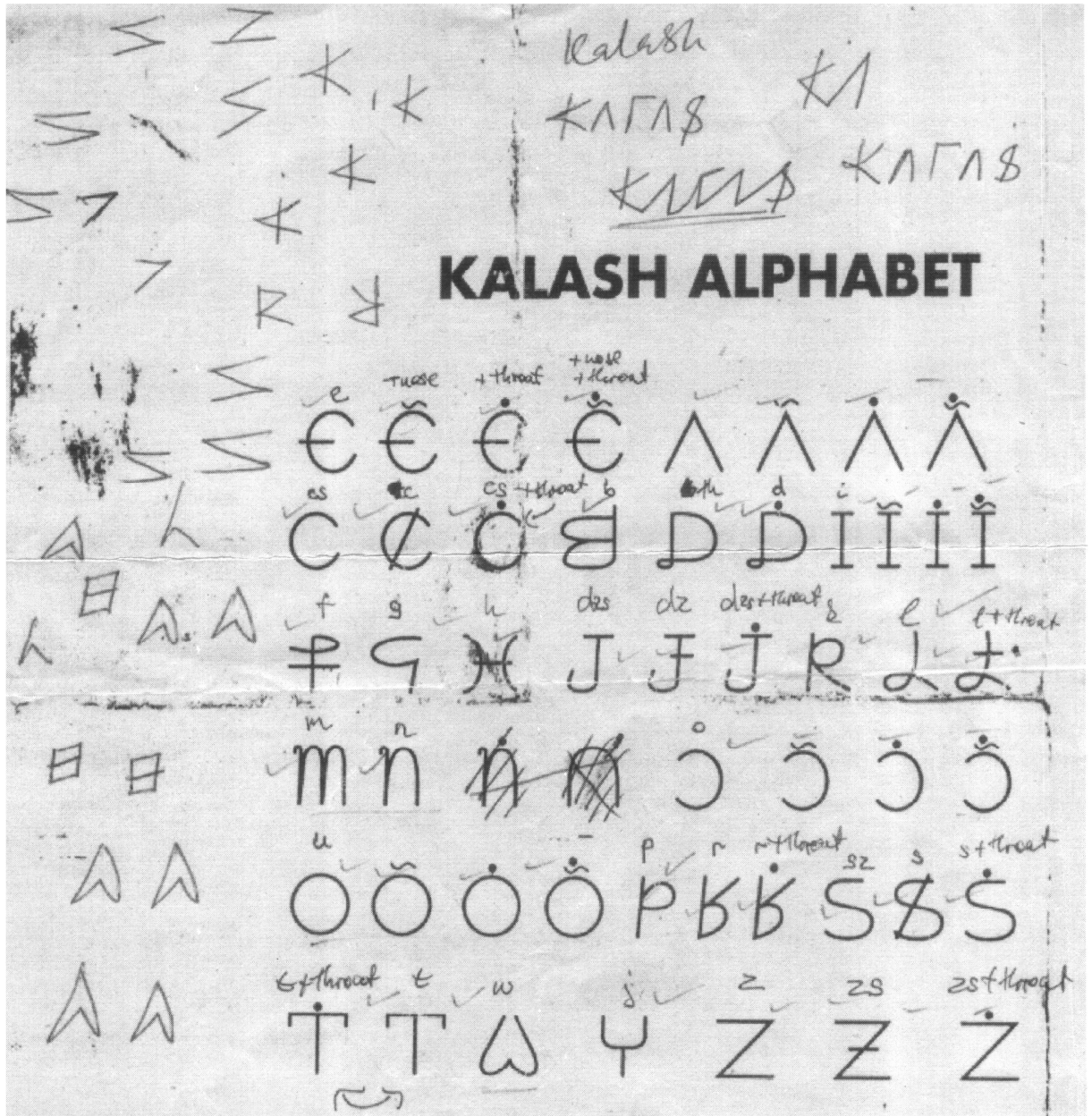
mh	1	1
jh	1	1
dh	1	1
E	1	1
U~	0	0
I~	0	0
i~	0	0
A~	0	0
v	0	0
tsh	0	0
rh	0	0
nh	0	0
lh	0	0
dzh	0	0
dz	0	0
Zh	0	0
Z	0	0
Th	0	0
R	0	0
N	0	0
Lh	0	0
Jh	0	0
J	0	0
I	0	0

1962

8293 (av. 4.2 phonemes per word)

Appendix 9: Kalasha Alphabet chart – indigenous script

(designed by Injinier Khan)



چھیت	چھیترا	چھیتاوا	چھیتراو
چھو	چھووا	چھوواو	
ڈوم	ڈومب	ڈومیل	ڈومیل
مین	مین	مینج	مینج
بون	بوین	بونج	بونج
چھیتراو	چھیتراواے	چھیتراالے	
بیرپو	بیرپوے	بیرپے	
بیرپو	بیرپوے	بیرپے	
کاو سانگاو	کاو سانگاوونا	کاو سانگاونا	
	کاو سانگاونا		
ھاواو	ھاواوھاو	ھاوااٹو	
پھاو	پھاوونا	پھاونا	

Appendix 11: Kalasha orthography questionnaire – Roman script (sample page)

ukdok	ugdok		
ukuSik	uguSik		
krom karaw	krom garaw	kromkaraw	kromgaraw
payda karaw	payda garaw	paydakaraw	paydagaraw
istrizha kuAk	istrizha guAk	istrizhakuAk	istrizhaguAk
du kuAk	du guAk	dukuAk	duguAk
han	hand	hanuna	handuna
phon	phonduna	phonuna	phonduna
mon	mond	mondr	
monas	mondas	mondras	
akul min	akul mind	akul mindr	
akulmin	akulmind	akulmindr	
akulminas	akulmindas	akulmindras	
bit	bitr	bitani	bitrani
pit	pitr	pito	pitro
		pit o	pitr o
		pit-o	pitr-o

Appendix 12: Summary of script choice issues by relevant factors

KALASHA SCRIPTS

PHONEMES	Consonants	Vowels	Semi-vowels	TOTAL
Normal	18	5	2	25
Retroflex	8	5		13
Aspirated	9			9
Nasalised	3	5		8
Retro/Nasal		5		5
Retro/Asp	3			3
TOTAL	41	20	2	63

FACTORS OF CHOICE	ARABIC	ROMAN
<u>Sociolinguistic</u>		
Familiarity to students	yes	growing
Acceptance in wider community	local	international
Tradition - preserving an artform	yes	no
Tradition - preventing new initiatives	yes	no
Political benefit of conformity	yes	no
Legal access	inhibited	promoted
Educational opportunities	limited	extensive
Cultural distinction, sense of freedom and self-esteem	no	yes
Connotation of spelling	simple	complicated
Connotation of alphabet	complicated	simple
<u>Linguistic</u>		
Alphabetic variants	yes (based on word position)	no
Retroflexion	established (using <i>toy</i>)	unestablished (use caps or ' or dot below?)
Aspiration	established (using <i>/h/</i>)	unestablished but intuitive (use <i>/h/</i>)
Nasalisation (word-final)	established (using <i>nun-e-ghunna</i>)	unestablished (use ~ following or overtyped)
Nasalisation (word-medial)	impossible	possible
Nasalisation (multiple vowels)	only single usage is possible after last vowel	single or repeated usage possible
Velar nasal (intervocalic)	impossible	possible
Velar nasal + stop (word-medial)	established (using <i>/ng/</i>)	unestablished (use <i>/ngg/</i>)
Basic vowel repertoire	limited (vowel height usually undifferentiated by available diacritics, so i=e & u=o)	sufficient (a, e, i, o, u)

Semi-vowel repertoire	none (undifferentiated in alphabet, so e=y & o=w)	sufficient (w & y)
Digraphs	established (only for aspiration)	unestablished for aspiration, /zh/ & /dz/; established for /sh/
Joining rules	yes (relatively complicated)	no
Word-position environments available to teach letters	word-final 90% clear (but unconventional)	any (100% clear)
Word-initial vowels	require word-medial aleph introduction	simple
Numeral direction	opposite	continuing
Capitalisation (sentence/name initialisation or emphasis)	impossible	possible (unless caps used for retroflexion, in which case caps distinction impossible)
Text formatting (bold, italics)	impossible	possible
Cursive/printing distinction	no	yes
Font range available	less	more
Quotation	rare, but not necessary in Kalasha as speech is always directly quoted (therefore no need to distinguish from indirect speech)	possible
Diacritics (affect handwriting speed and efficiency)	many necessary	few necessary
Hyphenation	impossible (reserved for end of sentence marking)	possible
Loan words (possible confusion with other orthography)	yes	yes
Technological		
Software - availability, cost, usability, technical help	prohibitive	practical
Interface with the Internet (freedom of information)	inhibited	promoted

Appendix 13: Kalasha literacy materials

PRESENTATION OF DRAFTED EARLY LITERACY MATERIALS TO THE KIOC 2000

Presentation Aims

1. To Introduce to Kalasha teachers and elders the attempts that we have made so far to honour and value their language.

- The Past- a history of our preparations
- The Present-The draft books- intentions, possibilities and limitations.
- The Future-Their felt needs for literature, current proposals reviewed.

2. To invite a genuine critique by the Kalasha of these materials.

These aspects of the readers will

be examined:-

Content and style

Language

Script Details

Illustrations

Format

Materials

(The results will be useful for the future development of literature.)

3. To explore together future strategies for producing materials.

- Writers
- Illustrators
- Consultants
- Publishing and distribution
- Other relevant topics as they are raised.

4. To establish a protocol for the use of photographs of recognisable Kalasha persons in future publications.

5. To listen to, and endeavour to understand any matters raised by the teachers with regard to these or future materials, and to base future strategies faithfully on their input.

Materials

The following materials will be presented during these meetings.

Pre-reader : **Peruk Ne Peruk** **(Same and Different)**

A draft publication of visual discrimination exercises designed to prepare students for recognising and distinguishing between the shapes of the letters of the Persian style Kalasha Alphabet.

Teacher's Manual for the Pre-reader

If the final editing is finished on time this will also be presented for feedback.

Alphabet Book: KaLaSa Alibe **(Kalasha Alphabet - Know your own alphabet)**

Copies of this book (1.) were provided to some schools, but hardly used due to lack of time and training.

1. A Persian style script version
2. A Roman script version in Roman order
3. A Roman script version in Persian order

Teachers Manual for the Alphabet Book

If the final editing is finished on time, this will also be presented for feedback.

Sound Book: KaLaSa Awas (Kalasha Sounds)

An innovative approach to foster awareness of sound-symbol correspondence.

Persian style alphabets are comprised of the word final form of the letters. To increase awareness of the sounds of the letters, a picturable word is presented with the focus sound in word final position. A simple practise exercise follows the sound on each page.

1. A Persian style script version
2. A Roman style script version in Roman order
3. A Roman style script version in Persian order

Word lists:

Based on the Dictionary compiled by Ron Trail and Gregory Cooper, and prepared by Dale Hamilton, these lists will be in Roman script, and in Persian style Kalasha script (if the software allows this on time).

Trial early readers:

Phase one titles;	<u>DraC</u>	<u>(Grapes)</u>
	<u>Bira</u>	<u>(Male goats)</u>
	<u>NaTik</u>	<u>(We Dance)</u>
Phase two titles:	<u>Mos Ahu</u>	<u>(Meat Bread)</u>
	<u>Tai Som Ek mon</u>	<u>(A Word with You)</u>

It is anticipated that in the future there will be a phonemically oriented systematic **core primer** developed. The current early readers are intended to foster a sense of anticipation about reading. To introduce children to the capacity of books to reflect everyday life, and to give readers the experience of sharing a reader, being read to, and the concept that making books is within their own reach too. These readers represent a 'whole language approach' which can only serve as a supplement to a yet to be **developed core primer**.

Preparation time permitting, a **DRAFT Kalasha/Urdu/English glossary** for use in schools may be presented for discussion.

Please note: Greg Cooper will be preparing for discussions on issues concerning the use, and choice of an orthography for Kalasha. An outline will be available.

Elsa Cooper will be presenting the draft literacy materials as above.

E. COOPER SYDNEY 2/12/00

Appendix 14: Genre scheme

A taxonomic summary of each of Longacre's (1983) discourse types. Surface structure types are also indicated as a separate heading within each genre.

NARRATIVE Monologue

+ **Contingent temporal succession**

+ **Agent orientation**

– **Projection**

REPORT

Reality: real

Situation: informal

Nature: serious

Characters: people

Audience: Kalasha people

Surface structure type: conversation

Topics: news, current affairs, weather

Purpose: to inform

Emotivity rating: 4

Possible literary formats: newspaper articles and columns

ANECDOTE

Type: real

Situation: informal

Nature: serious or humorous

Characters: people

Audience: people

Surface structure type: conversation

Topics: personal affairs and anecdotes

Purpose: to share experiences

Emotivity rating: 5

Possible literary formats: newspaper articles, primers

STORY

Reality: real or fictional

Situation: casual, relaxed

Nature: humorous, anecdotal

Characters: people (Kalasha people, tourists, anthropologists)

Audience: peers

Surface structure type: conversation, jokes

Topics: people in funny situations

Purpose: to entertain

Emotivity rating: 7

Possible literary formats: newspaper columns, anthologies, primers

FAIRYTALE

Reality: real

Situation: trusted relationship

Nature: serious

Characters: fairies and people

Audience: close friends or relatives

Surface structure type: report

Topics: encounters with fairies and evil spirits, occult phenomena

Purpose: to inform and convince

Emotivity rating: 2

Possible literary formats: newspaper or magazine articles; letters to editor; dissertations

MYTH

Reality: real

Situation: informal

Nature: serious

Characters: supernatural beings (demigods, spirits, giants, monsters, supermen)

Audience: younger generation

Surface structure type: explanation

Topics: mythology, supernatural history

Purpose: to educate

Emotivity rating: 2

Possible literary formats: textbook

FABLE

Reality: fictional OR real

Situation: informal

Nature: casual

Characters: animals and people

Audience: children

Surface structure type: recital

Topics: stories and life situations

Purpose: to entertain and/or to convey morals/values

Emotivity rating: 5

Possible literary formats: magazine article; illustrated anthology or booklets, primers

LEGEND

Reality: real

Situation: formal (in festival)

Nature: serious

Characters: ancestors, royalty

Audience: public

Surface structure type: song

Topics: ancient history, tales of former times

Purpose: to educate

Emotivity rating: 2

Possible literary formats: textbooks

HISTORY

Reality: real

Situation: formal occasion

Nature: serious, venerating

Characters: ancestors

Audience: public

Surface structure type: song, oration

Topics: heroes and heroic acts, merit feasts, mighty men of valour, war, kingdoms, major political allegiances, regional control, political and royal relationships and tactics

Purpose: to educate

Emotivity rating: 1

Possible literary formats: textbooks

PROCEDURAL Monologue

+ **Contingent temporal succession**

– **Agent orientation**

+ **Projection**

INSTRUCTION

Reality: didactic

Situation: informal

Nature: practical

Characters: family and colleagues

Audience: family and colleagues

Surface structure type: conversation

Topics: practical aspects of arboriculture, agriculture, construction, goat herding, cooking, weaving, religious rites, trekking

Purpose: to instruct others in methods for development of local technology

Emotivity rating: 4

Possible literary formats: instruction sheets; illustrated manuals; newspaper columns

BEHAVIOURAL Monologue

– **Contingent temporal succession**

+ **Agent orientation**

+ **Projection**

EXHORTATION

Reality: hortatory

Situation: formal

Nature: practical

Characters: family and colleagues

Audience: family and colleagues

Surface structure type: counsel, pep talk, advice, admonition

Topics: relationship advice, marriage guidance, behavioural expectations, ceremonial rites, festival participation

Purpose: to encourage, advise, instruct, admonish, train, enthuse

Emotivity rating: 1

Possible literary formats: booklets

BEHAVIOURAL Monologue

- **Contingent temporal succession**
- + **Agent orientation**
- **Projection**

EULOGY

Reality: real

Situation: funerals

Nature: serious

Characters: dead person and ancestors

Audience: dead person (and mourners)

Surface structure type: oration

Topics: acts of courage, enterprise, generosity and heroism

Purpose: to eulogize

Emotivity rating: 2

Possible literary formats: booklets

EXPOSITORY Monologue

- **Contingent temporal succession**
- **Agent orientation**
- **Projection**

INFORMATION

Reality: real

Situation: informal

Nature: serious

Characters: people

Audience: Kalasha people

Surface structure type: conversation

Topics: apiculture, astronomy, biology, environmental issues, foreign customs and culture, health, hygiene, linguistics, literacy, medical (diagnosis, treatment options), nutrition, technology

Purpose: to inform, educate

Emotivity rating: 4

Possible literary formats: newspaper feature articles, booklets, primers

DESCRIPTION

Reality: real

Situation: informal

Nature: educational

Characters: Kalasha people

Audience: Kalasha people, visitors

Surface structure type: exposition

Topics: customs and culture, anatomy, botany

Purpose: to educate

Emotivity rating: 4

Possible literary formats: illustrated magazine articles, textbooks, primers

RIDDLE

Reality: poetic

Situation: intimate

Nature: fun, humorous

Examples: 'My and your riddle: On a round stone, a horse's hoof print' (navel)

'My and your riddle: In the river, a white plate.' (moon)

Audience: children

Surface structure type: recital

Purpose: to entertain

Emotivity rating: 7

Possible literary formats: illustrated booklets; newspaper or magazine columns, primers

REPARTEE Dialogue**DISCUSSION**

Reality: real

Situation: informal

Nature: conversational, sometimes casual

Characters: people

Audience: family and friends

Surface structure type: drama

Topics: aid and development projects, local construction projects, cottage industries (honey, dolls, souvenirs) current affairs, employment opportunities, girls, hotels, legal cases, romances, schools, social issues, tourists and tourism, technology,

Purpose: to share knowledge and ideas

Emotivity rating: 4

Possible literary formats: newspaper feature articles; letters to editor

DEBATE

Reality: real

Situation: informal

Nature: emotionally charged

Characters: people

Audience: family and friends

Surface structure type: drama

Topics: disputes, land rights, legal cases, current affairs, family relationships, Kalasha-Muslim relations, Kalasha-Muslim conversions, menstrual hut issues, money and debts, ceremonial purity, practical arrangements

Purpose: to share opinions, publicize

Emotivity rating: 5

Possible literary formats: newspaper feature articles; letters to editor

BANTER

Reality: imaginary, vigorous and rousing raillery,

Situation: festive

Nature: fun, risqué to vulgar, very humorous

Characters: people

Audience: teenagers

Surface structure type: repartee and chorus

Topics: sexual situations

Purpose: to entertain

Emotivity rating: 6

Possible literary formats: lyric sheets to accompany audio cassettes

Appendix 15: Kalasha literature types

What follows is a list of potential literature types in Kalasha. Texts for many of the following categories of literature have already been written in Kalasha, using either an Arabic script or an interim Roman orthography, both of which I developed in the 1980s. Those marked with an asterisk (*) were either transcribed from recordings of indigenous speech or drafted by Kalasha individuals, while those marked with a circumflex (^) were drafted by my colleagues and me (in close collaboration with, and extensive checking and testing by, indigenous speakers).

1. LITERACY (all genres)^

- Pre-readers (to help potential readers get used to the handling of books, pages, symbol shapes, matching and differentiation, direction of text, both in the book and on the page, etc.)^
- Alphabet and orthography (books, charts, cards)^
- Phonics, spelling, handwriting^
- Numbers, counting, arithmetic, time
- Primers (phonic based, structured learning)^
- Transitional materials (to and from Arabic script)
- Dictionaries and glossaries (alphabetic and semantic)^
- Literacy training (teachers' manuals)^
- Language education (English and Urdu language learning materials)^

2. LOCAL INDIGENOUS LITERATURE*

- Performance literature (transcribed from unelicited recordings of popular speech acts of any genre performed in the literary media of songs, poetry, stories, rhymes, riddles, proverbs, prayers, etc.)*
- Prosaic literature (elicited spoken and written texts of current topics of Report, Anecdote, Story and Exhortation genres)*
- Cultural literature (elicited spoken and written texts of Description genre)*
- Archival literature (transcribed from unelicited and elicited recordings of speech acts of Myth, Fable, Legend, History and Eulogy genres performed by specialist orators)¹*

3. PERSONAL COMMUNICATIONS*^

- Correspondence (Report, Anecdote and Exhortation genres)*^

4. PUBLIC MEDIA (Report, Anecdote and Information genres)*

- Newspaper articles
- Web pages

5. FORMS AND RECORDS^

- Birth, death and marriage certificates and registers^
- Calenders and journals of events
- Weather records

¹ The specialist's role as an orator needs to be protected.

- Maps, land boundaries

6. INSTRUCTIONAL/FUNCTIONAL GUIDES (Instruction genre)

- Telephoning
- Finances (Currency, debts, IOUs, banking)
- Business administration and management
- Travel arrangements

7. COMMUNITY DEVELOPMENT (Instruction and Information genres)^*

- School education syllabus
- Health (fever management, worm and parasite infestations, eye diseases, goitre prevention and treatment)
- Hygiene (pathogens, cleanliness, food preparation, toileting)*
- Medicine (anatomy, antibiotics, medicinal treatment regimes)
- Family Planning (issues, options, methods)
- Environmental issues (forest protection, waste disposal, insecticides)
- Legal issues (procedures, advice)

8. ENCYCLOPEDIC KNOWLEDGE (Description and Information genres)^

- Natural science (botany, zoology, meteorology, seismology, astronomy)
- Geography (rivers, rocks and minerals, erosion, mountains, mapping)
- Animal science (sheep and goat husbandry, cattle raising, beekeeping)
- Agriculture, arboriculture, viniculture, silviculture
- Cottage industries (dairy production, winemaking, honey production, doll making)
- Politics

9. WORLD LITERATURE*

- Stories, poetry, fairytales, myths, legends, fables, history, proverbs, etc. from other cultures*.

Appendix 16 Kalasha discourse samples

following Longacre's (1983) discourse types

Narrative discourse

Se moc to gordok khojika desha paraw. Bo desha pai sapraw.
/se motʃ to gordok k^hodʒik-a deʃa paraw. bo deʃa pa-i sapraw.
'that' 'man' 'that' 'donkey' 'searching-to' 'far' 'went' 'very' 'far' 'Go-ing' 'found'
'That man went far to search for the donkey. Going very far, he found it.'

Saprai o gordok amaraw. Mari o tasa mos ashaw. Ashi abomaw.
sapra-i o gordok amaraw. mar-i o tasa mos aʃaw. aʃ-i abomaw./
'Find-ing' 'then' 'donkey' 'killed' 'Kill-ing' 'then' 'its' 'meat' 'ate' 'Eat-ing' 'vomited'

'Finding the donkey, he killed it. Having killed it, he ate its meat. After eating he vomited.'

Procedural discourse

Rajuk shēhē kai karin ki awel jac c'hilin. Phato to l'a'hay karin.
/radʒuk ʃēhē ka-i karin ki awel dʒats tʃ^hiʃin. p^hato to ʔaʰhaj karin.
'rope' 'thus' 'do-ing' 'do' 'that' 'firstly' 'goat hair' 'card' 'then' 'it' 'rolled' 'do'

L'a'hay kai to o sawzai kond'in. Kond'i to du lishen.
ʔaʰhaj ka-i to o sawza-i kondʒin. kond-i to du liʃen.
'rolled' 'make-ing' 'it' 'then' 'make-ing' 'spin' 'spin-ing' 'it' 'two' 'twist'

Du lishai to o l'en. Phato tre lishai rajuk karin. Shēhē thi
du liʃa-i to o ʔen. p^hato tre liʃa-i radʒuk karin. ʃēhe t^h-i
'two' 'twist-ing' 'it' 'then' 'ply' 'then' 'three' 'ply-ing' 'rope' 'make' 'thus' 'become-ing'
ek rajuk sawz hiu.
ek radʒuk sawz hiu./
'one' 'rope' 'made' 'becomes'

'This is how rope is made. Firstly, they card the goat hair. Then they make it into a roll.
Having made it into a roll then they spin it. Having spun it they twist it into two strands.
Having twisted them they ply them. Then twisting three they make a rope. That's how a rope
is made.'

Behavioural discourse

S'abas' o tay hatya ey may baya Mutimire nawaw!

/ʃabaʃ o taj hatja ej maj baja mutimire nawaw/

'bravo' 'oh' 'your' 'sake' 'hey' 'my' 'brother' 'Mutiimir' 'grandson'

'Bravo to you oh my brother, grandson of Mutimire'

Expository discourse

Biriuay moc bo shishoyak. Istrizha asta bo shishoyak. Biriuay ek asta

/biriu-ay motʃ bo ʃiʃojak. istriʒa asta bo ʃiʃojak. biri-u-ay ek asta

'Birir-in' 'people' 'very' 'beautiful' 'women' 'also' 'very' 'beautiful' 'Birir-in' 'one' 'other'

bo mashur ishnehari shiau. Se o d'a. Sawin pi shia ishnehari bo

bo maʃur iʃnehari shiau. se o ɖa. saw-in pi ʃia iʃnehari bo

'very' 'famous' 'thing' 'is' 'that' 'then' 'wine' 'all-'s' 'from' 'this' 'thing' 'very'
prus't'.

pruʃt.

'good'

'The people in Birir are very beautiful. The women are also very beautiful. In Birir there is one other very famous thing. That is wine. Of all things this is very good.'

Appendix 17: Kalasha alphabet chart – Arabic script (with Roman annotation)

آلیے		ا	ا	کالاشا	
ط	ت	پ	ب		
ش	چ	چ	ج	ج	
ڑ	ژ	ز	ر	د	د
گ	ک	ش	ش	س	
ن	م	ل	ل		
ھ	و	و	و	و	
ٹ	ے	ٹی	ی		

کالاشا آلییونا 35 نیشان شیان -

The Kalasha alphabet has 35 letters.

پاے آمپھون سوم وال پارین ژے تان
داستور کارین.



کھاتراناک شواں دُو سودا ژولُ

ایک آدوا ژے ماے دؤست مؤشتیاک ایک
داریا شویونا آؤزؤ مؤزیمان آیس. شاتارا
واتؤنا آبی بو چھوتیاک آیس. هوما
گرومانی ایک موچاس پاے آستا داریا
سینداے چاریمان آینی. تاسا موچاس سوم
ایک او ایک گاڈا پپھلاک هیکاس باش
کاکبے موند شواں آیس. سے شواں
آلاے آندلے کای جاگای رایسان آیس. ماگام
سے شواں هوما ہی کابھی دیشا آیس.

آژے ماے دؤست بو مازآس تھارا ایک او

موچی آستا کیریک دییک کھوشان.
کیریک دییک ہے شورؤک هاواو هاو،
سودا گاڈا پینڈ گال کاریکاس باتی
ٹھیت ٹھیتراے دراشنی گال کارین.
رات ہے. هاواو هاو دؤر دؤرلے پای آٹھاپنک
دیین ژے آنگؤشتے مؤزین. سوچا رات
لائے هین اوچے مول هالیوا پاچای ژون.

هیمان کھول تھی باسؤن ہے شورؤک
هاواو هاو، مؤمؤریت مولکی گھیری
شیشویاک هیکاس دیل. مؤٹھپک آزار
نؤان زینداگی شورؤک کارین. لاشنا
پیران تا نیچان گھیری نؤان پیران او
سامبی شیشویاک کای پائیؤ دیین.
مؤٹ مؤٹلے پاچھپتیاک تان شیشویاک
آواز گری چوٹی کھودایاس سیپھات
کارین. سودایاک آستا شیشویاک
هرونیکای پای لایپتیاک هین. شایا
واتؤنا ساک نیھاسالاک کارپؤ.

مؤمؤریتاے موج شیهیں تھی تان
زینداگی تیرین ژے کوشان تھی تان

[illegible]

ماہیگران بنی حاجت مومن

دیلی تھی مایکس بائی تیاری کاری

نیمیا کاس
تھاریکا

شاہ کیا پونا دو تھار کا شل -

1 شامسا پٹاناس نوؤنا جیس نیٹان ۰ کیا کی شیوؤنا

شیان ترے بیٹھان پی پھداک شیو -

2 شامسا بیٹھاس نونوا حصیں پوندیوراک بیٹھاس خانوائے ۰ کیا کی آویلو
مولنی بیٹھاس پھروک شاد - ایسا جیلے کورلے وارا کونا نا آویلو

تھارکا شیو کوریے وارا کونا او ہنو تھارکا شیو -

ادیک وات تلمای پھوننا لہاے۔ کسی دوفی ساسی جاپاس نوونا پنٹان ٹاپیل
کورے کی ساسی جاپاپ شاپو تہا نوونا ساہک ماہاو می پنٹان ٹاپیلوری۔

اس کتاب میں طالب علم حروف کو دیکھنے اور پہچاننے کی مشق کرتا ہے اور حروف کی بناوٹ اور فرائض کو دیکھتا ہے۔

یہ سب باتیں میرا مسکینہ کیلئے بہت ضروری ہیں۔

خود اعتمادی بڑھانے کا طریقہ

استعمال کے اصول :-

ان پڑھ بانی سے ہمیشہ فارغ وقت میں ملیں۔

۲ پڑھنے کو نہ کہیں بلکہ مسروف کی سپہان کر وائیں۔

۳۔ کتاب میں دو طریقے ہیں (۱) اس حرف کے نیچے نشان (۵) لگائیں جو کہ لام میں ایسے مُرُور

پانچوں مردوں مختلف ہو۔ (2) اس مرد یا مردان لگائیں جو کہ میلے خانے کے حوز

ملتا جلتا یا (dam) ہو : کسی صفحہ پر طریقہ اور کسی صفحہ پر دوسرا طریقہ۔

۴۸ مس کو عقوقاً وقت دے کہ وہ خود سوسے اور کیر صمغ حرام کے شمع نشان لگائیں

جو صمیم جواب ہے اُس کے نیمہ طالب علم خود نشان لگائے۔

Appendix 20: Kalasha alphabet – Roman script
(interim version as a result of the conference)

Kal'as'a Abe

a a' b c c'

d d' e e' g

h i i' j j' k

l l' m n o o'

p r r' s sh s'

t t' ts u u'

w y z zh z'

nis'an pashi awaz maan

Appendix 21: Kalasha orthographic alternatives – Roman script

(sample page tested at orthography conference)

- A~
- 1 RETROFLEXION: CAPS, NASALISATION: TILDES FOLLOWING
 bribo bishai jA~ kara ne e? mos au karik. she~he~ pe haw pruST. ko ne karik day? mos au o kas trupel? pari de, pai angguzhi oni. parim day la, khe~ haLi day. aLa bha~cuni anday kari de la. LO~ zhe cruka asta mishaa. o may chu, LO~ ta wEkhal huLa. ticak asta mishai e khosh. kura cruka dAim onika para de. gonj dur tada penyaka shian. da o chet, ama asta mishaa. mos au bo zaw huLa. istek gilās Da asta pe haw to bo maza araw dyaa. Da pe haw dyaa dada putra ne dik baS hawaw dyaa.
- a~
- 2 RETROFLEXION: APOSTROPHES, NASALISATION: TILDES FOLLOWING
 Bribo bishai ja'~ kara ne e? Mos au karik. She~he~ pe haw prus't'. Ko ne karik day? Mos au o kas trupel? Pari de, pai angguzhi oni. Parim day la, khe~ ha'l'i day. Al'a bha~cuni anday kari de la. Lo'~ zhe cruka asta mishaa. O may chu, l'o'~ ta we'khal hu'l'a. Ticak asta mishai e khosh. Kura cruka da'im onika para de. Gonj dur tada penyaka shian. Da o chet, ama asta mishaa. Mos au bo zaw hu'l'a. Istek gilās d'a asta pe haw to bo maza araw dyaa. D'a pe haw dyaa dada putra ne dik bas' hawaw dyaa.
- ā
- 3 RETROFLEXION: DOTS, NASALISATION: TILDES OVER
 Bribo bishai ja kara ne e? Mos au karik. Shēhē pe haw prušt. Ko ne karik day? Mos au o kas trupel? Pari de, pai angguzhi oni. Parim day la, khē haļi day. Aļa bhācuni anday kari de la. L̥ō zhe cruka asta mishaa. O may chu, l̥ō ta wēkhal huļa. Ticak asta mishai e khosh. Kura cruka dāim onika para de. Gonj dur tada penyaka shian. Da o chet, ama asta mishaa. Mos au bo zaw huļa. Istek gilās ḍa asta pe haw to bo maza araw dyaa. Ḍa pe haw dyaa dada putra ne dik baṣ hawaw dyaa.
- ā š
- 4 RETROFLEXION: DOTS, NASALISATION: TILDES OVER, FRICATIVE: WEDGES
 Bribo bišai ja kara ne e? Mos au karik. Šehe pe haw prušt. Ko ne karik day? Mos au o kas trupel? Pari de, pai angguži oni. Parim day la, khē haļi day. Aļa bhācuni anday kari de la. Lo že čruka asta mišaa. O may čhu, l̥ō ta wēkhal huļa. Tičak asta mišhai e khošh. Kura čruka dāim onika para de. Gonj dur tada penyaka šian. Da o čhet, ama asta mišaa. Mos au bo zaw huļa. Istek gilās ḍa asta pe haw to bo maza araw dyaa. Ḍa pe haw dyaa dada putra ne dik baṣ hawaw dyaa.
- ā š āya/aya
- 5 RETROFLEXION: DOTS, NASALISATION: TILDES OVER, FRICATIVE: WEDGES, STRESS: APOSTROPHES
 Bribō bišái ja kára ne e? Mos aú kárik. Šehē pe haw prušt. Ko ne kárik day? Mos aú o kas trupél? Pāri de, pāi angguži oni. Parīm day la, khē hāļi day. Aļa bhācuni andāy kári de la. L̥ō že črúka ásta mišáa. O may čhu, l̥ō ta wēkhāl hūļa. Tičak ásta mišhai e khošh. Kúra črúka dāim ónika pára de. Gonj dur tādā pēnyaka šian. Da o čhet, ama ásta mišaa. Mos aú bo zaw hūļa. Isték gilás ḍa ásta pe haw to bo maza áraw dyāa. Ḍa pe haw dyāa dāda pútra ne dik baṣ hāwaw dyāa.
- ā š āya/aya
- 6 RETROFLEXION: DOTS, NASALISATION: TILDES OVER, FRICATIVE: WEDGES, STRESS: UNDERLINE
 Bribō bišai ja kara ne e? Mos au karik. Šehe pe haw prušt. Ko ne karik day? Mos au o kas trupel? Pari de, pai angguži oni. Parim day la, khē haļi day. Aļa bhācuni anday kari de la. L̥ō že čruka asta mišaa. O may čhu, l̥ō ta wēkhal huļa. Tičak asta mišhai e khošh. Kura čruka dāim onika para de. Gonj dur tada penyaka šian. Da o čhet, ama asta mišaa. Mos au bo zaw huļa. Istek gilās ḍa asta pe haw to bo maza araw dyaa. Ḍa pe haw dyaa dada putra ne dik baṣ hawaw dyaa.

Appendix 22: Kalasha-English glossary sample – Roman script

a~a~ Yes!	adhe Some of a larger group, a few	al'el'a Over there (out of view)
a~c'ak Prickers, small thorns	adiik To eat	al'gana The goal in a game of fv:gaL
a'i Duck	adina Friday	-al'i Kinship suffix, 2nd person's or persons' plural kin, all cases
a'in Millet	adrak Forest, woods	al'uk Pumpkin, squash
a'nggu Finger, toe	adua Day	ala Up there, uphill
abad'o'i Childless (of both married and unmarried people)	adyaik To run	alagul Bustle, rush, activity
abadi House	adyek adyek hik To race	-alak hik Desire or inclination to do something
abas Useless, bad	agarak Same, equal	alapala hik To be greedy and avaricious (especially about eating)
abat To have children	agazer dek To bother, bore	alat Difficulty, trouble
abatha Wrong, inaccurate, unlawful, bad, off target	ageylik Ill-behaved, bothersome, insistent	albat Probably, perhaps
abayo Wow! (an exclamation of surprise)	agraik To become tired	-ali 2nd person plural, past tense, Class 4 verbs
abe' A fenced off place in a goat shed for the young goats and lambs	agranu Manger, feeding trough	alibe Alphabet
abi We (nominative)	ainda The future	aliphik To follow and catch, keep up with someone
abi You (plural nominative)	aj This, this present, current	alizanggi Large seed about the size of a walnut used as medicine for boils
abuji e Good morning!	ajap Remarkable, wonderful	almari Cupboard, shelves
abut parik To make a mistake	ajat Need	alu Potato
ac'ongki Crooked, bent	ajel Death, the time for one to die	aluca Plum, plum tree
ac'u A key made of a long wire	ajis Well-behaved, helpful, gentle, polite, courteous	aluok Potato curry
ac'u Wild celery	ajo Now	-am 1st person singular, present-indefinite tense, for Class 2 and 3 verbs
ac'uaga The valley of Achuaga between Rumbur and Bumburet	ajona Guest	ama Him, her, it (accusative, present, near)
ac'uik To dry up and become hard from evaporation and age	ajtu After, from this time on	ama Unripe, uncooked, raw
ach'o~ik Story, fable	ak karik To defecate (as said to a child)	amal Equinox (vernal and autumnal)
ach'o~ik The dry prickly leaves of hollyoak	-ak Diminutive suffix	amal, hamal Habit
achambi Ceremony celebrating birth	akaliat Minority	aman Peace, lack of war, subsiding of an event
achardini The wedge which holds the ploughshare in place	akhabir Old (of people)	amanat Entrusted
aci Back, again	akhe'ik To get firm, hard	amat Perhaps
acikret'i Lower body, from waist to feet	akher End, final time	amatak saras Ceremony at end of Chawmos (winter festival) to end sexual abstinence
aco'yak Watery, thin	akherat Trouble, hard work	ambak Very filling (of food)
ad'uek To blow up, cause to swell, inflate	-aki Adjectivizing suffix	ambo'i Lazy, ineffective, weak
ad'uik To swell, bloat	aksariat Majority	ambur geci dyek To be very dark
adalat Court	aktiar Authority (of owner over owned)	ambur Pliers, forceps
adam Body	aktiat Care	amch'ek To fit together
adap karik To be hospitable to someone, treat with respect	akul Wisdom, character	ame~a Sheep
adat Habit, custom	al'a Him, her, it (accusative, present, remote)	ame~ost Sheepskin
adem Adam (first man)	al'ahak hik To be impatient	ame~yak nisikayn The womb of a ewe
ademzat Very good, true	al'as'ing To elope with a man	
	al'ay There	
	al'ayhak Aside, away	
	al'e A place where salt is found	
	-al'e Possessing large size of something	

amin		anzek To send something with someone	as' Shoulder
amishaw The house for shepherds in the high pastures		apac'u Ouch!	as'ghar Very bitter
amishtyonu A cheese dish of fv:sac0 topped with hot butter		apaco'~i dyek To hop on one foot	as'is'a Widowed
amitra Disobedient		apak Leather pocket for stone in slingshot	as'ka dek To sigh
amondra Disobedient and disrespectful		apaw dek To stay, live	as'kek To dislike
amur karik To command, order		apaw diek To cause to stay, live, wait	as't' Eight
-an 3rd person plural, past tense, for Class 1, 2 and Causative I verbs		aphat Evil, disaster, trouble, misfortune, sickness, spiritual harm	as'ung Forked branch used for spinning goathair
-an Because		aphiat Comfort, peace, rest	asa He, she, it, (nominative, present, remote)
-an In, into, within		aphsus Regret	-asa In, during, with
an Mountain pass, boundary		apil A legal appeal	asakal Village headman
-an Nominative 3rd person plural suffix		aprasta Lazy, weak, incompetent	asar Effect
-an Present-indefinite tense, 3rd person plural, for Class 2 and 3 verbs		aprati Dull, cloudy, dirty, unclear, faded	ash Past tense root of fv:zhuk, 'to eat'
anacu'i The impure area in graveyard where women who die in childbirth are buried		apsar Officer	ashdihar Dragon
anday Here, in here		ar Awl	ashek Lover
andaz karik To aim something at something		ar Past tense root of fv:ka5rik, 'to do'	ashi Mouth
andazai Average size or number of, some--not much, not little		araba Tire, wheel	ashini They (inanimate) were
andena From here, local		aram mos Meat from animals which have been killed by decapitation or by Kalasha	ashis It (inanimate) was
andial'ay Around this area		arami Rest, relief	ashkara karik To speak openly, plainly
anena		arman Desire, longing, spell, enchantment	ashmaish Test, trial, temptation
angga Awake, alert, aware		aroik To ripen or get fully formed	asho Rag
anggar Fire		aru Peach	ashraphi Gold coin, treasure
anggarishiti Fireplace in a goat shed		-arum Ability	ashupi karik To disfigure, mess up
anggarwat Flintstone		aruz'ik To be in heat (of a goat)	-asi Kinship suffix, 3rd person's or persons' plural kin, all cases
anggris Western person		arwa Spirit, soul (of a dead person)	asi Their (present, remote)
anggu Vine		arzan Cheap	asik To be (of animate beings)
anggus't'ye muzhik To play hide the ring		arzi Case, complaint	asil Factual, real, genuine
anggushtyar Ring		-as dek To begin to do something	asingga Cooked thigh piece of a goat or sheep
anggut'i Spare room for guests		-as hik To be about to do something	asiru People of one's house, household
angguzhi Spatula for turning bread		-as mon dek To talk about someone or something	asis I was, he was
-ani Adjectivizer, singular		-as 2nd person singular, imperative, for Class-2 and 3 verbs	askan Easy
-ani From, with		-as 2nd person singular, present-indefinite tense, for Class 2 and 3 verbs	asman Sky
anish The village of Anish in Bumburet		-as Genitive, 3rd person singular	asmun Test, trial
anjarum Clothing		-as Kinship suffix, 3rd person's or persons' singular kin, all cases	aspa'~i The Aspani clan
anjil Measure of one handful			aspak Crotch of pants
anora Hunger			aspa'a Door header
anpal' Illiterate			aspandur A kind of seed which is burned for its smoke to counter evil influences
			aspap Possessions (tools, foodstuffs etc., which are stored)

aspar The village of Aspar in Birir Valley	awarek To cause a cow to begin to give milk	b fv:be. This is the third letter of the Kalasha alphabet
asta Also, too	awarik To give milk	ba'a' Lazy
asta He was, they were (hearsay)	awarik To take off the fire	ba'hik To roast, fry
astru Tears	awas Noise, voice	ba'i Payment in kind for use of something
asuoga Juicy, of fruit	awat Place	ba'kba'k Partly cloudy and partly sunny
at Flour	awatheri A place near the edge of something	ba'mba'i Tassel of corn bearing pollen
at Past tense root of fv:pa5Lik, 'to fall'	awazur Tamarisk, a bush with white flowers	ba'ng Opening
at'alak Small plateaus among the mountains	awe'yak Narrow, small	ba'tsak Short of average stature
at'e'~ The floor of a house	awel First	baba Sister (daughter of one's mother and father)
atek To bring in, take in, cause to enter	aweri hik To go without eating	babayuk Stream, creek
ath'i Bone	awicik To take or receive from someone	baca King
ath'igar May I be cursed!	awio thek To marry within one's clan	bacat karik To rescue, save someone
atik To enter	awizan hik To be startled, frightened	bach'e'~a Small bronze bells sewn onto the fv:kupa5s, 'women's headdress'
atok Flour water left from cleaning the bread bowl	awizha Family, people of one's household, relatives	bach'o'a Calf, one-year-old, male or female
atra There, (nearby, in sight)	awlat Descendants	bad- Bad
atrili The day before yesterday	awsı karik To graft a tree	bad'a chak Shadow of an animate thing
-au 3rd person singular, present-indefinite tense, for Class 1 and 4 verbs	awta Week	bad'acha A large winged red insect
au Bread, food	-ay In, at, on, to	bad'arak A kind of long whitish lentils
-aw 3rd person singular, past tense, for Classes 1, 2, 4 and Causative I and II verbs	-ay Kinship suffix, 1st person's or persons' plural kin, all cases	bad'u Yellow
aw Both (modifies numbers)	aya Here (nearby speaker)	bad'uaw Golden
-aw Doer of an act specified by a verb, agent	aya Mother	bad'ul'a Stout, fat
-aw From, with (plural)	aynak Eyeglasses	bad'una' A kind of bush bearing yellow flowers
-aw In, on (plural)	ayp Fault, blame, mistake	badan Body
-aw Kinship suffix, 2nd person's or persons' singular kin, all cases	ayukun Egg	badel To change
awa Air, breath	ayun Opium	badela Revenge, repayment, (good or bad)
awa Grandmother (father's/mother's mother)	az'a'i Apricot	badhek To shave, shear
-awa To cause an event specified by the verb	az'a'inja An apricot tree bearing fruit with bitter pits (too bitter to eat)	badian Sweater, jumper, jersey
awacek To tease, fool, hassle someone	az'inggik To ripen so as to break open	badik To grow
awacik To scold someone	az'onggyak Slender, thin	badmash Rascal, bad fellow, mischief-maker
awaday karik To do useless, unnecessary work	azap Trouble, suffering	badniat Base, corrupt
awagaw mastruk A month of spring	azat karik To divorce	badok Ax
awala karik To give into the care or custody of someone	azat Free	badri A leather strap, belt
awan Front panel of a man's shirt or kamiz	azec A large clay jar for cheese storage	badshikil Bad-looking, ugly
	azhash Cumin	badsul'uk Estranged
	azhe Fill-in word	badzek To stack wood to build a fire
	azheli Children, offspring	badzerak Young ewe, one to one and a half year-old
	azhl'i Real, pure	
	azho'yak Wart	
	azis The metal lead	

Appendix 23: Kalasha text sample – Roman script

Astek Kal'as'a Gho'~

Gamburi chinikas shawak ne ghoi~ o gurzenay
us't'i chinik tan te waljos'ia. (Chaumos)

Shehe~ pe maas haw tay bayao
tay bati mut'ani po'~a's rau paci chomay s'is'irim.
Gak zhe ame'yak thi o lec'hen zhe pakmod'en.

O suryak o menjas pa'go'e'ya tye, tay asta
gri al'a basthoni kandajaluna pay basim. (Surikas gho'~)

May chu ta mastrukas ja dem o
mastrukas som may khaltabari o.
Jamol'ika onjes's't'a bazayak nasenao o rom
batyak ame'yak. Bihot'as ey may kia zian o
s'ing suiras som geri uchundes o. (Sariek gho'~)

S'ungs'ungao day may kay ishlakan tyes.
May pi o kia chal'es bas'ara cong cuem.
Suda gad'a may kay ishlakan tyen.
May pi o kia chal'es a o tay babak him.

O dari o beruao dura pari.
Janik karin day, sariek karin day,
bira mor karin day.

Jangal bathanay mo kasi may khan o. Tay
khuray o suhorum j'anj'er o.
Asmanani hur goan day la o janatani Hur o
tay raphek o. (Chaumos)

Miuma'u awa al'a ta tu mi zhu may o warek de.
Phasat'uki awa al'a ta tu mi zhu may o warek
de. (Zhoshi)

Brangasulak ga'ey dyey Chetraw pat'r'ang E~E~ pat'r'ang.
(suda l'abe' gh'o'~)

Sudayak de da la ba'ba'o. Day aais te ta khe~
ari. Pasti dramiani kirik histim day
ghoi~ ric' d'ad'ol'a no'~ thi paron. Garamas't'en
day el'i kura. El'i o bachaa moas sudayak.
Pari shatra aau shal'a ni.

a

abé

a

- a₁ Ā Pron. I (Nominative); میں. [See the Personal Pronoun Chart on page 475 below for other personal pronouns.] Oblq: may 'me'. Variant: haiksa (Urtsun). Etym: ahám 'I' T-992.
- a₂ Ā V. Came (past tense root of ik 'to come'); آیا. se á-aw. He came. See also: ik.

Past tense forms of ik

Singular	Plural
áa 'I came'	ómi 'we came'
áa 'you came'	áli 'you came'
áaw 's/he came'	on 'they came'

- a- Ā Pfx. Not. a-móndra súda mon kō ne káriu day. A disobedient child does not listen to instruction. [In our data this prefix occurs in two words only.] See also: aşısa.
- abádi آبادی N. House; گھر. tanté abádi sawzél pútraso. His son will build a house by himself. Syn: dur; khatumán. From: Persian.
- abadōi آباڈوی Adj. Childless (whether married or unmarried); بے اولاد. Syn: istóra; haktá; hindáw.
- N. Bachelor, spinster; کنورا.
- abás آباس Adj. Useless, bad; بیکار. abás mon may ne khoš. I don't like useless language. Syn: gaybaná; bey-šáy; phusúl.
- abás kárik آباس کاریک V. To waste something; ضائع کرنا. wat abás ne karéli. One should not waste time.
- abát
- abát hik آبات حیک V. To have children; مویشی. ábi abát thi áik. We have children. From: Persian.

abát kárik آبات کاریک V. To give children, to cause to have children; اولاد دینا. hóma khodáy abát kái áau. God has given us children.

abathá آباتها Adj. Wrong, inaccurate, unlawful, bad, off target; گمراہ، غلط. abathá mon mo de. Don't say things that are wrong. Syn: rat₂; galát.

abathá hik آباتها حیک V. To become lost, go astray by accident; بھٹک جانا. may pay kawáy paráw haw abathá háwaw. a ne saprés. My goat went somewhere and got lost. I didn't find it.

abathá kárik آباتها کاریک V. 1) To miss, fail to find; کھو دینا. a dur abathá áris. khóji ne saprés. I missed the house. I looked but I didn't find it. 2) To misrepresent someone's words: (بات) تبدیل کرنا. se may mon abathá káda. He changed what I said.

abathá parík آباتها پاریک V. 1) To miss a target; نشانہ خطا جانا، خطا ہونا. bō abathá paráw. The bullet missed (the target). Syn: níšun parík. 2) To lose one's way; راستہ کھو دینا.

abáyo آبایو Intj. Wow! (an exclamation of surprise); اوہ (حیرانی کی کیفیت).

abáyo dek آبایو دیک V. To shout; چیخنا. madrák abáyo dáí máila ki, may píštuna kía kíbaw híu day? The frog cried out, "What is happening on my back?" Gen: mátrik 'to say'; ghóik 'to say'.

abé آبے N. Fenced-off place in a goat shed for the young goats and lambs; مویشی خانے میں میمنوں کو رکھنے کی جگہ. Syn: batyá ku. Variant: ambé.

ábi

ábi₁ آبی Pron. We (nominative); ہم. [See the Personal Pronoun Chart on page 475 below for other personal pronouns.] Oblq: hóma 'us'. Variant: ái. Etym: asmad- 'base of oblique cases plural of 1st person pronoun' T-986.

ábi₂ آبی Pron. You (nominative, plural); آپ. [See the Personal Pronoun Chart on page 475 below for other personal pronouns.] Oblq: mími 'your'. Etym: yuśmad- 'you plural' T-10511.

abúji e آجی اے Id. Good morning!; صبح بخیر. Lit: 'Did you wake up?'.

ábut parík پاریک V. To make a mistake; غلطی کرنا. ábut gála him. It seems I have made a mistake. Syn: khataí hik; galát kárik; galtí hik.

ačhámby آچھامبی N. Ceremony celebrating birth; بیجے کی پیدائش پر جشن. [This is performed five or six days after a birth. After purifying the mother with water and juniper smoke, girls take burning branches from the fire and run to the temple of Jestak, make a bonfire and jump over the flames shouting, "ačhámby a." It is called tuśulék in Bumburet. It is known by this name only in Rumbur.] Syn: tuśulék. See also: óndru učundék.

ačhardíni آچھاردینی N. The wedge which holds the plough-share in place; ہل میں پھانا کو سخت کرنے والا کیل. Whole: haw₁ 'plough'. See also: kyu. From: Khowar.

áči آچی Adv. 1) Back; واپس. áči i. Come back! Syn: pištyák. 2) Again; دوبارہ. From: Khowar.

ačikréti آچیکریٹی N. Lower body (from waist to feet); جسم کا کمر سے نیچلا حصہ. CPart: dhim 'upper body'; Whole: jhan 'body'.

ačoyak آچویاک Adj. Watery, thin (liquid); پتلا. kaí ačoyak húla, adhek at hístik haś. The gravy is too watery, we need to

ačúik

put a little flour in it! Syn: líga; Ant: čhúča; zak. Variant: ačhoyak.

ačhōik آچھوینک N. Story, fable; کہانی. may kay ačhōik dye. Tell me a story! [The story may be true or fabricated.] Use: dyek 'to tell'; Syn: khisá. Variant: ačhāik.

ačhōik₂ آچھوینک₂ N. The dry prickly leaf of holly-oak; شاہ بلوط کا کانٹے دار سوکھا پتا. ačōŋki آچونکی Adj. Crooked, bent; ٹیڑھا (چیز).

— N. Hook; کنڈا، آنکڑا. [Among other things these are used to lift the griddle off the fire and hang it up. It is commonly made from a forked branch.] Syn: áśuŋ.

áču₁ آچو N. Key made of a long wire; دروازہ کھولنے کے لیے ٹیڑھی سلاخ.

áču kárik آچو کاریک V. To make a key for a wooden latch lock; دروازہ کھولنے کے لیے ٹیڑھی سلاخ بنانا. a sim áču kái áam. I have made a wire key.

áču naí آچو نای N. Wooden lock, latch; لکڑی سے بنا دروازہ بند کرنے کا کھٹکا. Syn: kulpí naí 'padlock'.

ačúyak آچویاک N. Key for a metal lock; چابی.

áču₂ آچو N. Wild celery; ایک قسم کا جنگلی پودا جسے کھایا جاتا ہے.

ačuagá آچواگا N. The valley of Achuaga between Rumbur and Bumburet; بمبوریت اور رمبور کے درمیان کی ایک وادی.

ačúik آچویک V. 1) To dry up and become hard from evaporation and age; سوکھ کر سخت ہو جانا. payóst ačúis. The goatskin shriveled up. Syn: šanjúik; šupúkik. 2) To close up from old age; اکڑ جانا. bazá ačúi šían (His) hands are closed up. 3) To sleep; سو جانا. Syn: šanjúik; đúđik; prasúik. 4) To hibernate; سردی کے موسم میں سو جانا. its ačúin. Bears hibernate. Prdm: Class 1 (kárik).

adalát

ačúna آچونا *Adj.* Dried and hard; اکرئی هوئی. خشک (چیز). [Things that can become dry are food and goatskins.] *Ant:* phlaš; *Syn:* šaraw; daǰé; kaǰsé; kirkót.
adalát آدالات *N.* Court; عدالت. *From:* Arabic.

adalát kárik آدالات کاریک *V.* To file a case against someone in court; مقدمه کرنا. dawá kái, may som adalát áraw. Claiming that I had something of his, he filed a case against me in court. *Restrict:* The object is marked by som 'with'. *See also:* kameṭí kárik.

adám آدام *N.* Body; جسم. *Syn:* jhan, jhand; kalíp; badán; khamíl; čhēa; jísom. *Variant:* ča (Urtsun). *From:* Arabic.

adáp kárik آداپ کاریک *V.* To be hospitable to someone, treat with respect; ادب کرنا. ábi kaḷaša moč mími sáwin hátya adáp že izát kárik day, kóki taléy móndrani khodáy khošán híu. We Kalasha people treat you all hospitably and with respect because God is pleased with this kind of thing. *Restrict:* The object is marked by hátya 'for'. *Syn:* parwaríš kárik. *From:* Arabic.

adát آدات *N.* Habit, custom; عادت. húpala máila ki, mič a tan adát púra káriman áis. The scorpion said, "I was only doing what is habitual to me." *Syn:* amál₂. *From:* Arabic.

adát hik آدات هیک *V.* To become accustomed to, to become used to, to become habitual; عادی هونا. tay šiṣ adát thi šíau. Your head is used to it.

adát kárik آدات کاریک *V.* 1) To cause an act to become habitual; عادی کرنا. 2) To break in for riding; سدھانا. hášas bačhoáas thára nisí, moč to adát kárin. Men train a young colt by riding him.

aǰuék

adém آدم *N.* Adam (first man); آدم. *CPart:* bibiawá 'Eve (first woman)'. *From:* Arabic.

ademzát آدمزات *Adj.* Very good, true; نیک، آدمزاد، (انسان کا بچہ). *Lit:* 'born of Adam'. se bo ademzát moč. He is a very good man.

adhé آدھے *Pron.* Some (of a larger group), a few; کچھ. adhé ásta may de. Give me a few more!

— *Adj.* Some; کچھ. adhé moč íta áan. Some men have come. *Syn:* tíčak.

— *Adv.* Some, a little bit. šatará pái kóčik šurúk káda. adhé kóčiḷa čhak dyíta. Going there they began to dig. After they had dug a bit, the sun went down. *Etym:* ardhá- 'half' T-644.

adiná آدینا *N.* Friday; جمعہ. *From:* Persian.

adrák آدراک *N.* Forest, woods; جنگل. tára dawránuna đáran ne íman ašís kóki adrák bo ašís. At that time mud slides did not use to happen because there were lots of wooded areas.

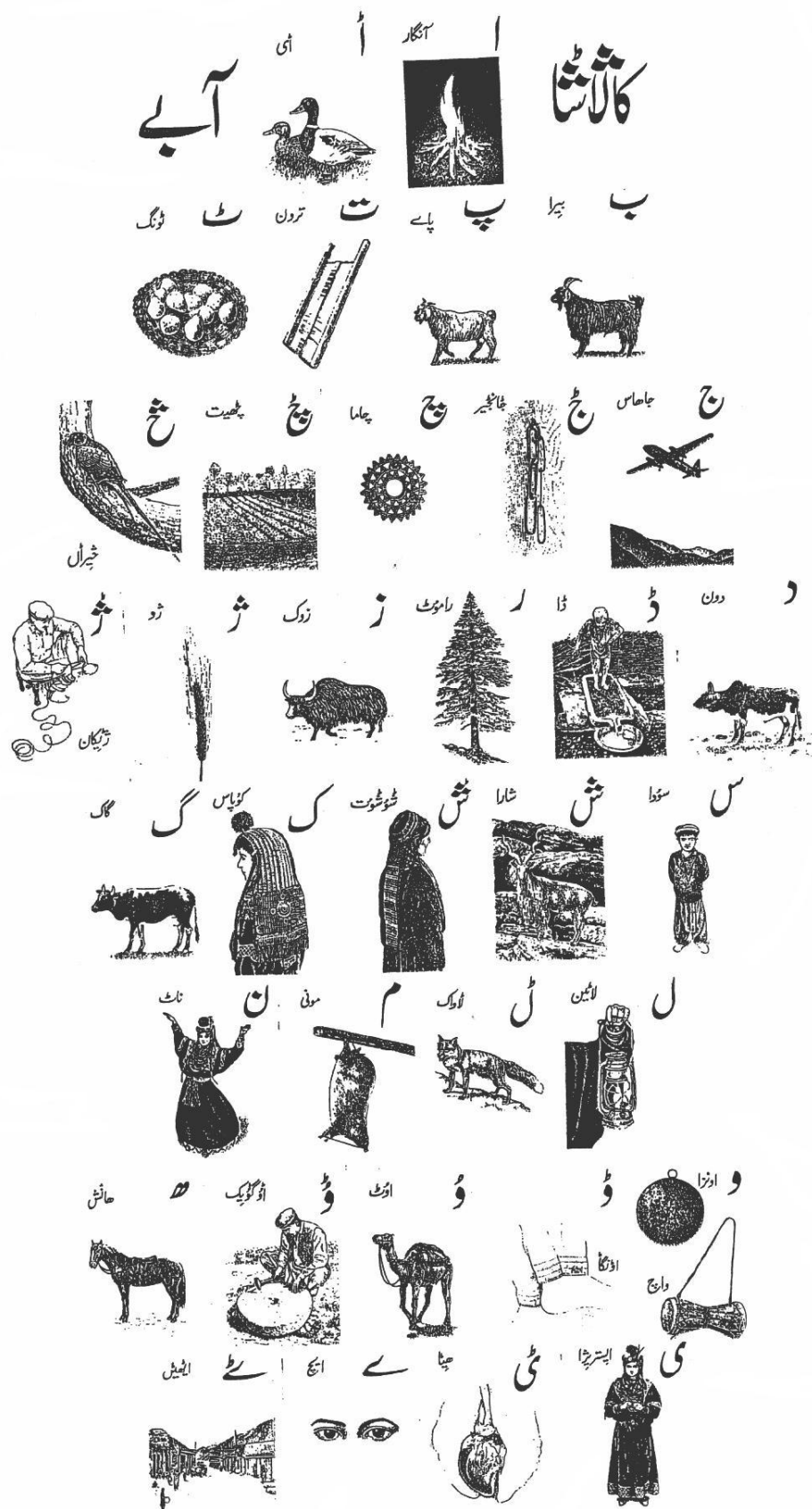
ádua آدوا *N.* 1) Day; دن. se ek ádua hóma dur áaw. One day he came to our house. *Syn:* bas₁. 2) Daytime; دن کا وقت. ádua háwaw e, se paráw. When day came, he left. *Etym:* ardhadivasa- 'noon' T-654.

adyék آدییک *V.* To run; دوڑنا. te mandawjáwani adyáini. They ran from the graveyard. *Past(hearsay):* adyána 'he/she ran'. *Etym:* *uddhāvati 'runs away' T-2020?. *Prdm:* Class 3 (ásik).

adyék adyék hik آدییک آدییک هیک *V.* To race; دوڑ لگانا. sūda adyék adyék thi, labé hin day. The children are having a race. *Gen:* labé 'game'.

aǰuék آڈویک *V.* 1) To blow up, cause to swell, inflate; پھلانا. 2) To spill; پانی لگانا. tu may nōay uk aǰuái e, thái. You have spilled the water under me, put it down! *Prdm:* Causative I (sawzék).

Appendix 25: Kalasha alphabet chart – Arabic script



ABBREVIATIONS AND SYMBOLS

asp	aspirated
dim.	diminutive
^h	aspirated (of previous consonant)
<i>italic</i>	text representing orthographic symbols: spelling (or occasionally word glosses for Kalasha sentences)
L ₁	first language, mother tongue
L ₂	second language, 'other tongue'
pl.	plural
sing.	singular
unasp	unaspirated
vd	voiced
vl	voiceless
V	vowel
V ⁿ	nasalized vowel
ˈ	stressed (of following syllable)
[]	enclosing phonetic symbols
//	enclosing phonemic symbols
{ }	enclosing a morpheme
~	nasalized (of vowel)
#	word boundary
-	morpheme boundary
=	clitic boundary
*	non-grammatical form

BIBLIOGRAPHY¹

- Adams, M 1990, *Beginning to read: thinking and learning about print*, Massachusetts Institute of Technology, Cambridge, Mass.
- Afghani, I 1973, *The Kalash society of Chitral: at a glance*, NWFP Studies Series No. 1, Pashto Literature Service, Peshawar.
- Ahmed, A 1986, 'The Islamization of the Kalash Kafirs', in A Ahmed (ed.) *Pakistan society: Islam, ethnicity and leadership in South Asia*, Oxford University Press, Karachi, pp. 23–8.
- Alaudin 1992, *Kalash: the paradise lost*, Progressive Publishers, Lahore.
- Alkire, W 1977, *An introduction to the peoples and cultures of Micronesia*, 2nd edn, Cummings, Menlo Park, Calif.
- Anderson, B 1991, *Imagined communities: reflections on the origin and spread of nationalism*, rev. edn, Verso, London.
- Awde, N & Samano, P 1986, *The Arabic alphabet*, Lyle Stewart, Kensington Publishing, New York.
- Baart, J 1995, 'The tones of Kalam Kohistani', paper presented at the third international Hindukush cultural conference, Chitral, Pakistan, 26–30 August.
- 1996, 'Some issues in developing a writing system for Kalam Kohistani', paper presented at the seminar on 'Languages of Pakistan: potential for education', Allama Iqbal Open University, Islamabad, 1 September, in *North Pak Technical Services Memo*, no. 12, Summer Institute of Linguistics, Islamabad, pp. 5–7.
- 1997, *The sounds and tones of Kalam Kohistani*, Studies in languages of northern Pakistan, National Institute of Pakistan Studies, Quaid-i-Azam University, & Summer Institute of Linguistics, Islamabad.
- 1999, 'Tone rules in Kalam Kohistani', *Bulletin of the School of Oriental and African Languages*, vol. 62, no. 1, pp. 87–104.
- 2003, 'Tonal features in languages of northern Pakistan', in J Baart and G Sindhi (eds), *Pakistani languages and society: problems and prospects*, National Institute of Pakistan Studies and Summer Institute of Linguistics, Islamabad, pp. 132–44.
- & Sagar, M 2002, *The Gawri language of Kalam and Dir Kohistan*, J Baart, viewed 5 February 2005, <http://www.geocities.com/kcs_kalam/gawri.pdf>.

¹ This is by no means an exhaustive bibliography of literature concerning writing systems or the Kalasha. References to most works on non-linguistic aspects of Kalasha society have been excluded, as have references to most literature to do with related and nearby languages. I have only included works that I consulted in the preparation of this thesis, and a few others of theoretical or areal significance, though not all are cited in the text. Several publications listed here contain articles or chapters about the Kalasha society or language by a number of other contributing authors; unless specifically cited in the text of the thesis, there is no separate reference to those articles or chapters.

- Backstrom, P 1992, 'Wakhi', in P Backstrom & C Radloff (eds), *Languages of northern areas*, Sociolinguistic survey of northern Pakistan, ed. C O'Leary, vol. 2, National Institute of Pakistan Studies, Quaid-i-Azam University, & Summer Institute of Linguistics, Islamabad, pp. 57–74.
- & C Radloff (eds), *Languages of northern areas*, Sociolinguistic survey of northern Pakistan, ed. C O'Leary, vol. 2, National Institute of Pakistan Studies, Quaid-i-Azam University, & Summer Institute of Linguistics, Islamabad.
- Barton, D 1994, 'Globalisation and diversification: two opposing influences on local literacies', in D Barton (ed.), *Language and education*, vol. 8, nos 1&2, *Sustaining Local Literacies*, Multilingual Matters, Clevedon, UK, pp. 3–7.
- Bashir, E 1983, 'Some areal characteristics of Kalasha', paper presented at the 12th Annual Conference on South Asia, Madison, Wis., November.
- 1986, 'Beyond split-ergativity: subject marking in Wakhi', in *Papers from the regional meetings*, vol. 22, no. 1, Chicago Linguistic Society, pp. 14–35.
- 1988a, 'Inferentiality in Kalasha and Khowar', in L MacLeod, G Larson & D Brentari (eds), *Papers from the 24th regional meeting of the Chicago Linguistics Society*, vol. 24, part 1, The General Session, Chicago Linguistic Society, pp. 47–59.
- 1988b, 'Topics in Kalasha syntax: an areal and typological perspective', PhD dissertation, The University of Michigan, Ann Arbor.
- 1990, 'Involuntary experience in Kalasha', in M Verma & K Mohanan (eds), *Experiencer subjects in South Asian languages*, The Center for the Study of Language and Information, Stanford, pp. 297–318.
- 1996, 'The areal position of Khowar: South Asian and other affinities', *Proceedings of the second international Hindukush cultural conference, Chitral, 1990*, Hindukush and Karakoram Studies, 1, University Press, Karachi, pp. 167–179.
- 1999, 'The Urdu and Hindi ergative postposition ne: its changing role in the grammar' in *Yearbook of South Asian languages and linguistics*, The University of Michigan, Ann Arbor.
- & Israr-ud-Din (eds) 1996, *Proceedings of the second international Hindukush cultural conference, Chitral, 1990*, Hindukush and Karakoram Studies, 1, Oxford University Press, Karachi.
- Beck, H 1964, 'Problems of word division and capitalization', in W Smalley et al., *Orthography studies: articles on new writing systems*, Helps for Translators, vol. 6, United Bible Societies, London.
- Bender, B & Rehg, K 1991, 'Micronesian languages in jeopardy?', paper presented at the sixth international conference on Austronesian linguistics, Honolulu, 20–4 May.
- Bendor-Samuel, D 1990, 'Factors affecting the successful development of spoken languages into written ones', in M Halliday et al (eds), *Learning, keeping, and using language 2: selected papers from the 8th world congress of applied linguistics*, Sydney, 16–21 August 1987, John Benjamins, Amsterdam, pp. 115–29.

- Bentin, S & Frost, R 1987, 'Processing lexical ambiguity and visual word recognition in a deep orthography', *Memory and Cognition*, vol. 15, pp. 13–23.
- Berry, J 1977, 'The making of alphabets revisited', in J Fishman (ed.), *Advances in the creation and revision of writing systems*, Contributions to the sociology of language, ed. J Fishman, vol. 8, Mouton, The Hague, pp. 3–16.
- Besner, D & Smith, M 1992, 'Basic processes in reading: is the orthographic depth hypothesis sinking?', in R Frost & L Katz (eds), *Orthography, phonology, morphology, and meaning*, Advances in phonology, eds G Stelmach & P Vroom, vol. 94, Elsevier Science Publishers, Amsterdam, pp. 45–66.
- Bhatia, T 2001, 'Punjabi', in J Garry & C Rubino (eds), *Facts about the world's languages: an encyclopedia of the world's major languages, past and present*, H W Wilson, New York, pp. 576–9.
- Biddulph, J 1986 (1880), *Tribes of the Hindoo Koosh*, Ali Kamran Publishers, Lahore.
- Bowers, J 1968, 'Language problems and literacy', in J Fishman, C Ferguson & J Das Gupta (eds), *Language problems of developing nations*, Wiley, New York, pp. 381–402.
- Bradley, D, Bradley, M, & Yongxiang, L 1999, 'Language maintenance for endangered languages of Central Yunnan', *Endangered languages and education, proceedings of the third FEL conference*, ed. N Ostler, The Foundation for Endangered Languages, Bath, pp. 13–20.
- Brenzinger M (ed.) (2001), *Language diversity endangered*. Mouton de Gruyter, Berlin.
- Bright W (ed.) 1992, *International encyclopedia of linguistics*, Oxford University Press, New York.
- Cacopardo, Alberto 1991, 'The other Kalasha. A survey of Kalashamun-speaking people in southern Chitral', *East and West*, no. 41, pp. 273–350.
- 1996, 'The Kalasha in southern Chitral, Part I: The eastern area', *Proceedings of the second international Hindukush cultural conference, Chitral, 1990*, Hindukush and Karakoram Studies, 1, Oxford University Press, Karachi, pp. 247–69.
- & Cacopardo, Augusto 1989, 'The Kalash (Pakistan) winter solstice festival', *Ethnology*, vol. 28, no. 4.
- Cahill, M 1999, 'From endangered to less endangered: case histories from Brazil and Papua New Guinea', *SIL Electronic Working Papers 1999–006, October 1999*, SIL International, Dallas, viewed 27 August 2001, <<http://www.sil.org/silewp/1999/006>>.
- 2001, *Endangered languages*, SIL International, Dallas, viewed 27 August 2001, <<http://www.sil.org/sociolx/ndg-lg-cahill.html>>.
- Calvet, L 1998, *Language wars and linguistic politics*, trans. M Petharam, Oxford University Press, New York.
- Cardona, G 1992a, 'Indo-Aryan languages', in W Bright (ed.), *International encyclopedia of linguistics*, vol. 2, Oxford University Press, New York, pp. 202–6.
- 1992b, 'Indo-Iranian languages', in W Bright (ed.), *International encyclopedia of linguistics*, vol. 2, Oxford University Press, New York, pp. 212–13.

- Carello, C, Turvey, M & Lukatela, G 1992, 'Can theories of word recognition remain stubbornly nonphonological?', in R Frost & L Katz (eds), *Orthography, phonology, morphology, and meaning*, Advances in phonology, eds G Stelmach & P Vroom, vol. 94, Elsevier Science Publishers, Amsterdam, pp. 211–26.
- Clark, J & Yallop, C 1995, *An introduction to phonetics and phonology*, 2nd edn, Blackwell, Oxford.
- Clark, R & Ivanič, R 1997, *The politics of writing*, Routledge, New York.
- Colombo, L & Tabossi, P 1992, 'Strategies and stress assignment: evidence from a shallow orthography', in R Frost & L Katz (eds), *Orthography, phonology, morphology, and meaning*, Advances in phonology, eds G Stelmach & P Vroom, vol. 94, Elsevier Science Publishers, Amsterdam, pp. 319–40.
- Cooper, E 1996a, 'Teachers' manual for "The same, not the same"' (in Kalasha), Arabic-script edition unpublished pilot edition.
- 1996b, 'Teachers' manual for "The Kalasha alphabet"' (in Kalasha), Arabic-script edition unpublished pilot edition.
- 1996c, 'Teachers' manual for "The same, not the same"' (in Kalasha), Roman-script edition unpublished pilot edition.
- 1996d, 'The Kalasha alphabet' (in Kalasha), Arabic-script edition, unpublished charts.
- forthcoming, 'Teachers' manual for "The Kalasha alphabet"' (in Kalasha), Roman-script edition, Kalasha Kitap, Sydney.
- forthcoming, 'Kalasha sounds' (in Kalasha), revised Roman-script edition, Kalasha Kitap, Sydney.
- Hall, K & Cooper, G 1994a, *The Kalasha alphabet* (in Kalasha), Arabic-script edition, Kalasha Kitap, Islamabad.
- 1994b, *The same, not the same* (in Kalasha), Arabic-script edition, Kalasha Kitap, Islamabad.
- 2000, 'The Kalasha alphabet' (in Kalasha), Roman-script edition, unpublished pilot edition.
- 2003, *The Kalasha alphabet* (in Kalasha), Roman-script edition, Kalasha Kitap, Sydney.
- & Hamilton D 2000, 'Kalasha Sounds' (in Kalasha), Arabic- and Roman-script pilot edition.
- forthcoming, 'Eleven new sounds' (in Kalasha), Roman-script edition, Kalasha Kitap, Sydney.
- Cooper, G 1984a, 'An analysis of Kalasha verb classes', unpublished manuscript.
- c. 1984b, 'Topical (semantic domain) dictionary (Kalasha-English)', Roman-script edition unpublished manuscript.
- 1986a, 'Notes on orthographic conventions established by unanimous decision', Arabic-script, unpublished manuscript.
- 1986b, *The Kalasha Alphabet* (in Kalasha), Arabic-script chart, Kalasha Kitap, Rawalpindi.
- 1992, 'Kalasha sociolinguistic survey results', unpublished manuscript.
- 1996, 'Family trees of each Kalasha village from the early 20th century', unpublished charts.
- Coulmas, F 1989, *The writing systems of the world*, Blackwell, Oxford.

- 1992, 'Writing systems', in W Bright (ed.), *International encyclopedia of linguistics*, Oxford University Press, New York, pp. 253–7.
- 1994, 'Protestant ethics and profane language: economic aspects of language standardization', in G. Lüdi (ed.) *Sprachstandardisierung* (The standardization of language), Swiss Academy of Humanities and Social Science, Fribourg, pp. 161–78.
- 1996, *The Blackwell encyclopedia of writing systems*, Blackwell, Oxford.
- Crystal, D 2000, *Language death*, Cambridge University Press, Cambridge.
- Dalby, D 1967, 'A survey of the indigenous scripts of Liberia and Sierra Leone: Vai, Mende, Loma, Kpelle and Bassa', *African Language Studies*, vol. 8, pp. 1–51.
- Daniels, P & Bright, W (eds) 1996, *The world's writing systems*, Oxford University Press, New York.
- Darling, G 1979, 'Merit feasting among the Kalasha Kafirs of North Western Pakistan, MA thesis, University of British Colombia, Vancouver.
- Dauenhauer, N & Dauenhauer, R 1998, 'Technical, emotional and ideological issues in reversing language shift: examples from Southeast Alaska', in L Grenoble & L Whaley (eds), *Endangered languages: language loss and community response*, Cambridge University Press, Cambridge, pp. 57–98.
- Day, R 1985, 'The ultimate inequality: linguistic genocide', in N Wolfson & J Manes (eds) *Language of inequality*, Mouton, Berlin, pp. 163–81.
- Decker, K 1992a, Factors affecting language maintenance and shift in Chitral District, Pakistan, MA thesis, University of Texas at Arlington.
- 1992b, *Languages of Chitral*, Sociolinguistic survey of northern Pakistan, ed. C O'Leary, vol. 5, National Institute of Pakistan Studies, Quaid-i-Azam University, & Summer Institute of Linguistics, Islamabad.
- 1996, 'Some observations on language vitality in Chitral', *Proceedings of the second international Hindukush cultural conference, Chitral, 1990*, Hindukush and Karakoram Studies, 1, Oxford University Press, Karachi, pp. 159–66.
- DeFrancis, J 1989, *Visible speech: the diverse oneness of writing systems*, University of Hawaii Press, Honolulu.
- Denker, D 1981, 'Pakistan's Kalash: people of fire and fervour', *National Geographic*, vol. 160, no. 4, October, pp. 458–73.
- Diringer, D 1968, *The alphabet: a key to the history of mankind*, 3rd edn, rev. R Regensburger, vol. 1, Hutchinson & Co., London.
- Dixon, R 1997, *The rise and fall of languages*, Cambridge University Press, Cambridge.
- Dow, J (ed.) 1991, *Language and ethnicity: focusschrift in honor of Joshua A Fishman on the occasion of his 65th birthday*, John Benjamins, Amsterdam.
- Downing, P, Lima, S, Noonan, M (eds) 1992, *The linguistics of literacy*, Typological Studies in Language (TSL), ed. T Givón, vol. 21, John Benjamins, Amsterdam.

- Duerksen, J, Fluckiger, C, Hampton, R, Lojenga C & Loving, R 1989, 'The development, production, and distribution of reading materials', *Notes on Literacy*, vol. 59, pp. 3–16.
- Dutcher, N 1982, *The use of first and second languages in primary education: selected case studies*, World Bank staff working paper no. 504, January, The World Bank, Washington.
- Edelberg, L & Jones, S 1979, *Nuristan*, Akademische Druck-u. Verlagsanstalt, Graz, Austria.
- Edelman, D 1983, *The Dardic and Nuristani languages*, Nauka Publishing House, Central Department of Oriental Literature, Moscow.
- Edelsky, C 1991, *With literacy and justice for all*, The Falmer Press, London.
- Elson, B & Pickett, V 1964, *An introduction to morphology and syntax*, Summer Institute of Linguistics, Santa Ana, Calif.
- Emeneau, M 1939, 'The vowels of the Badaga language', *Language*, vol 15, pp. 43–7.
- Fairclough, N 1989, *Language and power*, Longman, London.
- Feldman, L, Kostić, A, Lukatela & Turvey, M 1983, 'An evaluation of the "basic orthographic syllable structure" in a phonologically shallow orthography', *Psychological Research*, vol. 45, pp. 55–72.
- Fentz, M 1996, *Natural resources and cosmology in changing Kalasha society*, NIAS Reports, vol. 30, Nordic Institute of Asian Studies, Copenhagen.
- Ferguson, C & Heath, S 1981, *Language in the USA*, Cambridge University Press, Cambridge.
- Fill, A & Mühlhäusler, P (eds) 2003, *The ecolinguistics reader: language, ecology and environment*, Continuum, London.
- Fishman, J 1966, *Language loyalty in the United States*, Mouton, The Hague.
- 1972, *The sociology of language: an interdisciplinary social science approach to language in society*, Newbury House, Rowley, Mass.
- 1977, 'Advances in the creation and revision of writing systems', in J Fishman (ed.), *Advances in the creation and revision of writing systems*, Contributions to the sociology of language, ed. J Fishman, vol. 8, Mouton and Co., The Hague, pp. xi–xxviii.
- 1997, 'Language, ethnicity and racism', in N Coupland & A Jaworski (eds) *Sociolinguistics: a reader and coursebook*, Macmillan, Houndmills and London, p. 329–40.
- Cooper, R, Ma, R, et al. 1971, *Bilingualism in the barrio*, Language sciences monograph 7, Indiana University, Bloomington.
- (ed.) 1985, *The rise and fall of the ethnic revival: perspectives on language and ethnicity*, Contributions to the sociology of language, vol. 37, Mouton, Berlin.
- Fitch, M & Cooper, G 1985, 'Report on a language and dialect survey in Kohistan district', *Journal of Central Asia*, vol. 8, no. 1, pp. 39–49.
- Freire, P 1970a, *Cultural action for freedom*, Harvard Educational Review, Monograph series, no. 1, Harvard Educational Review, Cambridge.
- 1970b, *Pedagogy of the oppressed*, trans. M Ramos, Herder & Herder, New York.
- 1974a, *Education for critical consciousness*, Sheed & Ward, London.

- 1974b, *Education for liberation and community*, Australian Council of Churches, Commission on Christian Education, Melbourne.
- 1976, *Education, the practice of freedom*, Writers and Readers Publishing Cooperative, London.
- 1985, *The politics of education: culture, power, and liberation*, trans. D Macedo, Bergin & Garvey, South Hadley, Mass.
- 1994, 'The adult literacy process as cultural action for freedom', in J Maybin (ed.), *Language and literacy in social practice: a reader*, Multilingual Matters, in association with the Open University, Clevedon, UK, pp. 252–64.
- & Faundez, A 1989, *Learning to question: a pedagogy of liberation*; trans. T Coates, WCC Publications, Geneva.
- Freire, P & Shor, I 1987, *A pedagogy for liberation: dialogues on transforming education*, Bergin & Garvey, South Hadley, Mass.
- Frost R 1994, 'Prelexical and postlexical strategies in reading: evidence from a deep and a shallow orthography', *Journal of Experimental Psychology: Learning, Memory and Cognition*, vol. 20, no. 1, pp. 116–29.
- & Katz, L 1989, 'Orthographic depth and the interaction of visual and auditory processing in word recognition', *Memory and Cognition*, vol. 17, pp. 302–11.
- (eds) 1992, *Orthography, phonology, morphology, and meaning*, Advances in phonology, eds G Stelmach & P Vroon, vol. 94, Elsevier Science Publishers, Amsterdam.
- & Bentin, S 1987, 'Strategies for visual word recognition and orthographical depth: a multilingual comparison', *Journal of Experimental Psychology: Human Perception and Performance*, vol. 13, pp. 104–14.
- Fussman, G 1972, *Atlas linguistique des parlers Dardes et Kafirs*, vols 1 & 2, Publication de l'Ecole Française d'Extrême-Orient, Paris.
- 1977, 'Languages as a source for history', in A Dani (ed.) *History of northern areas of Pakistan*, National Institute of Historical and Cultural Research, Islamabad.
- Gaur, A 1992, *A history of writing*, The British Library, London.
- Gelb, I 1963, *A study of writing*, 2nd edn, University of Chicago Press, Chicago.
- Ghafoor, A 1994, *Quest for adult literacy*, National Education & Training Commission, Ministry of Education, Islamabad.
- Gilbert, G 1981, 'French and German: a comparative study', in C Ferguson & S Heath, *Language in the USA*, Cambridge University Press, Cambridge, pp. 257–72.
- Goody, J & Watt, I 1963, 'The consequences of literacy', *Comparative Studies in Society and History*, vol. 5, no. 3, pp. 304–45.
- Grace, G 1993a, 'What are languages?', *Ethnolinguistic notes*, vol. 3, no. 45, pp. 1–17.
- 1993b, *Culture in language and linguistics: linguistics' role in monocultural imperialism*, unpublished manuscript, University of Hawaii, Honolulu.
- Graff, H 1979, *The literacy myth*, Academic Press, New York.

- Gregory, G 1991, 'Community publishing as self-education' in D Barton & R Ivanič (eds), *Writing in the community*, Sage Publications, Newbury Park, Calif., pp. 109–42.
- Grierson, G 1915, *Linguistic survey of India*, vol. 8, part 2, *Dardic or Pisacha*, Accurate Printers, Lahore.
- Grimes, B (ed.) 1978, *Ethnologue*, 9th edn, Wycliffe Bible Translators, Huntington Beach, Calif.
- 2000, *Ethnologue: languages of the world*, 14th edn (database), SIL International, Dallas, updated every four years.
- Grimes, J & Gordon R 1980, 'Design of new orthographies', in J Kavanagh & R Venezky (eds), *Orthography, reading and dyslexia*, University Park Press, Baltimore, pp. 93–103.
- Gudschinsky, S 1970, 'More on formulating efficient orthographies', *The Bible Translator*, vol. 21, pp. 21–5.
- 1973, *A manual of literacy for preliterate peoples*, Summer Institute of Linguistics, Ukarumpa, Papua New Guinea.
- 1976, *Literacy: the growing influence of linguistics*, Mouton, The Hague.
- 1979, 'The role of the linguist in the preparation of materials', in O Simpósio de São Paulo, janeiro de 1969, Atas, 229–36. Programa Interamericano de Lingüística y Enseñanza de Idiomas, Universidade de São Paulo, São Paulo.
- Hackin, J 1932, 'The mythology of the Kafirs', *Asiatic mythology*, London, pp. 57–60.
- Hahn, R 1991, *Spoken Uyghur*, University of Washington Press, Seattle.
- Hallberg, D 1992, 'The languages of Indus Kohistan', in C Rensch, S Decker & D Hallberg, *Languages of Kohistan*, Sociolinguistic survey of northern Pakistan, vol. 1, National Institute of Pakistan Studies, Quaid-i-Azam University, & Summer Institute of Linguistics, Islamabad, pp. 83–141.
- 1999, *Indus Kohistani: a preliminary phonological and morphological analysis*, Studies in languages of northern Pakistan, vol. 8, National Institute of Pakistan Studies, Quaid-i-Azam University, & Summer Institute of Linguistics, Islamabad.
- Hamers, J & Blanc, M 2000, *Bilinguality and bilingualism*, 2nd edn, Cambridge University Press, Cambridge.
- Harris, R 1979, *The language makers*, Duckworth, London.
- 1980, *The language myth*, Duckworth, London.
- 1990, 'Language as social interaction: integrationalism versus segregationalism', in N Love (ed.) *The foundations of linguistic theory: selected writings of Roy Harris*, Routledge, London.
- , Schwab, I & Whitman, L (eds) 1990, *Language and power*, Harcourt Brace Jovanovich, New York.
- Harrison, J 1996, 'Kalash buildings', *Proceedings of the second international Hindukush cultural conference*, Oxford University Press, Karachi, pp. 345–58.
- Haugen, E 1972, *The ecology of language*, ed. A Dil, Stanford University Press, Stanford, Calif.
- Hawley, A 1950, *Human ecology: a theory of community structure*, Ronald Press, New York.

- Heegård, J 1996, 'Et alfabet bliver til' (An alphabet comes into being), term paper delivered at Department of General and Applied Linguistics, University of Copenhagen.
- 1998, 'Variational patterns in vowel length in Kalashamon', in J Niemi, T Odlin & J Heikkinen (eds), *Language contact, variation, and change*, Studies in languages, vol. 32, University of Joensuu, Faculty of Humanities, Joensuu, Finland, pp. 125–35.
- 2000, 'Linguistic and political aspects of alphabet-making for a threatened language', in C-E Lindberg & S Lund (eds), *17th Scandinavian Conference of Linguistics*, vol 2, Odense Working Papers in Language and Communication 19, Institute of Language and Communication, University of Southern Denmark, Odense, pp. 161–76.
- & Mørch, I 1995, 'Variation in the pronunciation of Kalashamon as spoken inside and outside of present-day Kalasha society', paper presented at the third international Hindukush cultural conference, Chitral, Pakistan, 26–30 August.
- forthcoming, 'Retroflex vowels and other peculiarities in Kalasha sound system' in A Saxena & J Gvozdanovic (eds), *Synchronic and Diachronic Aspects of Himalayan Linguistics, Selected Proceedings of the 7th Himalayan Languages Symposium, held in Uppsala, Sweden*, Mouton, The Hague.
- Henze, P 1956, 'Politics and alphabets', *Royal Central Asian Society Journal*, vol. 43, pp. 29–51.
- Heyd, U 1954, *Language reform in modern Turkey*, Israel Oriental Society, Jerusalem.
- Hill, A 1967, 'The typology of writing systems', in W Austin (ed.), *Papers in linguistics in honor of Léon Dostert*, Mouton, The Hague, pp. 92–9.
- Hockett, C 1958, *A course in modern linguistics*, Macmillan, New York.
- Hoffmann, C (ed.) 1996, *Language, culture, and communication in contemporary Europe*, Multilingual Matters, Clevedon, UK.
- Hofstede, A 1998, 'The modern South Arabian languages' in 'Supplement: abstracts from the open forum', in N Ostler (ed.), *Endangered languages: what role for the specialist? Proceedings of the second FEL conference*, 2nd edn, The Foundation for Endangered Languages, Bath, p. 123.
- Holdich, T 1891, 'Origin of the Kafir of the Hindu Kush', *The Geographical Journal*, no. 7, pp.42–9.
- Holm, W 1980, 'Learning to read and write in Navajo', in J Kavanagh & R Venezky (eds), *Orthography, reading and dyslexia*, University Park Press, Baltimore.
- Hoyle, R 1986, 'Sindhi/Urdu programmes on the Sharp computer', unpublished edition.
- Husband, C & Khan, S 1982, 'The viability of ethnolinguistic vitality', *Journal of Multilingual and Multicultural Development*, vol. 3, pp. 193–205.
- Inozemtsev, I 2001, 'Pashto', in J Garry & C Rubino (eds), *Facts about the world's languages: an encyclopedia of the world's major languages, past and present*, H W Wilson, New York.
- International Phonetic Association 1999, *Handbook of the International Phonetic Association*, Cambridge University Press, Cambridge.
- Israr-ud-Din 1969, 'The people of Chitral: a survey of their ethnic diversity', *Pakistan Geographical Review*, Peshawar.

- Jacoby, F 1949, *Atthis: the local chronicles of Athens*, Clarendon Press, Oxford.
- Jahani, C 1989, *Standardization and orthography in the Balochi language*, Acta Universitatis Upsaliensis, Studia Iranica Upsaliensia 1, Doctoral dissertation, Uppsala University, Sweden.
- Jan, S 1995, 'Development and self-determination: a Kalasha point of view', paper presented at the third international Hindukush cultural conference, Chitral, Pakistan, 26–30 August.
- 1996, 'History and development of the Kalasha', *Proceedings of the second international Hindukush cultural conference, Chitral, 1990*, Hindukush and Karakoram Studies, 1, Oxford University Press, Karachi, pp. 239–45.
- Jettmar, K 1986, *The religions of the Hindukush: The religion of the Kafirs*, trans. A Nayyar, vol. 1, Aris & Phillips, Warminster, UK.
- 1990, *The religions of the Hindukush: The religions of the Chitralis*, vol. 3 part 1, Brill Academic Publishers, Leiden.
- & Edelberg, L (eds) 1974, *Cultures of the Hindukush: selected papers from the Hindu-Kush cultural conference held at Moesgård 1970*, Franz Steiner Verlag, Wiesbaden, Germany.
- Jones, P 2001, 'Endangered species', *Books and Culture*, Mar.–Apr., p. 26.
- Jones, S 1966, *An annotated bibliography of Nuristan (Kafiristan) and the Kalash Kafirs of Chitral*, Part 1, vol. 41, no. 3, Historisk-filologiske Meddelelser udgivet af Det Kongelige Danske Videnskabernes Selskab, Copenhagen.
- 1969, *A bibliography of Nuristan (Kafiristan) and the Kalash Kafirs of Chitral*, Part 2: selected documents from the secret and political records, 1885–1900, vol. 43, no. 1, Historisk-filologiske Meddelelser udgivet af Det Kongelige Danske Videnskabernes Selskab, Copenhagen.
- Kachru, B 1965, 'The Dardic branch or sub-branch of Indo-European', *Languages of the world: Indo-European fascicle one*, eds C Voegelin & F Voegelin, *Anthropological Linguistics*, vol. 7, no. 8, pp. 284–94.
- Kalash Indigenous Survival Programme 2005, *Bibliografia*, KISP, viewed 27 March 2005, <<http://www.kalash.it/bibliografia.htm>>.
- Kalash People Welfare Society, n.d. 'Constitution', KPWS, unpublished manuscript.
- Katsiavriades, K & Qureshi, T 2004, *The Nastaliq alphabet*, KryssTal, London, viewed 25 March 2005, <http://www.kryssstal.com/writing_nastaliq.html>.
- Katz, L & Feldman, L 1981, 'Linguistic coding in word recognition: comparisons between a deep and a shallow orthography', in A Lesgold & C Perfetti (eds), *Interactive processes in reading*, Lawrence Erlbaum Associates, Hillsdale, NJ.
- 1983, 'Relation between pronunciation and recognition of printed words in deep and shallow orthographies', *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 9, pp. 157–66.
- Katz, L & Frost, R 1992, 'The reading process is different for different orthographies: the orthographic depth hypothesis', in R Frost & L Katz (eds), *Orthography, phonology, morphology, and meaning*,

- Advances in phonology, vol. 94, eds G Stelmach & P Vroon, Elsevier Science Publishers, Amsterdam, pp. 67–84.
- Kavanagh, J & Venezky, R (eds), *Orthography, reading and dyslexia*, University Park Press, Baltimore.
- Khan, F & Cooper, E 2000, 'Early Reader' (in Kalasha), Roman- and Arabic-script edition, unpublished pilot edition.
- Khan, M Afzal 1975, *Chitral and Kafiristan: a personal study*, Lt. Col. M Afzal Khan, Drosh, Pakistan.
- Khan, M Ayub n.d., Natural and cultural resources of Bumboret Valley, vulnerabilities to threat, and strategies for conservation, draft MA thesis, University of Peshawar.
- Kindell, G 2001, *Endangered language groups*, SIL International, Dallas, viewed 17 August 2001, <<http://www.sil.org/sociolx/ndg-lg-grps.html>>.
- King, R 1997, *Nehru and the language politics of India*, Oxford University Press, New Delhi.
- Kloss, H 1966, 'German-American language maintenance efforts', in J Fishman, *Language loyalty in the United States*, Mouton, The Hague, pp. 206–52.
- Kolarz, W 1967, *Russia and her colonies*, Archon Books, Hamden, Conn.
- Krauss, M 2001, 'Classification and terminology for degrees of language endangerment', in M Brenzinger (ed.) *Language diversity endangered*. Mouton de Gruyter, Berlin.
- Ladefoged, P 1975, *A course in phonetics*, Harcourt, Brace Jovanovich, New York.
- & Maddieson, I 1996, *The sounds of the world's languages*, Blackwell, Oxford.
- Laiton, D 1992, *Language repertoires and state construction in Africa*, Cambridge University Press, Cambridge.
- Landry, R & Allard, R 1985, 'Choix de langue d'enseignement: une analyse chez des parents francophones en milieu bilingue soustractif', *La Revue Canadienne des Langues Vivantes*, vol. 41, no. 3, pp. 480–500.
- Lee, E 1982, *Literacy primers: The Gudschinsky method*, Summer Institute of Linguistics, Dallas.
- Leitner, G 1877, *The languages and races of Dardistan*, part I, Lahore.
- 1880, 'A sketch of the Bashgali Kafirs and of their language', no. 43, *Journal of the United Service Institute of India*, Simla.
- Lepsius, R 1855, *Standard alphabet for reducing unwritten languages and foreign graphic systems to a uniform orthography in European letters*, Seeleys, London.
- Lewis, G 1981, 'Oh! No! we never mention her (thoughts on Turkish language reform)', *Asian Affairs*, vol. 12, no. 2, pp. 155–66.
- Lieberman, A 1980, 'Orthography and phonemics in present-day Russian', in J Kavanagh & R Venezky (eds), *Orthography, reading and dyslexia*, University Park Press, Baltimore.
- Lie, M, Baker, P, McEnery, T & Sebba, M 1999, 'Building a corpus of spoken Sylheti', *Endangered languages and education, Proceedings of the third FEL Conference*, ed. N Ostler, The Foundation for Endangered Languages, Bath, pp. 71–6.

- Lièvre, V 1996, 'The status of Kalasha women in the religious sphere', *Proceedings of the second international Hindukush cultural conference*, Oxford University Press, Karachi, pp. 337–43.
- Lindau, M 1978, 'Vowel features', *Language*, vol. 54, no. 3, pp. 541–63.
- Linde, C 1981, 'The organization of discourse', in T Shopen & J Williams (eds) *Style and variables in English*, Winthrop, Cambridge, Mass., pp. 84–114.
- Lines, M 1985, 'The Phoo festival: the Kalash of the Hindu Kush', *The Traveller*, vol. 15, no. 3, Autumn, pp. 50–5.
- 1995, 'The exploiters', paper presented at the third international Hindukush cultural conference, Chitral, Pakistan, 26–30 August.
- Longacre, R 1983, *The grammar of discourse*, Plenum Press, New York.
- Loude, J 1996, 'The Kalasha shamans' practice of exorcism', *Proceedings of the second international Hindukush cultural conference*, Oxford University Press, Karachi, pp. 329–335.
- Loude, J & Lièvre V 1980, *Kalash, les derniers "infidèles" de l'Hindu-Kush*, Berger-Levrault, Paris.
- 1984, *Solstice Païen*, Presses de la Renaissance, Paris.
- c. 1986, *Kalash solstice: winter feasts of the Kalash of north Pakistan*, trans. G Romaine & M Intrator, Lok Virsa, Islamabad.
- Lukatela, G, Popadić, D, Ognjenović, P & Turvey, M 1980, 'Lexical decision in a phonologically shallow orthography', *Memory and Cognition*, vol. 8, pp. 124–32.
- Macdonald, R & Thomas, H (eds) 1997, *Nationality and planning in Scotland and Wales*, University of Wales, Cardiff.
- MacDougall, R 2000, 'Individuals, cultures and telecommunication technology: An inquiry into another one of the effects and defects of e-mail', *Endangered languages: what role for the specialist? Proceedings of the second FEL conference*, 2nd edn, ed. N Ostler, The Foundation for Endangered Languages, Bath, pp. 91–8.
- Mace, J (ed.) 1995, *Literacy, language and community publishing: essays in adult education*, Routledge, London.
- Mackenzie, D 1987, 'Pashto', in B Comrie (ed.) *The world's major languages*, Oxford University Press, New York, pp. 547–65.
- Mackey, W 1968, 'The description of bilingualism', in J Fishman (ed.) *Readings in the sociology of language*, Mouton, The Hague, pp. 554–84.
- Macquarie University 1974, 'The analysis of languages', course handouts.
- Maddieson, I 1984, *Patterns of sounds*, Cambridge studies in speech science and communication, Cambridge University Press, Cambridge.
- Maggi, W 1995, 'Don't cry my daughter: lullabies as a keyhole into Kalasha culture', paper presented at the third international Hindukush cultural conference, Chitral, Pakistan, 26–30 August.
- Man, J 2000, *Alpha beta: how our alphabet changed the western world*, Headline, London.
- Mann, W 1999, 'Discourse across languages and cultures: The 24th UWM linguistics symposium, Milwaukee, Wis., September 1998', *Notes on Linguistics* 2: 83–87.

- Martinez Cobo, J 1987, *Study of the problem of discrimination against indigenous populations*, United Nations, New York.
- Maruyama, J 1982, 'Demographic survey of Bumboret valley', unpublished manuscript.
- 1983, 'An introduction to the Kalash people in Chitral', ed. G Cooper, unpublished manuscript.
- 2003, 'An introduction to the Kalasha people in Chitral', Site Shara, viewed 27 March 2005, <<http://www.site-shara.net/kalasha/eflm-kalasha.html>>.
- & Kojima, R 1987, *Jun and Reiko's bibliography: titles translated in English*, Site Shara, viewed 27 March 2005, <<http://www.site-shara.net/aboutus/article/e-biblio.html>>.
- Masica, C 1976, *Defining a linguistic area: South Asia*, University of Chicago Press, Chicago.
- 1991, *The Indo-Aryan languages*, Cambridge language surveys, Cambridge University Press, Cambridge.
- 1992, 'South Asian languages', in W Bright (ed.), *International encyclopedia of linguistics*, vol. 2, Oxford University Press, New York, pp. 38–41.
- Matthews, D 1995, 'A genre-based approach to literacy', *Notes on Literacy*, vol. 21, no. 3, pp. 35–46.
- Mattingly, I 1992, 'Linguistic awareness and orthographic form', in R Frost & L Katz (eds), *Orthography, phonology, morphology, and meaning*, Advances in phonology, vol. 94, eds G Stelmach & P Vroon, Elsevier Science Publishers, Amsterdam, pp. 11–26.
- May, S 2001, *Language and minority rights: ethnicity, nationalism, and the politics of language*, Longman, Harlow, UK.
- Maybin, J (ed.) 1994, *Language and literacy in social practice: a reader*, Multilingual Matters in association with the Open University, Clevedon, UK.
- Mithun, M 1992, 'Orthography planning', in W Bright (ed.), *International encyclopedia of linguistics*, Oxford University Press, New York, pp. 141–4.
- Mock, J 2004, *Dards, Dardistan and Dardic: an ethnographic, geographic and linguistic conundrum*, J Mock and K O'Neil, Calif., viewed 30 January 2005, <<http://www.monitor.net/~jmko/karakoram/dard.htm>>.
- Modiano, N 1973, *Indian education in the Chiapas Highlands*, Holt, Rinehart & Winston, New York.
- Mohanan, K 1986, *The theory of lexical phonology*, Studies in natural language and linguistic theory, ed. J Maling, D Reidel Publishing Company, Dordrecht, Netherlands.
- Moore, H (ed.) 1990, *The alphabet makers*, Summer Institute of Linguistics, Huntington Beach, Calif.
- Mørch, I & Heegård, J 1995, 'Linguistic observations on Kalashamun as spoken inside and outside present-day Kalasha society: work in progress', paper presented at the third international Hindukush cultural conference, Chitral, Pakistan, 26–30 August.
- 1996, 'Vokaler i kalashamon' (Vowels in Kalashamon). Term paper delivered at Department of General and Applied Linguistics, University of Copenhagen.
- 1997, 'Retroflekse vokalers oprindelse i kalashamon i historisk og areallingvistisk perspektiv; Variation i sprogbeskrivelsen: vokallængde i kalashamon' (The origin of retroflex vowels in

- Kalashamon in a historical and areal linguistic perspective; variation in language description: vowel length in Kalashamon), MA thesis, University of Copenhagen.
- 2000, 'How fast will a language die when it is officially no longer spoken?' in C-E Lindberg & S Lund (eds), *17th Scandinavian Conference of Linguistics*, vol 2, Odense Working Papers in Language and Communication 19, Institute of Language and Communication, University of Southern Denmark, Odense, pp. 125–34.
- Morgenstierne, G 1932, *Report on a linguistic mission to north-western India*, Instituttet for Sammenlignende Kulturforskning, Oslo.
- 1961, 'Dardic and Kâfir Languages', in *The Encyclopaedia of Islam: New Edition*, Vol. 2, Fasc. 25, E J Brill, Leiden, pp. 138–139.
- 1965, 'Notes on Kalasha', *Norsk Tidsskrift for Sprogvidenskap*, vol. 20, Oslo, reprinted in Morgenstierne (1973), pp. 183–238.
- 1973, *The Kalasha language*, Indo-Iranian Frontier Languages, vol. 4, 2nd edn, Instituttet for Sammenlignende Kulturforskning, Oslo.
- Mühlhäusler, P 1992, 'Preserving languages or language ecologies: a top-down approach to language survival', *Oceanic Linguistics*, vol. 31, no. 2, pp. 163–80.
- 1995, 'The interdependence of linguistic and biological diversity', in D Myers (ed.) *The politics of multiculturalism in the Asia/Pacific*, Northern Territory University Press, Darwin, pp. 154–61.
- 1996, *Linguistic ecology: language change and linguistic imperialism in the Pacific region*, The politics of language, eds T Crowley & T Taylor, Routledge, London.
- 2000, 'Language planning and language ecology', *Current issues in language planning*, vol. 1, no. 3, pp. 306–67.
- 2001, 'Ecolinguistics, linguistic diversity, ecological diversity', in L Maffi (ed.) *On biocultural diversity: linking language, knowledge, and the environment*, Smithsonian Institution Press, Washington, pp. 133–44.
- Munnings, D 1997, 'Reduplication in Khowar: and autosegmental treatment', *North Pakistan Newsletter*, no. 20, National Institute of Pakistan Studies, Quaid-i-Azam University & Summer Institute of Linguistics, Islamabad, pp. 14–35.
- 1998a, 'Syllable structure in Khowar', *North Pakistan Newsletter*, no. 22, Jan., National Institute of Pakistan Studies, Quaid-i-Azam University & Summer Institute of Linguistics, Islamabad, pp. 6–35.
- 1998b, 'Khowar lexical phonology', *North Pakistan Newsletter*, no. 23, Feb.–Mar., National Institute of Pakistan Studies, Quaid-i-Azam University & Summer Institute of Linguistics, Islamabad, pp. 3–18.
- Murray, D 1990, 'Literacies as sociocultural phenomena', *Prospect*, vol. 6, no. 1, Sep., pp. 55–61.
- Nakanishi, A 1980, *Writing systems of the world*, Charles E Tuttle, Rutland, Vt.
- Nettle, D & Romaine, S 2000, *Vanishing voices: the extinction of the world's languages*, Oxford University Press, New York.

- New London Group, The 1997, *A pedagogy of multiliteracies: designing social futures*, rev. edn, National Language and Literacy Institute of Australia, NLLIA Centre for Workplace Communication and Culture, Sydney.
- Nida, E 1949, *Morphology, the descriptive analysis of words*, University of Michigan Press, Ann Arbor.
- Ohala, J & Jaeger, J 1986, *Experimental phonology*, Academic Press, Harcourt Javanovich, Orlando.
- O'Leary, C 1992, *Sociolinguistic survey of northern Pakistan*, 5 vols, National Institute of Pakistan Studies, Quaid-i-Azam University, & Summer Institute of Linguistics, Islamabad
- Ong, W 1988, *Orality and literacy: the technologizing of the word*, New Accents, ed. T Hawkes, Routledge, London.
- Ostler, N (ed.) 1998, *Endangered languages: what role for the specialist? Proceedings of the second FEL conference*, The Foundation for Endangered Languages, Bath.
- 1999, *Endangered languages and education, proceedings of the third FEL conference*, The Foundation for Endangered Languages, Bath.
- 2000, 'Language endangerment, research and documentation: setting priorities for the 21st century', *Ogmios*, vol. 14, viewed 13 February 2005, <www.ogmios.org/146.htm>.
- & Rudes, B 2000, 'Endangered languages and literacy', *Endangered languages and literacy, proceedings of the fourth FEL conference*, The Foundation for Endangered Languages, Bath, pp. 11–16.
- Paap, K, Noel, R & Johansen, L 1992, 'Dual-route models of print to sound: red herrings and real horses', in R Frost & L Katz (eds), *Orthography, phonology, morphology, and meaning*, Advances in phonology, vol. 94, eds G Stelmach & P Vroon, Elsevier Science Publishers, Amsterdam, pp. 293–318.
- Pacific Resources for Education and Learning n.d., *TEAMS: Teach ReSPCT (Teacher Resource on Selected Pacific Cultural Topics)*, Topic 5, Language and communication, PREL, Honolulu, viewed 30 January 2005, <www.prel.org/teams/Topic5.asp#1>.
- Padilla, A 1990, 'Bilingual education: issues and perspectives', in A Padilla, H Fairchild & C Valadez (eds), *Bilingual education: issues and strategies*, Sage Publications, Newbury Park, Calif., pp. 15–26.
- Fairchild, H & Valadez C (eds), *Bilingual education: issues and strategies*, Sage Publications, Newbury Park, Calif.
- Parkes, P 1975, 'The social role of historical tradition among the Kalasha Kafirs of Chitral', BLitt thesis, University of Oxford.
- 1990, 'Kalasha rites of spring: backstage of a "Disappearing World" film', *Anthropology Today*, vol. 6, no. 5, pp. 11–13.
- 1994, 'Personal and collective identity in Kalasha song performance: the significance of music-making in a minority enclave', in M Stokes (ed.), *Ethnicity, identity, and music: the musical construction of place*, Berg, Oxford, pp. 157–88.

- 1996, 'Kalasha oral literature and praise songs', *Proceedings of the second international Hindukush cultural conference, Chitral, 1990*, Hindukush and Karakoram Studies, 1, Oxford University Press, Karachi, pp. 315–28.
- & Stock, J n.d., 'Kalasha praise songs: panegyric recital in a non-Muslim community of northern Pakistan', unpublished manuscript with musical transcription.
- Pattanayak, D 1979, 'The problem and planning of scripts' in G Samabasiva Rao (ed.), *Literacy methodology*, Central Institute of Indian Languages, Manasagangotri, Mysore, pp. 43–59.
- Perez, A, Santiago, A & Liem (eds) 1978 *Proceedings from the conference on the standardisation of Asian languages*, Pacific Linguistics, ed. S Wurm, Series C, no. 47, The Australian National University, Canberra.
- Perfetti, C, Zhang, S, Berent, I 1992, 'Reading in English and Chinese: evidence for a "universal" phonological principle', in R Frost & L Katz (eds), *Orthography, phonology, morphology, and meaning*, Advances in phonology, vol. 94, eds G Stelmach & P Vroom, Elsevier Science Publishers, Amsterdam, pp. 227–48.
- Phillipson, R 1992, *Linguistic imperialism*, Oxford University Press, Oxford.
- (ed.) 2000, *Rights to language: equity, power and education*, Lawrence Erlbaum Associates, Mahwah, NJ.
- Pike, K 1947, *Phonemics: a technique for reducing languages to writing*, The University of Michigan Press, Ann Arbor.
- Pikkert, J 1996, 'Perception of eastern Indonesian community leaders regarding vernacular reading needs for development', *Notes on Literacy*, vol. 22, no. 2, Apr., pp. 6–33.
- Potts, D 1987, 'Literature promotion and distribution' *Notes on Literacy*, Special Issue 2, pp. 26–33.
- Prinsloo, M & Breier, M 1996, *The social uses of literacy: theory and practice in contemporary South Africa*, Sached Books, Bertsham, South Africa.
- Pütz, M 1995, *Discrimination through language in Africa? Perspectives on the Namibian experience*, Mouton de Gruyter, Berlin.
- Radloff, C 1992, 'Dialects of Shina', in P Backstrom and C Radloff (eds), *Languages of northern areas*, Sociolinguistic survey of northern Pakistan, vol. 2, National Institute of Pakistan Studies, Quaid-i-Azam University, & Summer Institute of Linguistics, Islamabad, pp. 85–203.
- 1996, 'Shina language project', *North Pakistan Newsletter*, no. 13, Oct., National Institute of Pakistan Studies, Quaid-i-Azam University & Summer Institute of Linguistics, Islamabad, pp. 8–9.
- 1997, 'An exercise in areal linguistics: the place of Gilgiti Shina in the South and Central Asian linguistic areas', *North Pakistan Newsletter*, no. 17, Feb., National Institute of Pakistan Studies, Quaid-i-Azam University & Summer Institute of Linguistics, Islamabad, pp. 3–11.
- 1999, *Aspects of the sound system of Gilgiti Shina*, Studies in Languages of Northern Pakistan, vol. 4, National Institute of Pakistan Studies, Quaid-i-Azam University, & Summer Institute of Linguistics, Islamabad.
- Rahman, T 1996, *Language and politics in Pakistan*, reprint 1998, Oxford University Press, Karachi.

- 2000, 'Language of the Kalasha', *The News International* (Islamabad).
- Read, A 1934, *Balti grammar*, London.
- Rhydwen, M 2000, 'Strategies for doing the impossible', *Endangered languages: what role for the specialist? Proceedings of the second FEL conference*, 2nd edn, ed. N Ostler, The Foundation for Endangered Languages, Bath, pp. 101–6.
- Ricento, T (ed.) 2000, *Ideology, politics, and language policies: focus on English*, John Benjamins, Amsterdam.
- Riley, G 1980 'Language loyalty and ethnocentrism in the Guamanian speech community: seven years later', *Anthropological linguistics*, vol. 22, no. 8, pp. 326–33.
- Robertson, G 1974 (1896), *The Kafirs of the Hindu-Kush*, Oxford University Press, Karachi.
- Rockhill, K 1994, 'Gender, language and the politics of literacy', in J Maybin (ed.), *Language and literacy in social practice: a reader*, Multilingual Matters in association with the Open University, Clevedon, UK, pp. 233–51.
- Romaine, S 1995, *Bilingualism*, 2nd edn, Blackwell, Oxford.
- Rose, C 1992a, 'Kalash rites of spring', Institute of Current World Affairs, CVR-25, unpublished manuscript.
- 1992b, 'Songs of the Kalash', Institute of Current World Affairs, CVR-26, unpublished manuscript.
- 1992c, 'Progress and culture: the Kalash struggle to survive', Institute of Current World Affairs, CVR-27, unpublished manuscript.
- Sebastián-Gallés, N 1991, 'Reading by analogy in a shallow orthography', *Journal of Experimental Psychology: Human Perception and Performance*, vol. 17, no. 2, pp. 471–7.
- Schmidt, R & Kohistani, R 1995, 'The mirror of writing', paper presented at the third international Hindukush cultural conference, Chitral, Pakistan, 26–30 August.
- Schmidt, R & Zarin, M 1981, 'The phonology and tonal system of Palas /kohistyō~/ Shina', *Münchener Studien Zur Sprachwissenschaft*, vol. 40, pp. 155–85.
- Schomberg, R 1938, *Kafirs and glaciers*, Martin Hopkinson, London.
- Shah, W A 1974, 'Notes on Kalasha Folklore', in K Jettmar & L Edelberg (eds), *Cultures of the Hindukush*, Franz Steiner, Wiesbaden, Germany, pp. 69–80.
- Shahidullah, M 1964, 'The languages of the north-western frontier of Pakistan', in A Dil (ed.), *Pakistani linguistics* (1963), Linguistic Research Group of Pakistan, Lahore, pp. 37–44.
- Shalmowski, W 1985, 'The linguistic importance of the Dardic languages', *Journal of Central Asia*, vol. 8, no. 1.
- Shand, J 1979, 'Suggestions for revision of phonemic analysis and orthography in Ilianen Manobo', in M Wendell (comp.), *Notes on literacy: selected articles, issues 1–19*, The Summer Institute of Linguistics, Dallas, pp. 16–21.
- Shorish, M 1984, 'Planning by decree: the Soviet language policy in Central Asia', *Language problems and language planning* vol. 8, no. 1, pp. 35–49.

- Siddiqi, T 2001, 'Urdu', in J Garry & C Rubino (eds), *Facts about the world's languages: an encyclopedia of the world's major languages, past and present*, H W Wilson, New York, pp. 780–5.
- Siiger, H 1956, 'Ethnological field-research in Chitral, Sikkim and Assam: preliminary report', vol. 36, no. 2, *Historisk-filologiske Meddelelser udgivet af Det Kongelige Danske Videnskabernes Selskab*, Copenhagen, pp 1–64.
- 1967, 'Shamanistic ecstasy and supernatural beings: a study based on field-work among the Kalash Kafirs of Chitral', in C-M Edsman (ed.) *Studies in shamanism*, Stockholm, pp. 71–81.
- SIL International 2001, SIL International, Dallas, viewed 6 July 2001, <www.sil.org>.
- Simons, G 1989, 'The computational complexity of writing systems' *The fifteenth LACUS forum 1988*, eds RM Brend and DG Lockwood, Linguistic Association of Canada and the United States, Lake Bluff, IL, pp. 538–53.
- Sjoberg, A 1964, 'Writing, speech and society', *Proceedings of the ninth international congress of linguists*, ed. H Lunt, Mouton, The Hague.
- 1968, 'Sociocultural and linguistic factors in the development of writing systems for preliterate people', in W Bright (ed.), *Sociolinguistics*, Mouton, The Hague, pp. 260–76.
- Skalmowski, W 1995, 'The position of Khowar in the Dardic group', paper presented at the third international Hindukush cultural conference, Chitral, Pakistan, 26–30 August.
- Skutnabb-Kangas, T 1981, *Bilingualism or not: the education of minorities*, trans. L Malmberg & D Crane, vol 7, Multilingual Matters, Clevedon, UK.
- 1990, *Language, literacy and minorities*, Minority Rights Group, London.
- 2000, *Linguistic genocide in education; or worldwide diversity and human rights?*, Lawrence Erlbaum Associates, Mahwah, NJ.
- & Phillipson, R (eds) 1995, *Linguistic human rights: overcoming linguistic discrimination*, Mouton de Gruyter, Berlin.
- Smalley, W 1964a, 'How shall I write this language?', in W Smalley (ed.) *Orthography studies*, North-Holland Publishing, Amsterdam, pp. 31–52.
- 1964b, *Orthography studies: articles on new writing systems*, Helps for Translators, vol. 6, United Bible Societies, London.
- 1976, *Phonemes and orthography: language planning in ten minority languages of Thailand*, Australian National University, Canberra.
- Smith, G et al. 1998, *Nation-building in the post-Soviet borderlands: the politics of national identities*, Cambridge University Press, Cambridge.
- Sperber, B 1995a, 'No people is an island: the Kalasha and their relationships to the surrounding world', paper presented at the third international Hindukush cultural conference, Chitral, Pakistan, 26–30 August.
- 1995b, 'Nature in the Kalasha perception of life', in O Bruun & A Kalland (eds) *Asian perceptions of nature: a critical approach*, Curzon, Richmond, Surrey, pp. 92–109.

- 1996, 'Kalash: dresses and textile techniques', *Proceedings of the second international Hindukush cultural conference, Chitral, 1990*, Hindukush and Karakoram Studies, 1, Oxford University Press, Karachi, pp. 377–408.
- Ståhle, C 1970, *1900-talsvenska*, Stockholm, p. 71, cited in Tauli (1977).
- Steinberg, D 1982, *Psycholinguistics, language, mind, and world*, Longman, London.
- Strand, R 1973, 'Notes on the Nuristani and Dardic languages', *Journal of the American Oriental Society*, vol. 93, no. 3, pp. 297–305.
- 2001, *Indo-Aryan-speaking peoples of the Hindu-Kush region*, Richard F Strand, viewed 30 Jan. 2005, <<http://users.sedona.net/~strand/IndoAryan/IndoAryas.html>>.
- 2002, *A Bibliography of the Languages and Cultures of Nuristân and Environs*, Richard F Strand, viewed 27 March 2005, <<http://users.sedona.net/~strand/bibliography.html>>.
- Street, B 1984, *Literacy in theory and practice*, Cambridge University Press, Cambridge.
- 1995, *Social literacies: critical approaches to literacy in development, ethnography, and education*, Longman, London.
- 2001, *Literacy and development: ethnographic perspectives*, Routledge, London.
- Stubbs, M 1980, *Language and literacy: the sociolinguistics of reading and writing*, Routledge & Kegan Paul, London.
- Summer Institute of Linguistics 1993, *SILDoulosIPA (TrueType)*, SIL International, Dallas.
- 1996, *The work of the Summer Institute of Linguistics in literacy and education*, SIL International, Dallas, viewed 13 August 2001, <<http://www.sil.org/literacy/sillit.html>>.
- Systems (Private) Limited & Cooper, G 1992, *Raakim Publishing System*, ver. 1.51, Nastaliq Urdu word processing software modified for Kalasha, Khovar, Kalami, Burushaski and Shina, Systems (Private) Limited, Lahore.
- Sztemon, L 2002, *Roman alphabets*, Libor Sztemon, Havirov, Czech Republic, viewed 19 March 2005, <<http://www.sweb.cz/ls78/alphabets.htm>>.
- Tauli, V 1977, 'Speech and spelling', in J. Fishman (ed.), *Advances in the creation and revision of writing systems*, Contributions to the sociology of language, vol. 8, ed. J. Fishman, Mouton and Co., The Hague, pp. 17–35.
- Tikkanen, B 1995, 'Areal isoglosses of the transitional zone between Central and South Asia', paper presented at the third international Hindukush cultural conference, Chitral, Pakistan, 26–30 August.
- Topping, D 1973, *Chamorro reference grammar*, University of Hawaii Press, Honolulu.
- Trail, G 1996, 'Tsyam revisited: a study of Kalasha origins', *Proceedings of the second international Hindukush cultural conference*, Oxford University Press, Karachi, pp. 359–76.
- Trail, R 1987, 'Kalasha verb forms', unpublished manuscript.
- 1996a, 'Kalasha verb phrase', unpublished manuscript.
- 1996b, 'Kalasha case-marking system', *Proceedings of the second international Hindukush cultural conference*, Oxford University Press, Karachi, pp. 149–58.
- & Cooper, G 1985a, 'Kalasha phonemic summary', unpublished manuscript.

- 1985b, 'Kalasha-English lexical inventory', unpublished manuscript.
- 1987, 'A Kalasha - English / English - Kalasha dictionary', unpublished manuscript.
- 1997, 'Kalasha – English dictionary', rev., unpublished manuscript.
- 1999, *Kalasha Dictionary—with English and Urdu*, Studies in languages of northern Pakistan, vol. 7, National Institute of Pakistan Studies, Quaid-i-Azam University, & Summer Institute of Linguistics, Islamabad.
- Trail, R & Hale, A 1995, *A rhetorical structure analysis of a Kalasha narrative*, South Asia Work Papers, no. 1, South Asia Group, Horsleys Green, UK.
- Trosterud, T 1999, 'How many written languages in the world?', in *Ogmios*, no. 11, 15 May, p. 16.
- Turner, R n.d., 'Notes on Dardic', *Bulletin of the School of Oriental (and African) Studies*, vol. 4, no. 3, pp. 533–41.
- 1966, *A comparative dictionary of the Indo-Aryan languages*, Oxford University Press, London.
- Turvey, M, Fieldman, L & Lukatela, G 1984, 'The Serbo-Croatian orthography constrains the reader to a phonologically analytic strategy', in L Henderson (ed.) *Orthographies and reading: perspectives from cognitive psychology, neuropsychology, and linguistics*, Lawrence Erlbaum Associates, Hillsdale, NJ, pp. 81–9.
- Underwood, G & Underwood, J 1986, 'Cognitive processes in reading and spelling', in A Cashdan (ed.) *Literacy: teaching and learning language skills*, Blackwell, Oxford.
- UNESCO 2003, *Language Vitality and Endangerment*, Ad Hoc Expert Group on Endangered Languages, Intangible Cultural Heritage Section, UNESCO, viewed 29 November 2005, <http://portal.unesco.org/culture/en/ev.php-URL_ID=9105&URL_DO=DO_TROPIC&URL_SECTION=201.html>.
- Unicode Inc 2003, *The Unicode standard: version 4.0*, Addison-Wesley, Boston.
- US Census Bureau 2002, *Population and housing profile*, US Government, Washington, viewed 30 January 2005, <<http://www.census.gov/Press-Release/www/2002/cnmistatelevel.pdf>>.
- <<http://www.census.gov/Press-Release/www/2002/guamstatelevel.pdf>>
- van Heuven, V 1980, 'Aspects of Dutch orthography and reading', in J Kavanagh & R Venezky (eds), *Orthography, reading and dyslexia*, University Park Press, Baltimore, pp. 57–76.
- Van Orden, G, Stone, G, Garlington, K, Markson, L, Pintt, G, Simonfy, C & Brichetto, T 1992, '“Assembled” phonology and reading: a case study in how theoretical perspective shapes empirical investigation', in R Frost & L Katz (eds), *Orthography, phonology, morphology, and meaning*, Advances in phonology, vol. 94, eds G Stelmach & P Vroon, Elsevier Science Publishers, Amsterdam, pp. 249–90.
- Venezky, R 1977, 'Principles for the design of practical writing systems', in J Fishman (ed.), *Advances in the creation and revision of writing systems*, Contributions to the sociology of language, vol. 8, ed. J Fishman, Mouton and Co., The Hague, pp. 37–54.
- 1980, 'From Sumer to Leipzig to Bethesda', in J Kavanagh & R Venezky (eds), *Orthography, reading and dyslexia*, University Park Press, Baltimore.

- Verhoeven, L 1994, *Functional literacy: theoretical issues and educational implications*, John Benjamins, Amsterdam.
- Voegelin, C & Voegelin F 1964, 'Contemporary language situations in the new world', *Languages of the World: Native American Fascicle One. Anthropological linguistics*, vol 6, no. 6, pp. 2–45.
- 1965, 'Dardic branch or sub-branch of Indo-European', *Languages of the world: Indo-European fascicle one, Anthropological Linguistics*, vol. 7, no. 8, pp. 284–94
- & Schutz, N 1967, 'The language situation in Arizona as part of the Southwest Cultural Area', in D Hymes & W Bittle, *Studies in southwestern ethnolinguistics*, Mouton, The Hague, pp. 403–51.
- Wachtel, A 1998, *Making a nation, breaking a nation: literature and cultural politics in Yugoslavia*, Stanford University Press, Stanford, Calif.
- Weinstein, B 1982, 'Noah Webster and the diffusion of linguistic innovations for political purposes', *International Journal of the Sociology of Language*, vol. 38, pp. 85–108.
- Wikipedia 2005, *Arabic alphabet*, Free Software Foundation, Boston, viewed 25 March 2005, <http://en.wikipedia.org/wiki/Arabic_alphabet>.
- Willson, S 1996, 'Verb agreement and case marking in Burushaski', *Work papers of the Summer Institute of Linguistics*, University of North Dakota, Grand Forks, Session 40, pp. 1–71.
- Wolfram, W & Schilling-Estes, N 1998, *American English: dialects and variation*, Blackwell, Malden, Mass.