Medieval Tibeto-Burman Languages: PIATS 2000

CHRISTOPHER I. BECKWITH, Editor

BRILL

MEDIEVAL TIBETO-BURMAN LANGUAGES

BRILL'S TIBETAN STUDIES LIBRARY

EDITED BY

HENK BLEZER
ALEX MCKAY
CHARLES RAMBLE

VOLUME 2/6

MEDIEVAL TIBETO-BURMAN LANGUAGES

PIATS 2000: Tibetan Studies: Proceedings of the Ninth Seminar of the International Association for Tibetan Studies, Leiden 2000. Managing Editor: Henk Blezer.

EDITED BY

CHRISTOPHER I. BECKWITH

BRILL
LEIDEN · BOSTON · KÖLN
2002

Publication of the Proceedings of the Ninth Seminar of the IATS was made possible through financial support from the Gonda Foundation (Royal Dutch Academy of Sciences—KNAW) and was facilitated by the International Institute for Asian Studies (IIAS)

This book is printed on acid-free paper.

Die Deutsche Bibliothek - CIP-Einheitsaufnahme

Beckwith, Ch. I.
Medieval Tibeto-Burman Languages. PIATS 2000: Tibetan Studies: Proceedings of the Ninth Seminar of the International Association for Tibetan Studies, Leiden 2000. Managing Editor: Henk Blezer / edited by Christopher I. Beckwith. – Leiden ; Boston ; Köln : Brill, 2002
(Brill's Tibetan studies library ; Vol 2/6)
ISBN 90-04-12424-1

Library of Congress Cataloging-in-Publication Data

Library of Congress Cataloging-in-Publication Data is also available

ISSN 1568-6183
ISBN 90 04 12424 1

Cover illustration: Robert Nix
Cover design: Cédilles / Studio Cursief, Amsterdam

PRINTED IN THE NETHERLANDS

CONTENTS

This page intentionally left blank

PREFACE

The papers in this volume represent the efforts of a small number of Tibeto-Burman linguists working under the assumption that, as in historical linguistics in general, records of a language from an earlier period generally preserve features that are lost or obscured in records from a later period, and the number of changes increases over time.[1] All but one of the papers were given in Leiden in the summer of 2000 at a special Medieval Tibeto-Burman Languages Symposium held in conjunction with the 9th International Seminar on Tibetan Studies.

The Symposium papers were: Christopher I. Beckwith: 'Toward the establishment of a scientific field of Tibeto-Burman historical linguistics'; Tsuguhito Takeuchi: 'Further remarks on the Old Zhangzhung language'; Rudolf Yanson: 'Reconstructing some grammatical rules in Burmese (a case of Pali-Burmese interference)'; Kashinath Tamot: 'Some characteristics of the Tibeto-Burman stock of Early Classical Newari'; Ksenia Kepping: 'The verb in Tangut'; Shobhana Chelliah and Sohini Ray: 'Discovering Tibeto-Burman linguistic history through 16th Century Meithei manuscripts'; Nicolas Tournadre: 'The extraordinary relationship between Literary Tibetan and the dialects'; and Christopher I. Beckwith: 'Pyu, Old Tibetan, and Old Burmese versus the other Tibeto-Burman languages: the problem of early Tibeto-Burman dialectology'. Upon my request, David Bradley has generously contributed a detailed survey of the Tibeto-Burman languages to the present volume, covering not only their subclassification but also their present status, and including valuable remarks on the early literary languages. The papers are printed in order, approximately, of the earliest attestation of the languages concerned, followed by the comparative surveys and glossaries (also in chronological order). The Tangut glossary is the result of work done by its authors at the time of the 9th International

[1] If there were any doubt that this principle applies even to contemporary languages, David Bradley noted in his paper at the International Conference on Sino-Tibetan Languages and Linguistics in Lund, 1998, that a morphological element in a language he had done fieldwork on twenty years earlier has now completely disappeared; if he had not recorded it, there would be no way to know it had ever existed.

Seminar on Tibetan Studies in Leiden.[2] I would like to thank the Office of International Programs at Indiana University for providing financial support so that I could undertake this joint research effort.

I would like to thank all of the contributors for the thought and hard work they have put into their papers and glossaries, and Henk Blezer, the Convenor of the 9th International Seminar on Tibetan Studies, who has encouraged and helped me at every step of the way from conception and organization of the Symposium to publication of the proceedings. I also am indebted to William King, who helped me by proofreading my own papers, and most especially to Michael Walter, who very kindly entered Kashinath Tamot's paper and glossary into the computer for me and also helped proofread the manuscript. All remaining errors are my responsibility.

Tokyo, November, 2001 Christopher I. Beckwith

[2] After the draft of the present Tangut glossary was completed, I unfortunately was unable to reestablish contact with Dr. Kepping. Accordingly, all errors in this work are my sole responsibility.

ABBREVIATIONS

adj.	adjective
adv.	adverb
af.	affix
aux.	auxiliary
Bax.	W. H. Baxter. 1992. *A handbook of Old Chinese phonology.* Berlin: Mouton.
BEFEO	*Bulletin de l'École Française d'Extrême-Orient*
Ben.	P. K. Benedict 1972. *Sino-Tibetan: a conspectus.* Cambridge: Cambridge University Press.
BMFEA	*Bulletin of the Museum of Far Eastern Antiquities*
BSOAS	*Bulletin of the School of Oriental and African Studies*
c./ca.	circa
CAUS	causative
CLF	classifier
CNew	Classical Newari
conj.	conjunction, conjunctive
COP	copula
CTB	Common Tibeto-Burman
CVC	consonant-vowel-consonant
dial.	dialect(s), dialect form
ECN	Early Classical Newari (Old Newari)
EMC	Early Middle Chinese
EOC	Early Old Chinese—language of the Shang oracle bone inscriptions
FUT	future
Got	Gothic
Grk	Greek
HJAS	*Harvard Journal of Asiatic Studies*
HSR	Historic Sinological Reconstruction
JAOS	*Journal of the American Oriental Society*
JBRS	*Journal of the Burma Research Society*
JRAS	*Journal of the Royal Asiatic Society*
Kar.	B. Karlgren 1957. *Grammata serica recensa.* Stockholm: BMFEA.
Lat	Latin
LBur	Literary Burmese
LCN	Late Classical Newari
LMC	Late Middle Chinese (T'ang Chinese)
LOC	Late Old Chinese—the language of the Classics
LTBA	*Linguistics of the Tibeto-Burman Area*
LW	loanword(s)
LZZ	Late Zhangzhung
Man	Mandarin
Mar.	S. Martin 1987. *The Japanese language through time.* New Haven: Yale University Press.
MBur	Modern Burmese
MChi	Middle Chinese—unperiodized reconstruction (marked with a *)

MMan	Middle Mandarin
MOC	Middle Old Chinese
n.	noun
NBur	New Burmese
NJpn	New Japanese (modern standard Japanese)
NMan	New Mandarin (modern standard Chinese)
NMei	New Meithei
NP	Noun Phrase
NS	Nepal Samvat
num.	numeral
NZZ	New Zhangzhung
OBur	Old Burmese
OChi	Old Chinese—the ancestor of MChi (unperiodized reconstruction)
OEC	Old Eastern Chinese (dialect of the northeast China coast)
OIri	Old Irish
OJpn	Old Japanese—the language of Japanese texts to ca. 1000 A.D. (forms transcribed from Man'yôgana are marked with a *)
OKog	Old Koguryo—an ancient language once spoken in Korea and vicinity
OMan	Old Mandarin
ONew	Old Newari (Early Classical Newari)
OPyu	Old Pyu
OTib	Old Tibetan—the language of Tibetan texts written ca. 650-950 A.D.
OZ	Old Zhangzhung text
PAN	Proto-Austronesian
PASTPRES	past-present tense
PChi	Proto-Chinese—the unattested ancestor of Old Chinese
PEI	Proto-Eastern Indo-European
perh.	perhaps
PIE	Proto-Indo-European
PJK	Proto-Japanese-Koguryoic—the language family including the Japanese-Ryukyuan and Puyo-Koguryo branches
PJpn	Proto-Japanese
PKog	Proto-Koguryo
PNew	Proto-Newari
Pok.	J. Pokorny 1959. *Indogermanisches etymologisches Wörterbuch.* Bern: Francke Verlag.
pro.	pronoun
prt.	particle
PTib	Proto-Tibetan
PTB	Proto-Tibeto-Burman
PTP	Proto-Tibeto-Pyu
PTok	Proto-Tokharian
Pul.	E. G. Pulleyblank 1991. *Lexicon of reconstructed pronunciation in Early Middle Chinese, Late Middle Chinese, and Early Mandarin.* Vancouver: UBC Press.
rt.	root
Skt	Sanskrit
SOV	Subject-Object-Verb
ST	Sino-Tibetan

Sta.	Sergei A. Starostin 1989. *Rekonstrukciya drevnekitayskoy fonologičeskoy sistemï*. Moscow: Nauka.
SVO	Subject-Verb-Object
Tak.	T. Takata 1988. *Tonkô shiryô ni yoru chûgokugoshi no kenkyû*. Tokyo: Sobunsha.
TB	Tibeto-Burman
Tgt	Tangut
Tok	Tokharian A and Tokharian B
TokA	Tokharian A—the East Tokharian language
TokB	Tokharian B—the West Tokharian language
var.	variant
v.	verb
V	Verb; Vowel
v.c.	causative verb
v.i.	intransitive verb
v.inf.	verb infinitive
VP	Verb Phrase
v.t.	transitive verb
Wat.	C. Watkins 2000. *The American Heritage Dictionary of Indo-European roots*. 2nd edition. Boston: Houghton Mifflin.
ZDMG	*Zeitschrift der Deutschen Morgenländischen Gesellschaft*

This page intentionally left blank

INTRODUCTION

CHRISTOPHER I. BECKWITH

I

In the past few decades, as areas where Tibeto-Burman languages are spoken have become more accessible for fieldwork, many linguists have naturally wished to take advantage of the possibility of working directly with living Tibeto-Burman speakers. These scholars have produced much excellent descriptive work on modern Tibeto-Burman languages, and it is hoped that they will continue to do so. In fact, research on Tibeto-Burman has shifted so completely from its earlier focus on the literary languages that today it is almost exclusively on the modern spoken languages. In itself this is not, of course, bad at all; it is natural and good. Moreover, historical linguistics includes diachronic study of the modern languages, and everyone benefits from good new data.

Most historical linguists working on reconstruction of a common ancestor of a divergent language family, however, are accustomed to think that the traditional method, based on the Indo-European model of careful study of the earliest attested material in each branch, is the only proven model for divergent (or 'genetic') linguistic relationships, and we should use it when working within such a theory. In historical linguistics, as in history generally, time is a critical factor. Accordingly, while the earliest sources are not the only useful data, they are nearly always the most important for reconstruction both of earlier stages of individual languages and of earlier stages of parent languages.

A vast amount of research has been published on Tibeto-Burman languages and Tibeto-Burman reconstruction since 1972, when Paul Benedict's landmark study, *Sino-Tibetan: A Conspectus* was published. The *Conspectus*, which to this day remains the standard reference point for all Tibeto-Burman historical linguistic work, has many good qualities, but one serious fault, the source of many of its errors: the only premodern Tibeto-Burman languages cited (unsystematically) in the book are what Benedict calls 'Written Tibetan' and

'Written Burmese'. These terms are in fact simply equatable with whatever is contained in, respectively, Jäschke's Tibetan-English dictionary and Judson's two Burmese dictionaries, both of which, though excellent, contain many modernisms alongside their numerous archaisms. Although the younger generation of Sino-Tibetan linguists, at least in Europe, has apparently accepted the necessity of actually learning a little Tibetan—the oldest well-attested Tibeto-Burman literary language—until very recently this has not been the case. The necessity of personally coming to grips with Chinese reconstruction (an extremely debatable subject, to be sure) has also been avoided far too often. Learning how to use Karlgren's pioneering *Grammata Serica recensa* (1957), still a standard reference point for Old Chinese reconstruction, was thought to be enough. Thus, nearly all Tibeto-Burmanist publications until very recently have followed Benedict in basing themselves on the same strange mixture of Jäschke, Judson, Karlgren, and a great number of different modern Tibeto-Burman languages. Surely this is not the ideal way to proceed. We do want as much data as possible on as many languages as can be described, but as a rule it is the earliest genuine data that is most important and must be utilized to the utmost to achieve reliable results when it comes to reconstruction.

It has very often been remarked that we do not have sufficient old linguistic data on Tibeto-Burman languages to follow the Indo-European model, and also that the languages we do have are not old enough. This objection is simply wrong. While we do not have anything as old as Hittite, or Minoan Greek, most Indo-European languages are actually attested only from the medieval period and are thus no older than most branches of Tibeto-Burman, which are also attested from the medieval period. For example, other than a few words known from foreign transcriptions, Germanic is attested only from the fifth century of our era; the Celtic, Slavic, and Tokharian languages are attested later still; and some languages are known only from modern times. The fact is that we do have a significant corpus of old Tibeto-Burman linguistic material, it is much earlier than is available for many language families, and as David Bradley points out (in this volume), there is at least one early literary language for each major branch of Tibeto-Burman. By contrast, neither Uralic nor Tungusic, both of which are well-established language families, is attested until well into the Middle Ages, long after the earliest Tibeto-Burman language data. Virtually the same is true of most

other established language families in eastern Eurasia, such as Austronesian, Turkic, and Mongolic. In short, if we combine what can be learned from the early linguistic material with what is being learned about modern Tibeto-Burman languages we should eventually be able to achieve results comparable to those attained in other areas of historical linguistics.

One reason that the reconstruction of Proto-Tibeto-Burman is not a simple matter is, partly, phonology. Syllable onsets are extremely complex in some languages, such as Old Tibetan, but comparatively simple in others, such as Old Newari (Early Classical Newari). The productivity of prefixing in many languages compounds the problem. At the same time, syllable codas are extremely restricted—Proto-Tibeto-Burman must have allowed few, if any, consonantal codas, and when we do reconstruct a given branch, the inevitable result is the pruning away of many segments from those same complex syllables, resulting in Proto-Tibeto-Burman forms that often consist of only two or three segments, such as *la 'moon'. However, it is thus even more—not less—important that we follow the proven methods of historical linguistics. Traditional analysis of each old literary language, and internal reconstruction of each branch, is absolutely necessary. Direct comparison of individual modern spoken languages with each other, without attention to the historical languages, can only give misleading results. Moreover, we should bear in mind that our methodological model, Proto-Indo-European, is traditionally reconstructed with overwhelmingly monosyllabic roots, most of which consist of only three segments.

II

The papers in this volume address various issues of central importance in the development of a field of Tibeto-Burman historical linguistics on the Indo-European model. In order to build as firm a foundation as possible, the traditional historical linguistic method requires that several steps be followed.

First of all, it is necessary to do the basic philology—to find the texts and decipher the languages. While that has been done for most early Tibeto-Burman languages, for a number of them virtually the only thing that may be said is that they are 'early literary languages'. Several, such as Nam and Zhangzhung, remain undeciphered, while

others, such as Pyu and even Tangut, remain only partly deciphered, despite the existence of fairly extensive texts. "The Old Zhangzhung Manuscript Stein Or 8212/ 188," by Tsuguhito Takeuchi, presents one of several Old Zhangzhung texts that are currently being analyzed in an attempt to decipher the language. The paper by Kashinath Tamot, "Some Characteristics of the Tibeto-Burman Stock of Early Classical Newari," presents a great deal of material drawn from philological study of early manuscripts and reveals recent discoveries about the Tibeto-Burman side of this relatively little investigated language. Shobhana Chelliah and Sohini Ray, in their paper, "Early Meithei," survey the significant surviving literature in the neglected early Meithei or Old Manipuri language, including a large number of manuscripts and inscriptions, some perhaps as early as the ninth century.

Secondly, the target languages must be analyzed grammatically. "Reconstructing Some Grammatical Rules in Burmese: A Case of Pali-Burmese Interference," by Rudolf A. Yanson, shows how some elements of Burmese grammar which appear to be fully Burmese actually are the result of convergence with literary Pali in the Old Burmese period. It is an exquisite demonstration of careful, philologically informed linguistics.

Next, lexicography must be put on a firm basis, with careful study of the lexical stock of each early language, publication of glossaries, and, eventually, the compilation of historically sensitive citation dictionaries. The glossaries of Old Newari, Old Burmese, Pyu, Tangut, and early Meithei contributed to the present volume are intended to be a first step in this process—one that has barely begun.

Until very recently, regardless of whose theory we might consult about the development of the Tibeto-Burman family of languages, all branches of the family have been treated as equally old, implying simultaneous splitting into four, seven, nine, or however many daughter languages. Because we are still in the dark about intermediate families, we are unable to reconstruct confidently back to Proto-Tibeto-Burman, not to speak of further than that. Moreover, Chinese and Tibeto-Burman have also been treated as twin sisters, contemporaneous daughter languages of a common ancestor.

The most current example of the traditional model essentially repeats the model of Benedict (1972), with the only significant structural difference being the lowering of Karen's status from a higher-order level to just another branch of Tibeto-Burman and the

addition of a Qiangic branch, which is thought to include Tangut (Matisoff 1997: x). In view of the development of better known language families, however, the simultaneous division model of daughter languages from Proto-Tibeto-Burman is highly unlikely. One is also compelled to ask why there should be such an extreme difference in homogeneity between the two supposed branches of the putative Sino-Tibetan family: a Sinitic branch consisting of a few closely related modern Chinese languages descended from a few ancient Chinese dialects of, basically, one language, and a Tibeto-Burman branch which apparently gave birth to several highly divergent families in early childhood.

The traditional model is now in the process of being reexamined, as shown by George van Driem's recent proposal to reorganize the Tibeto-Burman family. He proposes a radical reconceptualization of Tibeto-Burman that follows a theory suggested by Nicholas Bodman (1980), though the latter scholar did not follow this model himself. According to van Driem, Chinese and Tibetan are especially closely-related sister languages within a 'Sino-Bodic' sub-branch of an Eastern branch of a totally reorganized Tibeto-Burman family (van Driem 1997:463). In other words, in this model Chinese is ultimately just another Tibeto-Burman language. There is a certain poetic appeal to van Driem's proposal, which also makes a more natural-looking family tree, though in itself it does not address the issue of the relative chronology of the splitting off of the other daughter languages of Tibeto-Burman proper.[1]

The classification of Tibeto-Burman languages by David Bradley in this volume, "The Subgrouping of Tibeto-Burman," also attempts to place the daughter languages in a more realistic family tree. In his survey of the Tibeto-Burman family, he points out which early literary languages are known so far, and to which branch of the family they are thought to belong, noting that in some cases the texts or the living native scholars who can interpret the texts are languishing and disappearing as quickly as many of the endangered modern spoken Tibeto-Burman languages. The paper in this volume by the present writer, 'Two Pyu-Tibetan Isoglosses', approaches the problem of classification from the other direction, positing an

[1] However, he has also proposed an equally radical model of East Asian linguistic relationships that does address this issue in an archaeological context (van Driem 1998).

intermediate daughter of Tibeto-Burman that includes both Pyu and Tibetan.

Tibeto-Burman has been connected with Chinese (and, in earlier times, with Taic and other languages) in the hypothetical 'Sino-Tibetan' family of languages. While this genetic relationship hypothesis is still considered unquestionable by some, in the past few years an increasing number of linguists have begun questioning it anyway. From a purely typological point of view it must be noted that the distinctive SVO syntax of Chinese is attested from the very earliest texts, the Oracle Bone Inscriptions, which date back to the second millenium B.C. Tibeto-Burman languages, by contrast, have SOV syntax, as do Japanese-Koguryoic, the 'Altaic' group of languages, and early Indo-European. It is generally agreed that the only Tibeto-Burman exception, the Karenic branch, which has a significant non-Tibeto-Burman component, has developed its own variety of SVO syntax under the influence of Taic languages. This fundamental grammatical difference between Chinese on the one hand and Tibeto-Burman languages (as usually conceived) on the other must therefore be well over three millenia old. If there were any special genetic closeness between Tibetan and Chinese, it must have existed before that time, but at such an early date it is unlikely that a distinctive Tibetic subfamily existed within Tibeto-Burman. This appears to be a fatal problem for van Driem's theory. The very heavy, multi-layered Chinese loan influence on Tibetan, and on Tibeto-Burman languages in general, that is now being emphasized by some linguists, most notably Laurent Sagart (1999), probably accounts for most—if not all—of the perceived closeness of the two languages. But before this issue can be decided, it is imperative that all possible loanwords be eliminated from the data used to reconstruct the putative Sino-Tibetan proto-language. It will remain impossible to achieve reliable results, and come to a well-supported conclusion on the received Sino-Tibetan theory or any alternative theory such as van Driem's, without dealing with loanwords, which are fundamental to the reconstruction of anything. Some of the above problems are dealt with by the present writer in 'The Sino-Tibetan Problem', in this volume.

The papers presented here demonstrate that the medieval Tibeto-Burman languages are of central importance to Tibeto-Burman historical linguistics. Just as in Indo-European historical linguistics, the earliest attested languages are indispensible for the reconstruction

of earlier forms of the common language, including not only the intermediate stages but the proto-language itself. If we do focus on and study exhaustively the earliest historically attested data, the common effort will give rise to a scientific field of Tibeto-Burman historical linguistics that will not only answer questions about the early languages, it will provide a meaningful context for the diachronic study of the modern languages.

References

Benedict, Paul 1972. *Sino-Tibetan: a conspectus*. Cambridge: Cambridge University Press.

Bodman, Nicholas 1980. Proto-Chinese and Sino-Tibetan: data towards establishing the nature of the relationship. *In* Frans von Coetsem and Linda R. Waugh, ed., *Contributions to historical linguistics: issues and materials*. Leiden: E.J. Brill, 34-199.

Driem, George van 1977. Sino-Bodic. *Bulletin of the School of Oriental and African Studies*, 455-488.

—— 1998. Neolithic correlates of ancient Tibeto-Burman migrations. *In* Roger Blench and Matthew Spriggs, ed., *Archaeology and language II: correlating archaeological and linguistic hypotheses*. London: Routledge, 67-114.

Karlgren, Bernhard. 1957. *Grammata Serica recensa*. Stockholm: reprinted from the *Bulletin of the Museum of Far Eastern Antiquities* 29: 1-332.

Matisoff, James A. 1997. *Sino-Tibetan numeral systems: prefixes, protoforms, and problems*. Canberra: Pacific Linguistics.

Sagart, Laurent 1999. *The roots of Old Chinese*. Amsterdam: John Benjamins.

This page intentionally left blank

THE OLD ZHANGZHUNG MANUSCRIPT STEIN OR 8212/ 188

TSUGUHITO TAKEUCHI (KOBE CITY UNIVERSITY OF FOREIGN STUDIES)

INTRODUCTION

In addition to the Old Tibetan texts found in the Dunhuang cave, there are a considerable number of texts in non-Tibetan languages written in Tibetan script. They include phonetic transcriptions of Chinese, Sanskrit, Uighur, Khotanese, and Tangut. In addition, there exist a number of unidentified or only partly identified texts.[1] Their identification is difficult because the languages written in these texts seem to have become extinct without leaving any comparable records in other scripts.

F.W. Thomas studied one of these languages, which he called 'Nam', and published a manuscript in Nam from the Stein Collection (Thomas 1948). According to Thomas, Nam belongs to the Tibeto-Burman language group. Another language identified and studied by Thomas is Zhangzhung, though he published only part of the texts and results of his research (Thomas 1926, 1933, 1967), leaving the task of more detailed research as well as decipherment of these texts to the future.

In two consecutive research projects on Bon studies (1997-1999 and 2000-2002) headed by Professor Yasuhiko Nagano of the National Museum of Ethnology in Osaka, I was put in charge of the Zhangzhung texts, in collaboration with Professor Nagano and Dr. Sumie Ueda, an expert in computers and statistics, who did computer analyses of the data.

The purpose of our research is to decipher the texts—the oldest written records of what may be the Zhangzhung language—and reconstruct the linguistic forms behind them in comparison both with the Zhangzhung language corpus found in later texts as well as with other Tibeto-Burman languages. So far we have databased the major

[1] Pelliot tibétain 1242, 1243, 1246, 1248, 1250 (the same language as 1243), and 1255 (the same language as 1242). The Berlin Collection also contains one text in an unknown language written in Tibetan script, though the text is from Turfan, not Dunhuang (Francke 1927).

texts and made a preliminary analysis of the linguistic constructions, including phonological, morphological, and syntactic structures. A preliminary report on our research was presented at the international symposium, "New Horizons in Bon Studies," held in 1999 at the National Museum of Ethnology, in Osaka. I read a second report at the 9th Seminar of the International Association for Tibetan Studies held at Leiden in 2000. The two reports were combined and published in a paper (Takeuchi, Nagano, and Ueda 2001) that outlines our research and results up to 2000.

In this paper I wish to focus on one unpublished Zhangzhung manuscript, Stein Or 8212/ 188, and present a transliterated text to serve as data for further analysis.

I. What is Zhangzhung?

Before going into the text at hand, let us briefly explain what Zhangzhung is, and look at the materials regarding Zhangzhung.

Zhangzhung is an extinct Tibeto-Burman language once spoken in the powerful kingdom of Zhangzhung in western Tibet. The kingdom was conquered and incorporated into the expanding Tibetan Empire in the seventh century A.D. Subsequently, the Zhangzhung people were gradually assimilated by the Tibetans; their language was eventually replaced by Tibetan and died out. Other than a few personal names scattered in Old Tibetan texts, the Dunhuang texts with which we are concerned are the only contemporaneous written records of Zhangzhung which have come down to us from the time the language was still apparently alive.

However, the Zhangzhung language did not vanish completely. Its tradition seems to have been maintained for a long time by Bonpo priests, for whom it was their sacred written language. In the 1960s, Tibetan Bonpo scholars published a Tibetan-Zhangzhung dictionary, and a bilingual text in Tibetan and Zhangzhung, the *Mdzod-phug*. Although the *Mdzod-phug* is ascribed to an eighth century Bonpo scholar, the text seems to have been written later, and the language described is significantly different from that found in the Zhangzhung manuscripts from Dunhuang. We therefore call it 'New Zhangzhung' in contrast to the 'Old Zhangzhung' language of the Dunhuang texts.

A comparison of Old Zhangzhung and New Zhangzhung reveals that the difference between the two is greater than can be explained as a result of natural historical change. In other words, New Zhangzhung is likely to have developed as a result of Tibetan Bonpos combining their knowledge of Old Zhangzhung with that of Tibetan and Sanskrit (Takeuchi, Nagano, and Ueda 2001: 56-57). Thus, although New Zhangzhung provides important information for the decipherment of Old Zhangzhung, especially the identification of lexical items, we need to formally analyze the Old Zhangzhung data first.

II. Sources

Three manuscripts in Tibetan script were previously identified as (or alleged to be in) the Zhangzhung language. Here we call them OZ (= Old Zhangzhung Texts) 1, 2 and 3. OZ 1 and OZ 2 are in the Stein Collection, now preserved in the Oriental and India Office Collections of the British Library. OZ 3 is in the Pelliot Collection at the Bibliothèque Nationale in Paris.

In 1999, while examining manuscripts bearing unknown languages in Tibetan script at the Bibliothèque Nationale, I found two more manuscripts that appear to belong to the same linguistic corpus. We will call them OZ 4 and OZ 5.

Thus we have five Old Zhangzhung texts, all of which were found in the famous Dunhuang cave. They are written on the backs of scrolls with Chinese Buddhist texts on the other side. All of them seem to be medical texts. In spite of these common features, they are apparently different texts—in other words, they are not pieces of the same text.

The basic palaeographical features and approximate size of each text are given below. OZ 1 and OZ 2 contain a similar amount of text. OZ 3 is the most extensive. OZ 4 and OZ 5 are much smaller. Due to its size, OZ 3 is the only one that has not yet been databased. Although palaeographical evidence indicates that the Zhangzhung texts were written after the Chinese texts, the Zhangzhung sides are designated as *recto* in the following descriptions.

1. *OZ 1: VP 755 (Ch. Fragment 43)*

This text is catalogued by de la Vallée Poussin under the catalogue number 755. It bears the site number Ch. (= Ch'ien-fo-tung) Frag-

ment 43. The text is written on the back side of a scroll which has a Chinese vinaya text on the other side. It lacks the beginning due to paper damage. The remaining part measures 187 x 25 cm and contains 126 lines, in which we count 578 different syllables, for a total of about 2,000 syllables.

A transliterated text of OZ 1 was published by F.W. Thomas (1967). I checked his transliteration against the original manuscript and revised the reading, then databased it.

2. *OZ 2: Or 8212 / 188*

F.W. Thomas also made an unpublished transliteration of OZ 2, which was found among his unpublished drafts preserved in the Collection of European Manuscripts of the Oriental and India Office Collections. In 1999 I checked his transliteration against the original manuscript and revised his reading. I present a revised reading of OZ 2 in the present paper.

3. *OZ 3: P 1251*

This text lacks the beginning (the top edge is torn off). The remaining part measures 285 x 25.5 cm and contains 287 lines, with a total of *c.* 6,300 syllables. OZ 3 has not really been touched. Though we quickly went through it, it will take some time to database it.

4. *OZ 4: P 1247*

This text lacks the beginning (the top edge is torn off). The remaining part measures 72 x 25.5 cm, and contains 46 lines of most probably Old Zhangzhung, in total *c.* 790 syllables. I have examined and databased OZ 4.

5. *OZ 5: P 1252*

Written on the back of the Chinese sutra *Guan-yin-jing*. Although the bottom edge is torn off, the text is complete with 47 lines of most probably Old Zhangzhung, in total *c.* 850 syllables. Two more lines are written on the Chinese side, but their relation to the *recto* text is not clear. I have examined and databased OZ 5.

III. OZ 2 (STEIN OR 8212/ 188)

This text is found among the group of texts with the requisition number Or. (= Oriental Manuscript) 8212, in which it bears the number 188 together with another totally independent manuscript. The text is written on the back of a scroll with the Chinese *Buddhanâma sûtra*. It lacks both the beginning and the end. What remains measures 115 x 26.3 cm and contains 86 lines. Eight more lines are written on the Chinese side; they are probably a continuation from the *recto* side. There are a total of 94 lines with 593 different syllables, i.e., about 2,200 syllables in all.

TEXT

Recto

1.] [/ pa zur ŋag na gra tan]
2. [±12] rag zu [tsins?] tad mïn pï rhyelse graŋe gva re rhi laŋ run
3. bug rhïŋ yos mu chïl cib min chil cib na ?aŋ da phyend min da phyen nve run graŋ ŋog [ŋag]
4. luŋ rhyun nu cha tu blun blan kun bun ci tsis tad min pi bu shag ci pi rhyelas graŋŋe rdim le
5. len ga sub rchusi rgyiŋs go phal rag mu saŋ graŋ se kha than si lŋa graŋ yu run min lun saŋ ci ni[ŋ]
6. than saŋ we ma rgyïŋs khar ŋog ŋag gra rtho [pu=du] tse mu maŋ kye saŋ graŋ nï win shvin shvin mar myag ti tse
7. [ci=ri] sa [rh]yasaŋ graŋ run graŋ ŋog ŋag tiŋ ci rad khve su gu ŋog ŋag [zho=no] rtho zu bo run graŋ na lo daŋ
8. nyo luŋ mar bag yil yil graŋ ge chvi ye pi khhe khes ge khyun ryuŋ rhelse zva slig ci rhiŋ niŋ tsamed
9. [min=sni] sa gver gver mu tiŋ yaŋ ge tod lŋa mle ma tha min slig saŋ na bun ma lŋa mu zhï kye na ma lod di pi
10. lï ŋog ŋag rï khum tshugs mïn ri shil rkhya pra nve bun ma lŋa rï zhi rkhya na ma lo di yil yil baŋ baŋ bï sug
11. [fiaŋs] ge ryvege bï nu rthuls bun ma lŋa bizhi ryvad na ma lo di kvi po kya min mans saŋ cï bun ma lŋa kvï
12. [mans] na ma lo di / pi khe khe ge gyun ryuŋ rhyelse gva re rhi lŋa run bug rhïŋ yos mu chïl cïb min chil
13. [cib] na ?aŋ da phyend mïn da phyed nve run graŋ ŋog ŋag gu bag cï tsis tad min pi yig bu tiŋ liŋs tha

14. [nim] lem ga sub rchu sï pu zhi na gyiŋ chud laŋ skyiŋ wïng braŋ baŋ mu ge toye rho khug shud ge ryvege ze ma
15. to tags putse mu run kye run ryuŋ na lod no takhe khe no gyun ryuŋ nï / ti zhi nve gyiŋ chud lŋa ma sa rŋalde
16. la khram rme ge ryvege ?i di sŋal tshars ti tse sa run ?ï run ryuŋ na lod no [zhis] gekhe no gyun ryuŋ nï khve si nu
17. gyiŋ chud lŋa maŋ saŋ shid ge run saŋ shiŋge ryvege saŋ khve mar mans saŋ run ryuŋ na lod no mu
18. mum khe khe no gyun ryuŋ nI zu zhi ci gyiŋ chud lŋa pi mu tos shiŋg pham rhyaŋ rme ge ryve ge gluŋ glaŋ
19. bag mu thuŋ mïn [cha=tsa] zog ryuŋ se tad mïn zva tha pï khekhe no gyun ryuŋ nï rhiŋ ni mu rve rumb mïn rve ru
20. na zhï kye star gad mïn star gad nve lŋa ni shig mu ran rŋvï / pï khe khe ge gyun ryuŋ rhyel se gva re rhi lŋa rburï kho
21. ga tog dad du ryug tso ?an rya mar byun stes ryuŋ tïŋ liŋs gun tsa[b] ce skyiŋ chu nul nal ka pye
22. ras na ma run khlaŋg khlaŋg stes po kya mïn slig saŋ nve mu wiŋ tha mïn la gra nu shïm ŋog ŋag
23. gra rthas cï [tsïs] tad min pi yig ryuŋ tïŋ lïŋs [tha?] chud lŋa skyïŋ chu nul nal ka phyeras na tse ze ma to tag
24. tags mu run ryuŋ tog ra lŋa mu zhï kye sthe bard de rho yu ŋog mar / cond ryuŋ ŋe zva slïg ci rhiŋ
25. nïŋ tsamed go ma ran mïn la skams na ŋo mu ran mïn tiŋ gyuŋs nve druŋ mu ran mïn zhas sthaŋs
26. nu rgvil mu rane ca toŋ tab tse tade ryvan tse du rhyuŋ lumse pi yig mu ran byer re zva slig ci // ma run khlaŋ
27. khlaŋg stes po kya min slig saŋ nve tse ?i di sŋal tshars sa run ryuŋ tog nar lŋas zhï ?i ge bag ci
28. ga nul nal ?i run tshum me rho yu dru dra mar zhegs ryuŋ ŋe zva slïg ci rhiŋ tsamed sko fium sko fipab shaŋ
29. thal rlïŋ se rva fiu drin drin se skhyers na rmo ma nul du phyud kyere phaŋ yas nve tsha gya khan kar rnïl / purd
30. tab mïn khan ryams nu ?al da ge rha ryams rvaŋ rï ran mïn se skhros ci skyiŋ rjel braŋ brugs tï ri
31. phyud kyer min sar skhro tshe tsis tad min rbu ri khog tog dad du ryug lŋa ŋa yïg sa khla su ye zun lod di
32. [du?] tse saŋ sthe par [mans] saŋ run ryuŋ tog ra lŋa khyve khyve thrïm thrim saŋ pyam phod lŋa ken sig sa
33. ŋa run ryuŋ rhïŋ niŋ [min=sni] shig saŋ run rŋvï yi de lo ci tse gluŋ glaŋ bag mu thuŋe tsha zog ryuŋ tog ra lŋa chiri

34. ga zug du blun blan pï mu lab tog du ci dags ge ta ye gaŋ khus blun blan duŋ da ryuŋ ŋe zva slig
35. cï rhiŋ niŋ ham sïg spuŋs mu nam tog dad de la na hams bi shos zhag rag nam tog
36. dad de zhag nve lend / rhu spuŋs luŋ thad nam tog dad de pe luŋ shvïn shvin graŋ rha nu car spuŋs rme thad
37. nam tog dad de mu byun rimb min staŋ ryambs cï she spuŋs spru thad nam tog dade rhild go
38. grag min go grags tshe rhyvid spuŋs gyva thad nam tog dade gyva waŋs mi ŋva spuŋs
39. smyer thad nam tog dade sham luŋ dru dra smyer cerd taŋ tsis tad min duŋ ryuŋ ryïŋ yos [min=sni] shig duŋ
40. tiŋ lŋis na lodo // tha ye ne thod ryvun tsig du rphuŋ lummse pi yig mu ran byer shis mu zhi kye sthe barde
41. rho yu ŋog ŋag mar cond ryuŋ ge tsugs mïn ŋo mu pho braŋ gra rta gab ge zhi min tiŋ rhyun gluŋ gluŋ
42. mu pho braŋe phal ma nï so na ga gve nyen bin no dra rŋad ni ru ma tha sho byun rhyvid mar daŋ drum drum shiŋg
43. ge ru ye mu saŋ pho rmad ma ta sho khre ma fieg rhyels skar ge [ŋve de=rtsede] mu saŋ pho rbaŋ ma ta sho pi pe kha thur
44. yil yil snan ge rbaŋ ŋo mu saŋ pho mu saŋ ye kye ti taŋe ba min sa nï go mu chvide ma shï min gvas gvas
45. tïŋ ru pu tse mu saŋ pho ŋo mu chvid de sla paŋ gva gva kun lyam tï tse mu saŋ pho druŋ mu chvïde kha thur
46. yïl yil snan rphaŋ khve tse mu saŋ pho rgvil mu chvïde seg tsham bïs bas lï ryugs zu tsïs tad min
47. pï yig mu saŋ pho she saŋ pho gum gun rva len ?ag tha pï ma skhu mur zhur ca pho khyu waŋ logs
48. lŋa: slye tha summ zuŋ rman spaŋ lyag ga na braŋ traŋ la shï min tïŋ cond ŋog ŋag / ryuŋ mïskhye
49. ra ga rhyu gru nuŋ gyu kharce phludï ka wa yun thul bud bud bu khvï khye ga de se rvaŋ bag khar ce khvis khyerd
50. ni chi sa ŋus gun pho rhyam ni ma sa ŋas gun nal rgyï ni de gligs glïgs gun pho war nï ma gligs
51. gun pho war ni de prand prand tog pho spald nï de khrund: khrund tog pho rgvi ni de dvabs dvabs to
52. tog pho khvil ni ma dvabs dvabs tog pho khvil ni rha ru mar rjel shaŋg shaŋg pho ge chim wa ri ma se
53. bi phud nï ti bi gu saŋ shïŋ gun tsan ce gyim bi la rug mu saŋ kye saŋ shiŋge tho thvi ni ryun rhyid mar daŋ

54. drum drum shiŋge tshame kaŋ smur tha se ram rgvag nï phoge tsham zhï: shiŋge tsham lŋa nïs gyiŋ ku runo
55. tsham phar nï tiŋ spin gril no rm[e] sthabs nï maŋ star fiŋag mïn star fiŋag na tog phar ma sthab
56. mïn len ne yi lod nï lo dril drib min dril drab nve tog pan ga [ŋag=rŋag] mïn len ne yi lo di // ri ma zheg
57. zheg pham rbe nu gun rïŋ mïn len ne yi lod do / rï rul gyol gyal rphï skhros cï gun ran [gkhla=gakhla] min
58. lene yi lo di [sgo=sto] lŋa gva gva shaŋ thal [ryïŋs=rvïŋs] ga mu ran ni rmo ma nul nal du phyud kyer ga mu rani tsha gya
59. khan kar rnïl purd / tab ga mu ran nï thaye go mu: ran min la skams na tsis ran min mu ye spal gle glaŋ ŋe
60. ba mins ni go mu khlaŋ khlaŋg rhan rta na tse ti go mu yil tsuŋ phuŋg skar zhïb [zhib=zhig] byun byabs nve mi ryan
61. khvi rtis ri sho nutse chi go mu mi byus sa rlvas: rya sho la skams kyese: tad min pi yig chi go mu: ye: mu
62. ran ni: la skams kye ye kye khla ni / tha: ci mi ryand // khvi rtis mu: go ya gabs shu zu bras skar tsu ri chi go mu
63. go mu tur lŋa: nyas zhers zhi go mu rhïl min plan thal kar tha chi go mu ye mu rhil ni gabs kye sthal mïn
64. lo rho ma tha la ya kye ye kye sthal ni // ŋo mu ran min tiŋ gyuŋs nve tsis ran min mu ye sla gle glaŋ ŋe
65. ba mins nï lig yil go ye skar te si sa buld / ga nan de: grab kri skar se mu byun byun no kye gyvad
66. [khrïm?] shi min byun mu gluŋ glaŋ spal phaŋ pu tse ti ŋo mu mu yïl go ye skar te so: ze buld ga
67. nan de grab lïg skar se mu byun byun no kye gyvad nï yil tsuŋ mar skar ryaŋ ryaŋ byun byas
68. ti nyu tsha go ye shel [ŋla=laŋ] skar ze buld ga nan de grab gyuŋ skar ze mu byund byund no kye gyvad ni
69. ŋo mu gluŋ glaŋ braŋ [phŋa=phaŋ] khve ba tsha go ye re ba [ŋla=laŋ] skar ze buld nan de grab / rafi skar ze fiog ci mu
70. byun byun no kye gyvad nï zu zu ŋo mu spal gle glaŋ se tad go mar tï chas dus drïg min pi ŋo
71. mu ye mu ran nï rgvil mu ran mïn cï tiŋ tab tse ran min mu ye spal gle glaŋ ge ba mïns ni
72. bu yaŋ gluŋ gluŋ she ge mu bu yaŋ bya bya she ge skar re bu yaŋ gluŋ gluŋ spru yand pu tse lyu mu na
73. so kro kye na rho yu gluŋ gluŋ mar ge mu yer yu bya bya tod ge skar re rho yu gluŋ gluŋ mar

74. ?and tï tse lvi mu na so kyil kye na ma shï min gluŋ gluŋ tïŋ ge mu ye [tsha] chil bya bya
75. shel ge skare ma shï min tiŋ rhyun gluŋ gluŋ gra ?and khve rab lyo chun suŋ don dan guŋ mu
76. shugs su gru gluŋ gluŋ myur ge mu su gru bya bya myur ge skar re su gru gluŋ gluŋ myur ?ub zu tse ba
77. mu na so reb gye na / pï rgval mu zud lo di nu tsïs ran mïn mu ye spal gle glaŋe ba mïns ni
78. rma bya druŋ mu gluŋ glaŋ zhas sphaŋs [pu] [rthï=cvï] pi druŋ skar bya bya byun byab ti skhroŋs du blun blan ryva
79. na phaŋ khve zu tsïg druŋ mu thaŋ zhas kyese tad mïn pi druŋ mu se tad mïn [rtha=cva] trug ran mïn tiŋ mu
80. spal gle glaŋ ŋe zva slïg ci [zhuŋ=nyuŋ] ge bilse nyuŋ ge ryams ce mu ge to ye rho khug shurd ge ryvege ze ma
81. to tags mu run ryuŋ ze mu to ni ward mu lab tse ma sa rŋal de la khram rme ge ryege ?i di sŋal
82. tshars sa run ryuŋ / mu to ni tsva mu lab ce saŋ shïŋge run saŋ shïŋge ryege saŋ sthe mar mans
83. saŋ run ryuŋ se mu to ni tu da mu lab ce pï mu tos shiŋg pham rhyaŋ rme ge ryvege gluŋ glaŋ bag mu
84. thuŋe tsha zog ryuŋ se mu to ni guŋ gun gver gver mu srub pï mu srub tse mu sth[v]ub nï ryaŋ ryaŋ
85. kye rgyed ŋa kye rgyese kye rgyed nï guŋ gun she la mu khu zug mïn lig wer rhyelse gluŋ zho mu to ni zun
86. cï m[u] shig ce [chubs ma] ni khas ba ce [chab=chub] nï re [ram=rum] [ma-mu] nï [sas=sus] re [

Verso

1. [-] thod mu rve rumb mï
2. ci tsis ran mïn rva sa la bar spogs se ba mïns nï ŋo [rhil] zher zher tso tsvad pu ŋo
3. rvaŋ yen yen rha ryams ti ŋo rum byo bya rbe zhegs khve / ngo ti min mar braŋ brugs zu
4. [sï] tsig ŋo rva sa ye la bar spogse tad mïn ŋa yïg ŋe rva sa laŋ la bar spogs
5. [se]s thva slig cï che tsis ran mïn sa ye la bar spogse ba mins ni ri liŋ mer [zhe?]
6. pham thïŋg pu pho zur shvin shvin dad ryams ti ŋags dïl dal re zhegs khve rkhu

7. d[ï]l dal la skrags zu ri ma tha ye skrags po mïd tsïs tad mi zu tha rï liŋ pa
8. [-iŋg] myer myer nï sa ŋa yo rvam mïn rva sa la bar spogs na lodo //

References

Beckwith, Christopher I. 1987. *The Tibetan empire in Central Asia: a history of the struggle for great power among Tibetans, Turks, Arabs, and Chinese during the Early Middle Ages*. Princeton: Princeton University Press.

Chang, Kun 1960. On Zhang Zhung. *Bulletin of the Institute of History and Philology, Academia Sinica*, extra volume no. 4 (= Studies presented to Tung Tso Pin on his sixty-fifth birthday), part 1: 137-154.

A. H. Francke 1927. Ein Dokument aus Turfan in tibetischer Schrift, aber unbekannter Sprache. *Sitzungsberichte der preussischen Akademie der Wissenschaften, phil.-hist. Klasse*, 124-130 + 1 taf.

Haarh, Erik 1968. The Zhang-zhung language. *Acta Jutlandica*, 40.1: 7-43.

Hoffmann, Helmut 1967. Żaṅ-źuṅ, the holy language of the Tibetan Bonpo." *ZDMG*, vol. 117, no. 2: 376-381.

—— 1972. Several Żaṅ-źuṅ etymologies. *Oriens Extremus*, 19.1-2: 193-201.

Hummel, Siegbert 1974-1983. Materialien zu einem Wörterbuch der Żaṅ-žuṅ-Sprache." *Monumenta Serica* 31 (1974-1975): 488-520; 32 (1976): 320-336; 35 (1981-1983): 305-308.

—— 1986. Der Ursprung der Sprache von Zhang-zhung. *Journal of the Tibet Society*, 6: 3-16.

—— 1995. Neues Material zur Sprache von Zhang-zhung. *Acta Orientalia*, 59: 162-168.

Kvaerne, Per 1995. *The Bon religion of Tibet*. London: Serindia Publications.

Lalou, Marcelle 1939. *Inventaire de manuscripts tibétains de Touen-houang*, 3 vols. Paris.

Martin, Dan 1997. *Zhang-zhung dictionary*. http://www.comet.net/ligmincha/html/zzdict1.html; 1998 edition on computer diskette.

—— 1999. Zhang-zhung encodement and calque translations. Paper presented at the symposium 'New Horizons in Bon Studies'.

Nishida, Tatsuo 1982. *Ajia no mikaidoku-moji* ['Undeciphered Writings of Asia']. Tokyo: Taishûkan .

Orofino, Giacomella 1990. The state of the art in the study on the Zhang zhung language. *Annali Istituto Universitario Orientale*, vol. 50, fasc 1: 83-85.

Sims-Williams, Nicholas et James Hamilton 1990. *Documents turco-sogdiens du IX^{e}-X^{e} siècle de Touen-houang (Corpus inscriptionum iranicarum*, Part 2, Vol. III / 3). London: School of Oriental and African Studies.

de la Vallée Poussin, Louis 1962. *Catalogue of the Tibetan manuscripts from Tunhuang in the India Office Library*. London: Oxford University Press.

Stein, Rolf A. 1971. La langue Żaṅ-źuṅ du Bon organisée. *BEFEO*, 58: 231-254.

Takeuchi, Nagano, and Ueda 2001. Preliminary analysis of the Old Zhangzhung language and manuscripts. *In* Y. Nagano and R.J. LaPolla (eds.), *New research on Zhangzhung and related Himalayan languages* (=Bon Studies 3: Senri Ethnological Reports 19). Osaka: National Museum of Ethnology, 45-96 + 18 pls.

Thomas, Frederick William 1926. Two languages from Central Asia. *JRAS*: 505-507.
—— 1933. The Zhang zhung language. *JRAS*: 405-410.
—— 1948. *Nam: an ancient language of the Sino-Tibetan borderland.* London: Oxford University Press.
—— 1967. The Zhang zhung language. *Asia Major* 13: 211-217.
Yoshida, Yutaka forthcoming. Sogdian. *In* G. Windfuhr (ed.), *The Iranian languages.* London: Curzon Press.

This page intentionally left blank

SOME CHARACTERISTICS OF THE TIBETO-BURMAN STOCK OF EARLY CLASSICAL NEWARI

KASHINATH TAMOT (TRIBHUVAN UNIVERSITY)

INTRODUCTION

Since the establishment of the Nepal Bhasa (Newari) Dictionary Committee (NBDC) on January 26, 1980, work on some classical Newari texts has commenced, and remarkable data has been collected. A concordance file of the Newari *Amarakośa* has been posted on the Internet (Alsop et. al., 1996), and a Classical Newari dictionary from about 40 source manuscripts has been published (NBDC 2000). However, aside from preliminary research by several native linguists, no analysis of this data has yet been made. There is much valuable data in Classical Newari for students of Tibeto-Burman languages to unearth.

The term 'Classical Newari' (CN) was given by Hans Jorgensen (1936b: 3) to denote "the language of the MSS." It is equivalent to 'Old Newari'. Jorgensen's Classical Newari covers the period of 1675-1859 AD. His earliest source, the Hitopadesa of Nepal Samvat (henceforth NS) 481 (AD 1361) is now considered to be a manuscript not older than the sixteenth century. The reading of NS 481 was checked and while examining the script closely was found to be NS 691 (Sthapit 1999).

Classical or Old Newari was previously considered a unified language. While compiling the Classical Newari Lexicon (1983-1989) from eleven *Amarakośa* manuscripts covering a period between 1381 and 1711, with approximately fifty-year intervals between each manuscript (A1 to A8), this writer realized that there were at least two stages of Classical Newari, i.e., Early and Late (Tamot 1992). This is approximately equivalent to the division of Nepal's history into Early (879-1482) and Late (1482-1768) Medieval periods. After the Gorkha conquest of the Nepal Mandal, the Modern Period of Nepalese history begins. We thus can follow the Newari language in transition from Old to Modern also. We see this transition in Kirkpatrick's (1811: 221-249) Newari vocabulary of

some 500 words collected in 1793. Jorgensen's 1936 *Dictionary* extends from the Late Classical Period (1675-1768) to the Early Modern Period (1768-1859). It does not cover the Early Classical Period, though there are interesting data in it. This paper presents some characteristics of the TB stock of Newari in the earlier period.

In an article on the location of the Licchavi capital of Nepal, Michael Witzel (1980: 326) noted the lack of our knowledge of Early Classical Newari was a gap in identifying the non-Sanskritic vocabulary of ancient (Licchavi) inscriptions of Nepal: "The syllabary structure of the many names of localities and persons mentioned in the Licchavi inscriptions does not conform with that of early Newari, but we do not know, of course, if the structure of Newari had changed considerably or not between the 8th and 14th centuries."

1. *Source materials*

There are some inscriptions, many palmleaf deeds, and several manuscripts which are very important for the study of Early Classical Newari. They were written on an Indo-Aryan model, and have heavily used Indo-Aryan vocabulary. Newari was little used in Sanskrit texts, and translations were mostly provided to help Newari people get the message. Although Newari has since developed as a language, using its verbs in sentences from the eleventh century on (Kölver and Sâkya 1985: 106 and 190), we find complete texts in Newari only from the twelfth century. The palmleaf deed of NS 235 (AD 1114) from Ukûbâhâl, Patan, is the most important document of Early Classical Newari (Malla 1990). There are a few inscriptions, and a few more palmleaf documents, from this period in which Newari is used. Mahes Raj Pant (1990 and 1993-1996) has published many such documents which are important for this study. Following one-leaf documents or inscriptions, and palmleaf deeds, we find the first Newari book, the *Haramekhalâ* of NS 494 (AD 1374). In the Newari of the Early Medieval Period of Nepalese history, the Early Classical Newari Period, we find only technical literature, or *vârṁaya*. Purely literary works appear only with the *Bhâgavata-purâṇa* of NS 625 (A.D. 1505).

The first Newari book is a translation from Sanskrit. Almost all books of this period are either translations, commentaries, abridgments, or free recensions from that language. The *Gopâlarâjavaṁśâvalî* (ca. NS 509/AD 1389) is the only text composed in Newari. It

has two parts: the *Vaṃśâvalî* (Chronicle) and the *Vrtânta* (Annals). The second part is composed freely in Newari. The following are some important texts in Newari composed up to the fourteenth century on the subject of medicine, law, lexicography, history, and astrology:

Haramekhalâ (NS 494/AD 1374), a Tantric medical treatise
Nâradasmṛti (NS 500/AD 1380), an Aryan law book, with a commentary by Manika
Amarakôśa (NS 501/AD 1381), an abridged commentary on a Sanskrit lexicon (A1 of Alsop, et al., 1996; Tamot 1983b)
Amarakôśa (NS 506/AD 1386), the *Balabodhinî*, a commentary on the Sanskrit lexicon (A2 of Alsop, et al., 1996)
Gopâlarâjavaṃśâvalî (ca. NS 509/AD 1389), a Sanskrit-Newari chronicle of Nepalese history
Daśâphala (NS 519/AD 1399), an astrological textbook
Auṣadhasarvasva (NS 520/AD 1400), a general medical book

Once a book was written, it was copied over the years, and this copying became a tradition. Hence we find the *Nâradasmṛti* (deposited in the National Archives of Nepal) of NS 500 copied in NS 527 (Cambridge University), 631 (Kaiser Library), 670 (National Archives), 714 (private copy in Bhaktapur), 721 (National Archives), and 820 (British Library in London, a source for Jorgensen 1936b).

2. *The Relationship of Tibetan to Newari*

Nepal was the principal trade center on the route between India and Tibet in ancient times, and Nepal has had regular contact with Tibet since the sixth century. Within the Tibeto-Burman family, Newari is probably the living language with the third oldest written tradition, the first being Tibetan and the second Burmese. The Nepal-German Manuscript Preservation Project has microfilmed some bilingual documents in Tibetan and Newari. And, recently, a Tibetan-Newari lexicon and phrase book, which had been brought from Lhasa, was discovered in New Delhi. The manuscript, which was perhaps composed in the early twentieth century, is written in Tibetan script, and contains modern colloquial Newari in Tibetan transcription (Cüppers, et al. 1996).

Only a few Newari words have been used in the comparative study of Tibeto-Burman languages. Shafer (1952) compared some 100 Newari words in his paper, "Newari and Sino-Tibetan," but Benedict used fewer Newari words in his book *Sino-Tibetan: a*

conspectus (1972). Glover (1970) found only 22% of Modern Newari words to be cognate with Tibetan when using a Swadesh list. We find about 40% Newari cognates with Tibetan when we compare Medieval Tibetan and Newari. The cognates in Table I below have been collected mainly from (Benedict 1972) for Tibetan and Tibeto-Burman and from Alsop, et al. (1996) for Early Classical Newari.

English	Early Classical Newari	Tibetan	Tibeto-Burman
bird (20)	kha 'fowl, bird' (bound form)	bya	*rak
breast (51) (milk)	ŋe-mo 'suckling'	nu-ma	*nuw
cold (94)	khoŋu	graŋ-ba	*kyam
die (61)	si-ca/ya	śi-ba /fitśi-ba	*siy
dog (21)	khi-câ	khyi	*kwiy
drink (54)	toṃn~e (< *ṭoṃ-ja)	fithuŋ-ba	-
dry (99)	gaṃ-n~e (< gaṃ-ja)	skam-po	*kaŋ
earth (79)	câ 'soil'	sa	-
fingernail (45)	lo-si	sen-mo	-
fish (19)	ŋa	ña	*ŋya
flesh (29)	lâ	śa	*sya
fly (64)	bo-ya	fiphur-wa	*pur/byer
grease (32)	sau 'oil'	tsho-ba	*sa.w
green (88)	vâŋu	sŋon-po, ljaŋ-gu	*ŋow
hand (arm) (48)	lâ	lag-pa	*lak/g-lak
hear (58)	ŋeṃ-ja/ñe tâva 'heard'	snyan-pa	*r-rgya
heart (52)	luṃ-gvaḍa	snyiŋ	*m-luŋ/s-niŋ
kill (62)	syâ-ca/ya	gsod-pa	*g-sat
knee (47)	pu-lṛ/pu-le	pus-mo	*put
know (59)	se-ye (< *se-ca)	śes-pa	*syey
leaf (25)	lap(a)te	lo-ma	*lap/pak
liver (53)	seṃ	mtśin	*m-sin
louse (22)	si	śig	*s-rik
man (17)	mi(ṃ) (LCN mi-jana)	mi	*r-mi(y)
moon (73)	la-ni	zla-ba	*s-la
night (92)	câ	źag 'day'	*ryak/ya
nose (41)	nhasa	sna	*s-na
not (8)	ma-	ma/mi	*ma

one (11)	chi	tsik	*t(y)ik
person (18)	mi	mi	*r-mi(y)
seed (24)	pu	fibru/sa-bon	*mruw
sleep (60)	mha-ŋa 'dream' nhiṃ-ḍa 'sleep'	rmi-ba nyal-ba	*mwiy *mye
sun, day (72)	ni(ṃ) 'sun, nhi 'day'/-nhu (bound form)	nyi-ma	*niy
tongue (44)	me (me-ca LCN)	ltśe	*m-lay/s-lay
tooth (43)	vâ	so	*s-wa
tree	siṃ	śiŋ	*siŋ
two (12)	ne, ni (LCN nase, nasi)	nyi	*g-nis
water (75)	ti (LCN) 'juice'	chu	*ti(y)
who (6)	su	su	-
woman (16)	mi-sa	mo/skyes-dman	*mow

Table I

4. *Some characteristics*

If we look closely at some Early Classical Newari words, we find some of the characteristics relating to that time. Among others, we see that the lack of both animacy concord and a classifier system appear to be characteristics of that time. These are part of the grammatical system of the later Newari language.

4.1. *Use of Tibeto-Burman words*

We have some 250 non-Sanskritic nominals that were used in the Licchavi inscriptions of the Ancient Period (A.D. 185-879) of Nepalese history. These are assumed to be proto-Newari words. Some are, certainly, Tibeto-Burman words, e.g., *siṃ* 'wood'; *co*, 'peak', *guṃ* 'hillock', *ku/khu* 'river', *cho* 'wheat' (Tamot 1980; Malla, 1981), *mhum* (ECN *mula*) 'sesame' (Tamot 1989), etc.

When Newari texts began to include verb forms in the eleventh century, words of Tibeto-Burman stock began to be used in inscriptions and palmleaf deeds. While we find only a few words in the twelfth to thirteenth centuries, by the fourteenth to fifteenth centuries Tibeto-Burman Newari words began to be used widely in documents and manuscripts, and the Tibeto-Burman stock of Early Classical Newari can be recorded for this period. Some of the Tibeto-Burman Newari words were later replaced by Indo-Aryan borrowings, affixed

with other forms, or expanded by synonymous compounding with Indo-Aryan words. Examples of borrowing are given in examples (1) through (7), of affixation in examples (8) through (16), and of synonymous compounding with Indo-Aryan borrowings in examples (17) through (19).

(1) *mi* 'man' replaced with *manukha* (Skt *manusya*)
(2) *niṃ* 'sun' replaced with *suraja* (Skt *surya*)
(3) *sau* 'oil' replaced with *cikkana* (Skt *cikvana*)
(4) *mhaca(mo)* 'wife' replaced with *kalâta* (Skt *kalatra*)
(5) *kemmo* 'husband' replaced with *bhâlata* (Skt *bhartr*)
(6) *phupa* 'elder brother' replaced with *dadâ* (Mundaric *dada*)
(7) *dhuṃgva* 'finished' replaced with *siddho* (Skt *siddha*)
(8) *khâ* 'door' affixed as *lva-khâ*
(9) *lo* 'quarrel' affixed as *lva-pu*
(10) *mo* 'child' affixed as *mo-câ*
(11) *su* 'cloud' affixed as *su-pâca*
(12) *nhasa* 'ear' affixed as *nhasa-pata*
(13) *aṭa* 'brick' affixed as *aṭa-pa*
(14) *si* 'lips' affixed as *mhuthu-si*
(15) *lhâye* 'hand over' affixed as *lava-lhâye*
(16) *poye* 'cover' affixed as *tvaka-puye*
(17) *lâ* 'hand' synonymized as *lâ-hâtha* (Skt *hasta*)
(18) *ku* 'father-in-law' synonymized as *ku-bâpa* (Skt *vaptr*)
(19) *lṛ/lai* 'foot' synonymized as *pata-lai* (Skt *pada*)

4.2. *Loss of Consonant Clusters*

In the non-Sanskritic vocabulary of the Licchavi inscriptions there are several words with consonant clusters, which are also found in Tibetan and other Himalayan languages. In these texts we find *priŋ* 'locality', which developed into *brumâ* (as in Lalitabrumâ) in early medieval city names and as *pi* (mhaypi) and *pa* (khopa) in particular names of places. We find *bru* or *bu* 'field' in many place names (Malla 1981), etc. But we also note that some such clusters have been lost in some words even in the Licchavi period, as *cho* < *chro 'wheat' (cf. Literary Tibetan *gro* 'wheat' in Nagano 1982:48) and *khu* < *kruŋ 'river' (cf. Tibetan *kluŋ* in Benedict 1972:39, *luŋ* in Cuppers, et al. 1996:19).

Sometimes insertion of vowels occurs within the cluster, as *barâ* 'arrow' from Tibeto-Burman *bla. If we compare Tibeto-Burman cognates, we find clusters in Proto-Newari and the lack of them in Early Classical Newari, as shown in Table II.

ECN	PNEW	OTIB	PTB	ENGLISH
pi/pe/pai	*pri/pre	bźi	*b-liy	four
khu	*khruk	drug	*d-ruk	six
cyâ	*cryat	brgyad	*b-r-gyat	eight
me	*mre	ltśe	*m-lay/g-lay	tongue
bi	*brul	sbrul	*b-ru.l	snake
vâ	*sva	so	*s-wa	tooth
saṃ	*sra	skra	*s-kra	hair

Table II[1]

4.3. *Open syllabicity*

We also find some closed syllable words in the non-Sanskrit vocabulary of the Licchavi inscriptions, for example, *guŋ* 'hillock', *mhum* 'sesame', *gval* 'place of a divinity; houses', *kuther* 'revenue office', and so on. If we compare Newari words with their Tibeto-Burman cognates, we find that Newari must have had closed syllable proto-forms. These were changed into open syllables in the Early Classical Newari period. This open syllabicity has caused some Early Classical Newari words to become homonymous, as the examples in Table III show.

ECN	PNEW	OTIB	PTB	ENGLISH
mi	*mi(y)	myi	*rmi(y)	man (homo)
mi	*mik	myig	*mik/myak	eye
mi (also me/mai)	*mi(y)	mye	*mey	fire
si-(ca/le)	*sil	bsil-ba	*(m)s(y)il	wash
si-(ca/ye)	*sis	śes-pa	*syey	know
si-(ca/ye)	*si(y)	śi-ba	*siy	die

Table III

We also have some examples of open syllabic Early Classical Newari words which we may compare with their Tibeto-Burman cognates, such as the numerals in Table IV and general vocabulary in Table V.

We see from the above that the open syllabic character of Early Classical Newari caused the loss of Tibeto-Burman affixes and final consonants, and vowel insertion also occurred, making the syllable open, as in *mi* or *mikha* from TB *mik, *ni/ne/nai* or *nasi/nase* from

[1] See Benedict (1972) and Tamot (1980, 1998).

TB *g-nis, the metathesized form *nhasa/nhâsa* from TB *s-na, and so on.

ECN	PNEW	OTIB	PTB	ENGLISH
chi	*chik	gtśig	*t(y)ik	one
ni/ne-/nai	*nis	gnyis	*g-nis	two
svaṃ	*svam	gsum	*g-sum	three
nhasa	*nhas	*snyis	*s-nis	seven
ŋaiyu[2]				seventy

Table IV

ECN	PNEW	OTIB	PTB	ENGLISH
mi/mikhâ	*mik	myig	*mik/myak	eye
cheṃ	*chim/chem	khyim	*kim	house
nhi	*n(h)ip	snabs	*s-nap	snot
nhasa/nhâsa	*n(h)as	sna	*s-na	nose
phâ	*phâk	phag	*pak/pwak	pig
lâ	*lâk	lag-pa	*lak	hand

Table V

4.4. *Polysystemic Phonology*

While studying modern Kathmandu and Bhaktapur Newari side by side, R.K. Sprigg (1983) realized that Newari was "a language without a vowel system." He concluded that the basis of Newari orthography is more prosodic than phonemic. His conclusion was based on J.R. Firth's (1890-1960) prosodic analysis, which proposes a polysystemic phonology, as opposed to the essentially monosystemic nature of phonemic analysis.

Early Classical Newari exhibits polysystemic phonology. We find different forms for the very same word, such as *mi/me/mai* 'fire', *thusâ/thosâ/thausâ* 'bull', and *ŋamha/ŋâmha* 'five persons', etc. These may be classed into three groups:

i. A palatal vowel group – i, e, ai
ii. A labial vowel group – u, o, au
iii. A velar vowel group – â, a

[2] A.D. 1114 (Tamot 1998).

Data demonstrating the prosodic (polysystemic) phonology of the palatal vowel group (*i, e, ai*) are given in examples (20) through (28).

(20) *ni/ne/nai* 'two'
(21) *mi/me/mai* 'five'
(22) *pi/pe/pai-tâ* 'four types'
(23) *lṛ/le/lai* 'foot'
(24) *thithi/thethe/thaithai* 'each other'
(25) *ci/ce/cai-ye* 'tie, bind'
(26) *to-phi/phe/phai* 'broom'
(27) *du-si/se/sai* 'millet'
(28) *thu-ti/te/tai* 'these'

Examples showing polysystemic phonology of the labial group (*u, o, au*) are given in (29) through (38).

(29) *thu/tho/thau-sâ* 'bull'
(30) *ku/ko/kau-nhu* 'particular day'
(31) *dâ-thu/tho/thau* 'middle'
(32) *ju/jo/jau-ye* 'go'
(33) *pu/po/pau-le* 'pay'
(34) *thu/tho/thau-lva* 'owner'
(35) *tha-ku/ko/kau* 'hard'
(36) *du/do/dau* 'is'
(37) *ma-khu/kho/khau* 'not'
(38) *pu/po/pau-lṛ* 'knee'

Examples showing polysystemic phonology of the velar group (*â, a*) are given in (39) through (48).

(39) *ŋâ/ŋa-mha* 'five persons'
(40) *ŋha/ŋha-tho/thau* 'former'
(41) *câ* 'night'
(42) *ca-chi* 'the whole night'
(43) *lâ* 'month'
(44) *la-chi* 'one month'
(45) *bâ* 'half'
(46) *ba-chi* 'one half'
(47) *me lâ* 'moon-light'
(48) *la-ni* 'moon'

Low front *â* is one of the main vowels of Tibeto-Burman. It is generally changed to the neutral vowel *a* in a front environment. We do not know how and why the main high vowels (*i, u*) differ, and in which environments, from the mid-vowels (*e, o*) and diphthongs (*ai, au*).

4.5. *Use of verb formatives*

It has long been discussed whether Newari has closed syllabic verbal roots, or open syllabic only. Conrady (1891) established the theory of the closed syllable based on the evidence of *wan* 'go', *dat* 'exist' and *jur* 'be'. Jorgensen (1936) developed the theory and divided the endings of the root into the classes I (n), II (t), III (l, unstable), and IV (l, stable). Some scholars have followed his schematization: Hale (1973), Kölver and Kölver (1978), Shresthavarya (1981), Kansakar (1982), and so on. Hargreaves and Tamot (1985:10) tried to make a generalization regarding Newari verb classes and Proto-Tibeto-Burman final consonants, proposing at least two classes:

A. If a Proto-Tibeto-Burman root that had a final nasal is found in Newari, it will be a Class I verb;
B. If a Proto-Tibeto-Burman root that had a final *-t or *-k is found in Newari, it will be a Class II verb.

On the other hand, native scholars believe strongly that Newari verbal roots are basically open syllabic. Sâgar (1952), Tamot (1990), and Sharma (1999) in particular have supported this theory. The Newars have a system of classifying verbs by the various formative suffixes: *-ye, -ne, -le, -pe, -te,* and *-ke*. Here, *-ye* corresponds to Jorgensen's Classes II and III, while *-ne* is Class I and *-le* is Class IV. The others are compound verbal suffixes.

Newari verbs are cited in their infinitive forms in dictionaries. While composing *Amarakośa* No. 2 (A2) in 1985 for the Nepal Bhasa Dictionary Committee, this writer realized that there are varieties of infinitive forms which repeat the same glosses (Tamot 1983). We have *bvaṃja, bvaṃñe* and *bvane* 'invite', and *yaca/yâca, yâya,* and *yâye* 'to do', and so on. When we examine the Early Classical Newari verb, we find that its root is basically open syllabic. Newari verb roots could not be used without formatives; the root form itself is equivalent to a nominal form. We mentioned above that closed syllabic proto-Newari words changed into open syllables in the Early Classical Newari period. In the course of time, Class I (-n) and Class IV (-l) verbs appeared as with consonantal endings, and the present morphophonemic system of modern Newari appeared. But what scholars thought were closed syllabic stem finals were in fact formatives. We should not forget that, like Tibeto-Burman in general, Newari is not basically an inflectional language; rather, it is an af-

fixing language. Formatives, or affixation, enable Newari to work as if through inflection.

Table VII shows a group of Early Classical Newari verb formatives, which changed in the later period.

Infinitive/Citation	ca/ja	ya/ne	ye/ne
Perfective/Future	cu/ju	yu/ñu	yi/ni
Infinitive of Purpose	ta/da/ra	ta/ña/la	ta/na/la

Table VII

On the basis of the forms *ta/da* of the infinitive of purpose, it is believed that *ta/da* should be the original formative suffixes from which the infinitives *ca/ja* developed (Tamot 1989:20).

In a recent seminar organized by the Central Department of Nepalbhasha (Newari), Tribhuvan University, native linguists and scholars realized that Newari verb roots are basically open syllabic, but create stems with formatives and are inflected in finite forms. Thus, in *pune* 'to wear' *pu˜* is the root, *pun-* is the stem, and *-e* is the non-past conjunct morphophoneme. Similarly, for *sile* 'to wash', *si* is the root, *sil-* is the stem, and *-e* is the non-past conjunct morphophoneme (Shrestha 2000). This differentiation between the root and the stem explains the development of Newari verb forms.

5. Conclusion

Scholars have not until now understood that Classical Newari has two stages of development. Some of them have rejected this idea. However, this writer has been floating the idea around since 1990 (Tamot 1992). In order to develop it, we have tried to discover some of the characteristics of Early Classical Newari. This is a preliminary effort to distinguish two stages of Classical Newari. There is also data for this on the internet (Alsop, et al., 1996) and in the new Dictionary of Classical Newari (NBDC 2000). However, this approach is currently limited to only a few scholars. Texts of the relevant period also have not been critically studied and published, and only a few of these are available.

Though Jorgenson has published a *Classical Newari Dictionary* (1936) and edited some old texts, his work does not cover the Newari language of the Early Classical period. One of the problems of ana-

lyzing data from later texts is that there may be residue from older periods due to the copying tradition. This creates confusion in identifying contemporary characteristics. Jorgensen (1936b) and *Nâradasmṛti* (NS820) bear features of the original text of NS 500 (AD 1380) from the copying tradition. Western scholars have not, to this point, shown an interest in editing texts and analyzing the language of Early Classical Newari. It is hoped that they will do this after data from the Lexicon becomes widely available.

As medieval Tibeto-Burman literary languages have not been much studied, the reconstruction of Proto-Tibeto-Burman has been done through modern Tibeto-Burman cognates. Sometimes this can be deceptive. Hence, we should now emphasize the study of medieval Tibeto-Burman texts so that a real reconstruction of Proto-Tibeto-Burman can be made. Such a comparative study would be a great tool for identifying the characteristics of a particular language as well as of its family. Let us now proceed to establish Tibeto-Burman historical linguistics as its own scientific field of inquiry.

References

Alsop, Ian, Kashinath Tamot, and Mimi Church 1996. Newari lexicon. *Asian Arts:* http://www2.pair.com/webart/mysqllex.

Benedict, Paul K. 1972. *Sino-Tibetan: a conspectus.* Cambridge: Cambridge University Press.

Conrady, August 1891. Das Newari: Grammatik und Sprachproben. *ZDMG* 45: 1-35.

Cüppers, Christoph, Kashinath Tamot and Philip Pierce 1996. *A Tibetan-Newari lexicon cum phrase book.* Nepalica 10. Bonn: VGH Wissenschaftsverlag.

Glover, Warren W. 1970. Cognate counts via the Swadesh list in some Tibeto-Burman languages of Nepal. *Occasional papers of the Wolfenden Society on Tibeto-Burman linguistics,* Vol. III: 23-26.

Hale, Austin 1973. On the form of verbal bases in Newari. *Issues in Linguistics* (papers in honor of Henry and Renee Kahane). Chicago: University of Illinois Press: 279-299.

Hargreaves, Henry, and Kashinath Tamot 1985. Notes on the history of some Newari verbs: preliminary evidence. Paper presented at the 18th ICSTLL, August 27-29, Bangkok, 17 pp. (mimeograph).

Jorgensen, Hans 1936a. Linguistic remarks on the verb in Newari. *Acta Orientalia* 14: 180-185.

—— 1936b. *A dictionary of Classical Newari.* Copenhagen: Levin and Munksgaard.

—— 1941. *A grammar of Classical Newari.* Copenhagen: Levin and Munksgaard.

Kansakar, Tej R. 1982. Morphophonemics of the Newari verb. *Occasional Papers in Nepalese linguistics.* Kathmandu: Linguistic Society of Nepal: 12-29.

Kirkpatrick, Colonel. 1811. *An account of the kingdom of Nepal.* London: William Miller.

Kölver, Ulrike, and Berhnard Kölver 1978. Classical Newari verbal morphology. *Zentralasiatische Studien* 12: 273-316.

Malla, Kamal P. 1981. Linguistic archaeology of the Nepal Valley: a preliminary Report. *Kailash* 8: 5-20.

—— 1990. The earliest dated document in Newari: the palmleaf from Uku Bahal, NS 235 (AD 1114). *Kailash* 16: 1-2, 15-25.

Nagano, Yasuhiko 1982. A preliminary report of the three Tibetan dialects in the northern Gandaki Valley. Bista, et al., *Anthropological and linguistic studies of the Gandaki area in Nepal.* Monumenta Serindica, no. 10. Tokyo: ISLCAA: 81-157.

Nepal Bhasa Dictionary Committee 1987. *Lexicon of the Classical Newari drawn from traditional kosa sources.* (Computer printout)

—— 2000. A dictionary of Classical Newari compiled from manuscript sources. Kathmandu: Cwasâ pâsâ.

Pant, Mahes Raj 1990. Tâḍapattramâ lekhiekâ, ahilesamme prakâsama naâekâ vikramako caudhauṃ satâbdîkâ 15 vatâ likhatapattra. *Pûrnimâ* 77 (VS 2047 Asvina): 1-14.

—— 1993-1996. Nevârarâjya ra Brahman. *Pûrnimâ* 84 (VS 2050 Vaishâkha): 49-62 and 90 (VS 2050 Caitra): 1-53.

Sâgar, Puspa Ratna 1952. *Subodh Nepâlbhâsâ vyâkaraṇa.* Kathmandu: Thauṃkanhe Prakâshan, NS 1072.

Shafer, Robert 1952. Newari and Sino-Tibetan. *Studia linguistica* 6: 92-109.

Sharma, Ramâpati Raj 1999. Nepâlabhâsâyâ dhâtu prakârarthaka taṃsâgvah. *Pâsukâ* 3.5: 8-9 and 3.6: 3-5.

Shrestha, Iswarananda 1981. *Newari root verbs.* Kathmandu: Ratna Pustak Bhandar.

Shrestha. Omkâreshvar 2000: Nepâlabhâsâyâ dhâtu kvaḥchîgu samasyâ: Chagû adhyayan. *Nepâlamaṇḍala Bvaṃgmâḥ* 2 (NS 11120 Caulâ): 1-24.

Sprigg, R.K. 1983. Newari as a language without vowel systems: a Firthian approach to the Bhaktapur and Kathmandu Newari verb. *Nepalese linguistics* 2: 1-22.

Sthapit, Jvala (ed.) 1999. *Hitopadaśa (N.S. (6?)91).* Yala (Lalitpur): Nepal Mandala Ghaḥcâḥdhukû, NS 1119.

Tamot, Kashinath 1980. Licchavi abhilekhayâ Saṃskṛta makhugu khaṃgvaḥyâ lyâḥcâḥ. *Kheluita* 2 (NS 1100): 131-136.

—— 1983. Kriyâ kriyârthakay dutinegu machinâ. Paper presented at the occasional meeting of the NBDC. (Typescript).

—— 1983b. *Putrapautrâdibodhinî: Amarakosayâ Nepalabhâsâ tîkâ, NS 501.* Kathmandu: Pâsâmunâ, NS 1103.

—— 1988. Notes on Proto-Newari numeral. Paper presented at the 9th Annual Conference of LSN, November 26-27, Kathmandu. (Handwritten manuscript).

—— 1989. Nepalbhâsâyâ kriyarthaka pratyaya ye, ne le yâ vyutpatti. Paper presented at the symposium of Classical Newari dictionary making. (Mimeographed.)

—— 1989. Etymological notes on the words of ritual seeds of an ancient inscription. Paper presented at the 10th annual conference of the Linguistic Society of Nepal, November 26-27. (Handwritten manuscript.)

—— 1990. A proposed model of Classical Newari verb stem formatives. Paper presented at a weekly meeting for creating a model for verb citation in the Classical Newari dictionary, held May 22nd.

—— 1992. *Pulâṃgu Nepâlbhâsâ nithî duguyâ chumâdhâra.* Yala (Lalitpur): Kulâṃ Sâhitya Pâḥlaḥ, NS 1113.

—— 1998. Prâkṛtakâlina Nepâlabhâśây chapulu. Abridged by Yogesh Raj. *Pasuka* 2.11 (NS 1118 Dillâ): 8-9 and 2.12 (Gumlâ): 8-9.

Vajracarya, Dhanavajra and K.P. Malla (ed.) 1985. *The Gopâlarâjavaṃśâvalî*. Wiesbaden: Franz Steiner Verlag.

Vajracarya, Shantaharsh (ed.) 1987. *Nyâyavikâsinî*. Kathmandu: Nepalbhasa Parishad, NS 1107 (in Newari).

Witzel, Michael 1980. On the location of the Licchavi capital of Nepal. *Studien zur Indologie und Iranistik* 5-6: 311-337.

TWO PYU-TIBETAN ISOGLOSSES

CHRISTOPHER I. BECKWITH (INDIANA UNIVERSITY)

Pyu, the language of the ancient Pyu Kingdom of Burma (Stargardt 1990), is attested in texts dating from Late Antiquity into the High Middle Ages. It has a strictly monosyllabic, open-syllable word structure, with most syllables having a simple C or CC onset, contrasting with the often highly complex Old Tibetan onset. This paper examines two lexemes in Pyu, Tibetan, and other early Tibeto-Burman languages in the light of correspondences in Chinese, with the aim of clarifying the relationships among the early Tibeto-Burman daughter languages.

I. WATER

Pyu *tdû* and Old Tibetan *cu* are the only early Tibeto-Burman forms of the word for 'water' with rounded vowels. In fact, since the Pyu digraph initial *td-* apparently represents an affricate—whether [ʧ] or [ʤ] is uncertain—the two words are virtually identical.

In *Sino-Tibetan: A Conspectus* (Benedict 1972), still the most ambitious published attempt to reconstruct Proto-Tibeto-Burman, not one but three forms are given for the Proto-Tibeto-Burman word for 'water', as shown in examples (1) to (3). These reconstructions are based on a number of modern languages, and reconstructed Chinese, without reference to any early Tibeto-Burman language.[1] The three different forms were created to satisfy the needs of arguments being made at each location in the book, not to explain the data on 'water'.

(1) PTB *ti(y) 'water' (Benedict 1972: 26)
(2) PTB *twiy 'water' (Benedict 1972: 45)[2]
(3) PTB *twəy 'water' (Benedict 1972: 169)

---

[1] On the methodological problems with this approach see the Introduction to this volume; see also the remarks by David Bradley (in this volume).

[2] Inspection of the data cited to support this reconstruction shows that the word means, simply, 'egg'. Benedict himself notes (1972:45, n. 149), "Dhimal distinguishes between *tui* 'egg' (TB **twiy*) and *tsi* (Toto *ti*) 'water'..."

The Pyu and Old Tibetan data do not contradict the reconstruction of a dental initial and high nucleus vowel, while the Late Zhangzhung and Tangut words for 'water' and the Newari word for 'juice' support it, so we can follow Benedict's reconstruction in (1) in a first attempt at a revised reconstruction, in (4) and (5).[3]

(4) PTB *tî (or *tiy) 'water' > LZZ *tiŋ* ~ *ti* 'id.', LCN *ti* 'juice', Tgt *gzi* 'water'

(5) PTB *tî (or *tiy) 'water' > Pyu *tdû* [ʧu:] (or perhaps [ʤu:]) 'id.', OTib *cu* [ʧu] 'id.', *bcud* 'juice, sap' (< √*cu*)

However, Pyu and Old Tibetan, representing two branches of Tibeto-Burman (Bradley, in this volume) presume an underlying form with a labial off-glide or coda. Benedict obviously noticed the labial element, and included it in his second and third reconstructions, in (2) and (3). However, his reconstruction in (1) would have resulted in Old Tibetan *ci, while those in (2) and (3) would have resulted in *ci and *ce (or perhaps *ti and *te) respectively, not *cu* [ʧu], the form actually attested in Old Tibetan and Pyu. Moreover, if the Old Tibetan honorific word for 'water', *cab* ([ʧab]), is related to the other words, the final labial element is even clearer. So, a reconstruction *tîw or *tîəw, as in example (6), is indicated by these two languages, but not by Zhangzhung, Tangut, and Newari.

(6) *tîw or *tîəw > Pyu *tdû* *[ʧu:] 'id.', OTib *cu* [ʧu] 'id.', OTib *bcud* [bʧud] < √*cu* 'juice, sap'

It is significant that the Newari word *ti* actually means 'juice'. The usual words for 'water' in Newari, Burmese, and Meithei do not seem to be related to each other or to the Tibetan and Pyu words, as shown in examples (7) through (9).

(7) NNew *la:* (*lakha*) 'water' < ONew *laṃkhva*

(8) NBur *yé* < LBur *re* < OBur *riy 'water'; NBur *yi* < OBur *rañ* 'juice, liquid'

(9) NMei *isiŋ* 'water' (perh. related to LZZ *tiŋ* 'water' < PTB **ti*, or to Tgt *gzi* 'water')

It is clear from the above data that generally speaking the Tibeto-Burman languages maintain a formal distinction between 'juice' or

[3] I am indebted to my colleagues Shobhana Chelliah, Ksenia Kepping, Kashinath Tamot, and Rudolf Yanson for some of the data on Meithei, Tangut, Newari, and Burmese respectively. Any errors are my own.

'liquid' and 'water', the former apparently being the simplex form, the latter an extended or complex form (or an unrelated word).

The Proto-Tibeto-Burman word *tî may be provisionally identified as the simplex word for 'juice' or 'liquid'. In light of the Pyu form *tdû* [ʧu:] and the Old Tibetan form *cu* [ʧu], it appears that their common early Tibeto-Burman ancestor, which we may provisionally call Proto-Tibeto-Pyu, had a complex word for 'water', as shown in examples (10) and (11) below. The problem now must be to discover if the distinction between simplex 'liquid' and complex 'water' goes back to or beyond Proto-Tibeto-Burman. In at least one case, Old Tibetan *cab* [ʧab] 'water (honorific)', there would seem to be a good possibility that the extension is an intrusive loan element, as suggested in (12).

(10) OTib *cu* 'water' and *bcud* 'juice, sap < PTib *cu [ʧu] < *tiu < PTP *tîw or *tíəβ < PTB *tî + -*w or *-əβ
(11) Pyu *tdû* < *tiu < PTP *tîw or *tíəβ < PTB *tî + -*w or *-əβ
(12) OTib *cab* [ʧab] 'water [honorific]'[4] < PTib *tiáb < PTB *tî + *ab ← PIE *ab, *ap 'water'

The external evidence in examples (13) through (17) supports the preceding line of thought. The word for 'juice' in Chinese, 汁 *zhî*, is from Middle Chinese **ʧip*,[5] from a theoretical Old Chinese *tiəp, though the word is not attested until Late Old Chinese. The Chinese word for 'water', in (14), is more problematic, but it seems to go back to the same root: 水 *shuǐ* from Early Old Chinese *tîwér. The doublet 㳄 *zhuǐ* 'water'—probably the regular reflex of *tîwér—indicates that *shuǐ* should probably be reconstructed with the same initial as *zhî* 'juice'.[6] Also relevant in this connection are the Japanese words *chi* [ʧi] 'milk' and *chi* [ʧi] 'blood' from Old Japanese **ti* 'milk, blood [i.e., body-liquid, juice']', and Old Japanese **tu* 'liquid, spit'.[7]

---

[4] Cf. OTib *caŋ* [ʧaŋ] 'liquor, beer', but this word is possibly unrelated.

[5] Middle Chinese, Old Japanese, and Old Koguryo words attested in 'phonetic' Chinese character transcriptions, are marked with a star (*); reconstructions and hypothetical forms are marked, as customary, with an asterisk (*).

[6] The irregularities in the Chinese word for 'water' and its cognate 川 *chuân* 'stream, river' remain problematic; the usual word 水 *shuǐ* is probably a dialect form. See the recent study by Sagart (1999:157-158), who gives a different explanation.

[7] The similarity of OJpn **tu* 'liquid, spit' to 吐 Mandarin *tǔ, tù* 'spit' < MChi **tʰɔ²*, **tʰɔ³* (Pul.) might suggest a loan, but **tu* appears to be a genuine Japanese word, occurring in very early compounds, including NJpn *mizu* 'water' < OJpn

(13) 汁 *zhî* 'juice' < MChi ⋆*ʧip* (Pul. 405) < LOC *tiəp < *ti-
(14) 水 *shuǐ* < MChi[8] ⋆*ʃwi*[2] < MOC dial. *ʃwêr < EOC *tîwêr < *tî + *wêr (cf. Karlgren 1957: 154 < OChi *s̬iwər)
(15) 沝 *zhuǐ* 'water' < MOC *ʧwêr < EOC *tîwêr < *ti- + *wêr (cf. Sagart 1999:158 < OChi *[b]t-lu[r]ʔ)
(16) 川 *chuân* 'stream, river' < MChi ⋆*ʧʰwian*[1] (Pul. 60). < MOC[9] *tîwer < *ti- + *wer
(17) NJpn *chi* [ʧi] < OJpn ⋆*ti* 'milk, blood' < PJpn *ti '(body-) juice, liquid'
(18) NJpn *tsu* 'spit' < OJpn ⋆*tu* 'liquid'

It may be objected that 'water' is part of the 'basic vocabulary', which is not supposed to be borrowed, but it has long been noted that the word for 'water' in many unrelated languages is the same, indicating that it is actually a common loanword.[10] Thus, contrary to

---

⋆*midu* < PJK *mir 'water' + (PJpn) *tu 'liquid' (cf. Martin 1987: 483). This suggests a common origin for both ⋆*ti* and ⋆*tu*. There is plentiful, clear, and largely ignored evidence of a close connection having once existed between Chinese and Japanese-Koguryoic—the latter language family consisting of Japanese (including the Ryukyu dialects) and Koguryo, an extinct language once spoken in Korea and southern Manchuria but unrelated to Korean (Beckwith 2000b). The relationship between Chinese and Japanese-Koguryoic was probably convergent rather than divergent, but there can be no doubt of its existence. Because this early, Chinese-related Japanese material is often extremely valuable for historical linguistic work not only on Chinese but on Tibeto-Burman, I have included references to it here.

[8] Pulleyblank's tone marks for Middle Chinese (his 'Early Middle Chinese') are converted to superscript numerals throughout. Middle Chinese *píng shêng* 'even tone' (unmarked in Pulleyblank 1991) is marked with the numeral '1'; *shǎng shêng* 'rising tone' (marked with an apostrophe in Pulleyblank 1991) is marked with the numeral '2'; and *qù shêng* 'departing tone' (marked with a superscript 'h' in Pulleyblank 1991) is marked with the numeral '3'.

[9] The origin or phonological motivation of the very large set of distinctions in the velars—aspirated stop : unaspirated stop : voiced stop : unvoiced fricative : voiced fricative—is unclear. This unusually large set of phonemes is also found in the dentals, though most proponents of HSR have recently opted for laterals, leaving the phonemic inventory unbalanced. Since the labials have the same large set, which however is generally thought not to have included fricatives until Old Mandarin (or at the earliest Late Middle Chinese), one is hard put to explain the size of these sets, not to speak of justifying the reconstructions that have been proposed for the individual phonemes. Because the evidence of borrowings, both external and internal (as reflected in the script) argue against the existence of many of these distinctions in earlier stages of Old Chinese when the characters were created, it is unclear if a phonemic distinction in aspiration should be reconstructed for any stage of the language earlier than Early Middle Chinese. Aspiration is accordingly not indicated in the Old Chinese reconstructions given here.

[10] Cf. Old Turkic *suβ* 'water' (*su* in most modern Turkic languages), a language no one thinks is related to Chinese or Tibeto-Burman. Despite the similarity to the

popular wisdom, words for 'water' are very frequently borrowed, and it is possible that, as in the case of the Tibeto-Pyu word, the Chinese word for 'water' is not a derived form of a native word (i.e., root plus an extension—of unknown origin and meaning), but a compound of two words, *tî and *wer, one (or even both) of which may have been borrowed. One of several possible loan scenarios is shown in example (19).

(19) NMan *shuǐ* < MChi ⋆*fui*^{2} < MOC *ʃwêr < EOC *tîwêr < *tî + *wêr ← PIE *wêr (Wat. 100) 'water'; cf. TokA *wär* TokB *war* < PTok *wär 'id.'

The Early and Middle Old Chinese forms must be reconstructed with final *-*r* because of the closely related words 泉 *quán* 'spring, source', from Middle Chinese ⋆*dzwian*^{1} (Pul. 262) from Middle Old Chinese *dzwer (cf. Sta. 579), and 川 *chuân* 'river, stream', from Middle Chinese ⋆*tʃʰwian*^{1} (Pul. 60). The latter, though placed by Starostin in an 'irregular' rhyme category (1989: 580: *ĉhun) due to the dictates of his method, in its only rhyming occurrence in the *Shih ching* (*Book of Odes*) rhymes with 焚 *fén* 'to burn', from Middle Chinese ⋆*bun*^{1} (Pul. 94), an obvious relative of 燔 *fán* 'to burn', from Middle Chinese ⋆*buan* (Pul. 89) from Middle Old Chinese *ber (Sta. 579: *b(h)ar),[11] which is widely compared to Old Tibetan *fibar-* 'to burn', from Tibeto-Burman *ber or *bar 'to burn' (Benedict 1972: 50).[12] Thus all three of the most common Chinese words for 'water' and closely related concepts such as 'river' and 'spring' (which in some languages, such as neighboring Old Koguryo, are not distinguished) include both the syllable *tî and the syllable *wer.

It is often said that anything can be borrowed, so we should not be too surprised if the word for 'water' has been borrowed by some of these languages—though from whom, it is not always easy to say. Moreover, even within Tibeto-Burman several major languages do

---

Middle Chinese and Tibeto-Pyu forms, it is unlikely to be a traceable borrowing. See Beckwith (2000) on the 'non-basicness' of most so-called 'basic vocabulary'.

[11] These words are clearly also related to 然 *rán* 'to burn' < MChi *ñian*^{1} (Pul. 264) < MOC *n^{y}er < *m^{y}er; cf. 'The Sino-Tibetan Problem', in this volume. The explanation for the differing forms requires more research.

[12] Note also OKog **măi* (dial. **mej*) 'water, river, spring' (Beckwith 2000b) < *mer, cognate with the OJpn root **mi* 'water' (Martin 1987: 483), from PJK *mîr ~ *mêr, which is undoubtedly the same word as wêr 'water'. Regarding the latter, rather than *tî, being the basis of the rhyme of most Chinese words relating to 'water', cf. 門 *mén* 'door', the phonetic of the character 聞 *wén* 'to hear', which in the *Book of Odes* rhymes with 川 *chuân* 'river', from EOC *tîwer.

not share the putative Proto-Tibeto-Burman word *tî, having perhaps replaced it either by borrowing or by internal shift. The Proto-Tibeto-Pyu word is just as likely to be a loanword as any of the other examples cited. Thus, another possibility is that a Late Proto-Tibeto-Pyu word *tîəw or *tîw was simply borrowed whole from an Old Chinese *tîwer with canonical loss of final *-(V)r, as generally in Tibeto-Burman.[13]

Nevertheless, it is clear that Tibetan and Pyu do share this isogloss, whether it is a special development of the reconstructible Tibeto-Burman root *tî, or a loanword. Although the Chinese word for 'water', 水 *shuǐ*—the character for which is attested in the earliest Chinese linguistic material, the Oracle Bone Inscriptions—could be argued to derive from the simplex root *tî, but with a different extension, it is unclear if the word for 'water' in Chinese is directly related to the extended Tibeto-Burman forms because most of its derivatives include (and may well be based on) the root *wer. Moreover, though the *ti element is undoubtedly derived from the same original word as Tibeto-Burman *tî, since that word is found also in Japanese it seems to be an areal *Kulturwort*. It is thus as likely that the Proto-Chinese and Proto-Tibeto-Burman forms were inherited from a common ancestor as it is that the Proto-Chinese, Proto-Tibeto-Burman, and Proto-Japanese forms were inherited from a common ancestor.

II. Ten

Pyu *sû* 'ten' corresponds to the root √*cu* [ʧu] of Old Tibetan *bcu* 'ten', minus the affricate element in the initial. Within Old Tibetan, the root *cu* 'water' is identical to the root *cu* 'ten'. Similarly, the Chinese character used to write the word 十 *shí* 'ten' is the phonetic in 汁 *zhî* 'juice', indicating that the two words were pronounced the same when the second character was created in Late Old Chinese times. One might therefore be inclined to see a simple relationship here—not only an etymological one among the three languages, but also a phonological one between the two etyma. However, the matter is not so simple.

---

[13] See 'The Sino-Tibetan Problem', in this volume.

The early Tibeto-Burman languages have radically differing forms of the word for 'ten', as shown in examples (20) through (24); the modern languages, such as Modern Meithei, in (25), are also often highly divergent.

(20) Pyu *sû* 'ten'
(21) OTib *bcu* 'ten' < √*cu* [ʧu], -cu ~ -*śu* [ʃu]'-ty (ten)' (the latter form only in *ñiśu* 'twenty')
(22) OBur *c^{h}ay* 'ten' > modern LBur *s^{h}ay* > NBur *s^{h}æ*; cf. LBur *kyip* 'ten'
(23) ONew *jî* 'ten'
(24) Tgt *dgafi* 'ten'
(25) NMei *təra* 'ten'

While it is generally believed that the form -*śu* in Old Tibetan *ñiśu* 'twenty' is a reduced form of the usual Old Tibetan decade marker (*b*)*cu* '-ty, ten'—which is the same as the affricated full form *bcu* [bʧu] 'ten'—due to assimilation with the final *–s* of *gñis* 'two', comparative evidence suggests that the decade root of Middle Old Chinese was probably unaffricated, and may be preserved in this form in Old Tibetan *ñiśu* 'twenty'. Examples (26) through (28), and possibly (29), show that there were apparently two roots in Old Chinese: a form with an alveodental stop or affricate initial and a form with a fricative initial. The Japanese forms indicate that the distinction was between free form and decade marker, and that there may also have been a vocalic difference between the two.

(26) Pyu *sû* 'ten'
(27) OTib -*śu* [ʃu] '-ty (ten)' in *ñiśu* 'twenty'
(28) OJpn **töwo*, **tö* (*[təwo], *[tə]) 'ten', **so* (*[su][14]) '-ty (ten)'
(29) Thai *sip* 'ten; -ty' ←LOC dial. + *śip

Despite the significantly different forms of the word for 'ten' in examples (20) through (25), and for some of the other numerals in Tibeto-Burman, Matisoff argues in a gloss in Benedict's *Conspectus*, "It now appears that all the Ch[inese] numerals, including '100', are cognate with the TB set" (Benedict 1972: 161, n. 435). He accordingly reconstructs an Old Chinese form *g(y)ip to correspond to *gip, one of Benedict's two reconstructions for Tibeto-Burman 'ten',

[14] Old Japanese vowels are still a hotly debated issue. The usual reconstruction of the decade form, **-swo* (Mar. 529), is problematic because the independent forms for 'ten', *töwo* (or perhaps *töwö*) and *tö*, have a different vowel (the actual phonetic value of which is also uncertain, but is currently thought to have been close to [ə]); the modern Japanese pronunciation of the three forms is [to:], [to], and [so] respectively.

which he says is "poorly represented;" see example (30). However, the Middle Chinese form had a voiced alveodental affricate initial, implying an Old Chinese voiced alveodental stop initial. Moreover, the Tibeto-Burman forms with voiced velar initials that are the basis of Benedict's reconstruction *gip, and Matisoff's two forms *gip and *gyap are found only "in composition" (Matisoff 1997:25), not in free or citation form. In short, this particular word should be reconstructed for Tibeto-Burman with an unvoiced initial, as shown in example (31), and no longer looks so much like Old Chinese.

(30) OChi *d'i̯əp/źi̯əp 'ten' < ? PST *g(y)ip (Matisoff in Benedict 1972: 175, n. 464; cf. Matisoff 1997: 25) > TB *gip: Limbu *gip* (in comp.), Miju *kap* ~ *kyep*, Mikir *kep* < *gip*, Maring *tśip* < *kyip*, Yawdwin *gyip* (in comp.), Bur *ăkyip* '10'[15] (TB **gip*)" (Benedict 1972: 19); cf. Benedict's proposal to derive OTib *bcu* from a TB *gip (1972: 94).

(31) PTB *kêp 'ten' > LBur *kyip*, Miju *kap*, Mikir *kep*, etc.

Benedict's other reconstruction, which he calls a "Kachin-Konyak-Bodo-Naga root," is given first as *ts(y)i(y), then, in an editorial note by Matisoff, as *tsyay (Benedict 1972: 94), as shown in examples (32) and (33). In a recent monograph, Matisoff now argues that it is necessary to reconstruct two related forms "at the proto-level," *ts(y)i(y) and *tsyay (Matisoff 1997: 25), which gave rise to Tibetan *bcu*, via "a quasi-regular (dissimilatory?) development after palatal affricate initials" (Matisoff 1997: 25 n. 28), as in (34).

(32) PTB *ts(y)i(y) > B *ătshai* 'ten' (Ben. 94)

(33) PTB *tsyay > B *ătshai* 'ten' (Matisoff in Ben. 94, n. 272)

(34) PTB *ts(y)i(y) ~ *tsyay > OTib *–cu* 'ten'

However, the forms in Pyu and Old Tibetan, the earliest Tibeto-Burman languages attested in segmental scripts, unambiguously support a late Proto-Tibeto-Pyu reconstruction *śiw or *śiəβ.[16] The latter corresponds very well to the Late Old Chinese form which was loaned to Taic and (Proto-)Japanese, as shown in (35).

(35) PTP *śíəβ > Pyu *sû* 'ten' (cf. *tpû* 'twenty') ~ OTib √*śu* [ʃu] > *śu* 'id.' : Thai *sip* <**śi̯əp* 'id.' : OJpn ✯*-so* [su] < **śu* 'id.'

[15] This special Burmese word for 'ten' is given in the dictionary as *kyip* and is said to be "used for *s^{h}ay* in numbering rational beings" (Judson 1866:745).

[16] Pyu *s-* corresponds regularly to OTib *ś-*. Pyu *h-* corresponds to OTib *s-*.

This raises some serious questions about the hitherto accepted reconstruction of the Old Chinese word for 'ten', which has been based on the Middle Chinese form **dźip*. If we look again at Old Japanese **töwo* or **tö* and the decade compounding form **-so* [su], while Old Japanese did not have a distinction between voiced and unvoiced initials, the full form does support an alveodental stop reconstruction for the initial of the Old Chinese free form. However, the initial consonant of the compounding form **-so* (supported by Thai *sip* and Old Tibetan *–śu*), indicates something else. We should probably reconstruct two roots for Middle Old Chinese, namely a combining form with *ś- initial and a free form with *d- initial, as shown in example (36). The final of the Chinese is also problematic. In light of the comparative evidence, it was surely voiced, and apparently not a stop; it cannot be reconstructed as *-p for Middle Old Chinese or earlier periods. By Middle Chinese times, or in the dialect that gave rise to Middle Chinese, the distinction was lost, so we only have a single Middle Chinese form.

(36) OJpn **töwo* [təwo], **tö* [tə] '10', **-so* [su] < **śu* '-ty': OEC *dêb ~ *śêb ~ MOC *dʲêb ~ *dśêb < EOC *deśeb

The remaining Tibeto-Burman forms raise further questions. While the initial and vowel of the Old Burmese form *chay* is perhaps reconcilable with Proto-Tibeto-Pyu, especially considering the New Burmese development of the initial, the final is not easily reconcilable with the Proto-Tibeto-Pyu final. The other Burmese word, *kyip*, which is relatable to other Tibeto-Burman words for 'ten', is not easily relatable to either of the Proto-Tibeto-Pyu forms.

However, it is agreed by everyone that the Chinese and Tibeto-Burman numerals are related, as a set, though it is not agreed if they are related by divergence or by convergence (borrowing). If Literary Burmese *kyip* and Tangut *dgafi* are related to the Chinese word or words for 'ten', these forms would descend not from Middle Old Chinese *dśêb from Early Old Chinese *deśéb, but from an Early Old Chinese *dekeb, as shown in (37).[17] This word could then have been simplified through reanalysis of the initial alveodental as a separable prefix, producing *kêb, the ancestor of Burmese *kyip*.

---

[17] The latter is highly evocative of PIE *dekm̥ 'ten'. See the discussion in 'The Sino-Tibetan Problem', in this volume.

(37) LBur *kyip* 'ten' < PBur *kyap ~ Tgt *dgafi* 'id.' < Proto-Burmic-Qiangic *dkaβ ~ *dkeβ ← early MOC dial. *dkeb < EOC *dekeb 'ten'

If we assume an ancestry similar to that for Burmese *kyip*, we can perhaps explain the Pyu and Old Tibetan forms as in example (38).

(38) Pyu *sû* ~ Old Tibetan *bcu* 'ten', *-(b)cu*, *-śu* '-ty'< PTP *śiβ or *śiəβ ~ *dśiβ or *dśiəβ ← early MOC *śêb ~ *deśêb < EOC *đekeb

The forms in Newari and other languages entail other problems. Matisoff reconstructs eight distinct 'Tibeto-Burman' roots for 'ten', with even more alternate and isolated forms (1997: 24 et seq.). This is unlikely in the extreme.

It is also unlikely that the Proto-Tibeto-Burmans had a full native numeral system which not a single branch of the vast Tibeto-Burman family retained. Rather, it appears that they simply did not have any higher numerals, and the daughter languages individually borrowed or innovated extended numeral systems fairly late in their history. That would explain Matisoff's inability to reconstruct a unitary Proto-Tibeto-Burman root for 'ten' and other problematic numerals.

At any rate, whatever may be decided ultimately about the origin of the numerals in the Tibeto-Burman languages as a whole, the fact that Pyu and Tibetan share a distinctive form of the word for 'ten' indicates that it constitutes an additional isogloss, another sign that the two languages descend from the same early daughter of Tibeto-Burman.

Conclusion

If we take Meillet's admonition to heart and focus on the anomalies, as has been attempted here, it is likely that we will develop a different picture of Tibeto-Burman, and for that matter, of the Sino-Tibetan theory, than the currently dominant one. Looking at some of the best-attested words in the earliest Tibeto-Burman languages and comparing them to Chinese, we are often struck by the usual simplicity of the Proto-Tibeto-Burman word compared to the Chinese word (if a comparable form exists). This simplicity is often retained even in phonologically more complicated-looking languages such as Old Tibetan. For example, in addition to 'water, juice', discussed above, a surprisingly high percentage of the Tibeto-

Burman words in the small Pyu corpus are shared by Tibetan. Consider examples (39) through (43).

(39) PTB *la (Ben. 42: *la[18]) 'moon, month; spirit' > Pyu *la* 'month', PTib *la (> OTib *zla* 'moon, *sla* month', *lha* 'spirit', *bla* 'soul, mind', etc.), OBur *la* 'moon', Tgt **li* 'moon, month'

(40) PTB *ru (Ben. 16: *rus) 'bone' > Pyu *ru* 'bone, relic', OTib *rus* 'bone', *ru* 'horn',[19] OBur *aruiw*^{4}

(41) PTB *da (Ben. 19: *day) 'that' > Pyu *ḍʰau* 'id.', OTib *de* < PTib *day* 'id.', OBur *tʰuiw* 'id.', Tgt *tʰa* 'id.' Note CTP *da la 'thereupon' > Pyu *ḍʰau lo:* 'id.'~ OTib *de la* 'id.'

(42) PTB *bî (Ben. 102: *biy) or *pî 'to give' > Pyu *på* [pi] 'id.', OTib *sbyin* (√*byi*) < PTib *bî 'id.', OBur *piy*^{4} 'id.'

(43) PTB *na (Ben. 31: *na) 'to be ill' > Pyu *hnî:* (< *s-na-) 'id.', OTib *na* 'id.', OBur *nâ*^{3} 'id.'

Pyu and many other Tibeto-Burman languages have been viewed as 'already simplified' if compared to languages such as Old Tibetan, which is considered to be 'conservative' or even 'archaic' because of its complex phonology. But Pyu shares much more with Tibetan than might have been expected, especially in view of its location and its apparent membership in the Southwestern branch of Tibeto-Burman (Bradley, in this volume). If these observations are supportable by other lexical (and morphophonological, etc.) isoglosses, as they seem to be, Pyu and Tibetan—and accordingly, the sub-branches of Tibeto-Burman to which they belong—should perhaps be classed together at a higher node in the family tree. In other words, they may well be descendants of the same early Tibeto-Burman daughter language.

References

Baxter, William H. 1992. *A handbook of Old Chinese phonology.* Berlin: Mouton de Gruyter.

Beckwith, Christopher I. 2000. Defining 'basic vocabulary': the correlation between frequency and rate of retention. Paper given at the Linguistic Society of America, Chicago.

[18] Benedict actually reconstructs two forms, *s-la and *g-la (1972: 42), but the index gives them as equivalent to *s-gla (1972: 216).

[19] Benedict reconstructs two separate etyma sets, *rus 'bone' (1972: 16) and three words for horn: *kruw = *krəw, *ruŋ = *rwaŋ, and *rwa-t (1972: 215).

—— 2000b. Toward Common Japanese-Koguryoic: A reexamination of Old Koguryo Onomastic Materials. *In* J.J. Nakayama and Charles Quinn, Jr., ed., *Japanese/Korean linguistics*, Vol. 9. Stanford: CSLI, 3-16.

—— 1996. The morphological argument for the existence of Sino-Tibetan. *Pan-Asiatic linguistics: proceedings of the Fourth International Symposium on Languages and Linguistics, January 8-10, 1996*, Vol. III. Bangkok: Mahidol University at Salaya, 812-826.

Benedict, Paul K. 1972. *Sino-Tibetan: a conspectus*. Cambridge: Cambridge University Press.

Blagden, C. Otto. 1911. A preliminary study of the fourth text of the Myazedi inscriptions. *JRAS*: 365-388.

Judson, A., ed. Robert C. Stevenson 1893. *Judson's Burmese-English dictionary*. Rangoon: Governnment Press (reprinted New Delhi: Asian Educational Services, 1993).

Karlgren, Bernhard 1957. *Grammata Serica recensa*. Stockholm: reprinted from the *Bulletin of the Museum of Far Eastern Antiquities* 29: 1-332.

Martin, Dan 1997. *Zhangzhung dictionary*. http://www.comet.net/ligmincha/html/zzdict1.html.

Martin, Samuel G. 1987. *The Japanese language through time*. New Haven: Yale University Press.

Matisoff, James A. 1997. *Sino-Tibetan numeral systems: prefixes, protoforms, and problems*. Canberra: Pacific Linguistics.

Pulleyblank, Edwin G. 1991. *Lexicon of reconstructed pronunciation in Early Middle Chinese, Late Middle Chinese, and Early Mandarin*. Vancouver: UBC Press.

Sagart, Laurent 1999. *The roots of Old Chinese*. Amsterdam: John Benjamins.

Shafer, Robert 1943. Further analysis of the Pyu inscriptions. *HJAS* 6: 313-366 + 5 plates.

Starostin, S.A. 1989. *Rekonstrukciya drevnekitayskoy fonologičeskoy sistemï* ['Reconstruction of the Old Chinese phonological system']. Moscow: Nauka.

Stargardt, Janice 1990. *The ancient Pyu of Burma, Vol. 1: early Pyu cities in a man-made landscape*. Cambridge: PACSEA.

Watkins, Calvert 2000. *The American Heritage dictionary of Indo-European roots*. 2nd ed. Boston: Houghton Mifflin.

ON PALI-BURMESE INTERFERENCE

RUDOLF A. YANSON (ST. PETERSBURG STATE UNIVERSITY)*

The problem of Pali-Burmese interference is of special interest for several reasons. The main one is, of course, that from the very first contacts of the Burmese with Pali texts Pali was considered to be a model language, and for several centuries the attitude that Pali was the only 'correct language' prevailed among the Burmese. What is correct for Pali is correct for Burmese—such was the motto of the first Burmese grammarians and the composers of early inscriptions. The significance of Pali in Burmese society from ancient times to the present was perfectly described by John Okell in his 'Nissaya Burmese' (1967). Such a situation naturally led to unusually wide scale Pali influence on Burmese.

Another reason that the problem of Pali-Burmese interference can be considered a pressing one is obvious: up to now it has not been given proper attention by scholars. Except for attempts to identify Sanskrit-Pali loans in Burmese—the latter being really abundant —there are no publications on the subject.

One more incentive for the following observations lies in the fact that it is impossible to come to convincing conclusions on this problem without fully taking into consideration the language of the early Burmese inscriptions. To the best of my knowledge, at present there are no scholars doing systematic research in this field.

Roughly three trends in Pali-Burmese interference can be determined. The first and most obvious one is Pali loans. In most cases they are easily identified in speech and texts, and the problem here is to explain the sometimes confusing phonetic and graphic form of some loans. Recently I explained the reason for the respelling of some Pali words in Burmese, and also the reason why some final syllables of several Pali loans, which were structurally closed,

* I take this opportunity to express my grateful thanks to Prof. Ch. Beckwith, who took the trouble to edit my paper, not an easy job at all, and who did it with great patience and carefulness.

evolved as if they had open syllables (Yanson 1997). But of course the problem calls for further research.

The second trend is the problem of implications of the Nissaya texts for the Burmese language. Nissaya texts, which are word-by-word or phrase-by-phrase translations of Pali texts, started to appear in Burma from the 15th century, and the tradition of creating Nissaya texts or of following Nissaya patterns in writing commentaries, or translating or retelling Pali texts, has been retained up to the present day. Okell (1967) has succeeded in showing how the Burmese had arranged their set of grammatical markers into a rather strict system adapted for reflecting Pali flexions. But he does not try to trace what specific constructions or grammatical markers used in Nissaya or Nissaya-like texts might have originated from Pali or under Pali influence. His reticence is justified, because unlike loans this sphere of Pali interference forms what may be called a concealed trend of Pali influence on Burmese. Below I am going to point out some cases of how alien grammatical peculiarities have become a part of Burmese grammar. It should be understood that I no longer maintain my previous view that Pali influence on Burmese was restricted to borrowings of lexical items (Yanson 1993).

There are two types of verbs in Burmese: functive and stative. Though they have very much in common, they differ in the way they function when used as attributes to NPs. Stative verbs usually follow the verb immediately, without the involvement of grammatical markers, and form a construction 'stative verb - NP', e.g.[1], *lu*^{2} *kong*^{3} (man-good) 'good man'. Functive verbs, unlike stative ones, precede the NPs if used as attributes, forming constructions equivalent to participle-NP constructions in European languages. In this case the

[1] All examples are given in transliteration. The meaning of some graphic symbols is different in the OBur inscriptions and in MBur. For instance, the letter which in the inscriptions indicated *c* is in MBur *s*, and the old combinations *ky, khy* are now pronounced *c, ch* respectively. There are some other divergences between the meaning of the same symbols in OBur and MBur. My transliteration will follow the old value of Burmese writing for both old and modern examples. Tones in the inscriptions are marked only from time to time. Here I mark the tones of OBur words according to the tone they have in MBur. Some syllables in MBur are pronounced with schwa and have no tone. Such syllables were present in OBur too. They are transliterated here with *a* in the place of schwa and no tone is marked. Syllables with final surds have no special graphic tone marker; all are pronounced with a final glottal stop, which sometimes is denied the status of tone marker and may be treated as a final consonant. When transliterated, such syllables do not require mention of the tone.

verb must be followed by the grammatical marker *so*^{3}, e.g.: *pye*^{3}*so*^{3}*lu*^{2} (run—*so*^{3}—man) 'running man'.[2] Stative verbs can also be used prepositionally, but in this case they also must take the marker *so*^{3}, e.g., *kong*^{3} *so*^{3} *lu*^{2} 'good man'. The construction V*so*^{3} can function only if followed by NPs.

There is absolutely nothing in the shape or functioning of the marker *so*^{3} which would make us suspect that it is not a native word, all the more so because it is only by using the intervening *so*^{3} that it becomes possible to express the relationship equivalent to the participle-NP construction in European languages. Only the pronoun *su*^{2} 'he, she; the one who...' can follow verbs directly without an intervening *so*^{3}, thus forming a construction V*su*^{2} 'the one who V'. Unlike V*so*^{3}, the construction V*su*^{2} can not precede NPs, i.e. the construction *V*su*^{2} NP is not possible. The construction V*so*^{3} *su*^{2} is however possible, the difference between it and V*su*^{2} being purely stylistic.

The Burmese inscriptions present a wholly inconsistent picture of the utilization of both elements, i.e. *so*^{3} and *su*^{2}. Along with constructions of the type V*so*^{3} NP, we find constructions of the type V*su*^{2} NP, which are unacceptable in Modern Burmese. The construction V*sa* NP is also possible, and all three constructions are interchangeable and correspond to Modern Burmese V*so*^{3} NP. On the other hand, along with the usual Modern Burmese construction V*su*^{2}, we meet in the inscriptions V*so*^{3} and V*sa*.

It is tempting to try to treat this case as one relating to phonology: the interchange of *a~o~u*. Since the inscriptions represent the earliest attested stage of the writing tradition of the Burmese, it is but natural that they contain numerous different spellings for the same words, and one of the discrepancies there is the use of the vowel *o* in the place of standard *u*, e.g., *hlo*^{2} instead of *hlu*^{2} 'to donate', *klañ*^{3}*jo*^{3} instead of *klañ*^{3}*ju*^{3} 'gratitude'.

---

[2] The description of the difference between the two types of verbs given here might seem to contradict what is written on the subject by J. Okell (1969: 43). He writes that functive verbs, used attributively, simply precede the noun, and gives the example *sok re*^{2} (drink-water) 'drinking water'. What Okell writes is true, but only for gerund constructions, which formally can be distinguished from participle-noun constructions by the possibility of inserting the marker *bhiu*^{1} 'for, in order to', between the components of the construction: *sok bhiu*^{1} *re*^{2} 'drinking water'.

Interchanging them with *a* could be explained as reflecting the hesitations of scribes who could not make up their mind in favour of either *o* or *u* and thus used just the consonant symbol, i.e., without adding any symbols to change the default vowel *a* into another vowel.

From the phonological point of view an explanation of the case might be as follows. The phonological system of Old Burmese did not contain the vowel *o* in open syllables, but in Modern Burmese such syllables are common. So when we meet in the inscriptions the vowel *o* in the place of standard *u* suspicion arises that the discrepancy reflects the process of formation of the new vowel.

If we accept such an approach to the problem, it will be difficult to explain what morpheme the interchanging so^3 ~ su^2 in the inscriptions could represent. If we accept the phonological explanation, then we have to accept that a single morpheme has over time split into two, and developed not only different vowels and tones for each of the new variants, but also produced new morphemes so distant in meaning that a common origin seems improbable.

In fact, the problem has nothing to do with phonology, but exemplifies the interference of Pali in the field of grammar.

The way the marker so^3 is spelled reveals its Pali origin. In Old Burmese the vowel *o* was never used in open syllables, unless by mistake or under the influence of Pyu, which is peculiar in using *o* in the place of *u* in Pali loans. Numerous modern Burmese words containing the vowel *o* were spelled with *aw* in the inscriptions. At the same time, spellings with *o* were common in Pali loans, and in Modern Burmese they retain their old spellings. So there is nothing strange in the fact that the Pali word so^3 got mixed with Modern Burmese words—being monosyllabic, unlike the main bulk of loanwords, it is difficult to identify it as something alien.

In Pali, so^3 is a third person pronoun, i.e. a kind of synonym of the Burmese word su^2. Along with Burmese su^2, the marker so^3 is also interchanged with *sa*. But *sa* is also a Pali pronoun, a variant of so^3, and this fact also supports the assumption that so^3 is of Pali origin. So, the interchanging words so^3 ~ su^2 ~ *sa* in Old Burmese represent three third person pronouns, one of which is Burmese and the other two Pali.

Now the constructions Vso^3 and V*sa* in the place of standard Burmese Vsu^2 become transparent—they are absolutely synonymous with Vsu^2, though probably of a higher order since they contain Pali

components. But since tones were not regularly marked in the inscriptions, and we spoke about the inconsistency in representing the vowel *u*, for which sometimes *o* was substituted, it may be that the construction *Vso*^{3} was not intended by the scribes to be a combination of Pali and Burmese words, but standard Burmese V*su*^{2}.

Though the construction V*so*^{3} (*su*^{2}, *sa*) NP in the place of Modern Burmese V*so*^{3} NP poses a more complicated problem, it too can be explained as a result of Pali interference.

There are no participles in Burmese, but they are common in Pali and are often used as prepositional attributes to NPs, forming constructions like English 'running deer' or 'praying monk' and so on. It appears that there were no means in Old Burmese to express such a relation, and to cope with the situation the Burmese introduced a new construction intended to fill this gap in Burmese grammar, viz. V*so*^{3} NP and synonymic V*sa* NP. The choice of Pali *so*^{3} and *sa* is clear—although Burmese V*su*^{2} corresponds to the Pali participles, it can not be used attributively, i.e. the construction *V*su*^{2} NP is inadmissible according to Burmese grammar. By using a foreign word in the place of Burmese *su*^{2} the scribes 'cheated' their grammatical rules in the sense that the construction contained a foreign word so it did not have to be in accordance with the usual rules. Thus the constructions V*so*^{3} NP and V*sa* NP were introduced into Burmese. The construction V*su*^{2} NP, inadmissible within the framework of Burmese grammar, became possible on the grounds of analogical extension—after all, the three elements are synonyms.

Another reason for introducing constructions containing Pali words was the semantics of the latter. Unlike Burmese *su*^{2}, which is applicable only to animate nouns, Pali pronouns have no limitation in this sense and can be applied to both animate and inanimate nouns. So suppose the Burmese, against their grammatical rules, started to use the native Burmese construction V*su*^{2} in the function of prepositional attributes, like Pali participles. Such a construction would be of limited use since it would be applicable attributively only to animate nouns and as such would not be able to cover the whole range of usage of Pali participles. For instance, such constructions as 'floating island' or 'moving cart' would be impossible. By introducing the Pali pronoun the Burmese overcame this limitation. Thus, logically the choice of *so*^{3} and *sa* in this case seems justified.

Gradually the pronoun *sa* was ousted from the construction (see below for the reason), and *so*^{3} became grammaticalised. That *so*^{3} very soon became grammatical is shown by the following facts. In the course of time, the construction V*so*^{3}*su*^{2} started to appear in texts. If *so*^{3} by this time had not been grammatical, the construction would be difficult to analyse—two synonymous pronouns with one verb. What is still more significant, the order of the two pronouns in the construction is fixed: *V*su*^{2}*so*^{3} is not possible. The fixed position of the two elements means that they differ grammatically.

It is worth noting that the process of grammaticalisation of Pali pronouns was very quick. In the earliest Burmese inscriptions, dated to the 12th century, the utilization of all three pronouns seems to be absolutely inconsistent, but by the end of the 13th century the situation had become standardized from the point of view of Modern Burmese.

The Pali pronoun *sa* was not just ousted from the construction under analysis, however. It came to be utilized in the finite form of Burmese verbs, with very interesting implications for the whole verbal syntagma.

There is no generally accepted approach among Burmanologists as to the role of the finite verb markers. Some treat them as markers of realised—non-realised action (Allott 1965), others as markers of past-present (non-future) and future tense respectively (Kassevitch 1990). There are two corresponding sets of markers, which for convenience I will call tense markers.

Of the modern set of past-present tense markers only two were in use in the earliest inscriptions, namely *i*^{1} and *satañ*^{3} (in the earliest inscriptions the latter was also spelled *sate*^{3}). In Modern Burmese the marker *satañ*^{3} is different from *i*^{1} and other past-present tense markers—which are neutral as to emphasis—in being emphatic.

In the inscriptions the marker *i*^{1} seems to be used in most cases in the same way as it is now. Unlike *i*^{1}, *satañ*^{3} is used very controversially. On the one hand, it is used in sentence final position, which is typical for tense markers, and as such is interchanged with the marker *i*^{1}. Taking this into consideration, it is tempting to treat both markers as belonging to one paradigm. On the other hand, the marker *satañ*^{3} can follow another marker, *am*^{1}, which has a clear future sense, and thus the construction V*am*^{1}*satañ*^{3} is quite common in the inscriptions and it has obvious future time semantics. By contrast, the marker *i*^{1} cannot combine with *am*^{1}.

It is remarkable that, according to Okell (1967: 105), in Nissaya texts the present tense of Pali is reflected by the markers i^1 and, rarely, $sata\tilde{n}^3$, and the future tense is reflected by both am^1 and $am^1 sata\tilde{n}^3$. It appears that the marker $sata\tilde{n}^3$ was used to express both present and future tenses, forcing us to accept that in fact $sata\tilde{n}^3$ had nothing to do with expressing tense. Accordingly, the question arises: what was it and how did it become a tense marker in Modern Burmese?

Unlike other Burmese grammatical markers, which are monosyllabic, $sata\tilde{n}^3$ is a complex unit. In Modern Burmese it is treated as consisting of a reduced form of one of the past-present tense markers—namely $sa\tilde{n}^2$—and the emphatic particle $ta\tilde{n}^3$. The reason for such treatment is as follows. Interrogative sentences in Burmese are formed by adding the interrogative particles la^3 or lai^3 to the finite form of the verb constituted by the tense markers $sa\tilde{n}^2$ and $ma\tilde{n}^2$.In this case the markers are used in reduced form, viz., in the form of *sa* or *ma*. Consider examples (1) through (4).

(1) lu^2 swa^3 $sa\tilde{n}^2$
man go PASTPRES
'A man goes (went).'

(2) lu^2 swa^3 $sa\text{-}la^3$
man go PASTPRES-INTER
'Does the man go?' (Or: 'Did the man go?').

(3) lu^2 swa^3 $ma\tilde{n}^2$
man go FUT
'The man will go'

(4) lu^2 swa^3 $ma\text{-}la^3$
man go FUT-INTER
'Will the man go?'

The above approach to the marker $sata\tilde{n}^3$ in Modern Burmese seems to be quite reasonable. But it is inapplicable to Old Burmese. It appears that in the earliest Old Burmese texts there was no marker $sa\tilde{n}^2$. In fact, it does not occur until sometime in the 13th century. Therefore, the component *sa* in $sata\tilde{n}^3$ cannot be treated as part of a tense marker. On the other hand, we have strong evidence that it is simply the Pali third person pronoun. Such confidence is based on the fact that three variants of the marker were actually used in the inscriptions, viz., $sata\tilde{n}^3 \sim so^3 ta\tilde{n}^3 \sim su^2 ta\tilde{n}^3$. As can be seen, three components, *sa*, so^3 and su^2, interchange, and we remember that all of them are third person pronouns. As to the component $ta\tilde{n}^3$ of the

whole marker, it can be used alone with nouns to express strong emphasis, as shown in example (5):

(5) *maŋ*^{3}*-tañ*^{3} *hlu*^{2}*-i*^{1}
king-EMPH donate-TENSE
'It is (was) the king, who donated'.

But when used in combination with *sa* (*so*^{3}, *su*^{2}) after verbs, it seems to be indifferent as to emphasis, and constructions V*i*^{1} and V*satañ*^{3} seem to be of the same emphatic value.

The discrepancy with *satañ*^{3} in Old Burmese can be explained from the point of view of peculiarities of the grammars of Pali and Burmese.

In Pali, participles are often used in a finite function, serving as the main predicate of a sentence. So such phrases as, for instance, 'I donate' and 'I (am) a donating one' are quite common in Pali and are actually synonymous in the sense that both can be used in the same context. By contrast, the Burmese language, as mentioned above, did not have participles and therefore had no means to construe sentences corresponding to Pali ones with participles in finite function. To fill the perceived 'gap' in their grammar, the Burmese followed the same course of action as they did with prepositional participles. They started to use the Pali pronouns *so*^{3} and *sa* with verbs as the equivalents of participles. Burmese grammar does not allow nominalised verbs, which in our case are the equivalents of Pali participles, to be used in finite function, so the original Burmese construction V*su*^{2} could not have been used in finite function. The constructions V*so*^{3} and V*sa* are, on the one hand, synonymous with Burmese V*su*^{2}, but, on the other hand, they are formed with a Pali component, which made it possible to treat them differently from the native construction. Thus nominalised verbs used in finite function were introduced into Burmese grammar to match the Pali participles in the same function. It is not surprising that in the end Burmese *su*^{2} also came to be used as a variant for *so*^{3} and *sa*—after all, the three morphemes are synonyms.

It is thus clear that the constructions V*satañ*^{3} and V*i*^{1}, which seem to be interchangeable in the inscriptions, are completely different from the grammatical point of view. The construction V*i*^{1} represents a finite form of the verb, expressing past-present tense, whereas V*satañ*^{3} is a nominalised verb, neutral as to tense. The two construc-

tions are related only as two ways of expressing the same proposition or idea.

As for the component *tañ*^{3} in the construction V*satañ*^{3}, most likely it functioned as a substitute for the copula. It seems that in Old Burmese, sentences consisting of NPs without copulas were not possible. Analysis of the inscriptions reveals that predicative sentences end either with the copula *phlac* 'to be' plus the tense marker *i*^{1}, or with *tañ*^{3}. This brings us to the following assumption. In Pali, predicative sentences consisting only of NPs are quite common. Suppose that Burmese grammar did not allow such sentences. In that case the Burmese were doomed to find ways to 'correct' their grammar to match the rules of Pali. They solved the problem by using the emphatic particle *tañ*^{3} in the place where, according to their grammar, a copula should be used.

We can't say that such a choice lacks logic. In a way, the semantics of *tañ*^{3}—underlining or stressing—are related to the semantics of a copula. The situation once more demonstrates a compromise between the rules of two grammars: on the one hand, there is no copula in sentences consisting only of NPs, thus matching Pali standards, while on the other hand there is something in the place where a copula should be according to native grammar; this 'something', due to its semantics, can be associated with the role of copula. With its new function, the emphatic particle *tañ*^{3} loses its original meaning and sentences ending with it bear none of the emphasis discussed above.

By having introduced *tañ*^{3} in the copula function, the Burmese didn't reject their own original copula *phlac* 'to be', but its use became considerably reduced. In place of the standard Burmese construction NP NP *phlac-i*^{1}, as in example (6), the construction NP NP *tañ*^{3} became prevalent.

(6) *su*^{2} *maŋ*^{3} *phlac- i*^{1}
he king COP-TENSE
'he is a king'

The reason the new construction became popular is probably not only that it resembled Pali, but is also due to pragmatic considerations. Since the texts were inscribed on stone, the need for simplicity in the graphic form of words and the desire to save space were of considerable importance. The real Burmese copula *phlac* with the marker *i*^{1} is definitely more complicated graphically, and occupies more space,

than *tañ*^{3}. To see how *satañ*^{3} became a tense marker in Modern Burmese, let us compare two sentences, in examples (7) and (8), which are possible in both Old and Modern Burmese:

(7) *ŋa*^{2} *kywan*^{2} *hlu*^{2}*-i*^{1}
I slave donate-TENSE
'I donate (donated) the slave (slaves)'

(8) *ŋa*^{2} *kywan*^{2} *hlu*^{2}*-satañ*^{3}
I slave donate-TENSE
'I donate (donated) the slave (slaves)'

From the point of view of Modern Burmese, the structure of both sentences is identical, the latter sentence being emphatic, and both can be analysed as NP(*ŋa*^{2}) NP(*kywan*^{2}) V(*hlu*^{2})-TENSE MARKERS. For Old Burmese, this analysis holds true for the first sentence too, but the second sentence should be analysed as having the following structure: NP(*ŋa*^{2}) NP(*kywan*^{2}) NP(*hlu*^{2}*sa*) COP(*tañ*^{3}), and the literal translation of the sentence should be 'I (am) the one, who donates (donated)', *hlu*^{2} *sa* being structurally a nominalised verb with the meaning 'the donating one'. As was said above, tense markers in Burmese occupy the final position in sentences. The complex unit *satañ*^{3} also always occupies final position in sentences and as such resembles tense markers. So it is not strange that in the course of time *satañ*^{3} was reanalysed as consisting of a reduced form of the tense marker *sañ*^{2} and the emphatic particle *tañ*^{3} by analogy with interrogative sentences, as described above. In the end, the Pali pronoun *sa* became grammaticalised, and the emphatic particle *tañ*^{3}, which served as a copula when the new construction was introduced, had its original meaning of emphasis restored. This is how the controversial Old Burmese unit *satañ*^{3} became a member of the tense marker paradigm of Modern Burmese.

Let us proceed to the peculiar Old Burmese construction V*am*^{1}*satañ*^{3} mentioned above. Its peculiarity seemed to be that two semantically contradictory grammatical forms were used with one verb—*am*^{1} has clear future time sense, and *satañ*^{3} was supposedly a past-present tense marker. Now we know that *satañ*^{3} in Old Burmese had nothing to do with expressing tense, and from this point of view the construction V*am*^{1}*-satañ*^{3} should not be looked upon as a controversial one. Yet it also is directly connected with Pali and therefore deserves detailed analysis.

The component *am*^{1} of the construction is in Modern Burmese a future tense marker, although of restricted use—it is considered to be archaic and is confined to Nissaya style texts. In the inscriptions the marker *am*^{1} is the only one associated with future time sense. Along with this, it has clear desiderative semantics. It is always followed by *satañ*^{3} when used with main predicates, and as a rule occurs in spells and curses, with which the inscriptions (most of which are dedications) usually end. Consider examples (9) through (11), which demonstrate the functioning of the construction in the inscriptions.

(9) *rahan*^{3} *khyi*^{2}*paŋ*^{1} *am*^{1}*-satañ*^{3}
monk extoll FUT
'the monks will extoll (my donation)'

(10) *ŋa*^{2}*-hnaŋ*^{1} *thap-tu*^{2} *ra*^{1} *am*^{1}*-satañ*^{3}
I-with equally receive FUT
'Equally with me (they) will be rewarded'

(11) *awici*^{2} *ŋray*^{3} *la*^{2} *am*^{1}*-satañ*^{3}
Avici hell come FUT
'(They) will fall to the hell Avici'.

The translation of the examples, reflecting the meaning of *am*^{1} as a future tense marker, holds good from the point of view of Modern Burmese. But in the inscriptions, with the contextual background, the sentences implicitly express also a desiderative sense, and therefore should be translated approximately as 'let the monks extoll in the future', 'let be rewarded in the future', 'let fall to hell in the future'. Besides the general context, one more peculiarity tells us that the desiderative sense is present in sentences with *am*^{1}. It appears that such sentences are used arbitrarily with causative constructions, and the choice between them is likely to be absolutely subjective. Thus, along with examples such as those cited in (9) through (11) above, those in examples (12) through (14) are quite common in the same context.

(12) *rahan*^{3} *khyi*^{2}*paŋ*^{1} *ciy*^{2} *satañ*^{3}
monk extoll CAUS *satañ*^{3}
'let the monks extoll'

(13) *thap- tu*^{2} *ra*^{1} *ciy*^{2} *satañ*^{3}
equally receive CAUS *satañ*^{3}
'let equally be rewarded'

(14) *ŋaray*^{3} *la*^{2} *ciy*^{2} *satañ*^{3}
hell come CAUS *satañ*^{3}
'let fall to hell'.

The only difference between the two constructions might have been that constructions with *am*^{1} were preferable in sentences with a positive sense: the monks will extoll, will be rewarded, and so on, whereas causative constructions were more frequent with a negative sense: let fall to hell, let not pray to Buddha and so on. Of course, some examples containing *am*^{1} in the inscriptions, such as those in (15) and (16), seem to express 'simple' future tense.

(15) *ŋa*^{2} *plu*^{1}*-la*^{2} *am*^{1}*-satañ*^{3}
I make-come FUT
'I will do (it, if my life is long)'

(16) *chu*^{1} *ma-lwai*^{3} *ra*^{1} *am*^{1}*-satañ*^{3}
reward for sure get FUT
'By all means (I) will get a reward'.

But from the following discussion it will be seen that even in such sentences we have to infer a desiderative sense.

The described functioning of *am*^{1} leads us to the conclusion that its real meaning was not purely temporal, but seems to be very close to the meaning of the Pali optative suffix, which is usually defined as expressing desire.

In addition to its Pali-like meaning, the spelling of *am*^{1} reveals its borrowed nature also. It is spelled with *anusvara*, which is used in the inscriptions with Pali loans, but not for Burmese words. If *am*^{1} were of Burmese origin, it would have been spelled with the letter *ma* with a 'killer' (*virâma*).

In Pali there are several suffixes for the optative, depending on the person of the verb. One of the suffixes is *am*. Verbs in Burmese do not differentiate person, and from this point of view all Pali suffixes for the optative should have been the same for the Burmese, so why they chose *am* remains unclear. It might have been that in some Pali text through which the Burmese first got acquainted with Pali, the suffix *am* was the most frequent.

We recall that in the inscriptions the marker *am*^{1} was the only grammatical means by which future tense could be expressed. This might lead us to the conclusion that by the time the first Burmese inscriptions were created, the Burmese language had not yet developed the grammatical category of tense, because *am*^{1} is not a pure tense marker. Of course, the absence of any other grammatical elements with future time sense might be explained by the peculiarity of the style of the inscriptions, which were, as mentioned above, dedicatory ones and almost all of the same composition, which generally did not

presuppose the mention of future time events. But even in such texts we do come across sentences where, according to the basic grammatical rules of Burmese, a future tense marker would be compulsory, as in example (17), where the verb *phyak chi*^{3} 'to destroy' is followed by the past-present tense marker *i*^{1}, although future time in the sentence is expressed lexically—*noŋ*^{2} *a*^{3} 'in future'. In Modern Burmese in such cases the marker of the future tense would be used instead of *i*^{1}.

(17) *noŋ*^{2}-*a*^{3} *sa*^{3} *mliy*^{3} *phyak-chi*^{3} *i*^{1} *hu*^{2} *mu*^{2}*ka*^{3}...
future-in son grandson destroy PASTPRES. Say if...
'If in future my children or grandchildren will destroy (my donation)...' (Lit.: in future my children, grandchildren destroyed. If say so...).

Examples similar to this are not rare in the inscriptions, and if there had been a native future tense marker in Old Burmese, it would be natural to expect it in such phrases.

It would be too hasty to conclude that the Burmese really did not have the grammatical category of tense in the period in question. But it is also impossible to deny that the marker *i*^{1} is sometimes used in inscriptions in a peculiar way. Analysis shows that in each case where the marker *i*^{1} is used in sentences with future time sense, the non-desiderative sense also is always present. It is obvious in example (17) above, as well as in examples (18) and (19).

(18) *noŋ*^{2}-*a*^{3} *kambha*^{2} *pyak ci*^{3} *i*^{1} *hu*^{2} *mu*^{2}*ka*^{3}
future-in world be destroyed PASTPRES. Say if..
'If in the future the world will be destroyed...' (lit.: in future the world is destroyed. If say so...);

(19) *noŋ*^{2} *ŋa*^{1} *ahlu*^{2}-*kiu*^{2} *lu*^{1}-*ca*^{1} *i*^{1} *hu*^{2} *mu*^{2}*ka*^{2}
future my donation-OBJECT infringe PASTPRES. Say if...
'If (someone) will infringe on my donation in the future...' (lit.: in future my donation someone infringed. If say so...).

Only if we accept the point of view that the marker *am*^{1} derives from the Pali optative suffix does the described use of the marker *i*^{1} become easy to explain—the semantics of the optative contradicts the non-desiderative semantics, so when the latter happened to be present in sentences with future time semantics, it was impossible to use *am*^{1}. Thus the special construction with *i*^{1} was introduced to cope with the situation.

The above arguments about *am*^{1} allow us to assume that it is of Pali origin and goes back to the Pali optative suffix.

We mentioned that in Modern Burmese *am*^{1} has become a future tense marker. It is also possible to trace the tendency of *am*^{1} to become a 'pure' tense marker in the inscriptions—at least some of the examples above allow such speculation. Tense markers occupy the final position in verbal syntagma, but *am*^{1} does not, and this is another problem to be solved concerning it.

Since *am*^{1} is a suffix in Pali, it does not occupy final position in Pali words; usually it is followed by endings. The Burmese, who were very careful to follow Pali patterns, faced an ambiguous situation when borrowing a Pali suffix: on the one hand, it could not have been used as the final element of verbs, because it had not been so used in Pali, while on the other hand, the semantics in which it was utilized in Burmese supposed the use of the new marker as a final element of verbal syntagma. The tone of the marker *am*^{1} helps to explain how the Burmese found a way out of this situation.

Usually short Pali loans in Burmese have the second, level tone. The marker *am*^{1}, however, has the first, creaky, tone. The two main tense markers in Modern Burmese, viz., *sañ*^{2} for past-present tense, and *mañ*^{2} for future tense, have the second tone. When a whole sentence is used attributively to an NP in Modern Burmese, the markers change their tone from second to first, as shown in examples (20) through (23).

(20) *ŋa*^{2} *swa*^{3} *sañ*^{2}
I go PASTPRES
'I go (went)',

(21) *ŋa*^{2} *swa*^{3} *sañ*^{1} *im*^{2}
I go PASTPRES house
'The house, to which I go (went)'.

(22) *ŋa*^{2} *swa*^{3} *mañ*^{2}
I go FUT
'I will go'

(23) *ŋa*^{2} *swa*^{3} *mañ*^{1} *im*^{2}
I go FUT house
'The house, to which I'll go'

It can be seen that under certain circumstances tense markers can change their position in sentences and lose their function as markers of the finite form of verbs. The construction V*am*^{1}-*satañ*^{3} is absolutely parallel to the above examples, in which whole sentences are used as attributes to NPs—the component *sa* from *satañ*^{3} is the NP (Pali third person pronoun), to which the preceding part of the sentence serves as an attribute and is connected to the NP by the

tense marker *am*^{1}. The example in (9) above, 'Let the monks extoll (my donation)' must formally be translated, literally, 'the monk let be extolling he'. As a matter of fact, it is exactly such a way of expressing the proposition 'the extolling (of my donation) by monks', that would be preferable in Pali.

The fact that *am*^{1} has the first, creaky tone perfectly confirms the proposed treatment of the analysed Old Burmese construction. Otherwise it would be impossible to explain why this particular Pali word has the creaky, and not the level tone, which would be natural for such a borrowed word.

The analysis of the Old Burmese construction V*am*^{1}-*satañ*^{3} once more demonstrates how the Burmese managed to achieve a compromise between the laws of Pali and those of their own language: on the one hand, *am*^{1} does not occupy final position in verbal syntagma, which corresponds to how it functions in Pali, while on the other hand in Burmese it functions in the same way as the native tense markers can function.

The introduction of the Pali pronoun *sa* into the Burmese verbal syntagma, and by this means converting the finite form of the verb into a nominalised one, led, in my opinion, to one more important implication for the grammar of Burmese.

In Modern Literary Burmese two markers of the past-present tense, viz., *i*^{1} and *sañ*^{2}, are the most common. In dictionaries both markers are usually treated as synonymous. The usual explanation of the difference between the two is that the marker *i*^{1} may be of higher rank, a bit archaic, and more often used in Nissaya styled texts. In fact, however, there is an essential difference between the two markers.

The marker *sañ*^{2} is widely used with predicates of embedded clauses in compound sentences. The marker *i*^{1} cannot be used in this function. But both can be used with the main predicate of independent sentences, and the choice between them is arbitrary. In compound sentences *sañ*^{2} is used with phrases functioning like NPs, and as such can be followed by grammatical markers used with nouns, as in example (24).

(24) *ŋa*^{2} *swa*^{3}-*sañ*^{2}-*kiu*^{2} *su*^{2} *si*^{1}-*sañ*^{2}
I go-PASTPRES-OBJECT he know-PASTPRES
'He knows (knew), that I go (went)' (lit.: that I go he knows).

In example (24) the marker i^1 can be used instead of $sañ^2$ with the main predicate si^1 'to know', but not with swa^3 'to go'. Practically all grammatical markers used with nouns can be used with $sañ^2$ in subordinate clauses. The marker i^1 cannot combine with grammatical markers, maintaining the function of nouns.

Let us try to discover why the two synonymous tense markers function so differently.

The marker $sañ^2$ has several homonyms, which form separate entries in dictionaries. Two of them bear further discussion. In the most comprehensive and reliable, although somewhat obsolete, Burmese lexical work, Judson's *Burmese-English Dictionary*, we find the following meanings for the two homonyms in question: 1. an owner, proprietor, e.g., $kun^2sañ^2$ (kun^2 'goods') 'a trader'; 2. affix, nominative; denoting the agent or subject, e.g., $miu^3sañ^2\ rwa^2i^1$ (rain-$sañ^2$-fall-i^1) 'rain falls'. In this example, $sañ^2$ could be substituted for the final marker i^1, but due to stylistic reasons i^1 is used (two homonyms should be avoided in short sentences). The meaning given in Judson's dictionary for the first morpheme $sañ^2$ 'an owner, proprietor' certainly does not account for all usages of it. In fact, its meaning is very broad and it cannot easily be translated correctly. Further examples demonstrating the meaning of this morpheme are: $thamaŋ^3sañ^2$ ($thamaŋ^3$ 'cooked rice') 'a cook', $haŋ^3sañ^2$ ($haŋ^3$ 'garnish') 'a cook', $eñ^1sañ^2$ ($eñ^1$ 'stranger, guest') 'a guest', $jhe^3sañ^2$ (jhe^3 'market') 'a trader', $dayaka^2sañ^2$ ($dayaka^2$ 'giver, benefactor') 'a giver, benefactor'.

The morpheme $sañ^2$ could be used not only within compounds, but also independently, in this case with the noun prefix *a*, which does not have any meaning—it is just a part of some Burmese nouns and is omitted when the latter are used in compounds. The functioning of $asañ^2$ is shown by examples (25) and (26).

(25) i^2 $asañ^2$ sum^3-*yok* ... $naŋ^3$-$satañ^3$
this specialist(?) three-CLF[+HUMAN]... entrust-$satañ^3$
'to these three specialists...entrusted';

(26) $rahan^3$ $sathe^3$ $asañ^2$ kra^3-$satañ^3$
monk rich specialist witness-$satañ^3$
'The monks, the rich, the specialists witnessed'.

The demonstrated functioning of $sañ^2$ ($asañ^2$) is common already in the earliest inscriptions. But we don't find at this time the two other homonyms of $sañ^2$, viz., the tense marker and subject (agent) marker. They appear by the end of the 13th century.

Having in mind how the Burmese utilised the Pali pronoun *sa* in verbal syntagma, it would not be unnatural to assume that the two grammatical elements *sañ*^{2}, viz., subject and tense markers, had originated from the above described morpheme *asañ*^{2}. The logic of the early Burmese might have been as follows.

In Pali the agent and the action are grammatically interdependent —the predicate must agree with the subject in person and number. In Burmese, by contrast, there is no formal concord between agent and action. But there is, and was already in the earliest inscriptions, a very strange peculiarity, which in all probability has no logical explanation and therefore may be treated as incidental: the familiar past-present tense marker *i*^{1} has a homonym, which is a possessive marker. It looks as if the same grammatical marker is functioning with both nouns and verbs. Consider example (27), which demonstrates the functioning of *i*^{1}:

(27) *aphe*^{2}-*i*^{1} *im*^{2} *pyak-ci*^{3}-*i*^{1}
father-*i*^{1} house be destroyed-*i*^{1}
'The father's house is destroyed'.

Since the possessive marker is positioned within the NP, the case might be treated as some sort of agreement between agent and action.

Now let us turn to the semantics of the nominalised verb formed by adding the Pali third person pronoun *sa* to Burmese verbs, which combination resulted in the above-described construction V*satañ*^{3}. The latter was used to represent Pali participles in finite function. Since *sa* is a third person pronoun, the construction V*sa* naturally would have been perceived as a third person participle even though the pronoun which defined it was alien. Burmese verbs are indifferent to person, but for learned monks who were the highest authorities in grammar, a situation where the agent would be expressed by the first person pronoun and the action by the third person participle should have seemed inadequate. Another thought that might have bothered pandits' minds could be that if it was possible to express the peculiarities of Pali by native means, why not do it. It is not difficult to treat the above described Burmese morpheme *sañ*^{2} as having the invariant meaning 'the one who...', which agrees with all persons. If used instead of Pali *sa* in the construction V*sa*, it would perfectly satisfy any desire to depersonify the equivalent of Pali participles. So gradually the construction V*sañ*^{2} started to be used along with V*satañ*^{3}, the two constructions being synonymous. The difference

between them was that the latter contained a Pali word, whereas the former was purely native, although it originated due to Pali interference.

In the course of time, the lexeme *sañ*^{2} became grammaticalised, the same as *satañ*^{3}, and became, along with *i*^{1} and *satañ*^{3}, a tense marker. Our assumption, that originally the construction V*sañ*^{2} constituted a nominalised verb is perfectly supported by the fact that it can be followed by grammatical markers used with nouns.

The agent or subject marker *sañ*^{2} also seems to have originated from the lexeme *sañ*^{2} (*asañ*^{2}). Maybe at the beginning the reason for using *sañ*^{2} with agents was the intention of the Burmese to follow Pali patterns according to which agents and actions have to agree. In this case, if *sañ*^{2} occurs with the predicate, it should also be used with the subject. Such logic might have been supported by the coincidentally homonymous shape of the possessive marker and the tense marker. It may also be that frequent utilization of the lexeme with its rather abstract meaning in sentences containing the enumeration of different agents—see example (26) above—played a part. Probably all these reasons influenced the practice of using *sañ*^{2} with NPs in other than its original function. It is natural that it got grammaticalised, and there is nothing in Modern Burmese that would make us suspect that the three homonyms *sañ*^{2} were once one lexeme.

In the above notes I have tried to show that Pali influence on Burmese was essential and manifold, and affected important spheres of Burmese grammar. Suffice it to say that the most usual way of expressing past-present tense in Modern Burmese owes its existence to Pali, while certain attributive constructions were introduced into Burmese grammar as a result of attempts to imitate Pali. Pali influence can be traced even in cases where no loans are involved. No doubt many more instances of Pali-Burmese interference might be discovered if research in this direction is given proper attention.

References

Allott, A. 1965. Categories for the description of the verbal syntagma in Burmese. *Lingua* 15: 283-309.

Judson, A., ed. Robert C. Stevenson 1893. *Judson's Burmese-English dictionary.* Rangoon: Governnment Press (reprinted New Delhi: Asian Educational Services, 1993).

Kassevitch,V. 1990. O kategoriyax vremeni i taksisa v birmanskom yazïke ['On the categories of tense and taxis in Burmese']. *Vostokovedenie* 16: 37-45.

Okell, J. 1967. Nissaya Burmese. *Journal of the Burma Research Society*, L.1: 95-123.

—— 1969. *A reference grammar of colloquial Burmese.* London: Oxford University Press.

Yanson, Rudolf 1993. Mon and Pali influence on Burmese: how essential was it? Uta Gärtner et al. eds. *Tradition and modernity in Myanmar: proceedings of an international conference held in Berlin from May 7th to May 9th,* 1993. Münster: LIT, 356-373.

—— 1997: O nekotorïx složnïx slučayax interpretacii grafičeskoy i foneticeskoy formï sanskritsko-paliyskix zaimstvovaniy v birmanskom yazïke ['On some complicated cases of interpretation of the graphic and phonetic form of Sanskrit-Pali loans in Burmese']. *Vostokovedenie* 19: 58-64.

This page intentionally left blank

EARLY MEITHEI MANUSCRIPTS[1]

SHOBHANA CHELLIAH (UNIVERSITY OF NORTH TEXAS)
AND
SOHINI RAY (HARVARD UNIVERSITY)

INTRODUCTION

We present a multifaceted philological study of early manuscripts in the Meithei language. Meithei, also known as *Manipuri* and *Meitheiron*, is a Tibeto-Burman language spoken in the Indian state of Manipur, which is bordered by Myanmar (Burma) to the East, Mizoram to the South, Nagaland to the North, and Assam to the West and Northwest. Early Meithei manuscripts and inscriptions are yet to be studied by linguists or anthropologists on a large scale. This is unfortunate, since the study of these manuscripts would reveal much about the linguistic, anthropological, and literary history of Meithei. Here we present information on eight aspects of the manuscripts:

(1) script
(2) number and genre
(3) dating
(4) authorship
(5) housing and ownership
(6) paper and condition
(7) status in the culture
(8) vocabulary and texts

1. *Script*

There are two scripts used in the writing of early Meithei literature: the Meithei Mayek and Bengali scripts. In the classification of scripts by K.S. Singh and Manoharan (1993: 26-29), Meithei Mayek is part of the Tibetan group of scripts, which originated from the Gupta Brahmi script. The original Brahmi script was modified to ac-

[1] This work was supported by an American Council of Learned Societies Fellowship (2000-2001) for Shobhana L. Chelliah and a Post-Doctoral Fellowship at the Harvard School of Divinity for Sohini Ray (2000-2001).

commodate the phonemic distinctions of Meithei. Like other Brahmi scripts, Meithei Mayek is syllabic. We refer the reader to Chelliah (1997a) for a description of this writing system.

The earliest evidence of writing in Meithei Mayek is 8th century inscriptions on a copper plate (N.K. Singh 1975: 10-11, C.M. Singh 1984: 23). Longer early attestations are stone inscriptions from the 16th and 17th centuries (N.K. Singh 1975: 10-18, M. Bahadur and P.G. Singh 1986).

The Bengali script was adopted in the late 1700s for the following reason. In 1729 A.D. the Meithei people, while never completely renouncing their original beliefs and practices, adopted Hindu practices. This change had enormous implications for the language and culture of the Meithei people. Many works about the pre-Hindu religion and other historical documents written in the original Meithei script, Meithei Mayek, were burnt at the time of the conversion. The Bengali script was adopted for the writing of Meithei. Indigenous literary productions were all but replaced by translations of Bengali religious and secular works. Due to its status as a sacral language and its literary and cultural ascendance in India during the 18th century, the Bengali language (used in original works as well as in works translated from Sanskrit) was accorded high prestige in Manipur. Representations of sacred scripture in Bengali, and contact with migrating Bengali and other Indo-Aryan speakers, led to the large-scale borrowing of lexical items into Meithei, which in turn led to significant structural changes in Meithei (Chelliah 1997b). As a result, 20th century Meithei is markedly different from 16th century Meithei.

Since early forms of Meithei Mayek have not been in use for over two hundred years, few people today can read the script. There is currently a move to revive Meithei Mayek in a modified form; while the old script contained 35 letters, the new one consists of 27. The reasons for this revision are discussed in Ray (2000). R.K. Sanahal Singh's *Catalogue* (see below) lists 437 pre-20th century manuscripts as written in Meithei Mayek and 176 in the Bengali script. Thus knowledge of the older script, Meithei Mayek, is crucial to understanding the wealth of available early Meithei literature.

The manuscripts in Meithei Mayek are all handwritten in a form that was not standardized among scribes, so that some idiosyncratic writing conventions do exist. For example, some scribes use a striking method of abbreviation with multiple diacritics on a single

consonant grapheme where adjacent syllables have the same consonant onset. Thus the word *pepupa* 'the carrying of an umbrella' could be written with three vowel diacritics on a single 'p' grapheme. In order to read the manuscripts written in Meithei Mayek, therefore, one must possess a knowledge of the more common writing conventions utilized by the scribes. Currently, there are only a few Meithei scholars who have this knowledge.

2. *Number and genre of manuscripts*

State-of-the-art documentation about the manuscripts can be found in *A Catalogue of Manipuri Manuscripts* (N.K. Singh 1984) and *A Catalogue of Old Manipuri Manuscripts* (R.K. Sanahal Singh 1992). R.K. Sanahal Singh's catalogue (henceforth *The Catalogue*), provides a list of all the early manuscripts currently known to exist in Meithei. The total number of different manuscripts listed in *The Catalogue* is 380; taking multiple versions of the same text into consideration, the number is 613. Table 1. lists the general category and the number of manuscripts per category listed in *The Catalogue*).

Table 1: Categories and Manuscripts per Category Listed in *The Catalogue*

Administration 22
Astrology and Astronomy 35
Charms and incantations 105
Craftsmanship 2
Creation Myths 17
Genealogy 61
History 82
Literature 65
Medicine and Physiology 7
Prophesies and predictions 11
Religion 130
Songs and Dances 13
Topography 28
Miscellaneous 35

What we do not have is a list indicating which manuscripts have been published and in what versions these published manuscripts are available. Furthermore, we do not know what transcriptions and/or

transliterations are available: are they in the old Meithei Mayek script, transliterated into modern Meithei Mayek, translated into Modern Meithei in the Bengali script, or translated into English, Bengali or some other Indian language?

Tables 2-4 provide sample titles of manuscripts from pre-20th century texts. Table 2 lists manuscripts on Manipuri history divided into the traditional categories of Immigration chronicles, Military chronicles, Royal chronicles, Family chronicles (i.e. genealogies) and Clan chronicles. Table 3 lists creation myths. Table 4 lists botanical and environmental treatises.

Table 2: Manuscripts on Manipuri History
Divided according to Traditional Categories

Kunthoklon (Immigration chronicles)

- *Poireiton Khunthokpa* (Immigration of Poireiton), an epic about the migration of Poireiton and his people to Manipur. He was invited to rule a small principality under the control of the Ningthouja clan.

Ningthou wari (Military chronicles)

- *Takhshenngamba*, an account of a war between the Manipuri King Garibniwaz and Tripura.
- *Ningthouron Lambuba* (Road taken by the king) which describes the military expeditions of the kings of Manipur. Although events are listed chronologically, no dates are provided.

Ningthourol (Royal chronicles)

- *Chainaron* (The art of war) outlines rules of chivalry and tells stories about the settling of conflict through deadly duels.
- *Cheitharol Kumbaba* (Account of the years) records astronomical events, epidemics, and the results of wars. It records events during the reign of King Garibniwaz between 1709-1748 (R.J. Singh 1998: 22, 85).
- *Loiyamba Shilyen* (Loiyamba oversees work), written during the rule of Loiyamba between 1074 and 1122 A.D., it lays out a plan for assigning occupations to families, the duties of the King, rules for rewards, rules regarding the wearing of costumes, and rules for the administration of justice (Sanajaoba 1993).

Yumdaaba Puyaa (Family chronicles)

• Each head of household was (and still is upon request) given a copy of the family genealogy as recorded by the Royal Archives. Approximately four hundred and seventy genealogies from the pre-20th century period are available at the Manipuri Sahitya Parishad (Manipur's National Academy of Letters) according to N.K. Singh (1984).

Salaai meihourol (Clan Chronicles)

• *Moirang Kangleiron* (The story of the Moirang Clan)
• *Chengleirol* (The story of the Chenglei chiefs). The Cheng-lei clan is one of the seven original clans of Manipur. The stories recorded here go back to the 4th and 5th century A.D.

Table 3: Legends and Creation Stories

• *Khongjongnubi Nongaron* (translation unknown), the legend of six girls who are ostracized by their parents for marrying men of their own choice. They are physically transformed into animals and transported into heaven.
• *Nungban Pombi Luwaoba* (translation unknown), the story of a lover who brings back the soul of his dead wife by overpowering death.
• *Leithak Leikharon* (The way of the world underground), a traditional creation myth influenced by Hindu creation mythology.

Table 4: Botanical and Environmental Treatises

• *Hidaklon* (Medicine language), a list of medicinal herbs and the ailments these can cure.
• *Salkau* (Cattle call), a treatise on cattle and the respect that should be shown to them.
• *Leiron* (Flower language), a description in verse of the many flowers that grow in Manipur.

3. *Dating*

Precise dating of the manuscripts is impossible because the colophons of most manuscripts have been lost or are illegible; of the

613 manuscripts listed in *The Catalogue*, only eight are dated.[2] Therefore, the dating of early Meithei manuscripts must be approximated through (1) mention of kings and their reign, (2) style of narration, (3) borrowings from Bengali, and (4) the script used.

For example, *Chainaron* (The art of war) is dated in *The Catalogue* as a 16th century manuscript probably because of the mention of king Mangyamba of Manipur who ruled at that time. Early texts also mention the recital of devotional poetry or long narrative verse during the coronation of particular kings.

Manuscripts that reflect Aryan influence are dated to the late 1700s through the 1800s, coinciding with the popularity of Hinduism. Narratives that rely heavily on Hindu mythology and show lexical borrowings from Bengali can also be dated as being written after the late 1750s. (R.J. Singh 1969: 6). It is also possible to date manuscripts as being pre or post conversion by looking at the script used, since the Bengali script is used after the adoption of Hinduism.

4. *Authorship*

The authorship of forty-six of the 613 manuscripts in *The Catalogue* is known. Writing was done under the patronage of the king, whose administration was divided into eight departments known as *loishang*s. The foremost among these was the Pandit Loishang (the Royal House of Scholars). Pre-18th century literature does not mention authors since the Pandit Loishang encouraged poets and scholars not to disclose authorship (N.K. Singh: 1984: i). In some instances, one scribe copied the works of many authors. Manuscripts were bound together by their owners, so that a single bound copy may contain works by a scribe or group of scribes and one author or several authors (R.K. Sanahal Singh: 1992: i).

5. *Housing and ownership*

The pundits, or traditional scholars, own the handwritten manuscripts, which are handed down to them from generation to generation. In 1984, the state government of Manipur founded the Manipur

[2] The method of dating, however, is curious. The manuscript Parikshit (an adaptation of an episode from the *Mahâbhârata*) written under the patronage of Garibniwaz, is dated to 1725. Each digit of the date is represented by a word that has numerical connections in Hindu mythology (R.J. Singh 1998: 28).

State Archives, and in the past years has successfully archived many of these valuable documents. They are thus available in microfiche at the Manipur State Kala Academy in Imphal, Manipur. Smaller collections are housed at the People's Museum at Kakching, a town southeast of Imphal.

In the late 1990s, under the auspices of the Indira Gandhi National Center for Arts in New Delhi, a large number of the original manuscripts were microfilmed and may be viewed there. Full copies of the manuscripts are not released by the Center.

6. *Paper and condition*

The earliest manuscripts are handwritten in Meithei Mayek on *agarbak*, a paper derived from the bark of a tree. Pens are made out of bamboo. Another technique used was to blacken the paper with charcoal and use a soapstone pencil for writing. *The Catalogue* indicates that some manuscripts were written on paper made of pressed wood pulp. There was an organized paper making industry in Manipur by the 1700s (R.J. Singh 1998: 11). Materials used must have been durable since the condition of most are listed as "good"; only a few are "old with pin holes."

7. *Status in the culture*

We have found that speakers view the pre-20th century literary corpus with great reverence. In recent years there has been a religious revival movement in Manipur, which aims to eliminate Hindu elements from Manipuri life and revive the ancient pre-Hindu religion, named *Sanamahism*. The effort to return to pre-Aryan times has placed even more value on the oldest manuscripts, as they have come to represent the time before the influence of Hinduism.

Younger speakers seem unaware of the number of older manuscripts still available for study. There is a certain reluctance to attempt translation of them because the archaic language is popularly perceived to be unintelligible. It is believed that such work can only be done by older Manipuri language experts. Younger speakers do, however, understand individual words from the texts or they can identify some contexts where they have heard the word used. One goal of our ongoing investigation is to determine whether the language used in the texts is markedly different from Modern Meithei or if the cultural value placed on the manuscripts has given

the texts this reputation. Further comments on this point are in section 8.

8. *Vocabulary and texts*

There are four main published sources for Archaic Meithei vocabulary. The first is N.K. Singh's *Manipuri to Manipuri and English Dictionary* (1964), in which he includes 254 words that he labels "archaic." It is not clear whether these are compiled from pre-20th century manuscripts or if they simply have restricted functionality—for example, perhaps they are used only in poetic or sacred texts in Modern Meithei. In the majority of cases the "archaic" words and the modern Meithei equivalents are not phonetically similar and cannot be related through sound change. In cases where the archaic and modern words are similar, the modern form is often a shortened version of the archaic compound, as shown in Table 5.

Table 5. Shortening of Archaic Meithei Compounds in Modern Meithei

ENGLISH	ARCHAIC MEITHEI	MODERN MEITHEI	ETYMOLOGICAL NOTES
button	punuŋ həyru	hayru	*punuŋ* 'shirt'
destination	waythuŋphəm	thuŋphəm	*way* 'whereabout', *thuŋ* 'reach' *phəm* 'place'
fishing trap	lolu	lu	*lon*, 'weave', *lu* 'trap'
fishing trap	luphui	lu	*phui* 'bear, give birth'
house	lol yim	yum	*yim* is presumably the same as *yum* 'house'
to call	pəw kəwbə	kawbə	*pəw* 'news' *kəw* 'call'
to fear	soŋ kiba	kibə	*soŋ* 'dense' *ki* 'fear'
to seat	phəmphəm bə	phəmbə	*phəm* 'seat' (noun) 'seat'
tusk	ləŋŋoy moysəyə	səmmu məya	*ləŋŋoy* 'elephant', *səmmu* 'elephant'.[3]

[3] Note that *moysaya* (possibly *moy* 'their' + *sa* 'body' + *ya* 'tooth') has changed to *məya* (*mə-* 'third person prefix' + *ya* 'tooth').

In order to present a consistent transcription system, we have adopted the following conventions. Aspiration is indicated with a following [h]. The palatal affricates and fricative are transcribed as [c]/[j] and [s], respectively. As neither of our secondary sources indicates tone consistently, tones are not indicated in the transcription.

The three other sources for Archaic Meithei words are dictionaries by N.K. Singh (1978), Shamkishore (1999) and N.D. Singh (1960). These three dictionaries list words in the Bengali script from manuscripts in their private collections and provide definitions and cultural commentary for these in Modern Meithei. N.K. Singh includes sample sentences from the manuscripts.

We have compiled a comparative list of basic Archaic and Modern Meithei words taken from N.K. Singh (1964 and 1975) and from three manuscripts that we have partially translated. This list is given in 'A glossary of 39 basic words in Archaic and Modern Meithei' (in this volume). As can be seen from the list, there is little difference in these basic words between Modern Meithei and the Meithei found in the manuscripts.

Similarly, if the language of the stone inscriptions is compared to Modern Meithei, the differences are minimal. To illustrate this, we provide the first seven lines of an inscription with English and Meithei equivalents. The spelling conventions, transliterations and translations are taken from M. Bahadur and P.G. Singh (1986:1-3). The top line represents the original inscription, the second line is a Modern Meithei equivalent, and the third line provides a word gloss. The selected inscription is not dated, but taking into consideration the mention of King Kaagingamba (who reigned between 1592-1652 AD.) and King Caraairongba (who reigned between 1697-1709 AD.), it is surmised that the inscription was made during one or each of their reigns. The inscription is on sandstone in a town ten kilometers to the west of Imphal, the capital of Manipur.

1. *Laailemaa* *chingthaalaalimaamubhu.*
   Laairemma *chingtharaklibabu.*
   Lairemma which was brought down.

2. *Khaakhaaikhingalbhaana khangdaadunaa. thak*
   Khaakhaaikhingalbhanaa khangdaduna. thak
   by Khaakhaaikhingalbha because he was unaware. on

3. *khudadaa* *kum.* *Khouthabhaakaa*
   khukaada *kumme.* *kouthabaga*
   the seat descended. as it was invoked

4. *aangaale.* *aakiba.* *kaakei.*
   ngaallee. *akiba.* *kaakei.*
   brightened. by a radiance of light. frightened.

5. *Laarembhaanaa* *thouchaalchaarambhaame.*
   Laairembina *thoujanbiriramme.*
   by the goddess graced with kindness.

6. *Kamthouchaasalkaa thounaadabha amarak ma Haaraak.*
   Konthoujamga *thounaaidaba* *amurak* *hannaa* *Haaorok*
   Konthoujam on the service once more Haaorok

FREE TRANSLATION: The likeness of the goddess, which was brought down by Kaagingamba because he was unaware, descended on the seat. As an invocation to the goddess was performed a bright radiant light frightened the king. Graced by the goddess with kindness, then Konthoujam Haaork once more continued in her service.

The consistent differences between the two forms of the language in these passages are spelling differences that may or may not indicate actual differences in pronunciation. We need to learn more about the spelling conventions of the 17th and 18th century before we can be certain about the significance of the spelling discrepancies. The spelling differences are: *bh* (rather than *b*) in the original inscription in lines 1, 3, 5, and 6. This is curious, as the voiced stop series entered the phonemic inventory after large scale borrowing from Bengali in the late 1700s; the vowel [a] is represented as *aa* in the original inscription. At times the Modern Meithei equivalent uses *a* (presumably [ə]), and at times it remains *aa*. Finally, *j* in Modern Meithei occurs as *ch* in the original inscription. There are few morphological or lexical differences between the two versions. The same can be said of the 12 inscriptions described in M. Bahadur and P.G. Singh (1986).

It is striking, therefore, to discover that manuscripts thought to be written before the 17th century, such as *Numit kappa* (The shooting of the sun), are barely intelligible to the Modern Meithei speaker. It may be that while some early manuscripts were written in older forms of Modern Meithei, other documents were written in some other variety quite divergent from Modern Meithei. There were, after

all, seven clans and seven corresponding sub-clans living in the Manipur plains in the 16th and 17th centuries. Promodini Devi (1989: 16) and N.K. Singh (1964: 10) list the following clan names: Ningthouja, Khumal, Angom, Luwaang, Moiraang, Khaabaa-Ngaanbaa and the Chenglei-Saraang. N.K. Singh lists two more names, Leisaangthem, and Haaorok Konthou. These may be sub-clan names. It is likely that some of the manuscripts were written in varieties spoken by these groups before the Ningthoujam clan, the speakers of present day Meithei, finally gained ascendancy over the others (Damant 1877, N.K. Singh 1964, Promodini Devi 1989). Damant observes that the difference between the language spoken in Manipur in the late 1800s and the language in the manuscripts of the 1700s is so great that it can hardly be accounted for through language change. Instead, he feels that the manuscripts may well be written in whichever variety was predominant at that time (1877: 37). N.K. Singh attempts to show that "the language spoken by the people in those days recorded in the old manuscripts or Archaic Meithei bears testimony of different vocabularies..." (1975: 5) by tracing words from pre-20th century manuscripts to different clans. See Table 6.

Compiling a word list for Old Meithei is a difficult task, since it is not clear yet whether what is called Old or Archaic Meithei is in fact one dialect, several dialects or several languages. Progress will only be made if each available text is studied and word lists are made on the basis of these texts. Similarities and differences between the vocabularies in the texts could then be used to determine the number and relative divergence of the varieties involved.

To date, only two pre-modern Meithei texts are available in print outside of Manipur. *Numit kappa* (The Shooting of the Sun) is printed with a free translation in English and a partial free translation in Modern Meithei in Hodson (1908). A short selection from *Samsokngamba* (a history of the war between the Manipuri king Charairongba and his son Pamhaiba against the kingdoms of Burma and Sumjok) is available in Damant (1877) and Grierson (1903-1928) in Roman transliteration with a free translation in English. There are also courses in the study of selected Old Meithei texts offered at the Manipuri Languages and Literature Department at Manipur University. We have heard that for the purpose of these courses, translations and commentaries into Modern Meithei of five of these texts, in whole or in part, are available. We are attempting to collect these documents.

Table 6. Meithei Words from pre-20th Century Manuscripts Compared to Modern Meithei Equivalents and the Clans Which Use These Words

English Gloss	Modern Meithei Word	Word from Pre-20th Century Manuscript	Clan Name
axe	iŋjəŋ	haypi	Moiraang
bridge	thoŋ	tiŋpi	Luwaang
cock	yenba	soypay	Khuman
drum	puŋ	khoŋ	Selloi Langmaai
fire	məy	yay	Selloi Langmaai
gold	səna	coynəw	Selloi Langmaai
house	səŋgay	yakon	Khuman
iron	yot	tetnaw	Selloi Langmaai
lake	pat	kon	Khuman
land	kəŋphal	ya	Khuman
land	kəŋphal	koŋ	Moiraang
oar	nəw	tawtek	Khuman
road, path	lambi	muŋpi	Luwaang
to die	sibə	totpə	Selloi Langmaai
umbrella	yempak	waykəw	Luwaang
water	isiŋ	loklaw	Moiraang

Conclusion

This short survey gives us hope that early forms of Meithei and perhaps related varieties can be accessed for further study. Obviously, data from Meithei literature will greatly contribute to our understanding of the historical development of the Meithei language in particular, and the Tibeto-Burman language family in general. The relationship of Meithei to geographically adjacent Tibeto-Burman languages is obscured by the extensive borrowing from Bengali and other Indo-Aryan languages that occurred after the adoption of Hinduism; therefore, study of pre-conversion manuscripts will help clarify the relationship of Meithei to other Tibeto-Burman languages. Additionally, the comparison of older and newer Meithei will contribute to the development of theories of language change, adding to our understanding of how language contact, cultural and political history, and language structure itself influence and effect language change. Since the required knowledge of the Meithei Mayek script

and the vocabularies and stories held in these manuscripts is known only to a few scholars in Manipur, the continued documentation of this knowledge is a pressing concern.

REFERENCES

Bahadur, Mutua and P. Gunindra Singh. 1986. *Epigraphical records of Manipur*, Volume 1. Imphal: Mutua Museum.

Chelliah, Shobhana L. 1997a. *A grammar of Meithei*. Berlin: Mouton de Gruyter.

—— 1997b. Competing language ideologies in Manipur. In Alice Chu, Anne-Marie Guerra and Chantal Tetreault eds., *Proceedings of the fourth meeting of the Symposium about Language and Society—Austin*, 288-297. Austin: University of Texas at Austin.

Damant, G.H. 1877. Note on the Old Manipuri character. *Journal of the Asiatic Society of Bengal* XIVI: part i, 36-39.

Devi, Nameirakpam Promodini. 1989. Social grammar of Meiteilon. Unpublished Ph. D. dissertation, Manipur University, Imphal.

Grierson, G.A. 1903-1928. *Linguistic survey of India*. Volume III, Part 3. Calcutta: Superintendent Government Printing, India. Reprinted 1967, Delhi: Motilal Banarsidass.

Hodson, T.C. 1908. *The Meitheis*. Delhi: B.R. Publishing Corporation.

Ray, Sohini. 2000. The sacred alphabet and the divine body: the case of *Meitei Mayek* in Northeastern India. Unpublished Ph.D. dissertation, University of California, Los Angeles.

Sanajaoba, Naorem. 1993. *Manipur: treaties and documents* (1110-1971). New Delhi: Mittal Publications.

Shamkishore, Soybam. 1999. *Ariba Məytəy/Məytəy lóngi lóngəy*. ['Old Manipuri to Manipuri language dictionary'] Imphal: Published by the author.

Singh, Ch. Manihar. 1984. The Meiteis: origin and affinity. *Journal of Manipur University* 1:1: 9-25.

Singh, K.S. and S. Manoharan. 1993. *Languages and scripts*. Delhi-Oxford: Anthropological Survey of India, Oxford University Press.

Singh, N. Dinachandra. 1960. *Khununglongi artha*. ['Meanings from the archaic language']. Imphal: Published by the author.

Singh, Ningthoukhongjam Khelchandra. 1964. *Manipuri to Manipuri & English dictionary*.

—— 1975. *Manipuri language status and importance*. Imphal: N. Tombi Ray Singh.

—— 1978. *Ariba Manipuri longei* [Old Manipuri to Modern Manipuri dictionary]. Imphal: Manipur State Kala Academy.

—— 1984. A *catalogue of Manipuri manuscripts*. Imphal: Manipuri Sahitya Parishad.

Singh, O. Bhogeshwar. 1977. *Numit kappa*. Imphal: Public Book Store.

Singh, R.K. Sanahal. 1992. *A catalogue of Old Manipuri manuscripts*. Imphal: The Manipur State Kala Academy.

Singh, R. Jhalajit 1998. *A history of Manipuri literature*, Volume II (1709-1819). Imphal: Published by the author.

—— 1969. *A brief survey of ancient Manipuri literature*. Imphal: Manipur Sahitya Parishad.

This page intentionally left blank

THE SUBGROUPING OF TIBETO-BURMAN[1]

DAVID BRADLEY (LA TROBE UNIVERSITY)

INTRODUCTION

There are three main tendencies in Tibeto-Burman and Sino-Tibetan reconstruction. One is a 'micro' approach using the traditional comparative method: internal reconstruction, reconstruction within closely-related subgroups, and then comparison of well-established reconstructed subgroups. Another is a 'macro' approach, called teleo-reconstruction by Benedict. This attempts to reconstruct the history of the entire family or large components of it from a direct comparison of forms in a variety of 'key' languages. A third is a 'megalo' approach, attempting to establish very remote relationships between generally-accepted language families. The third method is highly speculative, and the second has been productive in the past, but within Tibeto-Burman it is time to move on to more detailed studies at the micro level.

Beckwith (in this volume) canvasses some of the issues in the reconstruction of Sino-Tibetan, the relationship between Sino-Tibetan and Tibeto-Burman, and possible further links of Sino-Tibetan. Various megalo results have also attempted to extend Sino-Tibetan even further, initially notably by Sapir and Shafer, and more recently by Russian scholars. In all such cases the difficulty is to separate cognate material from contact material or chance resemblance. Some incorrect megaloclassifications have been proposed, and still persist in the literature. For example, the claim that Dai (in China, Zhuang-Dong) and Miao-Yao form part of Sino-Tibetan has been comprehensively disproved; the apparent similarities reflect long-standing contact relationships with extensive borrowings, mainly from Chinese. The difficulties are of course more subtle when the languages in contact are related to begin with. Conversely, when there is extensive contact between languages that are genetically and

[1] This is an extensively revised version of a paper originally presented at the 26th International Conference on Sino-Tibetan Languages and Linguistics in Osaka.

typologically distinct, like the Indic languages and some of the Indospheric Tibeto-Burman languages such as Newari or Mikir/Karbi/Arleng, some genetic characteristics may be obscured or lost.

Much macro-level work has gone into the subgrouping of Sino-Tibetan and Tibeto-Burman, but a lot of the basic descriptive and micro-level comparative work remains to be done. Some of the descriptive work is very urgent, as there are many Tibeto-Burman languages dying and others being heavily influenced by more dominant languages, some Tibeto-Burman or Sinitic and some not.

Some scholars, such as Beckwith (in this volume) and Sagart (1999), have suggested that part or all of the relationship between Sinitic and Tibeto-Burman is of a contact nature. Sagart has proposed to link Sinitic instead with Austronesian. Beckwith has noted typological similarities between modern spoken Tibetan and Altaic, as Hashimoto (1976) has seen between northern Mandarin and Altaic.

What follows is an attempt to place the various Tibeto-Burman languages into a likely classification schema. For a fuller presentation, see the maps and text by Bradley in Moseley (1994). The present classification is supplemented by observations on morpho-syntactic, phonological and lexical criteria, comments on some other classifications, and remarks on early orthographies.

There are two long-established and widely-cited classifications of Tibeto-Burman languages. One is Shafer's (1974), which splits it into four main parts: Bodic, Baric, Burmic and Karenic; the other is Benedict's (1972), with eight subgroups plus an 'other' category. These subgroups have been reclassified by Benedict (1976) into three groups. For a comparison of these and earlier classifications, such as Konow's in Grierson (still often used by scholars working with South Asian Tibeto-Burman languages), see Hale (1982).

With the recent addition of new data on languages of China and northeastern India, it has become clear that some modifications are needed. Specifically, some of Benedict's 'other' languages, the Xifan languages of western Sichuan, which were classified tentatively as Burmic by Shafer, form an additional group called the Qiang group by Sun Hongkai (1983) and here called Northeastern Tibeto-Burman. There is a somewhat problematic Central Tibeto-Burman group which includes languages that do not fit elsewhere; this includes the Adi-Mising-Nishi or Mirish languages, the Mishmi languages, and the Nungish or Rawang languages; also possibly Lepcha. The Central Tibeto-Burman languages and the Northeastern Tibeto-Burman

languages have been linked by Thurgood (1985) in his proposed Rung group, but this has not yet been demonstrated fully.

The classification of some languages is uncertain, in most cases due to very extensive contact: Bai and Tujia with Chinese, the Nungish and Naxi languages with Burmese-Lolo languages, Lepcha with a variety of Tibeto-Burman languages, Dhimal with Kiranti and possibly other languages, Newari due to its position at the crossroads of Nepal, and so on. The most thorough reconstructions done to date have been within the Kiranti branch of Western and the Burmese-Lolo branch of Southeastern, but some work has also been done on parts of all four main subgroups. The overall pattern may be summarized in Figure 1, which attempts to show both the geographical and the genetic distribution of the main branches of the Tibeto-Burman languages, as well as the placement of the early literary languages.

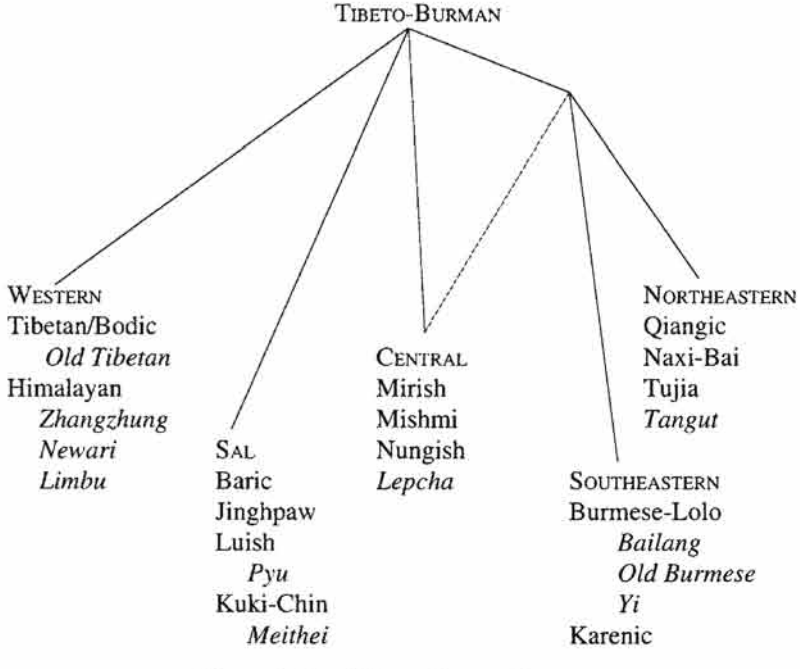

Figure 1. The Tibeto-Burman Languages

The Western group corresponds mainly to Shafer's Bodic group or Benedict's Tibetan/Kanauri plus Himalayan. Tibetan/Kanauri has four subgroups: Tibetan proper, the Gurung or Tamang subgroup, the

Eastern Bodic, Monpa or Tsangla subgroup, and the more distinct Kanauri subgroup, also known as West Himalayish or Western Pronominalizing from Konow. Himalayan, van Driem's Mahakiranti (2001), falls into a relatively homogeneous Kiranti, Rai-Limbu or eastern subgroup and a more disparate western subgroup including various languages classified elsewhere by Shafer including Newari, Chepang, Magar, Thangmi and Baram. Most of these languages have extensive morphology, especially on verbs; some, especially in the Gurung/Tamang subgroup, have word rather than syllable tone systems, often involving phonation as well, and many are nontonal. Most Himalayan languages have been heavily influenced by Indic languages, in lexicon and even to some extent in morphosyntax. For this reason the position within Tibeto-Burman of Newari, the most strongly influenced and longest-documented Himalayan language, has sometimes been regarded as uncertain. Apart from Tibetan, the main early literary language here is Zhangzhung, which is now agreed to have been a Kanauri or West Himalayish language.

The Sal group includes Shafer's Baric/Benedict's Bodo-Garo-Konyak groups, plus Jinghpaw (Kachin) and the Sak or Luish group. The subgroup and its name were proposed by Burling (1983). It has some lexical peculiarities not shared with other Tibeto-Burman languages; most languages have some morphology including parts of that reconstructed for Proto-Tibeto-Burman by Wolfenden (1929), but relatively simple tonal systems. Shafer's Kukish/Benedict's Kuki-Chin-(Southern) Naga is more distantly linked with Sal, though this also shows some lexical links with Burmese. Pyu appears to belong in the Luish group. Within the Sal group, the Kuki-Chin language Meithei is the longest-established living literary language.

The Central group may actually be a residual category, as the internal differences are very large. It comprises the languages of the border between northeastern India and Tibet, and also northern Burma and adjacent areas of China. The three subgroups are the Mirish or Adi-Mising-Nishi (formerly known as Abor-Miri-Dafla) in northwestern Arunachal Pradesh; the Mishmi (four languages) in northeastern Arunachal Pradesh and into southeastern Tibet; and the Nungish or Rawang group in northern Burma and northwestern Yunnan. The latter is lexically transitional to Burmese-Lolo. The position of Lepcha has long been debated in Tibeto-Burman studies; it may fit best here. Lepcha also has a long-established orthography.

The Northeastern group, also known as the Qiang group or Qiangic and formerly as Xifan or Horpa, is a cluster of seventeen languages spoken mainly in western Sichuan. The extinct Tangut language (also known as Xixia or Hsi-hsia) belongs here too. These languages have substantial shared verb morphology; most are tonal. Some of the southernmost languages are lexically transitional to Burmese-Lolo, but are phonologically and morphologically more typical of Northeastern. These include the Bai and Naxi languages; the latter has a pictographic orthography and a syllabary derived from it, of uncertain date.

The Southeastern group includes the Burmic or Burmese-Lolo subgroup and the Karen subgroup. These languages tend to have very little morphology and complex tonal systems typical of the northern Southeast and southern East Asian linguistic area. The oldest dated orthography in this subgroup is Burmese; however, there are also four Yi ideographic traditions of less certain date which may be earlier. Of all the Tibeto-Burman branches, the Karen subgroup is the only one to have SVO syntax. Bai also shows SVO as an alternative possibility; it has been heavily influenced for millenia by SVO Chinese. The position of Bai in Tibeto-Burman is debated; some scholars place it within Burmese-Lolo, and it has even been suggested by some that it could be regarded as a Sinitic language; but it appears to fit better in the Qiangic or Northeastern group. Even Karenic languages and Bai retain various verb-final syntactic characteristics. Apart from these, all Tibeto-Burman languages are typical SOV languages.

Matisoff (1986, 1991) uses an alternative classification which links the typologically similar and geographically adjacent languages of northeastern India in his Kamarupan subgroup: Adi-Mising-Nishi, the Mishmi languages, Kuki-Chin-Naga and Bodo-Garo but not Jinghpaw or Rawang/Nungish. That is, it corresponds to part of Central Tibeto-Burman and part of Sal. The Kamarupan subgroup is also recognised by DeLancey (1987) and by van Driem (1993), with the addition of Jinghpaw which Benedict regards as the crossroads of Tibeto-Burman.

Northeastern India is the area of greatest diversity within Tibeto-Burman; there are Western languages of several subgroups (Tibetan, Tsangla and so on), all branches of the Sal languages (Baric, Kuki-Chin, Jinghpaw, Luish or Sak), many Central languages including the Adi-Mising-Nishi (Abor-Miri-Dafla) languages and the various

Mishmi or Deng languages, and some Southeastern languages (Lisu and the Arakanese dialect of Burmese).

1. Western Tibeto-Burman

Western Tibeto-Burman includes Tibetan, the Tibeto-Burman language with the longest continuous literary tradition. There are several main varieties of modern spoken Tibetan, some of which have a separate modern literary version, but for all Buddhist Tibetans and for speakers of many other Western and Northeastern Tibeto-Burman languages Classical Tibetan serves as a liturgical language.

Also included are the Tamangic, Gurung Group, or West Bodish languages of north central Nepal, including Gurung, Tamang, Thakali, Manang, as well as Kaike and Ghale; these are quite close to Tibetan linguistically, especially the last two. Also fairly close to Tibetan but linguistically distinct are the Eastern Bodic 'Monpa' or Tsangla group of eastern Bhutan and adjacent areas of India and Tibet. Less similar, but still often referred to by the generic Indic pejorative term Bhotia or Bhote, are the various other West Himalayish or Kanauri languages of northwestern India. Still less similar are the Himalayan or Mahakiranti languages of Nepal and Sikkim.

1.1 Tibetan

Within 'Tibetan' itself there is a vast range of varieties, nearly all of which are linked by sharing Tibetan Buddhism and thus literary Tibetan. Scholars tend to divide this range into Western, Central (including Dbus or Ü including Lhasa; and Gtsang), Southern (also sometimes regarded as a subtype of Central, including Dzongkha, the national language of Bhutan, and Denzong, the Tibetan language of Sikkim), Amdo (northeastern) and Khams (mostly southeastern) subgroups. In India and Nepal Tibetans and other Western Tibeto-Burman groups are pejoratively called Bhotia or Bhote, and in China they are called Zang [tsaŋ^{51}]. Apart from its role as the language of Tibet, varieties of Tibetan are or were the official language of various kingdoms south of the main Himalayan range, from Ladakh in the west to Mustang in north central Nepal, Sikkim and Bhutan, and in the Sino-Tibetan marches in the east. In Nepal, the northern quarter of the country is inhabited mainly by Tibetan speakers, notably the

Sherpa in the northeast. Apart from the speakers of Nepali, much of the population of Bhutan and Sikkim speak some variety of Tibetan as a first or second language. Various Northeastern Tibeto-Burman groups including the Rgyarong and several others in western Sichuan also use Tibetan as a literary language and Amdo or Kham Tibetan as a lingua franca; most of these groups are included within the Tibetan nationality in China. The spoken Lhasa variety is widely used as a lingua franca, as is a variety of Kham sometimes known as 'Brogpa or 'nomad' language. In addition to the Buddhist Tibetans, there are two groups in Kashmir who have been converted to Islam: Balti and Purik. Their speech is phonologically conservative, apart from the many Islamic loanwords; '8' is [βgjat]. Balti is generally written in Urdu script; this is of course a relatively recent innovation.

1.2 Bodic and Bodish

The Tamangic or West Bodish languages are found in central Nepal. Many descendants of speakers have moved away from traditional areas, and now speak only Nepali. Within some groups such as the Gurung there is a cultural and religious continuum from north (Buddhism) to south (Hinduism). Languages in this sub-branch are Tamang, Gurung (each with substantial dialect differentiation), Ghale, Kaike, Thakali, Chantel, Rohani and Manang.

The 'Monpa' or East Bodic group lives along the borders of Arunachal Pradesh and Tibet, and especially in eastern Bhutan. Monpa is a pejorative Tibetan term for non-Buddhist Tibeto-Burman groups of this area. In eastern Bhutan Tsangla is a local lingua franca; this is also known as Sharchop ('eastern'). In northeastern Bhutan, northwestern Arunachal Pradesh in India, and adjacent areas of Tibet is the Dzalakha language, known in India as (Western) Monpa and in Tibet as Menba. Further east is another language, known as (Eastern) Monpa/Menba but distinct; and yet another language of central Bhutan, Olikha, is also known to outsiders by the same name. Van Driem has recently located some additional Eastern Bodic languages in Bhutan, and there are more just to the east in India: Sulung, Bugun, Dhammai and Hrusso. The Dhammai are better known as Miji, and are related to the Hrusso; they should not be confused with the Keman or Miju Mishmi to the east. The Hrusso (autonym) are better known as Aka or Angka; their language is poorly described but is also quite different from the Central Tibeto-

Burman languages nearby, as pointed out by Shafer. The Sulung (autonym Puroik) are a small group who were formerly slaves of the Central Tibeto-Burman groups around them; they are also found in adjacent areas of Tibet. Close to the Sulung are the Bugun or Khowa group who are under the cultural influence of the related Dzalakha 'Monpa'. It is possible that improved data on some of these languages will lead to their classification in one or more additional major subgroups of Tibeto-Burman.

The West Himalayish or Kanauri subgroup comprises a number of languages of northwestern India. Included (from west to east) are Chamba Lahuli, PaTani or Manchati, Tinan or Ranglo, Bunan or Gahri, Kanauri or Kinnauri, Kanshi, Rangkhas, Darmiya, and Chaudangsi/Byangsi. The term 'Lahuli' is sometimes used by outsiders to refer generally to the non-Tibetan languages of Lahul in northwestern Himachal Pradesh, including PaTani, Tinan, and Bunan as well as 'Chamba Lahuli', but is not used in this way locally. Zhangzhung is now generally agreed to fit here (Nagano & LaPolla 2001, Takeuchi in this volume), though the available materials are not as full as one might wish.

From its geographical position, Nam (Thomas 1948) may also belong in Bodic. Thangmi and Baram of central Nepal used to be classified here, but from new data it is now clear that they are Himalayan; see below.

1.3 Himalayan or Mahakiranti

These languages appear to form a distinct group within Tibeto-Burman, called Mahakiranti by van Driem (2001). They comprise the non-Western Tibeto-Burman languages of Nepal and western Darjeeling and Sikkim. In general Tibetan-related groups inhabit the northernmost areas of Nepal and Indic groups inhabit the southern part, especially the plains of the Terai. Between them, especially in the east, are the Himalayan or Mahakiranti languages.

Grierson's (1903-1909) division into pronominalised and nonpronominalised Himalayan languages has been shown by Bauman (1975) and Caughley (1982) to be based on secondary and independent morphological developments. Shafer (1974) divides these languages into West Himalayish, West Central Himalayish and East Himalayish sections of Bodic, with Newari less closely and less clearly linked. Benedict (1972) links the West Himalayish or Kanauri

languages more directly to Tibetan in his Tibetan-Kanauri, with Magar intermediate between this and his Bahing-Vayu, which comprises the rest of the languages here. Glover (1974) does not consider the West Himalayish/Kanauri languages, but the rest he divides on lexical grounds into East Himalayish (the Kiranti languages) and Central Himalayish which includes the rest as well as Tibetan and its outliers. This last classification accords with the traditional ethnic classification in Nepal, which groups together most of the Rai or Kiranti Tibeto-Burman languages of eastern Nepal.

The Magar are a numerous and widely scattered group found throughout the lower hills in western Nepal; this was formerly the language of a major kingdom in this area. Probably to be included in Magar are the Raute and Raji; but Kham Magar or Budha is quite distinct. Chepang, Bujheli or Gharti, and Bankariya form another cluster of distinct languages.

Newari or Newar is the traditional language of the Kathmandu valley, where it was the vehicle of a high civilisation using an Indic script. Newari (now known as Nepala bhasa) has had lower status since the Gurkha (Nepali-speaking) conquest of Nepal, but is still very widely used in the Kathmandu valley and in the low hills to the east. Nearly all speakers are bilingual in Nepali; there are substantial dialect differences and a literary tradition dating back to at least the twelfth century. As Tamot (in this volume) notes, until very recently the language was written with an independent Brahmi-derived script, but it is now switching to devanagari.

Recent work by van Driem and by Turin (1998) has shown that the severely endangered Baram and Thangmi languages of north central Nepal are in the same cluster—not core Kiranti or Rai languages, but closer to them than to any other part of Tibeto-Burman. This is based on partial sharing of the very extensive verbal paradigms reconstructed by van Driem for core Kiranti.

Nearly all the remaining Tibeto-Burman languages further east in Nepal are part of core Kiranti; the only exception is Dhimal in the southeast. Apart from Limbu and Yakkha in the east and Sunwar in the west, these languages are also grouped in the ethnic category Rai, also known as Kiranti (or Kirat) from the former kingdom of this area. Some Rai languages are nearly extinct, being replaced by Nepali or by other Rai languages. For example, Bantawa may have replaced some of the smaller adjacent languages. Linguistically Sunwar, Yakkha and Limbu also form part of the core Kiranti group.

A detailed survey was undertaken some years ago by the Linguistic Survey of Nepal, and overall results (Hansson 1991) are available. Most of the smaller Kiranti languages are endangered or moribund.

Languages included here are superbly-described but moribund Hayu (or Vayu), and Sunwar, Bahing or Rumdali, Chaurasia or Umbale, Jerung, Thulung, Khaling, Dumi, Koi or Kohi, Kulung or Kulunge and Sotang, Nachering (with its Parali dialect), Dimali, Chukwa, Sangpang or Sangpahang, Bantawa, Lambichong or Mugali, Dungmali and Khesang, Chamling or Rodong, Puma, Athapare, Belhare, Chintang, Chulung, Yakkha, Lumba, Mewahang or Newahang, Saam, Sambya, Bungla, Pongyong, Lohorong or Lorung, Yamphu, Yamphe or Yakkhaba, Tilung, and the numerous subtypes of Limbu. The use of the traditional 'Kiranti' (Limbu) script is being expanded, both by the Limbu elite in Nepal and in the education system of Sikkim. It is an Indic-derived alphabetic script of uncertain date but at least some centuries old.

2. Sal Tibeto-Burman

This subgroup was named by Burling (1983) from the distinctive etymon for 'sun' found in these languages; it is characterised by numerous other innovative etyma, such as *wal 'fire', *s-raŋ 'sky' and *nu 'mother'. It was first proposed as a subgroup in Benedict (1976), and comprises Shafer's Baric group (Benedict's Bodo-Garo-Northern Naga), which covers the plains of northeastern India as well as a large area of the hills to the east of this; Jinghpaw (Kachin), the central nucleus of Tibeto-Burman according to Benedict (1972), but a part of Burmic according to Shafer; Luish (also regarded as Burmic by Shafer) or the Sak group; and Kuki-Chin-(Southern) Naga (again considered to be Burmic by Shafer).

All these languages have fairly extensive prefix and suffix morphology; some have lost their tonal systems. The Luish or Sak group is scattered and moribund but formerly covered a much wider area; for a historical overview see Luce (1985). Jinghpaw is the core group in the Kachin cultural system, which also includes several Burmish and a few other groups which fit elsewhere linguistically. Baric includes Boro (Bodo or 'plains Kachari'), a millenium ago probably the main language of the upper Brahmaputra valley in northeastern India, with very closely related languages such as

Dimasha ('hills Kachari'), Kokborok (Tripuri), Lalung and so on covering the plains and low hill areas to the south, and also quite closely related Garo in the hills to the southwest; also the 'Koch' languages such as Atong, Rabha, Wanang and so on generally in the plains to the west, and the still rather closely related Northern Naga languages of northern Nagaland, Tirap District of Arunachal Pradesh, and adjacent areas of Burma to the east. Jinghpaw is spoken in a large area immediately to the east of the latter, with the Sak group scattered to the south, in an area mainly inhabited by speakers of other Tibeto-Burman languages. Löffler (1966a) has suggested that Chakma (spoken in Bangladesh and India) and its dialect Tongcengnya (Daingna, Daingnet, Doingnak, Dengnak, Tunzunnya, spoken in Bangladesh and Burma) should be included with the Sak group, but whatever the history of the speakers, linguistically these are now divergent dialects of Bengali spoken by a Buddhist group.

2.1 Bodo-Garo-Northern Naga (Baric)

2.1.1 Bodo-Garo

The speakers of Bodo-Garo languages are found mainly along the Brahmaputra in Assam, north and south of the river, and in adjacent states of northeastern India, with smaller numbers in Bangladesh, Bhutan and Burma. In some western areas they are known as Mech or Meche; this is now viewed as a pejorative name in India. Another general term is Kachari; the 'plains Kachari' are the Boro or Bodo. The language has a roman orthography and a devanagari orthography and is used as a medium in some primary and secondary schools. Closely related are Lalung, Dimasha or 'Hills Kachari', Hojai, and Kokborok or Tripuri. The latter is the traditional language of Tripura in northeastern India. Kok-borok means 'language of the people'; the word for 'people' is of course related to the name of the Boro. There are various dialects including standard Debbarma (western), also Riang or Reang (southeastern), Noatia or Tipra (eastern), plus various smaller dialects: Jamatia and Darlong (northeastern), Aiang, Dahula, Karpong, Koloi, Laitong, Muslung and Rupini. In Bangladesh, the Kokborok are known as Mrung, from the Arakanese name for this group. There are relatively recent Bengali scripts for Kokborok and Dimasha.

Garo is the language of the Garo Hills in western Meghalaya and nearby in Bangladesh. The standard dialect, Achik, covers two-thirds

of the area in the east, and the other main dialect, Abeng, is in the west. Between the two in the south is the Matabeng or Matjanchi dialect. Within Achik there are Gara, Ganching, Matchi, Dual, Chisak, and Awe or Akawe subdialects from south to north. Some Koch languages (Atong, Wanang, Hajong) are also officially (but incorrectly) regarded as dialects of Garo. There is a well-established nineteenth century roman orthography.

The Koch or Konch (Western Bodo-Garo) languages include Atong, Wanang, Rabha, Hajong, Ruga, Deori or Chutiya, and possibly others. Deori is the most divergent of the core Bodo-Garo languages according to Benedict.

Two closely-related languages of southeastern Nepal and nearby areas of India which may be linked to Bodo-Garo or to Central Tibeto-Burman are Toto and Dhimal. Published descriptions are very limited, but a wordlist of Dhimal, spoken in southeastern Nepal, is now available (Toba 1992).

2.1.2 Northern Naga

This is Shafer's Nagish portion of Baric, where Benedict and more recently Burling also place it. Its internal subgrouping has been the subject of French (1983); Marrison (1967) also independently separated it from other 'Naga' languages as his Naga A. The names of 'Naga' groups are notoriously confusing and confused; Assamese or other names of villages, rivers or towns where contact occurred, clan names for the very numerous subgroups of each group, names used by other 'Naga' groups to refer to a group, autonyms, and descriptive names are all used.

Khienmungan is known in the anthropological literature as Kalyokengnyu from their stone-roofed houses; this large group is the southernmost Baric 'Naga' group. About one-sixth of the group is in India but most are in Burma where parts of it are known as Nok-aw or Nauk-o (a clan name), Ponyo (a village name), Para or Paya (a Burmese name of uncertain origin) and so on. In India the Sema call them Tukhemmi and the Chang call them Aoshedd; many alternative representations of the autonym—Khiamngan, Khemungan and so on —also are seen.

The Chang are a small group of northern Nagaland; a roman orthography exists, but little recent linguistic data. The Ao and Konyak name for the group is Majung, Mojung or Manjung, and the

Sangtam name is Machongrr. Phom is another relatively small group with a roman orthography but little linguistic data. Their former autonym was Chingmengnu; in some older literature they are called Assiringia (a village name) and Tamlu.

Konyak is the largest 'Naga' group in India, with a roman orthography. It is at the northern tip of Nagaland, with a small number of speakers in adjacent areas of Burma as well. The 'standard' dialect is spoken at Wakching; in older sources this is also referred to as Tableng, Mulung, Kongon or Angwangku. Various sources list over thirty current clan or village dialects. In Burma to the east of the Konyak and south of the Wancho are the Htangan; Marrison (1967) suggests that this is closely related to Konyak, if not a dialect of it; but no linguistic information is available.

The Wancho are a substantial group, mainly at the southern tip of Tirap District of Arunachal Pradesh and extending into Burma. This group was formerly known as Banpara, Mutonia, Joboka or Jokoba, with subgroups Khulung-Muthun, Bor-Muthun and Horu-Muthun. It has two main subgroups: Changjan and Tangjan.

Haimi is a large group in Burma with nearly twenty named clan-dialects. Nocte is another large group, about half in India and half in Burma, and formerly known as Namsangia (a village), Borduria (another village) or Jaipuria (a town which many Nocte visit). It has six main dialect groups: Hawajap, Japejap, Kapajap, Lazujap, Photungjap and Tangjap. Tangsa again has a multitude of clan and other names; Moshang (Mawshang) and Shangge are two such. Tangsa, Rangpan, Haimi, Nocte and Wancho are similar enough that a shared roman orthography based on the Moshang clan dialect of Rangpan is now being implemented among Christians in Burma. This orthography is a considerable improvement on most such, as it indicates the tones.

2.2 Jinghpaw

Known in India as Singpho, in China as part of the Jingpo nationality, and in Burmese as Kachin, this group (whose own name is [tʃiŋ^{31}pʰɔʔ^{31}]) forms the core of the Kachin culture complex in northern Burma, with minor extensions into China and India. Of those who participate in the Kachin culture complex, more than half speak Jinghpaw as their first language. The rest, the Burmish-speaking Atsi, Maru, Lashi and Ngochang, speak it very fluently or

bilingually as a second language and use it as the medium of literacy. The late nineteenth century Jinghpaw roman orthography is very widely used, but unfortunately does not indicate tones. Separate roman orthographies are now being developed for some Burmish 'Kachin' languages. Within Jinghpaw there are some divergent dialects, most notably the Gauri (Kauri, Hkauri) dialect of the area just north of Bhamo in Burma. Sometimes other Jinghpaw clan names are also cited as if they were dialects.

2.3 Sak or Luish Group

The Kadu appear in the history of Burma as the former dominant group of the Tagaung kingdom in Upper Burma, under the name Kantu; they should of course not be confused with the eponymous Mon-Khmer group in Vietnam and Laos. Their language is moribund in the hills northwest of Mandalay; the group's autonym is [ɔsa?]. Ganan is closely related to Kadu and now spoken to its immediate west, in about twenty villages. Both Kadu and Ganan are poorly described.

Taman is spoken in one village north of Homalin in western Burma. There are several other extinct languages such as Malin, which was very close to Taman, and very small groups in adjacent areas of Manipur in India such as Andro or Phayeng; Sengmai or Sekmai; and Chairel (a village name) or Chakpa. None of these languages is well-described.

The Sak of Arakan and adjacent areas of Bangladesh are often known from the Burmese form of their name, Thet, or Arakanese That. Their autonym is [ətsa?]; they should not be confused with the Buddhist but Bengali-speaking Chakma.

Pyu or Tircul, the dynastic language of Sri Ksetra in Lower Burma from 638 to 830 AD, is most likely another Luish language; attempts to fit it into Burmese-Lolo are unconvincing; for more discussion see Beckwith (in this volume). Stargardt (1990) dates its Indic script to the 4th century AD. The last Pyu inscription is also the first Burmese inscription of 1112 AD.

2.4 Kuki-Chin

This subgroup is relatively cohesive, both geographically and linguistically, and has been extensively investigated by Shafer, who classifies it as part of Burmic. Benedict uses the earlier term Kuki-

Chin for the subgroup, and likewise links it to Burmese-Lolo. Both include here the 'Naga' groups apart from those in the Sal group (Northern Naga or Nagish languages). In this they differ from Grierson and Marrison who prefer to regard Naga as a subgroup of Tibeto-Burman. However, lexical and morphosyntactic similarities suggest that Kuki-Chin actually belongs in the Sal Group, but with some contact lexicon from Burmese.

Names for these groups are much more numerous than distinct languages. Firstly, there are overall names: in India those who live in Nagaland and northern Manipur are often called Naga, those who live in southern Manipur and adjacent areas of Assam are often called Kuki, in Tripura they are generally called Halam, while the largest group in Mizoram has renamed itself Mizo (formerly Lushai). In Burma all are linked under the category Chin. For example, the Thado (or Thadou) are usually called Kuki or sometimes Thadou Kuki in India, but form part of the northern Chin group and are thus called Chin or sometimes (more specifically) Thado Chin in Burma. One group of eastern Manipur, the Anal, decided recently to classify themselves as Naga rather than Kuki. Secondly, there are more specific names for subgroups—in many cases former names and names used by outsiders as opposed to autonyms. For example, the Arleng were formerly known as Mikir, and are now officially known by an alternative autonym, Karbi. Thirdly, some of these groups have recently been amalgamating and new names have been coined to refer to these larger groups; for example the Zeme/Nzeme (formerly Empeo), Liangmai (formerly Kwoireng) and Nruanghmei (formerly Kabui, including the Puiron dialect often referred to as a separate language) now often refer to themselves as Zeliangrong, although some Zeme prefer to remain Zeme; previously, before the addition of the Nruanghmei, the term Zeliang was coined to refer to Zeme plus Liangmai, and this is still used as well; not to mention the older cover term for these three groups, Kachha Naga ('bad Naga'), which for obvious reasons is no longer widely used. Fourthly, geographical names are sometimes used instead of the rather specific subgroup names; for example, Tiddim Chin instead of Kamhau Chin.

Some of the languages included are more divergent; the foremost example is Arleng (also known as Karbi, or formerly Mikir) which has long been in contact with Sal Group Tibeto-Burman languages as well as non-Tibeto-Burman languages; it has even been suggested that there may be a connection between Kuki-Chin and Lepcha via

this language (Bauman 1976). Also somewhat different is Meithei (Manipuri), which has long been the literary language of civilisation in the Manipur valley; now largely Hindu, it thus shows much Indic influence. Most of the other languages are spoken by hill groups, some extremely small, and have roman orthographies if any.

Linguistically the Kuki-Chin languages are characterised by tones (mostly not indicated in orthographies), extensive verb morphology often involving tonal alternations, and extensive suffixing with more limited prefixation. There is a widespread *ni ergative suffix on NPs; the basic word order is SOV. Many languages have developed a postpositional negative suffix derived from *lo which co-occurs with the Tibeto-Burman negative prefix *ma in some languages (such as Ao) and replaces it in others (such as Mizo).

2.4.1 Central and Southern Naga

The Ao language, with roman orthography based on the Chungli dialect and a large Mongsen dialect as well. also includes some eastern dialects such as those of the villages Yacham and Tengsa which show contact effects with Northern Naga (Baric, Sal Group) Phom and Chang. The Sangtam language is found in three main locations, with some dialect differences; the 'standard' basis for the roman orthography is the northern Lophomi dialect. The Rengma group, perhaps formerly more widespread, is also now scattered. Though regarded as one group, it probably includes three languages: Western, for which Tseminyu is the basis for the orthography; Northern or Ntenyi; and Eastern or Meluri (with the autonym Anyo); In fact Ntenyi and Meluri are genetically closer to Sema than to Western Rengma. For Lhota, Wokha is the 'standard' dialect and basis for the roman orthography. The Yimchunger language is spoken in two areas including one in Burma; the roman orthography represents the Yachumi dialect.

Tangkhul is also known as Luhupa ('savages') to the Manipuri. This large group is almost as numerous in Burma as in India. The standard written dialect is that of Ukhrul. The small Maring group lives to the south of the Tangkhul; until the Anal declared themselves 'Naga' Maring was the southernmost 'Naga' language.

The Sema are a large group with a roman orthography. Formerly included in 'Angami' were most of the Chakhesang and the Mao;

what is now known as Angami is only the western part of this larger group. Angami writing is based on the Tengima dialect.

Chakhesang is the new name for a composite group of eastern 'Angami': the Chokri and the Kezhama. It also includes some Sangtam, who live to the north and south of the eastern edge of this group.

The town of Mao (known to its inhabitants as Sopvoma) is the largest village of its group and hence the usual name for its language; it is close to Angami and Chakheseng and has a roman orthography.

Maram is again named from its main village. Linguistically it is similar to the Zeliangrong languages. The Zeliangrong group represents the amalgamation of three distinct 'Naga' groups: in the west and northwest, the Zeme, Mzieme or Nzeme, Liangmai in the northeastern area, and Nruanghmei to the east and south. The small Khoirao group is also included here. Roman orthographies exist for Zeme, Mzieme, Liangmai and Nruanghmei.

Also known as Manipuri, the Meithei language of the valley of Manipur has a very old native Indic orthography and a more recent Bengali-based orthography which does not indicate the tones. Dates as early as 799 AD have been proposed for the native orthography, based on dated inscriptions; for more details see Chelliah and Ray (in this volume). Since August 1992 it has been a Schedule VIII language of the Indian constitution, the only Tibeto-Burman language to achieve this official national language status in India. Bishnupriya Manipuri is an Indic language spoken by long-standing immigrants, with some Manipuri lexicon and reduced morphology; because of recent migrations, most speakers of Bishnupriya Manipuri now live outside Manipur. Many non-Meithei Kuki-Chin and other people in the valley also speak Meithei; some of their languages (especially Sak group and some Old Kuki languages) are being replaced by it.

2.4.2 Old Kuki

This term refers to a number of small groups around the Manipur valley and to the west and southwest. It includes some of the 'Kuki' groups of Tripura and nearby parts of Bangladesh and Cachar, for example Rangkhol and Bete; Halam; and Langrong. It also includes the strongly Mizo-influenced group Hmar; the Anal group of southeastern Manipur and adjacent Burma; the Chawte (Chote, Chaw, Kyaw) group of eastern Manipur and into Burma; the nearby

Mayol (Moyon, 'Mon') and Lamgang; the Kom of south central Manipur; and various other very small groups whose languages are nearly extinct, having been replaced by Meithei: Aimol, Kolhreng, Purum, Tarao and so on. Roman orthographies have been created for some 'Old Kuki' languages, but are not widely used. Some groups classified as 'Old Kuki' in Indian sources fit better in North Chin and are discussed there: Chiru, Gangte, Pawi, Simte, Thado(u) and so on.

2.4.3 North Chin

This subgroup includes what is perhaps the largest 'Kuki' group in India, the Thado(u), as well as such other 'Kuki' groups as Chiru, Gangte, Pawi and Simte. Its speakers are very widely scattered throughout Manipur and adjacent areas to the northwest, but are concentrated in southern Manipur; some also live in Burma in the Tiddim area. Some of the named subgroups in Burma are the Thado, Siyin, Paite and Vuite, Sokte, Kamhau and so on. There are recent roman orthographies in use for several of these.

2.4.4 Central Chin

The Central Chin are sometimes collectively known as Laizo ('central people') or Lai; this includes a very large number of dialects, some of which are so different as not to be mutually intelligible. In fact Mizo (formerly Lushai) is another variety of Central Chin, but with a separate literary tradition. The group includes the western varieties Bawm or Banjogi and Paangkhua or Pankho in Bangladesh; northernmost Zahao or Laizo in Burma and extending into India; also Tashon, Ngawn, Zanniat, Zophei, Lawtu, Lailen, Senthang, Tawr and many other groups. They inhabit the central area of the Chin State, including Falam, Haka and parts to the south, as well as adjacent areas of Bangladesh and Mizoram. Various roman orthographies are in use.

Mizo, formerly known as Lushai, is the largest single Chin group, with various dialects, concentrated in Mizoram but also nearby in Burma, including the Hualngo dialect. Its roman orthography is widely used in India. Some other Kuki-Chin groups are gradually becoming assimilated to the Mizo, culturally and linguistically; these include the Hmar ('Old Kuki') in the north and the Mara (Lakher or Miram) in the south. Some Chin nationalists use the new term Zomi to refer to the North and Central Chin including the Mizo.

2.4.5 South Chin

Again, this group has many different names. In Bangladesh there are a few speakers referred to as Khyang. In Burma the South Chin are very numerous, but some of those living in the plains no longer speak Chin. Subgroup names in the northern area of the range include Zolamnai ('on the trail'); Welaung and Matu (place names; the latter call themselves Ngala). In the central part are M'kang, Ng'men, Nitu (or alternatively the new name Daai), Hngisung, Utpu (a large subgroup, called Chinbon by the Burmese); and to the south the Chinbok or Saingbaung Chin (Burmese names), also known as Ashö Chin from the South Chin form of the word for 'person', cognate with Central and Northern Zo.

The Khami/Khumi group includes several diverse dialects which fall into two subgroups: Khami versus Khumi, both living mainly in Burma but also in Bangladesh. A roman orthography exists.

Mara, better known as Lakher from the Mizo name for the group, is also known as Miram from its Central Chin name. Most are in India but some live in Burma. There is a roman orthography, but the language is being replaced by Mizo.

To the northwest of the Mizo and Kuki in the southern hills of Assam are the Arleng, formerly known as Mikir and now officially called Karbi from another autonym. There are various dialects: Amri (western), Bhoi (southwestern, with Khasi influence), Rengkhang (southern) and so on. Many ethnic Arleng have become Assamese speakers. This is the most divergent language within the group.

3. Central Tibeto-Burman

The classification of these languages has not been finally determined, mainly because good lexical data has until very recently been lacking, and good morphosyntactic data is mainly still not available. This subgroup probably includes most of the languages spoken along India's northeastern border, the northern tip of Burma, and the adjacent border area of northwestern Yunnan Province, China.

There are three main subgroups: Adi-Mising-Nishi, Mishmi, and Rawang/Nungish. The Adi-Mising-Nishi (Shafer's Mirish, Benedict's Abor-Miri-Dafla) languages of much of Arunachal Pradesh, extend slightly into adjacent parts of Assam and Tibet. The rather diverse Mishmi languages are in India's extreme northeast and into

Tibet. Benedict classifies the Rawang or Nungish group of northern Burma and adjacent border regions of China as a link between Burmese-Lolo and the rest of Tibeto-Burman, while Shafer simply classifies it as Burmic. However it shows many lexical and morphosyntactic properties in common with the other Central Tibeto-Burman languages.

Another language which may fit here is Lepcha. Its genetic position is debated; while clearly Tibeto-Burman, it has been put with the Himalayan Tibeto-Burman languages of Nepal by the Linguistic Survey of India, with Ao and thus in Kukish and ultimately Burmic by Shafer, and recognised as aberrant for its geographical location by Benedict. Bauman (1976) notes some similarities with the aberrant (and geographically nearest) Kukish language, Mikir (Karbi or Arleng). Most recently, Bodman (1987) suggested a fairly close connection with Adi-Mising-Nishi and Nungish, and a more indirect one with Jinghpaw; he also casts doubt on the Austroasiatic connection suggested in Forrest (1962). However Jackson Sun (1993) has shown that it is not lexically closer to Adi-Mising-Nishi than to various other Tibeto-Burman subgroups.

Lepcha or Rong was the traditional language of southern Sikkim, Darjeeling and part of southwestern Bhutan, but not many speakers remain. It has a traditional Indic script in use since at least the early nineteenth century; see Sprigg (1983) for details. In 1977 it was made one of the official languages of Sikkim; since then a substantial effort has been made to teach the script in schools.

3.1 Mirish or Adi-Mising-Nishi

The Nishi, Bangni and Apa Tani or Mirish group was formerly known collectively as Dafla, a name now usually viewed as pejorative. They are now more commonly referred to by various subgroup names such as Bangni (Bengni) or Bangru in the west, Apa Tani (Apa Tanang) around Ziro, Tagen or Tagin in the northeast, and elsewhere as Nishi, Nyishang or Nishang. Nearly all are in India; the few in Tibet are included in the composite Luoba nationality whose name is derived from Tibetan *klopa* [lopa], 'savages'.

Hill Miri is a small and scattered group in central Arunachal Pradesh; their language is closely related to Nishi and less closely to Miri or Mising.

Adi, formerly known as Abor, is now divided into a large number of named subgroups all speaking very similar dialects. It includes Gallong (autonym Galo) in the southwest, Bokar (Bogar) in the northwest, Pailibo, Bori and Ashing in the north central area, Tangam and Shimong (Simong) in the northeast, Minyong at the centre (with Karko subgroup), Milang to their east, Pasi and Panggi at the south centre, and Padam in the southeast. Nearly all are in India, apart from a few Bokar in Tibet.

Formerly known as Miri, the Mising group is scattered over the upper Brahmaputra valley and into the hills to the north. This group is large in number, but not all speak the language. A romanisation has been developed but not yet approved for use in education. The language, with dialects, is quite closely related to Adi. Many languages of this group have roman orthographies.

3.2 Mishmi

Idu is one of the Mishmi groups; the Idu were formerly known as Chulikata (crop-haired) Mishmi. About two-thirds live in Arunachal Pradesh and the rest in Tibet. In Tibet they are included with the Luoba nationality. The Taraon Mishmi group, autonym [ta^{31}ɹuɑŋ^{55}], is also known as Tain, Taying or in China Darang Deng; they were formerly called Digaru Mishmi. Most are in India, some in Tibet and a village in northernmost Burma. The Keman, autonym [kə^{31}mɑn^{35}], is the third Mishmi group; other names include Kaman, Geman Deng and Miju. Two-thirds are in Arunachal Pradesh and the rest in Tibet. A fourth small Mishmi group has recently been located in India, but descriptive details to enable its classification are not yet available.

3.3 Rawang/Nungish

The Jinghpaw name for this group is Nung, which is the source for the terms used by Benedict and Shafer, and should not be confused with the Central Thai group of northeastern Vietnam and adjacent areas of China. Former Chinese names include Nuzi, Luzi and Jiuzi. It includes a very large number of subgroups with rather different languages and many subdialects. In Burma, where most live, the term Rawang (which fomerly referred only to the largest supergroup in Burma) is now used for these groups as a whole. In China speakers are included in two nationalities, Dulong and Nu, but some members of the Nu nationality speak the Loloish Tibeto-Burman languages

Nus or Raorou instead. In Burma the Mvtwang clan dialect of the Mvtwang clan cluster in the Rawang supergroup has been chosen as the standard, and a romanisation implemented among Christians; in China there is a romanised Dulong orthography and an Anung orthography used by Christians. These groups classify themselves in various ways: by clan, of which there are probably nearly two hundred; by clan cluster; by supergroups of several clan clusters; or most recently by the collective terms Rawang, Dulong or Nu.

The Dulong nationality in China, plus a portion of the northernmost members of the Nu nationality, form the Trung [tə^{31}ɹuŋ^{53}] clan cluster, the northernmost Nungish group. In Burma, the Zørwang clan cluster is also known as Jerwang or Tvluq; it forms part of the Gvnøng supergroup, which is quite distinct from the Rvwang (Rawang) supergroup. The clan cluster in Burma known as Dvru or Daru is also part of the Gvnøng supergroup. The Anung clan cluster, called after their autonym [a^{31}nuŋ^{53}], is in China and Burma. They and the Nusu (speakers of a Loloish language, see below) form most of the Nu nationality. The Dvngsar clan cluster, also known as Tangsarr, forms part of the Rvwang (Rawang) supergroup. The Mvtwang clan cluster is the most numerous; it is the main part of the Rvwang supergroup, and its Mvtwang clan dialect forms the basis for standard Rawang orthography. The Dvmang clan cluster is included within Mvtwang as well. Many Rawang in Burma who have a different first dialect can also speak Mvtwang as this has been used as the literary dialect. Lungmi (or Longmi), is the southernmost clan cluster, and has undergone considerable Jinghpaw influence. It is different from the Rvwang and Gvnøng supergroups.

4. Northeastern or Qiang Tibeto-Burman

This subgroup was proposed by Sun Hongkai (1983). It includes all the groups formerly called Xifan in Chinese, and some others. While all scholars agree that these languages are Tibeto-Burman, the exact grouping is not generally agreed upon. One proposal by Sun is that most of them (except Bai, Naxi, Baima and Tujia) form a subgroup which he calls Qiangic; Sun places Baima with Tibetan and Naxi, Bai and Tujia with Loloish. Nagano and Thurgood have suggested a further connection of Qiangic with Rawang/Nungish and the various Central Tibeto-Burman languages, called Rung by Thurgood (1985).

Ten Qiang Group languages are spoken by members of the Tibetan nationality, while the Qiang, Pumi, Naxi, Bai and Tujia nationalities are recognized as separate. Unlike the Southeastern Tibeto-Burman languages, which are lexically fairly close to them, the Northeastern Tibeto-Burman languages mostly have extensive verb morphology which can be reconstructed for this subgroup. Nearly all are tonal. It appears that extinct Tangut was a Qiang Group language, as it shares the characteristic sound change of *a to /i/, among other things. The extensive corpus of Tangut manuscripts from early in the second millenium AD also provides the earliest documentation for this branch of Tibeto-Burman.

4.1 Core Qiangic

All these languages fall into the Tibetan cultural orbit, most speak some Tibetan and many use Tibetan as the medium of literacy. Apart from the Qiang and the Pumi, who are recognised as separate nationalities, all are simply classified as Tibetan by the authorities, despite their languages. Qiang is another old Chinese name for an ethnic group of this area, but it is unlikely that it has always had only its current referent. These languages are quite diverse, but more closely related to each other than to anything else in Tibeto-Burman. Qiang includes two 'dialects' which are clearly distinct languages; the northern 'dialect' is nontonal, unlike the southern one, which also has substantial internal diversity.

From north to south, the languages are Baima in southern Gansu, Qiang (at least two languages), Rgyarong, Guichong, Zaba, Choyo, Ergong, Muya, Ersu, Shixing, Namuyi, and Pumi in western Sichuan and into northern Yunnan. Tangut was formerly spoken in the area to the northeast of the Baima in what is now Ningxia.

4.2 Other Qiang Group languages

Naxi, often seen as Nakhi in the western literature because of Rock's idiosyncratic transcription, appears to be transitional between Qiangic and Burmic; it shares lexical material with both subgroups, but like Bai and Tujia lacks the extensive morphology of core Qiangic. Its two main western varieties are mutually intelligible, but Moso, spoken further east in Yunnan and in Sichuan, is not. The traditional pictographic writing system, extensively studied by Rock, is hardly used now; it is actually a mnemonic for religious texts

known by heart. There is also a related syllabic system which can be read without knowing the text, and a new romanisation. Naxi has some Loloish lexical material, but Bradley (1975) shows that it does not share any of the most diagnostic Burmese-Lolo or Loloish sound changes.

Tujia is a very large nationality, with many millions of recognised members, but only about 170,000 of them speak Tujia. This nationality and language were 'discovered' after 1950, before which they were regarded as Chinese. The language has very many Chinese loanwords, and is mostly spoken by older people even in its remaining core area on the borders of Sichuan, Hunan and Hubei. It is not written.

By contrast Bai has a long and distinguished history as the language of the Nanzhao kingdom of western Yunnan. Due to nearly two millennia of Chinese contact, with extensive borrowings from various Chinese dialects at various stages in their development, it is extremely difficult to determine the exact position of Bai within Tibeto-Burman. Some linguists have even suggested that Bai has by now become a Chinese dialect. Even its syntax is partly Sinicised and some dialects have SVO order. Chinese scholars believe that it is a Loloish language, but this seems to be incorrect. A roman orthography based on the central dialect has recently been developed, but the speakers have long been accustomed to use Chinese for writing. The three dialects have substantial differences, nearly enough to lead to lack of mutual intelligibility.

5. Southeastern Tibeto-Burman

5.1 Burmese-Lolo

Burmese-Lolo (also more recently called Lolo-Burmese) is a large and diversified part of Tibeto-Burman; it can be subdivided into two main subgroups, Burmish (including Burmese) and Loloish. A third subgroup, Ugong, appears to be intermediate between the two. For details see Bradley (1979a). The Mru language of western Burma may also be a remotely related part of the Burmese-Lolo group, though some scholars suggest otherwise. All Burmese-Lolo languages have complex tonal and consonant systems but little or no morphology. The main phonological characteristics include the development of a third tone category, derived in part from *s and *ʔ

prefixes, additional to the two tone categories found in Sinitic and some other branches of Tibeto-Burman.

5.1.1 Mru

This language is spoken in western Burma and in Bangladesh. Its exact position in Tibeto-Burman is not certain, but it shows various layers of contact vocabulary from Kuki-Chin and from Burmese. According to Löffler (1966b) it is not Kuki-Chin as sometimes suggested, but may be remotely related to the Burmese-Lolo group; this is also Shafer's view; see also Luce (1985).

5.1.2 Gong

This language is spoken by a small and diminishing population of several hundred in western central Thailand. All speakers are to some degree bilingual in some variety of Thai; younger members of the group tend to be semi-speakers of Gong. In many locations where the language was formerly spoken it is now dead. The language is in some ways very conservative, e.g., in retaining medial /l/ in words such as 'four'; and in other ways quite innovative, for example changing initial *mr to /g/ as in 'horse' or 'high'.

5.1.3 Burmish

5.1.3.1 Burmese

Burmese is the national language of Burma. It has the largest number of speakers of any language within Tibeto-Burman and is second largest within Sino-Tibetan after Chinese: over 40 million speakers including members of other ethnic groups in Burma. It has a long literary history, with the earliest dated inscription from 1112 AD.[2] The script originated as an independent variety of brahmi borrowed from Mon, and has developed into its modern rounded form due to

[2] In my view, the first Burmese inscription, the so-called Myazedi (from the temple in Pagan where it was placed) or Rajkumar (from the donor who caused the inscription to be made; the meaning of 'crown prince' is obvious) was a political gesture from the half-Mon son and heir of Kyanzittha (ruled 1084-1113AD)—the general and later king who brought massive Mon influence with his Mon wife, daughter of the defeated Mon king, into the recently-established Burman kingdom. Kyanzittha's inscriptions were entirely in Pali or Mon; but the Myazedi is in Mon, Pali, Pyu and Burmese. It did not work; Kyanzittha's grandson Alaungsithu succeeded him.

being written mainly on palm leaves. Burmese script is also the basis for various orthographies for other groups in Burma, such as the Shan, most of the Karen, and so on.

In modern Burmese, there are diglossic High and Low varieties, with the former used mainly in written or other formal contexts. The name 'Myanmar' is the literary High term for the country preferred by the current government, while 'Burma' is derived from the spoken Low term. One major dialect is Arakanese in the west; this is also spoken in southeastern Bangladesh and adjacent areas of India, where it is usually known as Mogh or Magh. Arakan was reconquered by the Burmans about two centuries ago, and at that time much of the Arakanese court fled to what is now Bangladesh, where they now call themselves [məɹəma]. Other dialects are southeastern Tavoyan and Beik or Merguese; east central Intha around Inle Lake; Danu and Taungyo in the same area; and west central Yaw. All the regional dialects are more conservative phonologically than standard Burmese; e.g., Arakanese retains the distinction between /r/ and /j/, Tavoyan keeps medial /l/, and so on.

5.1.3.2 Other Burmish languages

To the northeast of the Burman area is a hill area with some inhabitants speaking closely related languages. There are four main groups, all to some degree integrated into the Kachin or Shan cultures of the surrounding larger groups. These Burmish groups go by various names:

OWN NAME	JINGHPAW NAME	BURMESE NAME	CHINESE NAME
Lawngwaw	Maru	Maru	Langsu
Tsaiwa	Atsi	Zi	Zaiwa
Lachik	Lashi	Lashi	Laqi
Ngochang		Maingtha	Achang

In China the Ngochang have separate Shan-like valley states and are recognised as a separate nationality, but in Burma they are mingled with the Lashi or live in Shan villages, often as the local blacksmiths. The Lashi are mostly in Burma. The rather more numerous Maru and Atsi are widely dispersed, but tend to be further south, in the Northern Shan State of Burma and nearby in China. Most Atsi, Maru, Lashi and some Achang operate as exogamous clans within the Kachin culture complex, using Jinghpaw as their literary language

and lingua franca and always marrying language-exogamously. Very small clan-based Burmish languages, Bola and Chintau, are found only in China among the Atsi and Achang respectively; there may be others. Roman orthographies for Atsi, Maru, Lashi and Achang exist but are hardly used. Within each group the regional differences are substantial, perhaps even as great as those between different groups which live together and intermarry, like some Atsi and Maru or some Lashi and Achang. Maru and Lashi are characterised by the addition of final stops to some syllables; these are absent in Atsi and Achang—a convenient shibboleth? In China, most members of the 'Jingpo' nationality are Atsi, with smaller numbers of Maru, Lashi and Bola as well as some speakers of Jinghpaw; in Burma the proportion of the 'Kachin' who speak Burmish home languages is smaller, but still substantial, especially in the Northern Shan State and the southeast of the Kachin State. There are also small communities in Thailand.

Another Burmish language is Phun (Hpun, Hpon, Hpön), formerly spoken in the gorges of the upper Irrawaddy north of Bhamo; it has two dialects, north and south, both moribund. It is phonologically more conservative than the 'Kachinised' Burmish languages, and of course lacks the Jinghpaw lexical material that can be found in them.

5.1.4 Loloish (Yi Group)

For details of the phonological and lexical subgrouping of these languages, see Bradley (1979a). Basically, all share an innovative two-way tonal contrast in original stop-final syllables; the Northern Loloish languages have mostly reversed phonetic values for these two tones compared to other Loloish languages. Central Loloish is characterised by extensive tonal splits leading to complex tonal systems usually including contour tones. Southern Loloish is more conservative for medial consonant, rhyme and tonal developments, while Northern Loloish is more conservative for some initial developments.

In China, the Yi nationality (a new post-1950 name to replace the former term Lolo, now regarded as pejorative in China) is classified into six language clusters, three of which (Nosu or Northern Yi, Nasu or Eastern Yi and Nisu or Southern Yi) are closely related to each other and form Northern Loloish. The remaining three, Southeastern Yi (including the Sani, Axi, Azhe and Azha), Central Yi (Lipho and

Lolopho) and Western Yi (including Lalo and others), form a part of Central Loloish. Most of these are only spoken in China. The Chinese classification of the Yi does not adequately reflect the position of the Pu groups in southern Yunnan whose nationality is Yi and whose languages are Southeastern Loloish (Shafer's Tonkin Lolo).

5.1.4.1 Northern Loloish

The three language clusters in the Northern Loloish group all use the autonym 'black people'; the form of the word 'black' of course differs: Nosu, Nasu and Nisu. Some of these languages preserve the initial prenasalized stop series reconstructible for all of Loloish—one of the other main phonological innovations of Loloish as a whole. Southern Nosu, some western Nasu and all Nisu have merged these into the oral voiced stop series. Confusingly, the southern Nosu and some northern Nisu call themselves Nasu, while some Nisu call themselves Niesu, due to differences in the reflex of the *ak rhyme within the three language clusters.

All three, as well as the Southeastern Yi group Sani, have a character-based script known to traditional religious practitioners. These are independent ideographic scripts, though the principle of the systems is Chinese and a very small number of the characters are borrowed from Chinese. Unlike Chinese, there is very frequent use of characters for homophones and near-homophones; and each lineage of traditional religious practitioners has somewhat different forms of the characters, even within one language cluster; characters are rotated, strokes omitted, added, or modified, or completely different pictographic representations used. The Sani orthography is the most distinctive; this is not so surprising as Sani is a Central Loloish language, unlike Nosu, Nasu, and Nisu.

Nosu is the largest group speaking a Loloish language, perhaps three million people mainly in southern Sichuan but also in northwestern and northeastern Yunnan. It has four main subvarieties: northern Tianba, northeastern Yinuo, western and central Shengza, and southern Sondi with subdialects Sondi and Adv; the last two are rather different from the first three. The largest group, over half of the total, is the Shengza, and the Xide local variety of Shengza has been selected as the standard; a new syllabic orthography based on 819 of the traditional Nosu characters as pronounced at Xide has

been widely used since 1978. Scholarly study of traditional manuscripts using the full range of characters and all regional variants also continues.

Nasu is quite closely related to Nosu but far from mutually intelligible, and has much larger internal differences than Nosu or Nisu. It is spread throughout western Guizhou and northeastern Yunnan, with a few in northwestern Guangxi. There are very numerous mutually unintelligible languages included, many of which are undescribed; doubtless some are still unknown. Chinese scholars divide it into three subgroups: southeastern (known also as Panxian from the county where they are concentrated, two subtypes); northeastern (most of the other Nasu in Guizhou, and some in extreme northeastern Yunnan and southeastern Sichuan; four main subtypes); and western (all in Yunnan, with five subtypes). Recent work by the author and others has located a number of previously undescribed and endangered Nasu languages, including Samei, Samataw, Sanyie, Ayizi and Gazhuo. Most of these smaller groups have no current literary tradition, but a number of the larger Nasu groups do, especially in Guizhou and northeastern Yunnan.

For the main variety of Nasu in north central Yunnan, a script based on the one devised by the missionary Samuel Pollard for Miao in Guizhou in 1905 was created and remains in limited use among Christian Nasu; confusingly, this language is called Nosu in the missionary literature but it is not the Sichuan Nosu; rather, it is identical to the Nasu described by the Chinese linguist Ma Xueliang in Wuding and Luquan counties, also called Hei Yi 'black Yi' in Chinese; before 1950 it was 'Hei Lolo'. But not all Hei Yi are Nasu; black refers to the aristocratic clans among other Northern Loloish groups, and so some Nosu and Nisu may also be Hei Yi. Non-aristocratic clans are called white, and Bai Yi refers to them, or also to various more Sinicized Yi groups who speak other languages, mainly of the Nasu cluster. To make matters worse, Bai Yi (with a different character for Yi) used to be used to refer to various groups speaking Thai (Zhuang-Dong) languages and now called Buyi in Guizhou, Zhuang in Guangxi, and Buyi or Zhuang in Yunnan.

The Nisu or Southern Yi are somewhat less diverse; nearly all are in south central Yunnan, but a few are also in Vietnam where they are still called Lôlô, as they were in China before 1950. The traditional Nisu writing system, which is considerably more unified

than those of the Nosu and Nasu, is no longer widely used, but a great deal of literature exists and is being analysed.

It is likely that the Yi traditional writing developed in eastern central Yunnan, near modern Kunming, during the Cuan kingdom (c. 5th to 7th century AD), and was dispersed and became diversified as a result of the Sui, Tang, Nanzhao and Yuan conquests of this area. Traditional migration stories and death chants trace all of these groups back to an origin around Dian Lake; dated and undated stone and copper bell inscriptions going back at least 700 years are scattered across northeastern Yunnan and into western Guizhou. Most manuscripts are undated, and are in any case the result of repeated copying each generation over a very long period. Some surviving dated manuscripts, including a few in Western libraries, are from the mid to late Qing period; but these are not the original dates, just the date of copying.

The traditional Nasu characters are being brought back into use in Guizhou, with each area allowed to use its local pronunciation; this follows the pattern formerly usual among the Chinese themselves. In Yunnan a new unified Yi script has been created and is intended to be used by the Nasu, Nisu and Southeastern Yi groups, as well as by those Yi without an orthographic tradition; this is not popular with the Nisu and Sani elite as there are more Nasu characters in the new script, which is also taught with Nasu as the default pronunciation at the Yunnan Institute of Nationalities. This composite script includes all shared characters (surprisingly few!) as well as many Nasu characters, some Nisu characters, and some Sani characters; again, it is meant to be pronounced according to the local speech form. In general, the various Yi literary traditions constitute a very rich and almost untapped resource on the history, religion and culture of these groups. Work is extremely urgent as the last of the fully literate traditional shamans are now old men. For further details and examples of these orthographies, see Bradley (2001a).

5.1.4.2 Central Loloish

Central Yi includes seven subgroups: the Southeastern Yi cluster, the Lisu cluster, Lalo, Lahu, Jinuo, Raorou and Nusu. It is characterised by the development of high or rising tones from certain prefixed *Low Stopped tones, among other phonological and lexical

innovations; languages of this cluster all have at least five tones, usually including some contour tones.

The Southeastern Yi cluster of Central Yi includes four named groups, Sani, Axi, Azha and Azhe, speaking fairly similar languages in southeast central Yunnan; the Chinese offical classification groups these as 'Southeastern Yi'. The best described is Sani, which also has an extensive literature in a traditional character orthography quite distinct from those of the Northern Loloish Nosu, Nasu, and Nisu language clusters. Azhe also has its own script, but this has not been investigated at all.

The Lisu are a large and widely dispersed group in China (most in western and northwestern Yunnan, but some in southern Sichuan as well); in Burma, mainly in the north; in Thailand; and a few in northeastern India, where they are known as Yobin from the former Burmese name Yawyin. Dialect differences are quite substantial; the Thailand dialect has extensive Chinese loans. Various orthographies exist; those now used include the 'Fraser' script, devised by a missionary of that name in 1914, which uses upper-case roman letters, upright and inverted, and punctuation marks for tones; and a new 1950s Chinese romanisation. Use of the latter is on the decline. There is also an indigenous syllabary invented in the 1920s in Weixi County in Yunnan, but now almost completely out of use. Some of the syllabic symbols resemble Chinese characters, but usually with different meaning, sound or both.

The names Lipo and Lolopo are the autonyms of a series of interrrelated groups classified by the Chinese as Central Yi. They speak languages which are linguistically very close to each other and to Lisu; some missionaries regarded one type of Lipo as an eastern dialect of Lisu. They are very numerous across northwestern central Yunnan. The Nosu or Northern Yi are to their north, Nasu or Eastern Yi are to their east, Nisu or Southern Yi are to their south, and Lalo or Western Yi are to their west; hence their Chinese designation as Central Yi. Some of the Lipo use a Pollard script, devised by missionaries for various languages, initially Miao; earlier, a few had adapted Nasu script; there was no traditional Lipo/Lolopo script. In addition to the large and relatively concentrated groups, there are other smaller Central Yi groups, some quite dispersed like the Micha and others, concentrated in some cases in one village or a cluster of villages. Little or no work has been done on any of these languages.

We have recently located some additional languages of the Lisu/Lipo/Lolopo cluster; they are Laemae in Bijiang County, Tanglang in Lijiang County, and Lamu in Binchuan County, all in northwestern Yunnan.

The Western Yi language cluster includes the Lalo or Lalu, Alo and various other groups centered around Weishan and Nanjian counties south of Dali in western central Yunnan. Björverud (1998) is a recent description of the syntax of one variety of Lalo. As in many other minority areas of China, the young people often have no knowledge of their group's traditional language. This group is also scattered as far south as Laos.

The Lahu are in southwestern Yunnan, northeastern Burma, northern Thailand and northwestern Laos, with a small group of refugees from Laos in the United States. The main dialect difference is between Black Lahu (Lahu Na) and Yellow Lahu (Lahu Shi), but there are extensive smaller differences within each; for details see Bradley (1979b). The Kucong (in Vietnam, Cosung) speak a variety of related but distinct dialects not mutually intelligible with Lahu and are scattered in south central Yunnan, far to the east of the main Lahu area; ten years ago they were officially amalgamated with the Lahu nationality in China, and their name also changed to Lahu in Vietnam. There are several roman orthographies; the most widely used is the Protestant Christian one for Black Lahu developed about 1920 and slightly revised in the early 1950s, though there are Catholic and Chinese romanisations as well, and two competing romanisations for Yellow Lahu. All of these mark the tones.

The compact Jinuo group in southern Yunnan was recognised as a separate nationality in China only in 1979, the only new nationality recognised in China since the 1950s. On its position in Central Loloish, see Bradley (1983).

The small Raorou group of about 2,500 in Lanping County, northwestern Yunnan, forms part of the Nu nationality; but their language is Loloish. Based on limited available data it appears to be Central Loloish.

Another part of the Nu nationality in Bijiang County to the west of Lanping is the Nusu, whose language (with three dialects) also appears to be Central Loloish. The other Nu in China speak one or other of the Central Tibeto-Burman Rawang/Nungish languages.

5.1.4.3 Southern Loloish

As for the Northern Loloish Yi nationality, there is a Southern Loloish nationality in China that includes a variety of distinct languages: Hani. Chinese linguists break this into three main subgroups: Hani-Akha, Haoni-Baihong, and Biyue-Kaduo. The first two are fairly closely related within the Hani-Akha or Akoid subgroup of Southern Loloish, but the third forms a separate cluster of languages.

In the Hani-Akha or Akoid group are the Akha, who are mainly in China and Burma, but also in Laos and Thailand. Dialect differences within Akha are fairly minor, apart from the Akeu dialect. According to their traditions, in which the ritual group name is [$za^{21}ni^{21}$], they migrated from the northeast about 500 years ago; to the northeast, the Hani [$ha^{31}ni^{31}$] in southern central Yunnan, Vietnam and Laos speak a cluster of very similar varieties, more or less mutually intelligible with Akha. To the northwest and north of the Hani are the Haoni [$xɒ^{31}ni^{31}$] and Baihong, speaking somewhat less similar dialects which extend into north central Yunnan. Within this cluster but more distinct from Akha/Hani/Haoni there are also several smaller groups: the Phana of Laos and the Sila or Sida of Laos and Vietnam, the Chepya and Muda of southwestern Yunnan, and probably others.

The core Hani/Akha languages share a striking transformation of the aspiration contrast from phonemic into allophonic: all originally stop-final syllables with voiceless stop or affricate initials become unaspirated creaky, while all originally non-stop-final syllables with voiceless stop or affricate initials become aspirated noncreaky. This does not happen in Haoni and other northern varieties of Hani; thus, for example, 'leaf' which is derived from *r-pak (low stopped tone) in all Loloish languages and has an aspirated [p^h] initial in nearly all Loloish languages, is [$p^h\underline{ɑ}^{21}$] in Haoni and other northern varieties of Hani, but [$p\underline{a}^{21}$] in Akha and Hani proper; see Bradley (2001b) for more details.

For Akha there are three competing romanisations: Catholic, Baptist, and a third based on the Hani romanisation, and for Hani in China there is another which was first proposed in 1952 but used mainly in 1958, then substantially revised in 1982 and used since.

The Bi-Ka subgroup is named from the Chinese term for its two main components: Biyue (Piyo) and Kaduo (Khatu). They are in southwestern Yunnan and are included in the Hani nationality; a

small number of Khatu live in northernmost Laos as well. There is a third related group, the Mpi, who live in two villages in Thailand; their tradition reports that they were brought as war captives from the north over 200 years ago—presumably from the Piyo and Khatu area. All Mpi are now bilingual in Northern Thai and culturally assimilated, but one village retains the language, at least for some purposes. The most striking sound change here is a partial merger of apical affricates into stops: *dz > /t/, *ts > /tʰ/.

The numerous small and mainly endangered Bisoid languages, including at least Laomian, Laopin and Sangkong of Yunnan, Hpyin in Burma, Bisu of northern Thailand, the five languages within the Sinsali (formerly Phunoi) ethnic group of northern Laos and the Côông of northwestern Vietnam, share the development of voiced stops corresponding to certain initial nasals in other Tibeto-Burman languages; for example, 'mother' is [ba^{33}] and 'fire' is [bi^{21}]. By contrast, they are the most conservative Loloish languages for final stops and nasals; for example, 'you' is [naŋ^{55}] and 'warm' is [lum^{55}]. Internal differences within Bisu and Sinsali are substantial; among the Bisu in Thailand there are distinct dialects for each of the four villages. Sinsali includes Laopan, Phongku, Lawseng and Pisu in addition to Sinsali or Phunoi proper. Few Laomian and Laopin still speak their language, Hpyin has been reported to be moribund since the 1930s (but is still alive), Bisu has now died in two of its four villages in Thailand, and Côông is one of the smallest nationalities of Vietnam. There is a new Thai-based writing system for Bisu, and a Lao one for Sinsali.

5.1.4.4 Southeastern Loloish

This subgroup includes a variety of languages whose speakers are called 'Pu' or servant in Chinese and Phula in Vietnamese. They are Shafer's 'Tonkin Lolo' subgroup of Loloish. A number of languages are included, including several just called Pu or Pula, and others such as Muang, Kathu, Laghuu and so on, scattered over southeastern Yunnan with a few in southwestern Guangxi and northernmost Vietnam. None of these languages has been adequately described.

5.2 Karen[3]

The Karen or Karenic languages are mainly spoken in Burma but extend into western Thailand, and there is a small community in India's Andaman Islands. The main distinctive characteristic is that all these languages are SVO in main clauses. Benedict (1972) treated Karen as a coordinate subgroup with Tibeto-Burman and Sinitic within Sino-Tibetan, but has more recently stated that Karen appears to be within Tibeto-Burman, and indeed close to the Burmic portion of Tibeto-Burman. Shafer (1974) places Karenic within Tibeto-Burman as a separate group coordinate with Burmic, Bodic and Baric. There is also considerable disagreement on the subgrouping within Karen, which includes a number of languages, many with several alternative names. Lexically, Karen shows various close links with Burmese-Lolo, as Benedict (1976) notes. Luce (1985) also gives substantial lists of Mon-Khmer lexicon in Karen (his Chart E) and uniquely Karen lexicon (Charts I and J). As several scholars have observed, innovative tonal developments in Karen are also extensive, with six tone categories reconstructed by Jones and eight by Luce. These have developed from the usual sources of tones in the region—syllable-initial and syllable-final characteristics—superposed on the overall Tibeto-Burman two-tone system found in all but Western Tibeto-Burman.

Jones (1961) suggests a subgroup which includes Pa-O, Pho and Lekeh versus the rest, which he divides into Sgaw and its dialects versus a central group with three subgroups: Padaung, Eastern Bwe, and Western Bwe. Solnit and Lehman classify Karen into Northern, Central or Bwe, and Southern, the latter including both Pho and Sgaw. Lehman differs from Solnit and Jones in grouping Gekhu with Bwe (Central) rather than Padaung (Northern).

Of the northernmost Karen groups, the Pa-O were formerly called Taungthu (Burmese for 'hill people') and Shan Tonghsu; these names are now regarded as pejorative. The Kayan (formerly known as Padaung) are the group whose women traditonally wear neck and knee rings. Also included here linguistically are Yinbaw (autonym Ka-ngan), Zayein (Latha), Gekhu ('upper', also sometimes written Gheko, Geko, Gekho, etc.) and probably Sawntung.

---

[3] I am grateful to David Solnit, F.K. Lehman, and R.B. Jones for personal communications on the subgrouping of Karen. They are of course not responsible for any errors in my understanding of their views.

'Kayah' (a new name invented by the Burmese to split them off from the other Karens) or Central Karen is still known as Karenni or Red Karen to its speakers; there are substantial dialect differences between east and west. It is a separate nationality with a separate state in Burma; there are some speakers in northwestern Thailand as well. There are various subgroups. For the small western subgroup Manu, the Burmese name is Manumanaw; the Kayah name is [pɯnɯ], which means 'western'. The name of the southwestern subgroup Yintale has been folk-etymologised into the Burmese Yin-Talaing or Mon Karen. The Blimaw subgroup of Karen, as described by Henderson, is sometimes simply known as Bwe or 'central' Karen. The group formerly known to its speakers as [brɛʔ], and hence Bre or Brek, now prefers the autonym [kəjɔ]. Geba is another Central Karen group.

Sgaw, in the Southern Karen subgroup, is the largest Karen language, known to its speakers as [sɣɔʔ] from the word for 'person'. It is spoken in the delta region of Lower Burma, in the Karen State of eastern Burma, and by the majority of the Karen in Thailand. There is a Burmese-based orthography using extra vowel and tone symbols, an older messianic movement's 'chicken track' orthography, and a new orthography developed by a Karen messiah more recently. Various divergent dialects have appeared in the literature as separate Karen languages; for example Mopwa (also known as Palaychi from one of the villages where it is spoken; studied by Jones) in the northwest, Paku [pakɯ] which is the Kayah word for Sgaw and also refers to a northeastern dialect of Sgaw, as well as Wewaw, Monnepwa and probably others.

Briefly reported by Jones, the Lekeh language also has a traditional orthography derived from Burmese but not widely known or used. The number of speakers is unknown; some at least live in the delta region around Rangoon.

The other major Southern Karen language is known to its speakers as Sho, to the Sgaw as Pwo or Pho, and in Thailand (from the word for 'person' there) as Phlong. Its speakers are mainly south of the Sgaw. This language extends quite far to the south, virtually to Burma's southern extremity, and onto some adjacent islands. There are very substantial dialect differences; the standard dialect has a Burmese-derived orthography with additional symbols and conventions for the vowels and tones of Pho.

Conclusion

The reconstruction of Tibeto-Burman from its subgroups has hitherto been mainly based on lexical similarities and sound correspondences in certain better-described languages (the 'macro' approach). While some attempts have been made to reconstruct inflectional morphology, starting from Wolfenden (1929), this remains an area where a great deal of work is needed. Selectively marshalled evidence has been used by some to suggest that there is no ergative marking and no verb morphology reconstructible for Tibeto-Burman; others have drawn the opposite conclusion using similarly restricted evidence. Neither case has been proven.

There has also been very little comparative study of historical morphosyntax in Tibeto-Burman, and little in the way of attempts to derive hypotheses about proto-culture and migrations from the distribution of etyma for flora, fauna, crops, and other vocabulary.

In a number of languages, there are traditional orthographies with substantial old text corpora that have hardly been tapped for comparative linguistic purposes, because the analysis of more accessible data has absorbed so much effort. Some such languages are extinct, like Zhangzhung, Nam, Tangut, and Pyu. These provide by far the earliest data on Northeastern Tibeto-Burman (Tangut) and on Sal Tibeto-Burman (Pyu), and very valuable information on Western Tibeto-Burman (Zhangzhung, Nam).

Another kind of source is embedded in Chinese historical materials. Such are the Bailang songs contained in the *Hou Han Shu* (Coblin 1978), which represent an early Loloish language of nearly 2000 years ago. There is also some Loloish and Bai lexical material from the Tang period in the *Man Shu* (Luce 1961). A further type is the extensive vocabularies prepared by the Ming and Qing translation bureau: a representation of the orthography of the language, a Chinese gloss and Chinese characters used to represent the phonetic value of the word. Nishida has used some of these to great advantage: the earliest available materials on Ersu (Nishida 1973) and the earliest glosses and phonetic representations of Nasu (Nishida 1980); the difficulty here is that unfamiliar scripts (such as Nasu) are usually somewhat garbled or at least deformed. Other such glossaries are equally interesting, as they provide attestation of the phonetics of some languages such as Burmese from intermediate periods long

after the development of the orthographies but still several hundred years ago.

But much richer are the materials in those languages with long continuous literary traditions, including some from every branch of Tibeto-Burman: Western Tibeto-Burman has Tibetan, Newari and Limbu, and if Zhangzhung is added, all four main branches are represented. Central Tibeto-Burman has Lepcha; Sal Tibeto-Burman has Manipuri and had Pyu; Northeastern Tibeto-Burman has Naxi and had Tangut; and Southeastern Tibeto-Burman has Burmese, Nosu, Nasu, Nisu and Sani. While none of these is as ancient as the continuous Sinitic literary tradition, they all have a crucial role to play in clarifying the history of Tibeto-Burman.

In addition, the situation with regard to the description of the modern languages is uneven, and every major subgroup awaits a full comparative treatment. Only when these gaps have been filled can the micro approach be fully used to evaluate competing macro hypotheses. Tibeto-Burman has a large contribution still to make to general historical linguistics, and Tibeto-Burmanists still have a lot of fascinating and challenging work to do.

References

Bauman, J.J. 1975. *Pronouns and pronominal morphology in Tibeto-Burman.* PhD dissertation, University of California, Berkeley.

—— (1976) An issue in the subgrouping of the Tibeto-Burman languages: Lepcha and Mikir. Paper presented at the International Sino-Tibetan Conference, Copenhagen.

Benedict, P.K. 1972. *Sino-Tibetan: a conspectus.* Cambridge: Cambridge University Press.

—— 1976. Sino-Tibetan: another look. *JAOS* 96/2: 167-197.

Björverud, S. 1998. *A grammar of Lalo.* Lund: Department of East Asian Languages, Lund University.

Bodman, N.C. 1987. The place of Lepcha in Sino-Tibetan—a lexical comparison. Paper presented at the International Sino-Tibetan Conference, Vancouver.

Bradley, D. 1975. Nahsi and Proto-Burmese-Lolo. *LTBA* 2/1: 93-150.

—— 1979a. *Proto-Loloish.* Scandinavian Institute of Asian Studies Monograph Series Number 39. London and Malmö: Curzon Press.

—— 1979b. *Lahu dialects.* Canberra: Australian National University Press.

—— 1983. The linguistic position of Jinuo. *In* Chu, C. et al. (eds) *Proceedings of the fourteenth International Conference on Sino-Tibetan Languages and Linguistics.* Taipei: Student Book Publishing Co., 21-42.

—— 2001a. Language policy for the Yi. In S. Harrell (ed.) *Perspectives on the Yi of Southwest China*, 195-213. Berkeley: University of California Press.

—— 2001b. Hani dialects. Paper to be presented at the International Sino-Tibetan Conference, Kunming.

Burling, R. 1983. The Sal languages. *LTBA* 7.2: 1-32.
Caughley, R.C. 1982. *The syntax and morphology of the verb in Chepang.* Pacific Linguistics B-84.
Coblin, W.S. 1978. A new study of the Pai-lang songs. *Tsing Hua Journal of Chinese Studies* n.s. 12.1-2: 179-216.
DeLancey, S. 1987. Sino-Tibetan languages. *In* Comrie, B. (ed.) *The world's major languages.* London: Croom Helm, 799-810.
van Driem, G. 1993. The Proto-Tibeto-Burman agreement system. *BSOAS* 56.2: 292-334.
—— 2001. *Languages of the Himalayas: an ethnolinguistic handbook.* Leiden: Brill.
Forrest, R.A.D. 1962. The linguistic position of Rong (Lepcha). *JBRS* 45.1: 41-7.
French, W.T. 1983. *Northern Naga: a Tibeto-Burman mesolanguage.* Ph.D. dissertation, City University of New York.
Glover, W.W. 1974. *Sememic and grammatical structures in Gurung (Nepal).* Norman: Summer Institute of Linguistics.
Grierson, G.A. 1903-1909. *Linguistic survey of India,* Vol. III parts 1-3. Calcutta: Superintendent of Government Printing.
Hale, A. 1982. *Research on Tibeto-Burman languages.* Berlin: Mouton.
Hansson, G. 1991. *The Rai of Eastern Nepal: ethnic and linguistic grouping findings of the Linguistic Survey of Nepal.* Kathmandu: Linguistic Survey of Nepal and Centre for Nepal and Asian Studies, Tribhuvan University.
Hashimoto, M.J. 1976. The agrarian and pastoral diffusion of languages. *In* M.J. Hashimoto (ed.) *Genetic relationship, diffusion and typological similarities of East and Southeast Asian languages.* Tokyo: Japan Society for the Promotion of Science, 1-14.
Henderson, E.J.A. 1965. *Tiddim Chin: a descriptive analysis of two texts.* London: Oxford University Press.
Jones, R.B. 1961. *Karen linguistic studies: description, comparison, and texts. UCPL* 25.
Löffler, L.G. 1966a. Chakma und Sak, ethnolinguistische Beiträge zur Geschichte eines Kulturvolkes. *Internationales Archiv für Ethnographie* 50: 72-115.
—— 1966b. The contribution of Mru to Sino-Tibetan linguistics. *ZDMG* 116: 119-160.
Luce, G.H. 1961. *Man Shu (Book of the Southern Barbarians).* Data Paper Number 44, Southeast Asia Program, Cornell University. Ithaca: Cornell University.
—— 1985. *Phases of Pre-Pagán Burma.* 2 Vols. Oxford: Oxford U. Press.
Marrison, G.E. 1967. *The classification of the Naga languages of North-East India.* PhD dissertation, University of London.
Matisoff, J.A. 1986. The languages and dialects of Tibeto-Burman: an alphabetic /genetic listing, with some prefatory remarks on ethnonymic and glossonymic complications. *In* J. McCoy & T. Light (eds), *Contributions to Sino-Tibetan studies, presented to Nicholas C. Bodman.* Leiden: E.J. Brill, 3-75.
—— 1991. Endangered languages of mainland Southeast Asia. *In* R. H. Robins & S. A. Wurm (eds) *Endangered languages*, 189-228. Oxford: Berg.
Moseley, C. (general ed.) 1994. *Atlas of the world's languages.* London: Routledge.
Nagano, Y. 1983. *A historical study of the rGyarong verb system.* Ph.D. dissertation, University of California, Berkeley.
Nagano, Y. & R. LaPolla (eds) 2001. *New research on Zhangzhung and related Himalayan languages.* Senri Ethnological Reports 19. Osaka: National Museum of Ethnography.
Nishida, T. 1973. *Tosu yakugo no kenkyû: shin gengo Tosugo no kôzô to keitô* [= A study of the Tosu-Chinese vocabulary Tosu I-yu]. Kyoto: Shokado.

—— 1980. Roro yakugo no kenkyu: rorogo no kozo to keito [=A study of the Lolo-Chinese vocabulary Lolo I-yu]. Kyoto: Shokado.
Sagart, L. 1999. *The roots of Old Chinese.* Amsterdam: Benjamins.
Shafer, R. 1974. *An introduction to Sino-Tibetan.* Wiesbaden: Harrassowitz.
Sprigg, R.K. 1983. Hooker's expenses in Sikkim: an early Lepcha text. *BSOAS* 46/2: 305-325.
Stargardt, J. 1990. *The ancient Pyu of Burma. Vol. 1: Early Pyu cities in a man-made landscape.* Cambridge: PACSEA and Singapore: ISEAS.
Sun, Hongkai 1983. Liu jiang liuyude minzu yuyan ji qi xishu fenglei—jian shu Jialing Jiang shang you, Yaluzangbu Jiang liuyude minzu yuyan. [=The Nationality languages in the six valleys and their language branches]. *Minzu xuebao* 3: 98-274.
Sun, Jackson 1993. The linguistic position of Tani (Mirish) in Tibeto-Burman: lexical assessment. *LTBA* 16/2: 143-198.
Thomas, F.W. 1948. *Nam, an ancient language on the Sino-Tibetan borderland.* London: Oxford University Press.
Thurgood, G. 1985. The 'Rung' languages: notes on their proto-morphosyntax and subgrouping. *Acta Orientalia* 46: 79-99.
Toba, S. 1992. Dhimal wordlist. Ms.
Turin, M. 1998. The Thangmi verbal agreement system and the Kiranti connection. *BSOAS* 61/3: 476-491.
Wolfenden, S.N. 1929. *Outlines of Tibeto-Burman linguistic morphology.* London: Royal Asiatic Society.

THE SINO-TIBETAN PROBLEM

CHRISTOPHER I. BECKWITH (INDIANA UNIVERSITY)

INTRODUCTION

The modern historical linguistic method is based on the fundamental, well-supported premise that phonological change is regular. The implication is that inherited forms in related languages exhibit *regular correspondences*, as in example (1), where initial **p-* in the proto-language corresponds regularly to initial *p, p*, zero, *p, f,* and *p* respectively in the daughter languages cited. Although Indo-European historical linguistics was initially based largely on comparative phonological and functional analysis of the morphological systems of the major ancient and medieval languages rather than on the lexicon per se, such correspondences, constrained by their phonological environment according to strict rules, nevertheless are among the foundation stones of Proto-Indo-European reconstruction.

(1) PIE **pəter* > Lat *pater*, Grk *patêr*, OIri *aθir*, Skt *pitár*, Got *fadar*, TokA *pâcar*

Another defining characteristic of a genetically related family of languages is that it has a *unique set* of forms and specific features not found in any other unrelated language or group of languages. When Tokharian was discovered early in the twentieth century, scholars immediately recognized that it possessed features typical of Indo-European, including the 'Indo-European set' of lexical items, and therefore that the language belonged to the Indo-European family despite the presence of unusual features not found in other languages of that family. Similarly, although Germanic, Greek, Indo-Iranian, and so on all have their own unique sets of elements not found in other Indo-European daughter language families, they are readily identifiable as belonging to Indo-European because of what they do share, such as pronominal, verbal, and other paradigms, the basic numeral system, and the 'Indo-European set' of lexical items, including, among other things, kin-terms such as **pətêr*. Semitic, Austronesian, and other well-established, uncontroversial families of

languages can be characterized in similar ways. The Sino-Tibetan theory, like the Indo-European theory, proposes to explain the similarities among the Chinese and Tibeto-Burman languages as the result of inheritance from a common ancestor, and the differences as the result of independent innovations after their divergence. It also contends that the relationship is exclusive—Taic, for example, despite many similar typological features and shared lexical items, is no longer considered to belong to Sino-Tibetan.

We should therefore expect to find regular phonological correspondences between Chinese and Tibeto-Burman, ideally between morphological markers, though since many of the languages involved are of the isolating type, the 'morphological markers' may be unbound or semi-bound function morphemes. There should also be a unique set of forms which are held in common (a positive correlation) but which are not found in any other language (a negative correlation). If tests for these two conditions fail, the theory must be considered untenable. If a common set of forms can be posited, but they are shared with another language or language family (for example, Indo-European, or Japanese-Koguryoic), there are two possible conclusions: either the Sino-Tibetan family should be widened to include or be included by the other language or languages, or Sino-Tibetan is simply a convergent grouping of languages that have borrowed a significant number of features from each other or from another language or languages, and share some other typological features due to their location in the same world area.

The lone putative Sino-Tibetan morphological system—first proposed by Conrady (1896), at a time when Taic was included as a member of the family—has recently been disproven (Beckwith 1996), leaving lexical correspondences and the phonological rules derived from them as the sole basis for the Sino-Tibetan theory (Beckwith 1996: 812-813). Unfortunately, those who are convinced of the rectitude of the theory claim that all the obvious similarities between Chinese and Tibeto-Burman, and especially between Chinese and Tibetan, are inherited. They claim that old loanwords from Chinese into Tibetan or vice versa are extremely rare (Benedict 1972, Matisoff 1991, Pulleyblank 1995b, Vovin 1997, Wang 1995), ignoring the fact that loanwords, too, once borrowed, obey the same phonological rules as inherited words. The common vocabulary of neighboring languages always includes loanwords (either monodirectional or bidirectional), whether or not the languages are genetically

related. Failure to separate out loanwords, especially early ones, from inherited vocabulary can be fatal for a linguistic relationship theory.

The resulting problem of widespread irregular correspondences threatens to falsify the Sino-Tibetan theory. For example, the unconditioned correspondence of Chinese *m- to Tibetan *m-, *^{m}b-, and *b- is not regular. Ignoring the problem will not make it go away, nor will pronouncements that there is no problem.

The dearth of regular correspondences between Chinese and Tibeto-Burman is perhaps the most critical single problem with the received Sino-Tibetan theory. Irregular correspondences are a feature even of the most widely accepted Sino-Tibetan etymologies, such as the words in examples (2) through (4), a lexical set—one of several often treated as fundamental proofs of the Sino-Tibetan theory—proposed already by Yakhontov (1965: 15). They are given here in accordance with the reconstruction systems of Pulleyblank (1991) for Middle Chinese,[1] and of Karlgren (1957), Starostin (1989), and Baxter (1992) for Old Chinese, in all of which reconstructions the words do look similar, though none of them explains the *l* in the onset of Old Tibetan *lŋa* 'five'. As will be seen, the apparent similarity among these words disappears in Old Chinese.

(2) 五NMan *wǔ* 'five' < LMC *ŋu < MChi ⋆*ŋɔ*^{2} (Pul. 325) < OChi *ŋo (Kar. 35) ~ *ŋaʔ (Bax. 795) ~ *ŋhâʔ (Sta. 563) : OTib *lŋa*

(3) 吾NMan *wú* 'first person pronoun' < LMC *ŋu < *MChi* ⋆*ŋɔ*^{1} (Pul. 325) < OChi *ŋo (Kar. 35)[2] ~ *ŋâ : OTib *ŋa*

---

[1] The Late Middle Chinese (LMC) examples here are my own reconstructions of the largely unattested nasal-initial dialect usually cited by Sinologists. LMC had prenasalized voiced stop initials rather than nasals (Pulleyblank 1984: 68); cf. the discussion in Section II. The other Middle Chinese forms in examples (2) through (4) are taken verbatim from Pulleyblank (1991), except that his tone symbols have been converted mechanically to numerals (1 = level, 2 = rising, 3 = departing). Subsequent examples in this paper have often been modified slightly; a source citation is provided when they are quoted exactly. The terms Middle Chinese (MChi) and Old Chinese (OChi) are used to refer to traditional—essentially unperiodized—reconstructions, as in the framework of Historic Sinological Reconstruction (on which see below). This article follows the usual pragmatic practice of citing Chinese words and the characters used to write them, regardless of period or dialect, according to their New Mandarin pronunciation, given in *pinyin* romanization. No theoretical, historical, or political conclusions should be drawn from this practice. The digraph "ng" has been converted mechanically to "ŋ" and the macron (¯) to circumflex (ˆ), except in direct quotations. Medieval attestations in Chinese character phonetic transcriptions or *fanqie* 'spellings' are marked with a star (⋆).

[2] This word does not occur in a rhyme in the *Book of Odes*, so it is not given in the rhyme tables of Starostin and Baxter; the form *ŋa is intended to represent

(4) 魚NMan *yú* 'fish' < LMC *ŋü < MChi ⋆ŋiă[1] (Pul. 380) < OChi *ŋi̯o (Kar. 40) ~ *ŋ(r)ja (Bax. 804) ~ *ŋha (Sta. 561) : OTib *nya*

Although recently Laurent Sagart (1995, 1999) has argued that the Sino-Tibetan corpus includes numerous loanwords, including some from Tibeto-Burman into Chinese, the problem of establishing what regular correspondences exist between the two language families, and determining whether they are due to divergence or convergence, remains as unsolved as ever.

The few modern linguists who have publicly doubted or rejected the received Sino-Tibetan theory have done so on the basis mainly of other criteria, not those of phonological correspondences strictly speaking (Beckwith 1987: 3-4, 1996; Miller 1988). Recent advances in Old Chinese reconstruction, from Pulleyblank (1962) and Yakhontov (1965) to Starostin (1989) and Baxter (1992), within the framework of Historic Sinological Reconstruction (HSR),[3] have not directly addressed the problem of irregular correspondences, which thus continues to be an obstacle to general acceptance of any Sino-Tibetan theory.

It is too early to attempt a complete, periodized reconstruction of Old Chinese, to which a complete, periodized reconstruction of Proto-Tibeto-Burman, or just Common Tibeto-Burman, could be compared. Even Middle Chinese remains problematic, despite the nearly universal and wholly unjustified practice—begun by Karlgren —of citing Middle Chinese reconstructions without asterisks. The main reason for the many problems with Old Chinese reconstruction (there is little point in talking about Proto-Chinese reconstruction, since it is rarely discussed in the field) arise from Sinologists' dependence on a traditional system of analysis based on the one hand on the 切韻 *Ch'ieh yun*, a pre-T'ang Chinese rhyme book, and the systematized post-T'ang rhyme tables derived and developed from it, and on the other hand on the rhymes of the 詩經 *Shih ching*, or *Book of Odes*, the received text of which dates to sometime in the middle of the first millenium B.C., all combined with information derived from comparison of the phonetic elements found in Chinese characters, as well as rebus usages, borrowings, and so forth, in early

---

Starostin's system. The word 語 *yŭ* 'speech, to speak', in which the phonetic is 吾 *wú*, does occur as a rhyme in the *Odes*, and is reconstructed *ŋi̯o (Kar. 35) ~ *ŋ(r)ja? (Bax. 805) ~ *ŋa? (Sta. 562).

[3] See the discussion below.

texts. Although this collection of information derives from every period of Chinese history from the Shang to the Sung Dynasty—a period of over two thousand years—this traditional method, which has been called Historic Sinological Reconstruction (HSR), has achieved some remarkable results, and is currently the only method actually in use by scholars (other than the present writer) for reconstructing Pre-Han period Old Chinese. Rather than criticize the method more than it has already been criticized, it will only be suggested here that many of the problems of HSR, including its tendency to create unnatural forms, arise from its relative isolation from other, harder, data. In better-established fields of historical linguistics it is normal to consider all relevant data, particularly data from other related languages and data from loanwords. The latter often constitute the only evidence of earlier forms of the target language, because much is lost over time and cannot be reconstructed purely internally. The Tibeto-Burman languages contain a large number of early Chinese loans, and the early Tibeto-Burman languages, especially, will play a great part in the eventual development of Old Chinese reconstruction whether or not the Sino-Tibetan theory is maintained.

However, since Tibeto-Burman data—which is nearly always considered to be 'cognate', due to premature acceptance of a crude model of the Sino-Tibetan genetic hypothesis—has long been used rather freely by Sinologists to provide templates to aid in the reconstruction of Old Chinese within the HSR framework, there is a pressing need for a careful examination of the interrelationships of some of this data. If some specific assumptions of the Sino-Tibetan theory are reconsidered, it is conceivable that it may be possible to outline a modified version of the received theory, to reject the received theory and propose a new one, or to propose an entirely new theory of the divergent and convergent interrelationships of the languages of eastern Eurasia. With these questions in mind, this paper examines some of the irregular correspondences between Tibeto-Burman and Chinese that have long been stumbling blocks both for Old Chinese reconstruction and for the Sino-Tibetan theory, and also investigates the hitherto largely overlooked but serious problem of non-uniqueness.

I. EAR • MAN

The ordinary Mandarin Chinese word for 'man, person, people', 人 *rén*, derived from Middle Chinese ⋆*nyin*^{1}, which is said to derive from Old Chinese *nyin, does not correspond to Proto-Tibeto-Burman *mî 'man, person, people', its semantic equivalent.[4] Faced with this difficulty, Sino-Tibetanists argue that the true Chinese cognate of the Tibeto-Burman word is 民 *mín* 'people', a form first attested in Middle Old Chinese, usually with the somewhat pejorative meaning 'the common people'. Since this word is not found in Early Old Chinese, whereas *rén* is, *mín* would not seem to be the direct cognate of Tibeto-Burman *mî.

Historical phonologists of Chinese have long known that the initial *r-* of New Mandarin syllables—including syllables that became *ri* [ʐï] in early Mandarin and have become New Mandarin *er* —derives from Late Middle Chinese initial *ź-*[5] from Early Middle Chinese ⋆*ny-*. There are now considered to be two regular Old Chinese ancestors of Early Middle Chinese ⋆*ny-*, namely *n- and *ŋ-. While a few words, including 兒 *ér* 'child, son', as in example (5), descend regularly via Late Middle Chinese *ź-* from Middle Chinese ⋆*ny-* but go back to *ŋ- in Old Chinese, most have normally been reconstructed with initial *n- or *ny-. Thus, 人 *rén* 'person, man; people' is traditionally reconstructed as in (6) for Old Chinese.

(5) NMan 兒 *ér* 'child, son' < LMC *źi* < *^{n}*źi* < MChi ⋆*nyi*^{2} < OChi *ŋye (Bax. 212)

(6) NMan 人 *rén* [ʐɛn] < LMC *źin*^{6} < *^{n}*źin*^{7} < MChi ⋆*nyin*^{1} < OChi *ńi̯ĕn (Kar. 110) ~ *nin (Sta. 581) ~ *njin (Bax. 784)

However, the traditional reconstruction fails to explain why 耳 *ěr* 'ear' (from Middle Chinese ⋆*nyi*^{2}), in example (7), was used as the phonetic in constructing 弭 *mǐ* 'ends of a bow', in (8), and other

[4] Benedict reconstructs **mi*(*y*) and **r-mi*(*y*) (1972: 107, 119, 158).

[5] Pulleyblank's presentation (1984, 1991) of T'ang dynasty Chinese ('Late Middle Chinese'), while undoubtedly a fairly accurate theoretical reconstruction of the ancestor of Old Mandarin, is not true to the *attested* data in its representation of LMC *r*, which is unambiguously clear in foreign transcriptions, but occurs only in coda position. The initial *ź* is regularly transcribed as *ź* in the extensive LMC corpus written in Tibetan script, which script also accurately and regularly transcribes LMC final [r] as *r*. In the present instance (among others) his reconstruction of **r-* for the initial of such words in LMC cannot be supported.

[6] Attested in the Tun-huang LMC texts (Takata 1988: 380-381).

[7] Attested in the Tun-huang LMC texts (Takata 1988: 292, 381).

characters with initial *m-*, or why 爾 *ěr* 'you' (from Middle Chinese ⋆*nyi*^{2}) in example (9) is the phonetic in 瀰 *mí* 'to fill, overflowing', in example (10), and in several other characters.

(7) NMan 耳 *ěr* 'ear' < MChi < ⋆*nyi*^{2} OChi *nyi (?)
(8) NMan 弭 *mǐ* 'ends of a bow' < MChi ⋆*myi*^{2} < OChi *myi (?)
(9) NMan 爾 *ěr* 'you' < MChi ⋆*nyi*^{2} < OChi *nyi (?)
(10) NMan 瀰 *mí* 'fill, overflow' < MChi ⋆*myi*^{1} < OChi *myi (?)

Pulleyblank long ago noticed this phenomenon in the word 柔 *róu* 'soft' and its etymological relatives: "柔 M.[8] *ńi̯u < *nūh ∶ 矛 M. *mi̯u*; but M. *ńi̯u 'soft' seems clearly to be related to words like 耎 M. *ńi̯wen 'soft', 懦 M. *ńi̯ou 'weak, soft' which must have original *n-" (Pulleyblank 1962:237). Unfortunately, he did not pursue this lead further, and although the normal reaction to the discovery of such anomalies in Chinese has been to propose a solution, in this case the response has been to ignore the problem.

The solution must be that these words, and others beginning with *r-* today, derive from an Old Chinese *m-* initial which was palatalized by assimilation. The same process occurred in the history of Tibetan from Old Tibetan to New Amdo Tibetan and—before non-front vowels—from Old Tibetan to New Central Tibetan, wherein earlier *my- regularly became *ny-*. A more thorough examination of the Chinese evidence reveals that many, perhaps most, Mandarin syllables now beginning with *r-* or pronounced *er* once began with a bilabial.[9]

The words 人 *rén* 'man, person, people' (from Middle Chinese ⋆*nyin*^{1}) and 民 *mín* 'people' (from Middle Chinese ⋆*myin*^{1}) belong to the same rhyme category in Middle Chinese and Old Chinese, have the same tone, and are obviously very closely connected semantically. In light of the derivation of Mandarin initial *r-* in the majority of cases from Old Chinese *my- via Middle Chinese ⋆*ny-*, it is clear that these two words go back to a single Old Chinese ancestor beginning with *m- and ending with *-in. Yet this ancestor must

[8] M. is Pulleyblank's abbreviation for 'Middle Chinese' in this quotation.

[9] In addition to the words of this type discussed in the present paper, a list of words with initial *r-* < *my-, with examples of relatives having initial *m-* or another bilabial (on which see the discussion below), would include 柔 *róu* (矛 *máo*), 而 *ér* (而已 = 耳), 二 *èr* (佴 *èr*, 倍 *bèi*), 乳 *rǔ* (孚 *fú*), 燃 *rán* (燔 *fán*, 焚 *fén*), and 任 *rèn* (in 任那 *Rèn-nà* < *mimna 'Mimana', a Korean P.N. in Old Japanese transcription), among others.

have split, because *mín* remained unpalatalized and did not become *rén*.

Although there are other possible explanations for this, the most likely answer is the simplest: the unpalatalized word is a dialect borrowing, like the variant readings of 日 *rì* 'sun' given in example (17) below and discussed further in Section VII, and like the later plethora of 'borrowings' of words that did not undergo—or did not fully undergo—the change of Middle Chinese **ny-* to *r-* (such as 你 *nǐ* 'you', from such a dialect; the phonetic of the character is a simplified form of 爾 *ěr* 'you'). In the case of the words for 'man' and 'people', the dialect that donated *mín* 'people' to the central dialect had obviously never palatalized its margins; the coda of both *rén* and *mín* was either *-n originally (and, as an alveopalatal already, not subject to further palatalization) or *-r, which changed to *-n in late Middle Old Chinese, as shown in examples (11) and (12).

(11) NMan 人 *rén* [ʐɛn] < LMC *źin* < *ⁿźin* < MChi **nyin^{I}* < late MOC *nyîn ~ *myîr < early MOC *mîr < *^{m}bîr

(12) NMan 民 *mín* 'people' < LMC *fibyin* [^{m}b^{j}in] < MChi **myin^{I}* ← late MOC dial. *mîn ~ *mîr < early MOC *mîr < *^{m}bîr

The same process explains the relationship between 耳 *ěr* 'ear' and 弭 *mǐ* 'ends of a bow' in (13) and (14), and other such pairs, such as 爾 *ěr* 'you' and 瀰 *mí* 'to fill' in examples (15) and (16), 日 *rì* (~ *mì*) 'sun' and 汨 *mì* 'the Mi-lo river' in (17) and (18),[10] and so on.

(13) 耳 *ěr* 'ear' < MChi **nyi^{2}* < OChi *myi

(14) 弭 *mǐ* 'bow ends' < MChi **myi^{2}* < OChi *mi

(15) 爾 *ěr* 'you' < MChi **nyi^{2}* < OChi *myi

(16) 瀰 *mí* 'to fill' < MChi **myi^{1}* < OChi *mi

(17) 日 *rì* 'sun' < MChi **nyit* < OChi *myič < *myik^{y}
日 *mì* 'sun' < MChi **mɛyk* < OChi dial. *mik

(18) 汨 *mì* 'the Mi-lo river' < MChi **mɛyk* (Pul. 213) < OChi *mik

The Middle Chinese initial **ny-* of the ordinary word for 'man, person, people', NMan 人 *rén*, is the regular reflex of the Middle Old Chinese initial *my-. This corresponds well to the bilabial nasal initial of the Proto-Tibeto-Burman equivalent—and putative cognate—of this word, *mî 'man, person, people'. The problem of the syllable finals (and the theory that the Chinese words earlier had velar nasal codas) is discussed below in Section IV.

[10] See the discussion in Section VI.

II. Not • You • Two

One of the most notable distinctions between the modern Chinese and Tibetan languages are the differences to be seen in the simple negatives, as shown in (19).

(19) NMan *bu-* 'not' : NTib *ma-* 'not'

Despite the apparently total phonetic non-correspondence of these two words, the existence of other negatives in Chinese with earlier initial *m- has been used as an argument in favor of relating the Chinese and Tibeto-Burman negatives (Pulleyblank 1995: 187). However, the existence of two series of negatives—a stop set and a nasal set—in Chinese, both with bilabial initials (as well as the existence of other related sets of words with both stop and nasal initials), and the parallel existence of alveodental initial negatives in other possibly connected languages, has never been explained (Pulleyblank 1995: 187-188). The non-correspondence of the Chinese stop-initial negative set to the solidly-reconstructed lone Proto-Tibeto-Burman negative *ma is one of the most enduring and widely ignored problems in Chinese historical phonology and comparative Sino-Tibetan linguistics. It should not be left under the rug, where it has been swept in the past.

The phonological problem underlying the non-correspondence of the negatives is particularly noticeable because 不 *bù* 'not' is the ordinary negative in Chinese from antiquity down to the present day. The bilabial stop initial is found also in several other negatives, such as 否 *fǒu* 'not so' (with Middle Chinese **p-*), while still other negatives, such as 無 *wú* 'not exist, not have, there is not', had bilabial nasal initials throughout antiquity (for example, 亡 *wáng*—the Early Old Chinese and early Middle Old Chinese semantic equivalent of 無 *wú*, both of which derive from an Old Chinese negative stem *ma-), the latter mostly shifting to bilabial glides in the late T'ang period.[11] Since this phenomenon is by no means restricted to one category within the lexicon, it is undoubtedly to be explained phonologically, not morphologically.[12] The task

[11] See the discussion of *wú* and other negatives by Schuessler (1987: 646).

[12] Sagart suggests that the nasal ~ stop relationship found in the negatives may be due to semantic extension of a phonetically similar verb meaning 'to eliminate' or the like, prefixed with a bilabial nasal prefix morpheme that subsequently fused with the etymological initial before Shang times (Sagart 1999: 84).

is to discover a phonological process at least as regular as that which changed the Early Middle Chinese initial palatal nasals into the initial *r-* of Mandarin.

The answer is to be found in the shifting of the prestige or official dialect, and interdialect borrowing, in the Old Chinese period. The process was fortunately not restricted to high Antiquity, as we also have excellent data attesting to it from the Early Middle Ages (the T'ang dynasty, effectively from the early seventh to the late ninth centuries A.D.), in the earliest segmental script texts in the Chinese language, the Late Middle Chinese texts from Tun-huang in Tibetan script (Takata 1988).[13]

These texts, supported by the *kan-on* layer of Middle Chinese loanwords in Japanese (cf. Pulleyblank 1991: 2), reveal perhaps the single most distinctive phonological characteristic of the Late Middle Chinese language: its prenasalized voiced stops. For example, 我 *wǒ* 'first person pronoun' is attested as *figa* [ᵑga] (MChi ⋆*ŋa*^{2}), 那 *nà* 'that' as *fida* [ⁿda] (MChi ⋆*na*^{3}), and 魔 *mó* 'demon' as *fiba* [ᵐba] (MChi ⋆*ma*^{1}) (Takata 1988: 304-306). These correspond to syllable-initial nasal consonants in the Chinese of the immediately preceding and following periods. Although this shift was, in part, temporary, it had some irreversible consequences. As seen in Section I above, certain syllables which had previously had nasal onsets moved into the fricative onset category, and when the capital moved out of the semi-oralized nasal onset dialect region into a dialect region with ordinary nasal onsets the shifted forms remained as a residue of the previous dialect shift.

The capital of unified China moved many times in Chinese history, but after the T'ang it never returned to the 'Western Capital', Ch'ang-an (modern Xi'an), giving some the impression that the linguistic change was simply a temporary fluke.[14] However, the city of Ch'ang-an and other cities nearby were often capitals of China at different periods in Chinese history from at least the Chou dynasty onward, lasting for very long intervals each time. This shift of capitals or cultural centers, and accordingly of the prestige dialect

---

[13] The importance of the shifting of the capital and with it the prestige dialect has been pointed out previously; cf. Pulleyblank (1984: 2-4).

[14] The prenasalized *stops* of LMC are indeed lost with the post-T'ang dialect shift, so the LMC evidence has generally been ignored. Since no major modern dialect group descends directly from the LMC prenasalized dialect the LMC forms are generally omitted here unless relevant to the discussion, as in Section I.

spoken by scholar-officials, soldiers, and others based in and sent out from the capital, not only affected the development of Middle Chinese, it had earlier affected Old Chinese—and even pre-Early Old Chinese—in a similar way.

In Proto-Chinese there was only one negative stem, *ma, but before Early Old Chinese a shift to a prenasalized stop dialect must have occurred, producing (*^{m}ba- ~) *^{m}pa-.[15] Generally, the connection of this dialect's prenasalized stop forms with the ordinary nasal forms of other dialects would have remained clear and not caused any ambiguity, as we know from the T'ang case, and normally the shift back to a nasal onset dialect would have taken place unnoticeably, with the exception of a few stranded 'loan' forms. However, the prenasalized stop articulation of a word with a *bilabial nasal* onset in the official dialect would have sounded to speakers of other (ordinary nasal onset) dialects the same as a contraction of two words—a negative plus a word beginning with a labial stop—in their own (non-prenasalized) dialect. Because the negative stem in the official dialect was pronounced with a prenasalized stop, it would thus have seemed to speakers of most Chinese dialects that there were two negative stems, *ma- and *ba- (< *^{m}ba-/), in the prestige dialect, producing *ma- and *pa- in the pronunciation of the prestige-dialect by speakers of other, non-prenasalized-onset dialects.[16]

The dialect stem-form with prenasalized stop initial was thus borrowed—as a simple stop initial—into the non-prenasalized dialect (the eventually prevailing mainstream dialect), in which the inherited negative stem with an ordinary nasal initial was maintained side by side with the borrowed prestige form. This change must have taken place before Early Old Chinese, since even in the Oracle Bone Inscriptions there are already several derived forms belonging to both nasal and oral initial stems. Because there are contemporaneous, semantically related pairs of negative words in which one word has a bilabial nasal initial, while its mate has a bilabial stop initial, it is also

---

[15] The same process produced forms such as *^{n}da > ^{n}ta, *^{ŋ}ga > *^{ŋ}ka, and so on, depending on the point of articulation of the underlying nasal initial, as happened again in LMC. The devoicing of the resulting stop in OChi appears to be due to the phonology of the borrowing dialect, which seems not to have had voiced stops. However, it has sometimes been argued that PChi itself did not have voiced stops, in which case the change was directly *ma- > dial. *^{m}pa- → *pa-.

[16] There is some textual support—involving the negative 不 *bù*—for this process having occurred at least twice before the T'ang, namely in the Western Chou and Western Han periods (Sagart 1999: 88-99, citing research by Wolfgang Behr).

clear that these prenasalized initial negatives were borrowed as a set. Even though semantic drift has since occurred and the new forms have taken on lives of their own, the borrowing contributed eventually to the development of new words derived from roots with bilabial stop initials. Yet pairs of derived forms continue to correspond in large part, as summarized in example (20), derived from Schuessler (1987).[17]

(20) 不 *bù* < late MOC *pré < EOC *pare 'not'
毋 *wú* < late MOC *mré < EOC *mare 'should not'
弗 *fú* < late MOC *pét < EOC *pate[18] 'not, cannot'
勿 *wù* < late MOC *mét < EOC *mate[19] 'should not, don't'
匪 *fěi* (< *fǐ*) < late MOC *prê < EOC *parê 'is not, un-'
靡 *mǐ* < late MOC *mrê < EOC *marê 'there is no, without-'

The prenasalization and deprenasalization of the dominant dialect brought about by physical movement of the political center had repercussions in other parts of the lexicon as well, including the pronouns. Reconstruction of the pronouns—in particular the second person pronouns—is dependent upon recognizing the process, since one consequence of the correct internal reconstruction of bilabial nasal initials for the second person pronouns in Chinese is that the Tibeto-Burman equivalents (as well as the Japanese) no longer seem to correspond, as shown in example (21).

(21) OChi *myî, *myâ 'you' : PTB *na 'you' : OJpn *na* 'you'

Because phonetic series based on *n-, a depalatalized dialect form of initial *ny- (which developed from *my- in late Middle Old Chinese or in some dialects much later[20]), spread throughout the language before the earliest surviving dialect split, traces of the earlier bilabial

---

[17] Schuessler gives the first four as two sets (Schuessler 1987: 48). His reconstructions are: 不 *pjə 'indicative objective', 弗 *pjət 'indicative volitional', 毋 *mjə 'injunctive objective', 勿 *mjət 'injunctive volitional'.

[18] A fusion of 不 *bù* and 之 *zhî*, considered to be of rare occurrence in MOC (Schuessler 1987: 175); there are also MChi readings of 不 with final *-t (Pulleyblank 1991: 43), but the ancestor of NMan *bù* is attested without a final in LMC (Takata 1988: 376-377).

[19] Apparently a contraction of 毋 *wú* 'don't, shouldn't' and 之 *zhî* (< MOC *tê) 'it', in some cases (Schuessler 1987: 650).

[20] See Section 1 above. There is good evidence that the palatalization did not happen at the same time all over China, and may not have happened at all in some dialects which have since disappeared, such as Ch'u (replaced by the Central dialect before the Han period) and the far northeastern dialect (a descendant of which was spoken in the Lo-lang Commandery and other areas of what is now Korea).

nasal initial are now found only in etymologically related words (such as *mín* 'people', discussed above) borrowed later from dialects that had not undergone the palatalization. Moreover, due to still later dialect shifting, irregularities have developed in Mandarin, such as the initial of the modern second person pronoun 你 *nǐ* 'you'. While such forms have provided ammunition for Sino-Tibetanists who want to compare a Proto-Chinese *na or *naŋ to a Proto-Tibeto-Burman *na or *naŋ, the purely internal Chinese evidence unequivocally indicates that the second person pronoun root had a bilabial initial in *early Middle Old Chinese*. The same applies to the numeral 二 *èr* 'two', which purely internal evidence indicates should be reconstructed with a bilabial initial.[21] However, the dialect shifting phenomenon described above also provides a solution to the non-correspondence of the Old Chinese pronouns to the Tibeto-Burman and Japanese forms.

The apparent relationship of certain Tibeto-Burman and Japanese words with initial *n- to Chinese words with initial *m-, as well as purely internal Chinese evidence, indicates a change of Proto-Chinese *w to *m in Early Old Chinese in all positions (except perhaps after palatal velar stops[22]), merging with original *m. When the change occurred, a form *Twe (where T represents any alveodental stop) would have produced *Tme in the plain nasal dialect, but *T^{m}be in the prenasalized stop dialect. Each had a different outcome.

Firstly, in the plain nasal dialect the stem-form *Tme would not have surfaced because it has a disallowed initial cluster which would have been regularly metathesized (as were certain other clusters that developed later when syncope reduced earlier CVCV words to the monosyllabic shape CCV, as discussed below), producing *mTe. The initial nasal of the latter form having assimilated to and voiced the stop, producing *nde—a form virtually identical to a typical semi-oralized nasal (prenasalized stop) dialect form—the word was analogically 'restored' to a plain nasal *ne in the unprenasalized dialect.

---

[21] Cf. 佴 *èr* 'to second, assist' (with 耳 *ěr* 'ear' < OChi *myi) as phonetic), *bèi* 倍 'double; times, n-fold', 並 *bìng* 'both, side-by-side', 朋 *péng* 'pair, two strands of cowries'; ccf. note 9 above.

[22] The exception is drawn partly from the evidence of putative early loanwords of Indo-European origin, including the most well-known example, the word 犬 *quǎn* < MChi *$k^{j}wen^{2}$ 'dog', which is probably related to PIE *k$^{\text{y}}$won 'dog' (Wat. 46; cf. Pulleyblank 1995: 179-180). The source of Tokharian *ku* 'dog'—and perhaps TB *kwê 'dog' as well—is said to be a PIE nominative singular form *kúwô (Adams 1999: 179).

The short vowel *e regularly changed to *a in Tibeto-Burman, resulting in *na, an exact match for the reconstructed second person pronoun of Proto-Tibeto-Burman and the attested second-person pronoun in Old Japanese, as shown in example (22). It would seem possible that some medieval Chinese dialects could either have inherited a descendant of a Middle Old Chinese lengthened-grade form—*nê— of this word, as in (23), or created the form *nî by analogy with the form 爾 **nyî*[2] 'you' inherited from Early Middle Chinese.[23]

(22) PTB *na, OJpn *na* '2p. pro.' < *ne < *^{n}de < *mTe < *Twe
(23) 你 NMan *nǐ* '2p. pro.' < *nê < *^{n}dê < *mTê < *Twê

However, the scenario in (23) is unlikely for 你 *nǐ*, the character for which has a simplified form of 尒 as its phonetic. The word is clearly a retention of the older Middle Chinese form by a non-prenasalized dialect, which never underwent the shift from nasal to fricative described in Section I above.[24]

Secondly, in the prenasalized dialect, a stem-form *T^{m}bê-, from *Twê-, would also have a disallowed initial cluster, but in this case the alveodental stop *T would have been analyzed as being outside the syllable entirely and would simply have been dissimilated, leaving *^{m}bê-. Having become identical to the prenasalized form of *mê-, it was analogically 'restored' to *mê- when the official dialect shifted out of the prenasalized area, because after the alveodental stop was deleted the connection with the etymological ancestor *Twê was no longer retrievable. This new stem *mê- is the ancestor of the earliest attested Old Chinese second person pronoun, 女 *rŭ* 'you', which is written with the pictographic character 女 'woman' as a rebus in Early Old Chinese and early Middle Old Chinese, revealing that the two words were homonyms at that time.[25] The derived vowel

---

[23] See the discussion of 爾 *ěr* 'you' in Section I.

[24] For another explanation see Pulleyblank (1984: 69).

[25] The development of a distinctive pronunciation of 女 in the sense of 'woman, female' (NMan *nǚ*) began well before LMC, in which it is attested as *dźi/fidźi* *[(n)dźü], explained by Pulleyblank (1984: 69) as representing a retroflex [ɳ] (though this explanation is highly unlikely for the northwestern dialect of the T'ang, based on the Tibetan transcription evidence and Japanese readings). In fact, the word for 'woman', reconstructible for early MOC as *mîna, from PChi *Cwêna (with an unknown initial consonant C), diverged from the second person pronoun no later than Late Old Chinese (see Section VIII below). Cf. OJpn **mina* 'woman', and OJpn **me* 'female' < PJpn *miCa, thought to be from *mina (Martin 1987: 474). The reading *rŭ* (attested LMC *źu*) in the sense 'you' is the expected form. The alternate character 汝 *rŭ* 'you' was developed in Antiquity to distinguish the two

(and other evidence for both words) tells us that we have to deal with an extended form for 'you',[26] as shown in (24).

(24) NMan 汝 (女) *rǔ* [ʐu] 'you' < LMC źu^2 < *^{n}źu^2 < *^{n}yu^2 < MChi ⋆*nyu*^{2} < ⋆*nyô*^{2} < late MOC *nyâ < *mnyâ < *m^{y}nâ < early MOC *mîna ~ *mêna < *^{m}bêna < *T^{m}bêna < EOC *Twêna

The pronoun 爾 *ér* 'you', in (15), a word with a clear bilabial nasal initial in early Middle Old Chinese, is undoubtedly related to 汝 (女) *rǔ* 'you' and gradually replaces the latter in late Middle Old Chinese times, as shown by Sagart (1999: 142-143). This shift is most probably due to the movement of the capital out of the prenasalized dialect area at the beginning of the Eastern Chou period in 771 B.C.

The Early Old Chinese second person possessive pronoun 乃 *nǎi* 'your' also belongs to a phonetic series (Karlgren 1957: 249) with a probable bilabial-initial member—namely 仍 *réng* 'repeat, as before, again and again'. The semantics of this word relate it to other words for 'twos', 'pairs', 'doubles', and so on, nearly all of which have bilabial initials, as noted above. This, together with the closely parallel phonology of the second person pronoun stem, suggests that the word for 'two' had a similar initial to that for 'you' in Early Old Chinese, i.e., something like *Twe. To determine what it might have been, we need to look at Tibeto-Burman.

The native Old Tibetan word for 'two'—usually ignored by Sino-Tibetanists in favor of Old Tibetan *gñis* 'two', a loan from Old Chinese *gnyis or *nyis—is *do*, a word with many etymologically related forms within Tibetan, unlike lexically isolated *gñis*. Old Tibetan /o/ generally derives from Proto-Tibetan (and in turn Tibeto-Burman) *wa or *we, so Old Tibetan *do* is from Proto-Tibetan *dwa or *dwe, as in (25). If we assume that the Chinese word for 'two' originally had a voiced stop initial like the Tibetan word, the Chinese form may be reconstructed, analogously to 女 *rǔ* 'you', as *dwê(s), as shown in (26).

(25) OTib *do* 'two' < PTib *dwa < *dwa ~ *dwe

(26) NMan 二 *èr* 'two' < LMC *źi* < *^{n}źi^3 < *^{n}yi^3 < EMC ⋆nyi^3 < LOC *nyis < early MOC *myi(s)[27] < *mî(s) < *mê(s) < *^{m}bê(s) < *d^{m}bê(s) < *dwê(s)

---

words, whether because they were still homonyms or because they were no longer homonyms remains to be clarified.

[26] The extended forms of the various OChi pronouns must have had different morphological bases. Cf. Schuessler (1987) and Sagart (1999).

III. Bitter • Shell

Old Tibetan *krab*[28] 'armor' has long been considered to be a relative of Chinese 甲 *jiă* 'armor, shell' from Middle Chinese *⋆kaɨp/kɛ:p* (Pul. 145) from Old Chinese *krap (Sta. 592: *⋆krâp* ~ Bax. 766: *⋆krap*) 'shell, armor'. That is about as far as we can go with this word alone. However, Old Chinese also has another word for 'shell, armor', 介 *jiè* from Middle Chinese *⋆kəɨjʰ/kɛ:jʰ* (Pul. 155) from Old Chinese *krac (GSR 327a), which is the phonetic in 芥 *jiè* (also read *gài*) 'mustard'. The latter two words have homonymous Middle Chinese and Old Chinese forms. The mustard plant was introduced to Japan from China, and in Japanese 'mustard' is *karashi* [karaʃi], from Old Japanese *⋆karaci* [karaʧi], so this is an obvious loanword from Old Chinese, but from an unsyncopated form, *karac. Old Japanese *⋆kara* 'shell' is thus apparently related to the common root of the two words for 'shell' in Old Chinese, *kara-, but more importantly, perhaps, Old Japanese *⋆karaci* [karaʧi] 'mustard' is related etymologically within Old Japanese to the adjective *⋆kara-* 'acrid, pungent, bitter', which corresponds to 苦 *kŭ* from Middle Chinese *⋆khɔ*[2] (Pul. 175) from Old Chinese *krâ.[29] Since the Old Tibetan word for 'bitter' is *ka*, from Tibeto-Burman *ka (Benedict 1972: 18), we have regular correspondences between Chinese and Japanese, but, as shown in examples (27) and (28), not between Chinese and Tibeto-Burman.

(27) 芥 Late MOC *krac 'mustard', 苦 Late MOC *krâ 'bitter', both < Early MOC *kara- : OJpn *karaci* 'mustard', *kara-* 'bitter, acrid, pungent' : OTib *ka* 'bitter' < PTB *ka 'id.'

---

[27] It has been suggested that the final *-s (attested to by the TB loanforms, including OTib *gñis*) may derive from the initial of the word for 'three'. If such a process of accretion occurred, it would account for nearly all of the 'prefixes' and some of the finals known from the Old Tibetan forms of the Chinese numerals.

[28] There is no phonemic distinction between the two unvoiced stop and affricate series in Old Tibetan, as is now well known (examples of both aspirated and unaspirated forms are found for most words). There are, however, two clear allophones that occur in complementary distribution. The aspirated form occurs as sole initial immediately before the nucleus (or glide and nucleus), or after a nasal, while the unaspirated form occurs in all other allowed onset positions. A phonemic system of transcription is adopted here to avoid deceiving anyone into thinking the initial aspiration of Old Tibetan words such as *kha* or *k a* 'mouth; bitter' (in Classical Tibetan written *kha* [kʰa]) and *khrap* or *krap* 'armor' (Classical Tibetan *khrap* [kʰrap]) can be used as evidence of early aspirated phonemes in Tibeto-Burman, as has been mistakenly argued in the past.

[29] The origin of the aspiration of the initial stop in Middle Chinese in this and many other words is unclear.

(28) 介 Late MOC *krac 'shell, armor', 甲 Late MOC *krap 'id.', both < *kra < Early MOC *kara- 'shell, armor': OJpn *kara* 'shell' : OTib *krab* 'armor'

The phonetic of 苦 *kŭ* [k^{h}u] 'bitter' is 古 *gŭ* [ku] 'old', the phonetic of which is in turn 口 *kŏu* [k^{h}ow] 'mouth', as in (29).

(29) 口 NMan kŏu 'mouth' < MChi **k^{h}əw^{2}* (Pul. 174) < LOC **krû* < late MOC **krô* (Sta. 560: **khô?* ~ Bax. 771: **kh(r)o?*) < early MOC **karo* < EOC **karâ*[30] : OTib ka 'id.'

Since the putative Old Tibetan cognate is *ka* 'mouth', from Tibeto-Burman **ka* or **mkha*, again we do not have a regular correspondence. We must reconstruct beyond late Middle Old Chinese to an earlier stage to propose regular correspondences.[31]

IV. Horse • Plum

The rare Old Tibetan word *rmaŋ* 'horse, steed'[32] and Literary Burmese *mraŋ* 'horse' have been accepted as being etymologically related (Coblin 1974), and by extension connected to the Chinese word for 'horse' in (30), despite the problem of the velar-nasal final in the Tibeto-Burman forms.

(30) 馬 NMan *mă* 'horse' < MChi **mai^{2}/mɛ:^{2}* (Pul. 206) < OChi **mra-* (Sta. 561: **mrâ?* ~ Bax. 775: **mra?*) OTib *rmaŋ* 'horse, steed' ~ LBur *mraŋ* 'horse'

---

[30] Or perhaps, late MOC **krû* < early MOC **karô* < EOC **karâ*. This is one of the problematic rhymes in the *Odes*. The character 口 'mouth' is the phonetic in words such as 古 *gŭ* 'old', which belong to the 'fish' rhyme in late MOC, so it is necessary to reconstruct final **a* ~ **â* for this word in EOC. Note also OKog **χurtsi* (or **kurtsi*) ~ OKog dial. **kutsi* 'mouth', a cognate of NJpn *kuchi* 'id.' < OJpn **kuti*, which appears to be related somehow to the OChi word for 'mouth'.

[31] A wider survey of data from modern Tibeto-Burman languages indicates a Proto-Tibeto-Burman reconstruction **ka* (Huang 1992: 28), but the evidence adduced by Benedict (1972: 120) for **m-ka* argues in favor of a Proto-Tibeto-Burman **mVkha*. The latter looks even less like any of the Chinese forms, though it does correspond closely to Sanskrit *mukha* 'mouth', which is etymologically obscure.

[32] The word *rmaŋ* generally occurs only in compounds. The initial of *rta*, the ordinary OTib word for 'horse', has yet to be satisfactorily explained. Matisoff's explanation, "epenthetic *t* after liquids" (Benedict 1972: 32 n. 102; cf. 43 n. 139) makes no sense in Tibetan internal reconstruction, since there are plenty of examples of liquid initials followed by vowels without any "epenthetic *t*". Although the origin of the initial **t-* of pre-OTib **tra*, presumably < **t-mra* < PTib **mra*, is still a problem, it is clear that these two Tibetan words for 'horse' were borrowed separately from Chinese, whether directly or indirectly is unclear.

Some form of the root syllable of this word, *mar-, is found in most eastern Eurasian languages (Mongol *mori-n*, Korean *mal* [mar], etc.). It has often been suggested that it is related to a Proto-Indo-European word *marko 'horse, which—since it is only attested in the Germanic and Celtic branches—has in turn been said to be a probable loan from an unknown Eastern Eurasian source. However, it is now known from archaeology and genetics that domesticated horses were introduced to China from the West around 2000 B.C. (Sagart 1999: 196). Due to the recent discovery of large numbers of mummified bodies of Europoid people, the oldest of which date to around 2000 B.C. (Mair 1998), in the vicinity of Lop Nor (an area that was then on the western border of the ancient pre-Chinese culture area), it may be considered probable that the introducers of the horses, who also brought wheat, barley, domesticated sheep, and other Western cultural artifacts to China, were the same Europoid people. The fact that all linguistically identified 'indigenous' languages spoken on the western borders of traditional China up to the end of the first millenium A.D. were Indo-European languages, the arrival of which is so ancient it predates Chinese historical sources, suggests that the mummified people were Indo-European speakers who introduced both the horse and the word for 'horse' to China. Though of course this does not *a priori* mean that the word in question was *marko, the possibility is not excluded.

Whatever the origin and early form of the eastern Eurasian word for 'horse', the Tibeto-Burman and other forms, as well as recent HSR reconstructions, suggest an originally disyllabic or longer word. After the period of syncope within the Central dialect of Chinese (in which disyllabic words were contracted to monosyllabic ones), by late Middle Old Chinese the word for 'horse' is said to have become *mra?.[33]

[33] The proposal to reconstruct a glottal stop coda for *all* words which in MChi had rising tone—because of some rhyme contacts with words thought to have a velar stop coda in late MOC, tonal distribution of related characters, and loanword and dialect evidence—has been adopted generally in the HSR framework (Baxter 1992: 320-324). This is problematic because in many cases the exact same characters also have MChi readings in the level tone, which is assumed to reflect no coda. Moreover, the necessary application (within HSR) of the rising tone hypothesis to words which had finals already (i.e., *rù shêng* 'entering tone' syllables) produces unnatural codas such as those in *smə̄ʔh (Sta. 549) and *hmi(k)ʔ(s) (Bax. 764) for 晦 NMan *huì* 'dark, obscure' (which has the same phonetic as that in the word 海 *hăi* 'sea', analyzed below). In the case of 'horse', Sagart implies that the correlation of the

The early Tibeto-Burman peoples did not all learn about horses at the same time, but they must have learned about them from the Chinese, long after the breakup of Proto-Tibeto-Burman. This is clear because if the word for 'horse' had been borrowed directly from its western Eurasian source, the first syllable would have been *mar- or *ma- and would have been retained as such in Tibetan. Instead, the early Tibeto-Burmans borrowed a syncopated form of the word from Middle Old Chinese[34] and subsequently from each other, developing the forms shown in (31), among others.

(31) OTib *rmaŋ* : LZZ *hraŋ* : LBur *mraŋ* : ONew *sarhaṃ* 'horse'

The usual Old Tibetan word for horse, *rta*, from pre-Old Tibetan *tra (*tr- being a disallowed sequence in Old Tibetan), evidently from an earlier form *t-mra (cf. Rgyarong *ᵐbro* 'horse'[35]), has no final. Taken together, the Tibeto-Burman forms might suggest that either the donor dialect of Chinese or the borrowing dialect of Tibeto-Burman added a final velar nasal to an original stem *mra, from *mara. However, the presence of a velar nasal coda in the word for 'horse' in several Tibeto-Burman languages, as well as the HSR reconstruction *mraʔ, may indicate that a velar was present in the donor dialect of Chinese when the Tibeto-Burmans borrowed it. If *mra were the earlier form, reflecting a Chinese donor which had lost any earlier final that the word might have had, the Tibeto-Burman languages with an open syllable would reflect the earlier form. However, this is generally thought to be the exact reverse of the historical process in Tibeto-Burman. Moreover, there is additional evidence to support the precedence of the final velar nasal.

It has long been accepted that the native word for 'horse' in Old Japanese, **uma*, is an early, preliterate loanword from 'Late Old Chinese', borrowed along with Japanese *ume* 'plum' (from Proto-

---

hypothetical glottal stop in Chinese and the velar nasal in TB is a coincidence, since he argues that the word was borrowed into Chinese "from an early Tibeto-Burman language, perhaps one in which Benedict's *m-rang 'horse' was reflected as [mrā] . . ." (1999: 196). However, further evidence indicates that in the word for 'horse', at least, the correspondence is not a coincidence, as shown below. The rising tone is likely to derive from disyllabic stress patterns in early MOC.

[34] Some TB languages have borrowed the word from even later forms of Chinese.

[35] The initial of Rgyarong *ᵐbro* 'horse' reflects an OChi dialect with oralized nasals, as in OTib *ḥbrug* [ᵐbrug] 'dragon'. The word *ᵐbro* thus could have been borrowed from Chinese independently of Tibetan. Rgyarong has regularly lost TB final velar nasals.

Japanese *umay) from 梅 Old Chinese *rmay, partly because neither horses nor plum trees are native to Japan and partly because of the fact that, except for the mysterious initial vowel, both words are virtually identical to the Middle Chinese and Modern Mandarin equivalents, *mǎ* and *méi*. For the same reasons it is also suspected that Japanese *umi* 'sea', from Old Japanese **umi*, has probably been borrowed from Old Chinese. The words for 'plum' and 'sea' rhyme in the *Book of Odes*, and have the same phonetic, 每 *měi* 'each' (the phonetic of which is, in turn, 母 *mǔ* 'mother') so 'plum' and 'sea' must have been homophonous with 'each' when the characters were created (apparently not at exactly the same time) and before the metathesis and other later changes, indicating a recon-struction *marê for Early Old Chinese, as in (32) and (33).

(32) NJpn *ume* 'plum' < OJpn **umɛy* < PJpn *umay ← 梅 LOC dial. *rmay < late MOC *mrê < EOC *marê[36]

(33) NJpn *umi* 'sea' < OJpn **umi* < PJpn *uməy ← 海 LOC dial. *rməy < early MOC *mrê < EOC *marê[37]

Considering the form of the word for 'horse' in Middle Old Chinese, it is clear that the Old Japanese word reflects an Old Chinese *rma, from *mra. However, although it is also believed that Japanese has canonically lost all final consonants (assuming they existed in Proto-Japanese[38]), the Hateruma Ryukyu dialect has final velar nasals in a good number of words for which other evidence supports the earlier existence of a final. This is the case for 'horse', Hateruma *ʔmaŋ ~ nmaŋ* [^{m}maŋ] (Martin 1987: 74; Janhunen 1998: 422). It must therefore be concluded that the peripheral Chinese dialects that contributed the word for 'horse' to Proto-Japanese and to Common Tibeto-Burman both had a final velar nasal.[39] Considering that the Middle Chinese word has the rising tone, which 'cross-rhymed' (i.e., rhymed) with words reconstructed with a velar stop, the Old Chinese

---

[36] LOC *mray > MChi **məy^{l}* > NMan *méi* [mej].

[37] NMan *hǎi* [χaj] < MChi **χəy^{2}* apparently developed from *mχəy < *χməy < *χ-rmay (from a compounded neologism *χV(r)-rmay?) or *s-rmay.

[38] Koguryo data supports the existence of final consonants in Proto-Japanese-Koguryoic (Beckwith 2000), but even in Koguryo some probable earlier finals have been lost, while in other cases the finals (especially those with syllable-final [n]) are almost certainly innovations.

[39] Thus we must conclude that the direction of borrowing was from Chinese into TB, not the other way around, as Sagart argues (1999: 195-196). Attempts to provide the word with an internal TB etymology—'tall dog', etc. (Matisoff, in Benedict 1972: 43, n. 139)—also must be abandoned

form *mraŋ should be reconstructed for Early Old Chinese as *mraga, derived from *marga, from a Proto-Chinese *marka. (Early Old Chinese words of the shapes CVrCV and CrVCV regularly became CrVC by Middle Old Chinese.[40]) The latter perfectly reflects an expected eastern Indo-European reflex (in *-a) of the Celtic and Germanic word *marko.

One thing that is incontrovertible in all this is that every early attested form of the word for 'horse' in Chinese, Tibeto-Burman, and Japanese has the nucleus vowel *-a-* in the root syllable (including the Middle Chinese loans into Japanese, *ma* ~ *ba* 'horse'), indicating that the ancestor of all of them had the vowel *a.[41] This casts doubt on the correctness of the reconstructions of the Middle Chinese form of this word with a fronted vowel *ɛ (Pul. 206) or *æ (Bax. 775).

V. Not • Mother

The phonetic of the character 梅 *méi* 'plum', in example (33) above, is 每 *měi* 'each', the phonetic of which is 母 *mǔ* 'mother'. The latter may be reconstructed as *mré for late Middle Old Chinese, as in (34).

(34) 母 *mǔ* 'mother' < MChi ⋆*məw*^{2} < LOC *mɨ̈ô < *mrá < late MOC *mré (Bax. 778: *m(r)o/i? ~ Sta. 548: *mə̄?) < early MOC *mare

Old Tibetan *ma* (from Tibeto-Burman *ma) 'mother, woman', while a 'linguistic near-universal', should of course also be cognate to Chinese *mǔ* 'mother'. Instead, we have yet another problematic non-correspondence.[42] However, it is also notable that the homophonous Tibeto-Burman negative *ma corresponds to several negatives in Chinese that now (ignoring tones) rhyme with the word for 'mother', such as 毋 *wú* 'don't, shouldn't', from Middle Chinese ⋆*mu*^{1}. In fact,

---

[40] Another good example is OTib *brgya* 'hundred' ← EOC 百 *^{m}berg^{y}a < *merg^{y}a < *merk^{y}a 'id.' (> late MOC *prak).

[41] Starostin (1989: 687) compares various OChi reconstructions of 馬 *mǎ*. If an OChi dialect had a lengthened-grade vowel in the first syllable, it would have become *morVŋ during the vowel shift, and could have been the source of the Mongol word *morin* 'horse' (cf. Janhunen 1998: 417-420, who reconstructs *morï).

[42] The present discussion does not take into consideration 媽 *mâ* 'mother', and 婆 pó < MChi ⋆*ba*^{1} '(old) woman, mother-in-law, wife', or NJpn *ha(-ha)* 'mother' < OJpn ⋆*pa* and NJpn (*o-*)*ba* 'aunt, (old) lady', (*o-*)*baa* 'grandmother'. The words with stop initials appear to derive from semi-oralized nasal onset forms, but in view of the semantic differences they may be etymologically distinct, like Russian *mat'* 'mother' and *baba* '(older) woman'.

as Baxter (1992: 467) notes, "The graph for 母 *mǔ* < **m(r)o?* 'mother' is regularly used in bronze inscriptions for the word now written . . . 毋 *wú* < *mju* < **m(r)jo* 'don't'."[43] Since the two were thus already homonyms at that time, we have to reconstruct *wú* 'don't' the same as late Middle Old Chinese *mré 'mother', although Tibeto-Burman supports a reconstruction *ma.

In fact, as shown in Section II above, all of the negatives in Chinese derive from a Proto-Chinese root *ma- 'negative', which does correspond perfectly to the Tibeto-Burman root *ma- 'negative' after all, as shown in (35) and (36).

(35) 母 *mǔ* 'mother' < late MOC *mré < early MOC *maré : OTib *ma* 'id.' < PTB *ma 'id.'

(36) 毋 *wú* 'don't' < late MOC *mré < early MOC *maré : OTib *ma* < PTB *ma 'negative'

The correspondence, however, is between the monosyllabic open CV Tibeto-Burman forms and the open first syllable of the disyllabic Chinese forms. In short, the second syllable of the Chinese forms seems to have no reflex at all in Tibeto-Burman. The same correspondence applies to numerous other examples, such as 苦 early MOC *kara 'bitter', given in (27) above. This regular correspondence allows us to establish the vowel *a for the first syllable[44] of the early Middle Old Chinese words involved, and at the same time supports the theory that the words must have been disyllabic in that period of Chinese.

IV. Firewood • People

After correct reconstruction of the initial of 人 *rén* 'man, person, people' as *m- in early Middle Old Chinese, it now corresponds to the initial of the Tibeto-Burman word *mî 'man, person, people', as shown in section I above. The rhymes, however, do not correspond in any hitherto published reconstructions of Old Chinese, since Old Tibetan *myi* 'man, person, people', from Proto-Tibeto-Burman *mî,[45] is an open syllable, while 人 *rén* 'man, person, people' and 民 *mín*

[43] The character 毋 'don't' is also sometimes written in the *Book of Rites* (禮記 *Li chi*) with the character 母 'mother' (Karlgren 1957: 48).

[44] In many words the EOC vowel is difficult to recover; it could have been *e, *a, or perhaps even *o.

[45] Reconstructed as *r-miy and *mi(y) by Benedict (1972: 107, 119).

'people'[46] have a final alveodental nasal in late Middle Old Chinese, at least as suggested by current interpretations of *Book of Odes* rhyme data. Since *rén* rhymes in the *Odes* with 薪 *xîn* [ɕin] 'firewood', which has been widely argued to be cognate to Old Tibetan *śiŋ* 'tree, wood', from Tibeto-Burman *siŋ (Benedict 1972: 55), we have another irregular correspondence, as in (37) and (38).

(37) 人 NMan *rén* [ʐən] 'man, person' < MChi ⋆*nyin*^{I} < OChi *myin : OTib *myi* 'man, person' < PTB *mî

(38) 薪 NMan *xîn* [ɕin] 'firewood' < MChi ⋆*sin*^{I} < ?OChi *siŋ : OTib *śiŋ* 'tree, wood' < PTB *siŋ

The problem has been recognized by scholars working in the HSR framework, who have generally argued that the word for 'firewood' and the words 人 *rén* 'man' and 民 *mín* 'people' had a velar nasal coda (Sagart 1999: 52, 135-136).[47] The word 氓 *méng* (now usually read *máng*) 'vagrants, people' has been cited in support of the argument for a velar nasal final, but the word has the phonetic 亡 *wáng*, from Middle Chinese ⋆*muaŋ*^{I} (Pul. 318), not 民 *mín*, and is reconstructed *mraŋ according to the rules of HSR (e.g., Sagart 1999:135-136); it could not have been a homonym of 民 *mín*.[48] Tibeto-Burman is thought to preserve final velar nasals—and as seen above in the word for 'horse', several branches of Tibeto-Burman preserve the final even where it has been lost in Chinese—but *mî 'man, people' has an open syllable, so the Chinese and Tibeto-Burman words still do not correspond.

It is now generally accepted that some[49] alveodental nasal finals of Late Old Chinese derive from a Middle Old Chinese liquid. However, the change occurred not only in the phonetic series proposed by

---

[46] Sagart rightly dismisses Benedict's attempt to explain the final nasal as a collective suffix, but unfortunately accepts the HSR theory of a final velar nasal for *xîn* 'firewood' (Sagart 1999: 135) in the *Odes*.

[47] As mentioned above, ST proponents, following Karlgren's incorrect reconstruction of the OChi initial of 人 *rén* with an alveodental or palatal nasal (see section I above), ignore the word and compare 民 *mín* 'people' alone to TB, but for the codas it makes no difference.

[48] Sagart explains the *-r- as an infix (1999: 135).

[49] Some alveodental nasal finals appear to be retentions of earlier forms, as seen in words such as 犬 NMan *quán* < MOC **kwen* 'dog'. This word is undoubtedly related to OTib *kyi* 'dog' < CTB **kwî* or **kwê*, but the Chinese word, at least, has long been thought to be a loanword from PIE **kwon* 'dog', indicating that the final nasal is original. The simplest explanation for the discrepancy between the Chinese and TB forms of this word is that the Tibeto-Burmans got it from PIE *kwô (or *kúwô; cf. note 22), while the Chinese got theirs from PIE *kwon.

Starostin (1989) and accepted by Sagart and others, but in other rhymes as well, as indicated by numerous supposedly irregular cross-rhymes between *-r and *-n (Starostin 1989), by numerous phonetic series containing words belonging to more than one rhyme—some of the members of which have early Middle Old Chinese final *-r—and by loanword data. In addition to the indirect evidence of the open syllable in Proto-Tibeto-Burman *mî 'man, person, people', indicating either an original open syllable or a lost final *-r in pre-Proto-Tibeto-Burman, the possibility exists that Benedict's alternate reconstruction *rmiy (1972: 107) is correct and indicates preservation of the final *-r from a syncopated form, via metathesis, as in the word *rmaŋ* 'horse' in Old Tibetan. However, there is also Chinese internal evidence for the rhyme presently under consideration having ended with *-r.

The word 新 *xîn* 'new', the phonetic of the character for 薪 *xîn* 'firewood' (both from Middle Chinese ☆*sin*^{l}), may plausibly be connected to Tibeto-Burman *sar 'new' (Old Tibetan *gsar-*) and Japanese *sara* 'new' (Old Japanese ☆*sara*), and thus had a final *-r. The vowels would appear at first glance to be an obstacle to this reconstruction, but we also have the word 鮮 *xiân* 'fresh', reconstructed by Starostin as *sar and compared by him directly with Old Tibetan *gsar-* 'new' (1989: 340; cf. 572: *shar), which is from Tibeto-Burman *sar 'new, fresh' (Benedict 1972: 207).[50] According to the reconstruction approach taken here, this pair of words reflects earlier *sêr (which became *sîr, *sîn, and eventually *xîn* 'new') and *ser (which became *sar and eventually *xiân* 'fresh') respectively.

The word 薪 *xîn* 'firewood', which is first attested in the received text of the *Book of Odes*, either did not exist in early Middle Old Chinese and thus never had a final *-r, or else it did exist then, unattested, and did have a final *-r which became *-n together with the other words in the same rhyme. There are two additional bits of evidence in favor of the latter argument. The ordinary word for 'firewood' in Old Chinese as a whole is not 薪 *xîn* but 柴 *chái*, from Middle Chinese ☆*dʑaɨj*^{l}/*dʑɛːj*^{l} (Pul. 47), which Karlgren reconstructs for Old Chinese as *dz'ăr (Kar. 103; cf. Bax. 811: *dzjejs, Sta. 571:

[50] The words 新 *xîn* 'new' and 鮮 *xiân* 'fresh' are both connected to TB *sar 'new, fresh' by Matisoff (Benedict 1972: 172), but he does not discuss the vowel problem, other than to remark (Benedict 1972: 189), "note Ch[inese] alternation: *si̯ĕn* 'new' ~ *si̯an* 'fresh'."

*ćejh), while the closest other word to *xîn* 'firewood' semantically is 炭 *tàn* 'charcoal', the character for which has as phonetic 山 *shân* 'mountain', a word which definitely had a final *-r in early Middle Old Chinese, as shown by the cross-rhymes listed by Starostin (1989: 578)[51] and other evidence.[52] Thus, in late Middle Old Chinese 人 *rén* 'person' and 薪 *xîn* 'firewood'[53] rhymed as shown in (39) and (40).

(39) 人 *rén* 'man, person' < LMC *źin* < *ʺźin* < EMC **nyin* < LOC *nyin < late MOC *nyîn ~ *nyîr < early MOC *mîr ~ *mêr : OTib *myi* < PTB *mî ~ *mê 'man, person' (or perh. early MOC *mîr > MOC dial. *mrî > *rmî → CTB *rmî > irreg. *mî)

(40) 薪 *xîn* 'firewood' < EMC **sin* < LOC *sin < late MOC *sîn ~ *sîr (perh. < early MOC *sîr ~ *sêr) : OTib *śiŋ* 'tree, wood' < PTB *siŋ

The attempt to relate the Chinese word for 'firewood' and the Tibeto-Burman word for 'tree, wood' has fatal problems. In light of the evidence against the comparison, the conclusion to be drawn is that they are accidental look-alikes and are simply unrelated. Both phonology and semantics suggest that the Chinese word 薪 *xîn* 'firewood' is related to the Chinese words 柴 *chái* 'firewood' and 炭 *tàn* 'charcoal', not to Tibeto-Burman *siŋ 'tree, wood'.[54]

Although the word 氓 *méng* 'vagrants, people' does not explain anything about the phonology of 民 *mín* 'people' (or 人 *rén* 'man, person') within the current HSR framework, the latter system's

---

[51] He nevertheless reconstructs 山 *shân* 'mountain' as **srân* (Starostin 1989: 576).

[52] One of the two OKog words for 'mountain', 達 **tar ~ *dar*, is clearly a loan from a LOC dialect form 山 *Tar < early MOC *Ter, calling for a reconstruction 炭 *Ter 'charcoal' and 薪 *Têr 'firewood'. Here "T" represents any alveodental; the cause of the initial variation in Chinese here (and in many other cases) is unknown.

[53] Because final *-*r*(*V*) in OChi corresponds to zero in TB, it appears that the only real possibility of relating the Chinese and TB words here is to assume borrowing. In such case, the Chinese word for 'firewood' would have been borrowed from the usual TB word for 'tree, wood', presumably in the LOC period, after the change of final *-*r* to *-*n*. The finals would still be problematic, but several scholars have proposed that in many cases the reconstructed old rhyme *-*iŋ* was palatalized by its high vowel and in late MOC became *-*in* (e.g., Sagart 1999: 51-52), thus 薪 *xîn* 'fire-wood' < MChi **sin* < LOC *sin < early LOC *siŋ ← TB (a particular TB language, not PTB) *siŋ 'tree, wood'.

[54] Internal TB evidence has been martialled in favor of the argument that a pair *sin 'firewood' ~ 'new' constitutes a PChi set cognate with a TB pair *siŋ, 'tree' ~ 'new', but the Lolo-Burmese evidence cited in favor of this etymology is ambiguous (Lolo-Burmese neutralized the distinction between *-ik, *-it, and *-ic), while other, unambiguous, TB evidence, especially that from Tibetan and Karen, falsifies the theory. None of the Chinese words for 'tree, wood', appear to be cognate with the TB word.

reconstruction of 氓 *méng* as *mraŋ for Old Chinese, if correct, does provide further support for reconstructing 民 *mín* as *mîr for Middle Old Chinese. The form *mraŋ could be explained as a monosyllabic reduction of an early Middle Old Chinese extended form of *mîr, i.e., *miráŋ (theoretically from an earlier *wirága < *wir-; see further below), the accent on the second syllable having caused the syncopation of the first in the change from early Middle Old Chinese to late Middle Old Chinese.[55] This is, moreover, virtually the only way the words 氓 *méng* and 民 *mín* could be related.

In short, there is no support for the reconstruction of a final velar nasal, but plenty of support for the reconstruction of a final *-r, in 人 *rén* 'man' and 民 *mín* 'people' in Early Old Chinese.

V. Five • I • Fish

In the *Book of Odes*, 五 *wǔ* 'five', and 吾 *wú* 'first person pronoun, I', among many other words, rhyme with 魚 *yú* 'fish'. Not only do these words belong to the same HSR rhyme class, the standard character for *wú* 'I' has 'five' as its phonetic. This seemingly unique set of three semantically unrelated but phonologically related words has long been cited as proof of the Sino-Tibetan theory because there seems to be a regular correspondence between the modern Mandarin forms and the modern Tibetan forms, and—so it has been argued—there is a regular correspondence between the Old Chinese forms as well, at least according to Karlgren's system: 五 *ŋo 'five', 吾 *ŋo 'first person pronoun', 魚 *ŋi̯o 'fish'. Moreover, in Middle Old Chinese bronze inscriptions, the then newly-coined word 吾 *wú* 'first person pronoun' is generally written with 魚 'fish' as phonetic instead of 五 'five' (Sagart 1999: 143-144; Karlgren 1957: 40),[56] indicating that the three words 'five', 'I', and 'fish' were at that (earlier) time homonyms in the dialect of those who wrote the inscriptions. However, there are actually serious problems with the putative regular correspondences between the Chinese and the Tibeto-Burman forms.

Because 魚 *yú* 'fish' is the phonetic in 魯 *lǔ* 'blunt, stupid; P.N.', the Old Chinese forms of 'fish', 'five', and 吾 *wú* 'I' all must have

[55] Assuming the first vowel was originally *î, the word should be from a non-palatalizing dialect such as the one which loaned 民 *mín* to the central dialect.

[56] This was first pointed out to me by Wolfgang Behr, p.c.

had a liquid in them, as confirmed by Old Tibetan *lŋa* 'five'.[57] The usual first person pronoun in Old Japanese, **wa*, also corresponds regularly[58] to Old Chinese 我 *ŋa, which is widely believed to be cognate with *ŋa, the putative Proto-Tibeto-Burman first person pronoun (Benedict 1972: 93).[59] As in Tibetan, there is no reflex of a liquid in this word, nor is there in Tibeto-Burman *ŋa ~ *ŋya 'fish', which does not correspond to the clearly reconstructible late Middle Old Chinese form *lŋâ 'fish', though Old Tibetan *lŋa* 'five' does correspond to the late Middle Old Chinese word *lŋâ 'five'.[60]

There are several first person pronouns in Old Chinese, including 我 *wǒ* (from Middle Chinese **ŋa*[2]), well attested in the Early Old Chinese Oracle Bone Inscriptions as a plural pronoun 'we'. The normal first person singular pronoun in Early Old Chinese is 予 *yú* (inscriptional form 余) 'first person pronoun, I', which may be reconstructed as *lâ for late Middle Old Chinese.[61] The neologism 吾 *wú* 'I' first appears in the Eastern Chou dynastic period, when the capital moved out of what was then northwestern China, and gradually replaced 余 / 予 *yú*. Since 吾 *wú* was then a homonym of 'five' (*lŋâ) and 'fish' (*lŋâ), but must have been pronounced something like the older first person pronoun 余 *yú* as well, 吾 *wú* must have had the same liquid in it. Its late Middle Old Chinese form *lŋâ would thus seem to be best explained as a regular, syncopated

---

[57] In the unique Japanese-Koguryoic system, the numerals of which have so far resisted connection to any other language, the numeral for 'five' is OKog **ütsi* (< PKog **utui* or **ituí*), OJpn **itu*.

[58] Cf. OJpn **wani* 'alligator or crocodile', borrowed from LOC **ŋwan* 'alligator sinensis' (there are no such animals in Japan), among other examples. In this connection ccf. 錢NMan *qián* 'money' < MChi **dzian* (Pul. 250) → NJpn *zeni*, accepted as having been borrowed from Chinese into pre-Old Japanese.

[59] In an incisive study of Chinese and Tibeto-Burman pronouns, Sagart concludes that "*ŋa is no more inherited in Tibeto-Burman than it is in Chinese" (1999: 145); cf. his excellent earlier discussion (Sagart 1995: 199-202).

[60] OJpn **iwo* 'fish' may reflect the same initial seen in OTib *lŋa* 'five'. It corresponds to an OChi form *lŋô, showing the middle stage of the Great Chinese Vowel Shift (**â* > **ô* > *u*), and preserves the original sequence of consonant segments that was metathesized within LOC (along with *rC- metathesized forms—most of which were from earlier *Cr- initials—when the official dialect shifted back to a more conservative dialect that had *Cr- and *Cl- sequences; see below), *lŋô becoming *ŋlô and eventually MChi *ŋyo*, LMC *ŋü*, and NMan *yú* [ü]. The compounding form *iwa-* in *iwashi* 'sardine' corresponds perfectly to late MOC *lŋâ 'fish', but it is uncertain if it is actually derived from a relative of **iwo* 'fish'.

[61] Sagart reconstructs this as *ᵇla (1999: 142-144); Starostin reconstructs it as *la? (1989: 562).

form of an earlier *laŋa, an 'eastern' cognate of the 'central' (i.e., western) dialect form, *laɣa. Both words (余/予 *laɣa and 吾 *laŋa) can thus be explained as local developments of an Early Old Chinese *laga 'first person singular pronoun'. If the word 吾 *laŋa had been inherited from Proto-Sino-Tibetan or loaned into Proto-Tibetan in the Early Old Chinese period, the expected Old Tibetan form would be *laŋ or *laŋa, or even *lŋa. But we do not have such forms, nor do we find them elsewhere in Tibeto-Burman. Although the Middle Old Chinese first person plural pronoun 我 *ŋa[62] 'we' corresponds perfectly by form to its putative Proto-Tibeto-Burman cognate *ŋa, there is no regular correspondence between 余 /予 *yú* or 吾 *wú* 'I' and Tibeto-Burman. In fact, with respect to the trilogy of words 'five', 'I', and 'fish', even 我 *ŋa 'we' is irregular. Moreover, because intervocalic *g regularly became *ŋ in the eventually dominant dialect of Middle Old Chinese, and because 余 /予 *yú* was *laga, 我 *wǒ* was probably *aga or *ega, etc. in Early Old Chinese. Thus the correspondence of the putative Proto-Tibeto-Burman form *ŋa 'first person pronoun' is with the second syllable of a Middle Old Chinese word in which the velar nasal is a secondary development.

Returning to 魚 *yú* 'fish', we must reconstruct it as *laŋa for early Middle Old Chinese, from an Early Old Chinese *laga, as shown in (41). It can hardly be cognate to Proto-Tibeto-Burman *ŋya 'fish'.

(41) PTB *ŋya 'fish' : 魚 *yú* [ü] 'fish' < LMC *ŋü* < MChi ⋆*ŋyu*^{l} < ⋆*ŋyo* < *Late Old Chinese *ŋyâ < *ŋlâ < *lŋâ 'fish' < Early MOC *laŋa < EOC *laga < PChi *laga ~ *laka

If the Tibeto-Burman word had been borrowed from Old Chinese, it might have come from a dialect in which late Middle Old Chinese *lŋâ metathesized to *ŋlâ and became Late Old Chinese *ŋyâ before the vowel shifted to Early Middle Chinese ⋆*ŋyo*, the ancestor of Late Middle Chinese *ŋü* and Mandarin *yú* [ü]. This might explain the Tibeto-Burman word for 'fish', but it would have to be, again, a loan from a very late form of Chinese. Since most Tibeto-Burman languages do have an obvious reflex of the Proto-Tibeto-Burman word *ŋya 'fish', either this word was individually borrowed very

[62] The word is usually reconstructed *ŋay in the HSR framework (Sagart 1999: 142). Such a proto-form would correspond regularly to an OTib *ŋe, but we do not have such a form. The 'elegant' word *ŋed* 'we', which is often compared to this HSR reconstruction of the EOC pronoun 我*wǒ* 'we', has developed the vowel *e* purely internally from *a, via regular umlaut caused by the suffix *-d*: *ŋa + *d > *ŋed*.

late into the Tibeto-Burman daughter languages or the Tibeto-Burman word for 'fish' is not directly related to the Chinese word.[63]

In short, of the three words in question, the Common Tibeto-Burman word for 'five' is definitely a loan from Chinese, the Tibeto-Burman word for 'fish' does not correspond to the Chinese word, and the remaining word—the Old Chinese first person pronoun 吾 *wú*—does not correspond to the putative Proto-Tibeto-Burman first person pronoun *ŋa. The latter, while possibly a late loanword from Chinese as well,[64] is a linguistic near-universal, being found in many unrelated languages all along the western Pacific rim from the Bering Straits to Australia. Since it is non-distinctive, however, it cannot be used to demonstrate the genetic relationship of anything.

VI. Sun • One

In the *Book of Odes*, the word for 'sun', 日 *rì*, pronounced *[nyit] in Early Middle Chinese, rhymes with syllables ending in *-t, as expected, but it also rhymes in a few instances with *-k. Since there is solid internal evidence that the word earlier had a final *-k itself, it appears that palatalization spread in some cases to the coda as well.

The word 即 *jí* 'then', which had a final velar stop in Middle Chinese, rhymes directly in the *Odes* with its phonetic series relative 節 *jié* 'knot, bamboo section; festival', as well as with a number of other words that also have dental stop finals, namely 室 *shì* 'room', 栗 *lì* 'chestnut', 漆 *qî* 'lacquer', 瑟 *sè* 'psaltery', 疾 *jí* 'feverish', 穴 *xué* 'cavern', and 實 *shí* 'fruit'. These rhymed in turn with other words, including 結 *jiê* 'form fruit' ~ *jié* 'knot', 七 *qî* 'seven', 厲 *lì* 'harsh', 滅 *miè* 'extinguish', 吉 *jí* 'auspicious', and 一 *yî* 'one'. A problem arises because it seems clear from external comparative evidence that some of these words—such as 七 *qî* 'seven'—should have had an alveodental final in Old Chinese, as they have in Middle Chinese, while others—such as 節 *jié*—should have had a velar final. Karlgren (1957: 113, 243) invents two phonetic series for 即 *jí* to take care of the problem, while more recently, Baxter (1992:768)

[63] The TB words could not be derived from *laga with deletion of the initial syllable because TB preserves initial *l-, as in TB *la 'dative-locative' (OTib *la*), *la 'moon (etc.)' (OTib *zla-* 'moon', *sla-* 'month', *bla* 'soul', *lha* 'god', etc.), *lak 'hand, arm' (OTib *lag-*), and so on.

[64] Sagart argues that the PTB first person pronoun was *ka, and that *ŋa was borrowed from Chinese (1999: 145).

gives alternate forms (**kit/k*) for some words, such as 結 *jiê ~ jié*; but these are hardly satisfactory solutions. The reason for these anomalies, it now is generally agreed, is that an earlier velar stop coda was palatalized after a long high vowel (**î* and/or **ê*) and merged with alveopalatal stop codas, producing a rhyme in the *Odes* dialect, but the change was either not as complete in the Chinese dialect that was the ancestor of Middle Chinese or there were later borrowings from unpalatalized dialects.

The mechanics of this change, which occurred by late Middle Old Chinese—the language of the *Book of Odes*—appear to be that the unstressed mora of the long high vowel **î* was reduced to a glide, which moved into one of the syllable margins, palatalizing it. While compensatory relengthening of the remaining stressed mora (Hayes 1989) regularly occurred in the *Odes* dialect, apparently with an accent shift that caused palatalization of the other margin, as in 日 *rì* 'sun', the latter development did not take place completely, or at all, in other dialects, where only one margin was palatalized.[65]

Thus, unlike 結 *jié* 'knot' < MChi **ket* < OChi **kêk*, where only one of the margins was palatalized in the dialect ancestral to Middle Chinese (assuming the word is inherited directly, not borrowed from another dialect, as is likely), in the case of 日 *rì* 'sun' both the bilabial onset (on which see section I above) and the velar coda were palatalized in that dialect. The latter development is discernible not only because of an alternate Middle Chinese reading (Sagart 1999: 159) but because of the word 汨 *mì* 'name of the Mi-lo River' (the character for which has 日 'sun' as its phonetic) from Middle Chinese **mɛjk* (Pul. 213), from a southern dialect Old Chinese pronunciation **mik*, and a northern dialect Han period name where the character for 'sun' is read **mik* (Pan 1944: 118; Giles 1962: 149).[66]

---

[65] This theory was first presented at the International Conference on Sino-Tibetan Languages and Linguistics held at Champaign-Urbana in 1999. Another way of looking at it might be that the vowels **î* and **ê* developed a palatal on-glide. Following compensatory lengthening (Hayes 1989), after the glide moved into the onset in certain cases an off-glide developed and palatalized the coda. However, this would not tell us why the latter change happened only in such cases.

[66] Cf. the rhyme of 日 *rì* with 節 *jié* 'joint' < OChi **tsik* (Bax. 596, 621). The supposed TB cognate **nî* (**niy* in Benedict 1972: 31) 'sun' appears to be another late loan from Chinese (cf. Sagart 1999: 160). It is not found in all branches of TB; some have forms that reflect **mî* (or unrelated words). I would like to thank my colleagues on the Warring States Working Group internet list for kindly checking Dubs for me.

We have, in other words, reflexes of two ancient dialect forms (comparable to those represented by 人 *rén* 'man, person, people' and 民 *mín* 'common people'), as shown in (42) and (43).

(42) *Central dialect:* 日 NMan *rì* 'sun, day' < LMC *źir* < *^{n}źir < *^{n}yir < MChi ⋆*nyit* (Pul. 266: ɲit) < LOC *nyit < late MOC *nyič < *m^{y}i^{y}k < *mîk < *mêk < early MOC *^{m}bêk < *wêk

(43) *Peripheral dialects:* 日 NMan *mì* 'sun, day' < MChi ⋆*mɛjk* (Pul. 213) < LOC *mîk < MOC *mêk < early MOC *^{m}bêk[67] < *wêk

The palatalization obviously never took place in the peripheral dialects, for one of several possible reasons: they retained an earlier vowel *ê; they shifted earlier *î to *ê or another still lower vowel; or—the simplest and most probable solution—those dialects did not experience vowel reduction in the unaccented mora and thus did not palatalize their margins. The borrowing took place after the palatalizing process in the central dialect had run its course. Out of eight rhymes in the *Book of Odes*, the word 日 *rì* 'sun, day' rhymes six times with words ending in *-t in Middle Chinese, and two times with words ending in *-k in Middle Chinese. In the same text, the word 一 *yî* 'one' occurs as a rhyme in only two poems, in both of which one of the other rhyme-words ends in *-t, the other in *-k. The evidence thus indicates we should reconstruct final *-k both for 一 *yî* 'one', and for 日 *rì* 'sun' in early Middle Old Chinese, and the latter word descends from Early Old Chinese *wêk as in (44).

(44) 日 NMan *rì* < LMC> *źir* (Takata 1988:380-381) < MChi ⋆*nyit* < LOC *nyît < late MOC (*Odes* dialect) *nyíč < *m^{y}ík^{y} < *mîk < early MOC *^{m}bêk < EOC *wêk 'sun'

In short, the late Middle Old Chinese palatalization spread not only from the nucleus to the onset but also in some cases to the coda. The seeming anomalies we see in the *Odes* rhymes are due to the fact that the lineal Old Chinese ancestor of Middle Chinese and its descendants either did not palatalize all the codas which were palatalized in the dialect of the *Odes* or—the more likely cause—the exceptions are borrowings from non-palatalized dialects, like the southern dialect word 汨 *mì* (from *mîk or *mêk) 'name of the Mi-lo River', and the northern dialect reading 日 *mì* 'sun' (from *mîk from earlier *mêk), as discussed above. It would appear that the coda of this rhyme was a

[67] 'Sun' in Tangut is ⋆*mbe*, according to Sofronov's system (Ksenia Kepping, p.c., 2000), reflecting a *me- < *mêC or *mîC if it is from Chinese.

palatal affricate in the dialect of the *Odes*; because the Old Chinese ancestor of Middle Chinese neutralized any distinctions that may have existed between *-t^{y}, *-k^{y}, and *-č, we cannot tell from Middle Chinese alone what the coda of a word in this *Odes* rhyme was before late Middle Old Chinese.[68]

Comparative evidence supports the preceding argument. Not only does the word 節 *jié* 'joint of bamboo, section', from Middle Chinese ⋆*tset* (Pul. 155: tsɛt), have a clear relative in Old Tibetan *tsigs* 'joint, metrical section', and Common Tibeto-Burman *tsik 'joint' (Benedict 1972: 27-28), as in example (45), the long problematic Chinese numeral 一 *yî* 'one', which has been omitted from most careful Sino-Tibetanists' comparative lists (on good phonological grounds), now —with its final velar restored, as shown in (46)—appears to have been loaned to Common Tibeto-Burman after all.

(45) 節 NMan *jié* 'joint of bamboo' < MChi ⋆*tset* < late MOC *tsič < *tsik (< ?) : OTib *tsigs* 'joint, metrical section' < PTB *tsik 'joint'

(46) 一 NMan *yî* 'one' < MChi ⋆*it* < late MOC *îč < EOC *îk ~ *êk : OTib gtśig ~ OBur *tac* [tit] (< *tik) ~ Pyu *tå* (*[ti]) ~ Tgt ⋆*tik*, all < CTB *tîk ~ *têk 'id.' < *te ('determiner'?) + *îk ~ *êk 'one' ← EOC *îk ~ *êk 'one'

VII. Dragon • Sheep

The preceding discussion clarifies several important problems in Old Chinese phonology. Leaving aside for the moment matters of etymology, the comparative and internal evidence provides solutions to problems in the history of vowel shifts, word structure changes, and possibly even the development of tones, during the long Old Chinese period.

Apparent Old Tibetan and other Tibeto-Burman reflexes of the Chinese words □ *kŏu* 'mouth', 苦 *kŭ* 'bitter', and 馬 *mă* 'horse', among others, establish that the Chinese words had the vowel *a ~ *â in the *first* syllable at some time during the Old Chinese period. Similarly, the Tibeto-Burman relatives of Chinese 人 *rén* 'man, person', and 民 *mín* 'people' suggest that the Chinese words had the vowel *î during the Old Chinese period, while the Tibeto-Burman

[68] Virtually the same neutralization happened in Lolo-Burmese; see note 54.

reflexes of Chinese 二 *èr* 'two' and 你 *nǐ* 'you' establish the vowel *e ~ *ê for the same period of Old Chinese.

Old Japanese reflexes of Chinese 介 *jiè* 'shell', 芥 *jiè* 'mustard', 苦 *kǔ* 'bitter', and 馬 *mǎ* 'horse', among others, together with Tibeto-Burman data, indicate that the complex initial clusters of late Middle Old Chinese, evidently a language with monosyllabic root structure, derive from simple initials in Early Old Chinese.

The history of Tibeto-Burman and Japanese reflexes of Chinese 馬 *mǎ* 'horse', 梅 *méi* 'plum', 甲 *jiǎ* 'shell, armor', and 五 *wǔ* 'five', among others, reveals that the complex initial clusters and monosyllabic root structure of late Middle Old Chinese are the result of extensive syncope and metathesis of earlier disyllablic or longer forms.

The fact that some loanwords in Tibeto-Burman and Japanese-Koguryoic which have final velar nasals correspond to words with the rising tone in Middle Chinese (for example, the attested loanforms of the word 馬 *mǎ* 'horse' in those languages) reveal that this tone may sometimes be the relic of an unknown final which in Old Chinese could rhyme in some cases with words that are later (in Middle Chinese) attested with a final velar stop. Since the unknown final is preserved as a velar nasal [ŋ] in foreign loanwords from peripheral Chinese dialects, it accordingly appears to be the eastern (and other peripheral) dialect reflex of a continuant *γ in the central dialect, both deriving from an earlier intervocalic voiced velar stop *g.[69] At the same time, some Tibeto-Burman loans from Chinese have final velar stops corresponding to final velar nasals in Middle Chinese, which might seem to indicate that the oralization of nasals discussed in Section II affected nasal codas as well in Early Old Chinese and Middle Old Chinese. For example, Old Tibetan *fibrug* [^{m}brug] 'dragon' corresponds perfectly to an HSR-type reconstruction of 龍 *lóng* 'dragon'[70] as Old Chinese *mroŋ or *mruŋ—except for the final, which has remained puzzling. The onset of the Tibetan word is a prenasalized voiced stop, corresponding to a plain nasal onset in the reconstructed Chinese word, possibly indicating that the oralization of the nasals in the donor Chinese dialect affected both

[69] Cf. Section V.

[70] Cf. 龐 *páng* 'lofty', also read *lóng* 'replete' < MChi baiwŋ/bœ:wŋ ~ ləwŋ (Pul. 233, 198) < OChi "^{A}b-roŋ" (Sagart 1999: 40), which word has 龍 'dragon' as phonetic, as does 寵 *chǒng* 'to favor' < MChi *$tr^{h}uawŋ^{2}$ (Pul. 58) < OChi *drâŋ ~ *trâŋ or according to Sagart "*^{A}hroŋʔ" (1999: 40).

onsets and codas.[71] But this also would suggest that the final *-ŋ of the Middle Chinese is earlier than the *-g of the Old Tibetan, implying a change that is the reverse of a linguistic universal tendency. Another possibility we must consider, therefore, is that the Chinese donor dialect retained a voiced velar stop *-g from Early Old Chinese, when the word would have had the form *mCraga, which became *^{m}râŋ, then *^{m}rung in the dialect ancestral to Middle Chinese **luŋ*^{1} and modern *lóng* [luŋ], while in the dialect that donated the word to Tibetan, *mCraga became *^{m}brâg, then *^{m}brug, as shown in (47). This and other examples, such as 'sheep', in (48), together with the *Odes* rhyming data, indicate that some of the final velar nasals of Middle Chinese derive from voiced velar stops in intervocalic position in Early Old Chinese. One may thus assume that what appear in the *Odes* to be 'cross-rhymes' such as the above are yet another phonological feature peculiar to the language of that text.

(47) 龍 *lóng* 'dragon' < MChi **luŋ*^{1} < OChi *^{m}ruŋ ~ late MOC *^{m}râŋ (Sta. 588: *roŋ ~ Bax. 774: *C-rjoŋ [C-roŋ]) < EOC *m-Craga > MOC dial. *^{m}brâg > *^{m}brug → OTib *fibrug* [^{m}brug] 'id.'

(48) 羊 *yáng* 'sheep' < MChi **yaŋ*^{1} < late MOC *laŋ (irreg.) (Sta. 587: laŋ ~ Bax. 800: *(l)jaŋ [laŋ]) < *Clâŋ < EOC *Claga > MOC dial. *lâg > *lug → OTib *lug* 'id.'

This conclusion then brings up the question of the large number of reconstructed *-r finals in Middle Old Chinese (usually *-n in HSR, though some are now reconstructed as *-r), and the abnormally large number of reconstructed laterals. It is possible that these anomalous numbers are due to intervocalic reduction of what were alveodental stops in Proto-Chinese (among other reasons such as later mergers). Many features of late Middle Old Chinese and Late Old Chinese that have been ignored, or explained as due to earlier prefixing, infixing, or suffixing, appear to be traceable to syncope, metathesis, dialect shifting, and interdialectal borrowing. This does not rule out the effect of morphological features that were clearly present in earlier periods of Chinese but have mostly been lost and are quite difficult to recover (Sagart 1999), but it does reduce the likelihood that certain rare processes, such as infixing, were ever productive morpho-phonological features of Old Chinese.

---

[71] Such a shift is not unknown. In this analysis, the final *-g* of the Old Tibetan form could be a reduction from *-^{ŋ}g or *-^{ŋ}k (as in some Eastern dialect pronunciations of Standard American English final *-ŋ).

VIII. WHAT • COW • EYE

One conclusion that may be drawn from the examples analyzed above is that despite a great deal of similarity, both typologically (Kiyose 1997, Sakakura 1993, Janhunen 1997) and etymologically, it is difficult to argue in favor of a simple relationship among any combination of the Chinese, Tibeto-Burman, and Japanese-Koguryoic languages. The apparent similarities among various groupings of these languages have long stimulated speculation about wider relationships—including, among the more recent proposals, connections with Austronesian, Caucasian, Na-Dene, or Indo-European—but so far there has been no general agreement on any of the proposals (Meacham 1995, Vovin 1997), except that they are mostly unsubstantiated by the evidence cited by their supporters.

The one proposal which has a greater degree of likelihood is that which argues for a connection between Indo-European and Chinese or Sino-Tibetan. While this is not a new idea, its tenacity is based on some structural features that appear to be held in common by the two groups, as well as on the obviousness of some lexical correspondences. E.G. Pulleyblank is the latest and most prominent Sinologist to argue in favor of relating Sino-Tibetan to Indo-European (Pulleyblank 1966, 1995). Despite the criticism his theory has received, it has not been disproven. Since recent archaeological discoveries in China have given it significantly more credence, it now deserves serious reconsideration.

In order for Sino-Tibetan to be considered a genetic family of languages in the same sense as Austronesian, Semitic, and Indo-European, it is necessary to demonstrate that there is a unique 'Sino-Tibetan set' of forms not found in any other language or language family. If Pulleyblank's theory is well-founded, therefore, it could potentially disprove the Sino-Tibetan theory itself. There are thus at least two good reasons to examine his proposal here in the light of the new reconstructions of Old Chinese proposed above.

Three "Old Chinese and Proto-Chinese" etymologies by Pulleyblank are especially interesting: the word 牛 *niú* 'cow' (1995: 167, 180-181), reconstructed by him as *ŋʷə̀ɣ (1995: 180); the two interrogative stems, which give words for 'what?', 'why?', 'how?', 'who?', and so forth, derived by him from a root *ăkʷ- (1995: 172-173); and the word 目 *mù* 'eye', derived by him from *mjkʷ-, where the initial *m- is a 'body-part prefix' as found (fossilized) in Old

Tibetan (1995: 175). The linguistic details of Pulleyblank's arguments in favor of connecting these words via a genetic relationship with Indo-European are sufficiently idiosyncratic as to make them easily dismissable by the casual reader. It is also difficult to discern in his presentation a regular, principled account of the sound shifts that should have taken place if the Chinese words are indeed related to the Indo-European words, whether by divergence or convergence. The explication of the fate of Proto-Chinese *w in Early Old Chinese given in Section II above appears to provide a simpler solution to some of the problems with his arguments,[72] as shown in examples (49), (50), and (51).[73]

(49) OTib *ba* 'cow' (Pulleyblank 1995: 180) [74] : ?OJpn ⋆*u-* 'cow' (< ?*ŋu) : NMan 牛 *niú* < MChi ⋆*ŋuw* (Pul. 227) < MOC *ŋʷe (Sta. 548: *ŋuə ~ Bax. 779: *ŋʷji [ŋʷi]) < *ᵐgʷe < gmô < PChi *gʷô : PIE *gʷou (cf. Lat *bos*, Skt *gauḥ-*, *go*, TokB ke_{u}, TokA *ko*)

(50) OTib *ga-* 'interrogative stem' : NJpn *–ka* 'interrogative suffix' : NMan 何 *hé* 'what' < MChi ⋆*ɣa* (Pul. 122), 胡 *hú* 'why' < MChi ⋆*ɣɔ* (Pul. 126), etc. < OChi *ga- 'interrogative stem' (何 *hé*: Sta. 565: *g(h)âj ~ Bax. 762: *gaj; 胡 *hú*: Sta. 561: *ghâ ~ Bax. 763: *ga) < *ᵐka < *kma < *kwa ~ *kwe : PIE *kʷe 'interrogative stem' (cf. TokB *kâ* 'why', $k_{u}ce$ 'whom, what, which', etc.)

(51) OTib *myig* 'eye' < PTB *mêk : OJpn ⋆*me*/⋆*ma-*, Ryûkyû Hateruma dial. *miŋ ~ mîŋ* < PJpn *miŋa ~ *maŋi : NMan 目 *mù* < MChi ⋆*muk* < early MOC (irreg.) *ᵐbâk ~ *ᵐbek < EOC *wêk < PChi *ôk : PIE *okʷ- (cf. TokA *ak*, TokB *ek*)

These examples would seem to indicate that the interrogatives in Chinese and Tibetan are not examples of a unique Sino-Tibetan set. Since this may accordingly indicate a relationship of some sort with Indo-European, whether divergent or convergent, let us further test Pulleyblank's proposal by attempting to reconcile a selection of the etyma discussed above—some of which have also been discussed by him, though with very different reconstructions—with Proto-Indo-European, in (52) through (61).

---

[72] Note these regular changes: PChi *w > MOC *m(y), PChi *-ô > EOC *we, PChi *o > EOC *e, EOC *ᵐgV- > *ŋV-, EOC *ᵐkV- > *gV-; on the latter two, see the discussion in Section II, and similar arguments by Pulleyblank (1995: 166, 172-173); his etymologies, however, are radically different from those suggested here.

[73] All forms in (49) – (51) have identical semantics unless otherwise noted.

[74] Rgyarong *mbo* 'cow' (Huang 1992: 87) suggests that a Chinese cognate of OTib *ba* would have begun with *m-, making Pulleyblank's suggested connection of the two more difficult.

(52) OTib *myi* 'man, person' < PTB *mî : (perh.) OJpn ✯*pi-* 'person' : 人 *rén* 'man, person' < LMC *źin* < EMC ✯*nyin* < late MOC *nyîn < *myîn ~ *myîr < early MOC *^{m}bîr < EOC *wîr ~ *wêr 'man, person' : PIE wîr- 'man'

(53) OJpn ✯*mina,* ✯*me* 'woman' : 女 *nǚ* 'woman' < MOC *mîna < *^{m}bîna < *^{m}bêna < *C^{m}bêna < EOC *C^{w}êna : PIE *g^{w}ena 'woman'

(54) PTB *na 'you' : OJpn ✯*na* 'you' : 女 (汝) *rǔ* 'you' < MOC *mîna < *^{m}bîna < *C^{m}bêna < PChi *Twêna : PIE *tu (nom.) ~ *twe- (oblique stem) 'you' ~ *twei-no 'thine'

(55) OTib *do* 'two' < PTib *dwa ~ *dwe : 二 *èr* 'two' < LOC *nyis < late MOC *nyis < *mî(s) < *^{m}bê- < *d^{m}bê- < EOC *dwê- : PIE *dwo-, *dwe- (f.) 'two'

(56) OTib *ma* 'negative' < PTB *ma : OJpn ✯*na-* 'negative' : 不 *bù* < late MOC *pré < EOC *pare 'not' < PChi *^{m}pa- < *ma-, NMan 毋 *wú* < late MOC *mré < EOC *maré 'should not', both < PChi *ma- 'negative stem' : PIE *mê (> Tok *mâ*, Skt *mâ*) (cf. Pulleyblank 1995: 187-188), *ne 'negative'

(57) PTB *ma 'mother' : 母 *mǔ* 'mother' < late MOC *mré < EOC *maré (< PChi *maðer, or *matre?) : PIE *mâtêr 'mother' (cf. Pulleyblank 1995: 186), *mari 'bride, young woman'

(58) OJpn ✯*umi* 'sea' < PJpn *umәy ~ *^{m}mәy : 海 late MOC dial. *rmay < *mray < EOC *marê : PIE *mori 'sea' (> Germanic *mari, Latin *mare* 'id.') (cf. Pulleyblank 1995: 188-189)

(59) OTib *rmaŋ* 'horse' < CTB *mraŋ ← 馬 OChi dial. *mraŋ : OJpn ✯*uma* ~ *^{m}ma, Ryûkyû Hateruma dial. *^{m}maŋ* < PJpn *umaŋ ~ *^{m}maŋ ← 馬 OChi dial. *rmaŋ < late MOC *rmaŋ < *mraŋ < EOC *mraga < PChi *marka : PIE *marko 'horse'

(60) OJpn ✯*mimi* 'ear' (written 三 三 ✯*mi-mi* 'three-three') < Pre-OJpn *mәi- < PJpn *mәri 'ear' < PJK *miri (cf. OKog 密 ✯*mir* 'three') : 耳 *ěr* 'ear' < MChi ✯*nyi*^{2} < late MOC *n^{y}rê ~ *n^{y}rî < *mîrê ~ *mîrî < early MOC *^{m}bîrê ~ *^{m}bîrî < EOC *wêrê ~ *wêrî < PChi *ôz-ê ~ *ôz-î < *ôs-ê ~ *ôs-î[75] : PIE *ous- 'ear'

(61) OTib *gcig, cig* < CTB *têk 'one' < *t- + *êk : 一 *yî* < MChi ✯*it* < late MOC *îč < *îk < EOC *êk : Skt *êka* (cf. Mitanni Indo-Aryan *aika*)[76] < PIE *oi-ko (cf. Pulleyblank 1995: 185)

---

[75] The same *Odes* rhyme includes 來 *lái* 'wheat, barley, to come' (Sta. 548: rə̄), which occurs in 'cross-rhymes' with words ending in a final velar, so perhaps *lái* < *^{m}braχ < *^{m}bras : OTib *fibras* [^{m}bras] 'rice, fruit, result' : PIE *bhares 'barley'.

PTB *appears* to have had the same root for 'ear' and 'nose', *na, with differentiating prefixes, producing OTib *rna* 'ear' and *sna* 'nose'. It is likely that *sna* reflects a pre-PTB form *(s)na, but in view of the OChi and OJpn forms of the word for 'ear', the reconstruction of the PTB word for 'ear' requires further study.

[76] The distinctive Indo-Aryan word for 'one', *êka*, is attested in Mitanni texts from the 16th century B.C. as *aika* (Drews 1988: 60). The agreement of the Chinese form with the Indo-Aryan could conceivably be taken to indicate the possibility that a subgroup of the *marya* (Skt 'young warrior') charioteers who invaded India in the mid-second millenium B.C. could have invaded China too. On the other hand, the

The fact that this brief sketch includes good correspondences between East Asian internally motivated reconstructions and Indo-European reconstructions for the first two numerals, the second person pronoun,[77] and the interrogative and negative stems, cannot be simply brushed aside, nor can many of the other correspondences that are very easily found.[78] Whether borrowed or inherited, the data suggest a relationship that was once quite close.

It may be wondered if there is any substantial *non-linguistic* evidence for believing in a possible connection of some sort between these two most representative languages of East and West. Others have long noted similarities between Chinese and Indo-European (or Sino-Tibetan and Indo-European) and so forth, but it has so far been difficult to convince anyone of the rectitude of such comparisons because of difficulties in traditional Chinese reconstruction that have prevented resolution of the irregular correspondences. Pulleyblank, following in the footsteps of other earlier Sino-Tibetanists, has long argued—on both lexical and structural phonological grounds—that the Sino-Tibetan and Indo-European families are genetically related, though the relationship in his view is remote in time. Recent archaeological discoveries in western China and East Turkistan have once again drawn attention to this theory and forced Sinologists to at least consider the possibility that the immense cultural impact of new Western imports—the domesticated horse, domesticated sheep, wheat, barley, and so forth, all of Near Eastern origin (Barnes 1993, Barber 1998: 653-654, Good 1998: 659), as well as the fully developed chariot, a Western invention (Drews 1988)—was reflected in loanwords, as is normal in the linguistic history of the rest of the known world. With the discovery of large numbers of Europoid mummies in western China, the earliest of which date to about 2000 B.C., and advances in Old Chinese reconstruction that have revealed previously overlooked lexical and perhaps structural relationships

---

apparent presence of sound changes and lexical items characteristic of Tokharian, Iranian, Greek, Germanic, and Slavic suggests that the immigrants spoke a language belonging to an otherwise unknown branch of Indo-European.

[77] And perhaps the first also—PTB *ŋa 'I' is a perfect reflex of PIE *egô 'I', parallel to PTB *na 'in' and PIE *(e)no 'in'.

[78] For example, TokB *mâ* 'negative' < PIE *mê, TokB *mâka* 'many, much' < PIE *megh-a 'great': OTib *ma* 'negative' (< *ma ~ me), *maŋ* 'many, much' (< *maga ~ *mega); TokB *nes* 'be, exist, dwell' < PIE *(e)no-s 'be, exist', TokB *ne-* 'locative postposition' < PIE *(e)no- 'id.' (Adams 1999: 345, 341) : OTib *gnas* 'be, exist, dwell' (< *na-s), *na* 'locative postposition'.

(Pulleyblank 1995), there can no longer be much serious doubt that the Westerners who brought the new technology to China spoke one or more early Indo-European languages. However, with the notable exception of Pulleyblank, Sino-Tibetanists by and large continue to reject any Indo-European linguistic relationship, whether convergent or divergent, with the exception of, literally, a small handful of loanwords—evidently much later borrowings—most of which are also still vigorously debated. Even Pulleyblank himself argues that his putative relationship between Sino-Tibetan and Indo-European is much earlier than 2000 B.C. (Pulleyblank 1995, 1998).

Yet if the early Indo-European peoples had a revolutionary cultural and linguistic impact on every area of the world into which they migrated, including places where an Indo-European language did *not* supplant the native language, as in the ancient Near East (Drews 1988), and reasonable arguments can be made for the presence of numerous early Indo-European loanwords in Chinese and other East Asian languages, we must consider the possibility that the Indo-Europeans had an equally great impact on the native peoples they encountered when they reached East Asia. In any case it does not make sense to completely deny any and all Indo-European linguistic influence when we know Indo-European speakers were living in contact with Chinese, even Proto-Chinese, for millenia.

It is now becoming accepted by archaeologists and historians that revolutionary cultural changes took place at several points in the second millenium B.C. in the area that became China, namely at the beginning, ca. 2000 B.C.; again in the 14th century, with the establishment of the Shang dynasty and the appearance of the Shang's westernmost semi-client state of Chou; and finally, at the turn of the next millenium, in 1027 B.C., with the Chou conquest of the Shang; both dynasties were heavy users of chariots. While the late third and early second millenium B.C. saw the arrival of the earliest known Europoid peoples in the Lop Nor area of western China, that was not the only time Europoid peoples arrived in the area. Another group arrived several centuries after them, closer to the middle of the millenium (Barber 1998: 653-654), and other groups arrived later still. All linguistically identified 'native' languages spoken up to the western borders of traditional China (i.e., present-day western Gansu) until nearly the end of the first millenium A.D. were Indo-European—Tokharian, Iranian, and Indic. The arrival of these languages predates any notice in ancient Chinese historical sources or,

so far, any concrete indication of the linguistic affiliation of any earlier dwellers, at least in East Turkistan.[79] We must conclude that the intrusive Europoid peoples were indeed Indo-European speakers, though we do not have any way of determining just what language they spoke unless we can discern the characteristic features of one Indo-European language from the loanwords present in Chinese.

The cultural record leaves little doubt that the Indo-Europeans had as powerful an impact on the native peoples of East Asia as they had on peoples elsewhere in Eurasia. This means that the Indo-European arrival resulted in a 'punctuated equilibrium' as postulated by Dixon (1997). If this view is correct, there are two possible explanations for the subsequent development of language families in the region.

The fact that a small number of immigrants can have an immense impact on a language spoken by a large population is well known in East Asia. It is historically attested twice in the history of Japanese, first during the early Middle Ages when only a tiny handful of Chinese went to Japan and a handful of Japanese went to China and came back, but the language was inundated and reshaped by Middle Chinese loanwords, and secondly with the Meiji Restoration in the nineteenth century, when a tiny number of foreigners came to Japan and a handful of Japanese went abroad, but the language was again inundated by loanwords, this time from English (long before the American occupation after World War II). One scenario, therefore, is that the Indo-European immigrants could have had a powerful impact on the local peoples, contributed some material and linguistic culture —loanwords—and then have been absorbed by the locals. This is what happened at about the same time in the ancient Near East, where primarily Semitic cultures were influenced by an influx of Indo-Europeans, but withstood and outlasted them (Drews 1988).

When intrusive languages encounter weaker or less numerous local populations, the tendency is for the newcomers' language to

---

[79] Pulleyblank has pointed out that the Han histories refer to the people dwelling south of the oasis cities in the Tarim as 羌 Ch'iang, who are considered by him and many others to have been early Tibeto-Burmans. However, we do not know if the Ch'iang were indeed Tibeto-Burmans. The first Ch'iang group to be linguistically identified, because of the preservation of a text translated into their language in the Later Han period (Coblin 1978), is the Pai-lang Ch'iang. The name Ch'iang is a Chinese word for western sheepherding people and probably referred to Indo-Europeans (who apparently introduced the domesticated sheep) as well as to Tibeto-Burmans, but we really do not know if the Tarim Ch'iang were Tibeto-Burmans, Indo-Europeans, a mixture of the two groups, or some other people entirely.

replace the natives' language without significant 'substratum' influence, but sometimes even a small intrusive group can successfully maintain and eventually impose their own language in the colonized territory. This happened in Turkey with the replacement of Greek and other languages by Turkish, in Egypt with the replacement of Coptic by Arabic, in Hungary with the replacement of German and other languages by Hungarian, and in Korea with the replacement of Koguryo, Paekche, and other languages by Korean. The degree of 'substratum' influence varies from case to case. The alternative scenario, therefore, is that despite their smaller numbers, the Indo-European newcomers maintained their language, which the local East Asian people or peoples adopted, changing its phonology to accord with their own phonological system or systems. In this scenario, the Indo-Europeans' language, especially the phonology, was changed by the adopting natives much as in recent times English has been changed by the Jamaicans in Jamaica, by the Indians in India, by the Nigerians in Nigeria, and similarly in other countries where the introduced English language is now a native language. Whether the disyllabic (and possibly longer) forms that must be reconstructed for Proto-Chinese, *contra* the prevailing view (Packard 1998: 6-7), are the residue of such an ancestry or not according to either scenario is a subject that calls for further investigation.

Conclusion

There are always at least two different explanations for similarity between two languages: divergence from a common ancestor, commonly known as 'genetic relationship', and convergence, also known as 'loan relationship'. Since both convergent and divergent forces are operating at the same time (at different rates) between any two languages in contact, in order to demonstrate the existence of a genetic relationship beyond a reasonable doubt it is necessary to show that the comparands are not simply loanwords. For example, it is true that virtually the entire world's modern languages share the same word for 'chocolate', but that does not mean they are all genetically related to Nahuatl, or to the languages that spread this loanword. Moreover, even though the words 'mama', 'papa', 'tea', and 'coffee', as well as 'chocolate' and probably many other words, are held in common by a large number of modern languages around the world, we cannot

therefore conclude that the languages are related (except by convergence). If one were to insist that the words in question are not borrowed, but genetically inherited in each language from a common ancestor, that would mean the chosen set of words could *not* be used to demonstrate a genetic relationship among the languages involved because it is a *non-distinctive set*. The discussion in Section VIII above, while hardly the last word on the matter, is sufficient to establish that much of the putative 'Sino-Tibetan set' is non-distinctive, being found also, arguably, in Indo-European and Japanese-Koguryoic (at least). If the explanation for this fact is that Chinese, Tibeto-Burman, and Indo-European are genetically related, as some have argued, it must be shown that the common lexical set is not also found in other nearby language families, such as Austronesian, Austroasiatic, Tungusic, and perhaps others still further afield.

While similar lexical items must be considered when dealing with the Sino-Tibetan theory, or any other theory of linguistic relationship, it is not possible to take them at face value. In fact, we must be just as careful when reconstructing only within Tibeto-Burman, or Chinese, or Japanese-Koguryoic. It is only now becoming accepted that the Tibeto-Burman numerals have been borrowed from Chinese—though they were apparently not borrowed into Proto-Tibeto-Burman, as seems clear from the most recent attempt to reconstruct a Proto-Tibeto-Burman system of numerals (Matisoff 1997), but into some of the early daughter languages, and were probably borrowed from one Tibeto-Burman language to another after that. This means that the numerals and many other lexemes cannot be used to reconstruct Proto-Tibeto-Burman, though they are very useful for determining features of later periods and daughter languages of Tibeto-Burman, and especially for reconstructing the donor languages—including both Chinese and the as yet unidentified Indo-European language spoken by the introducers of ancient Western technology.

In sum, it is certainly possible that Tibeto-Burman and Chinese are genetically related, but since the best-supported etymologies appear to be shared not only with Japanese-Koguryoic[80] but also with Indo-European, it is likely that the relationship is either one of shared

[80] For example, of the four words cited by Bradley (in this volume) from the Bisoid languages of the Southern Loloish branch of Tibeto-Burman (found in Vietnam, Laos, Thailand, and China), three could easily be cognates of the Japanese equivalents: ba^{33} 'mother' : OJpn **pa* 'id.', bi^{21} 'fire' : OJpn **pï* 'id.', naŋ^{55} 'you' : OJpn **na*.

loan influence from the same donor or one of common descent from the same intrusive ancestor. Further study is therefore needed to determine more precisely the history of the interrelationship of these four families.

References

Adams, Douglas Q. 1999. *A dictionary of Tokharian B.* Amsterdam: Rodopi.

Barnes, Gina L. 1993. *China, Korea, and Japan: the rise of civilization in East Asia.* London: Thames and Hudson.

Barber, E. J. W. 1998. Bronze age cloth and clothing of the Tarim Basin: the Kroran (Loulan) and Qumul (Hami) evidence. *In* Mair 1998, Vol. II: 647-655.

Baxter, William H. 1992. *A handbook of Old Chinese phonology.* Berlin: Mouton de Gruyter.

Beckwith, Christopher I. 1987. *The Tibetan empire in Central Asia: a history of the struggle for great power among Tibetans, Turks, Arabs, and Chinese during the early middle ages.* Princeton: Princeton University Press.

—— 1996. The morphological argument for the existence of Sino-Tibetan. *Pan-Asiatic Linguistics: Proceedings of the Fourth International Symposium on Languages and Linguistics, January 8-10, 1996.* Bangkok: Mahidol University at Salaya, pp. 812-826.

—— 1999. Defining 'basic vocabulary': the correlation between frequency and rate of retention. Paper given at the Linguistic Society of America, Chicago.

—— 2000. Toward Common Japanese-Koguryoic: a reexamination of Old Koguryo onomastic materials. *In* J.J. Nakayama and Charles Quinn, Jr., ed., *Japanese/Korean Linguistics,* Vol. 9. Stanford: CSLI, pp. 3-16.

Benedict, Paul 1972. *Sino-Tibetan: a conspectus.* Cambridge: Cambridge University Press.

Coblin, W. South 1974. An early Tibetan word for 'horse'. *Journal of the American Oriental Society* 94.1: 124-125.

—— 1978. A new study of the Pai-lang songs. *Tsing Hua Journal of Chinese Studies* n.s. 12.1-2: 179-216.

Conrady, August 1896. *Eine indochinesische Causativ-denominativ-bildung und ihr Zusammenhang mit den Tonaccenten.* Leipzig: Harrasowitz.

Dixon, R.M.W. 1997. *The rise and fall of languages.* Cambridge: Cambridge University Press.

Drews, Robert 1988. *The coming of the Greeks: Indo-European conquests in the Aegean and the Near East.* Princeton: Princeton University Press.

Driem, George van 1997. Sino-Bodic. *Bulletin of the School of Oriental and African Studies,* 455-488.

—— 1998. Neolithic correlates of ancient Tibeto-Burman migrations. *In* Roger Blench and Matthew Spriggs, ed., *Archaeology and language II: correlating archaeological and linguistic hypotheses.* London: Routledge, 67-114.

Giles, Herbert A. 1898. *A Chinese biographical dictionary.* London: Quaritch; repr. Taipei: Literature House.

Good, Irene 1998. Bronze age cloth and clothing of the Tarim Basin: the Charchan evidence. *In* Mair 1998, Vol. II: 656-668.

Hayes, Bruce 1989. Compensatory lengthening in moraic phonology. *Linguistic Inquiry* 20: 253-306.

Huang, Bufan, ed. 1992. *Zang-Mian yuzu yuyan cihui: a Tibeto-Burman lexicon.* Beijing: Zhongyang minzu xueyuan chubanshe.

Janhunen, Juha 1992. Das Japanische in vergleichender Sicht. *Journal de la Société Finno-Ougrienne* 84:145-161.

—— 1996. *Manchuria: an ethnic history* [= Mémoires de la Société Finno-Ougrienne 222]. Helsinki: The Finno-Ugrian Society.

—— 1997. Problems of primary root structure in Pre-Proto-Japanic. *International Journal of Central Asian Studies* 2:14-30.

—— 1998. The horse in East Asia: reviewing the linguistic evidence. *In* Mair 1998, Vol. I: 415-430.

Karlgren, Bernhard. 1957. *Grammata Serica recensa.* Stockholm: reprinted from the *Bulletin of the Museum of Far Eastern Antiquities* 29: 1-332.

Kiyose, Gisaburo N. 1997. Yamataikoku no gengo o ronji gen-Nihongo no kochi ni oyobu. *Gogen tankyû* 5: 202-244.

Mair, Victor H., ed. 1998. *The bronze age and early iron age peoples of eastern Central Asia.* 2 Vols. Philadelphia: Institute for the Study of Man.

Martin, Samuel E. 1987. *The Japanese language through time.* New Haven: Yale University Press.

Matisoff, James A. 1991. Sino-Tibetan linguistics: present state and future prospects. *Annual Review of Anthropology* 20: 469-504.

—— 1997: *Sino-Tibetan numeral systems: prefixes, protoforms, and problems.* Canberra: Pacific Linguistics.

Meacham, William 1995. "Deep-level reconstructions" of linguistic prehistory and the return to the nuclear area. *In* Wang 1995: 299-324.

Miller, Roy Andrew. 1988. The Sino-Tibetan hypothesis. *Bulletin of the Institute of History and Philology, Academia Sinica* 59.2: 509-540.

Packard, Jerome L. 1998. Introduction. *In* Jerome L. Packard, ed., *New approaches to Chinese word-formation: morphology, phonology and the lexicon in modern and ancient Chinese.* Berlin: Mouton de Gruyter.

Pan, Ku, trans. Homer H. Dubs 1944. *The history of the Former Han Dynasty*, Vol. II. Baltimore: Waverly Press.

Pulleyblank, Edwin G. 1962. The consonantal system of Old Chinese. *Asia Major*, 9: 58-144, 206-265.

—— 1996. Chinese and Indo-Europeans. *Journal of the Royal Asiatic Society* 1996: 9-39.

—— 1983. The Chinese and their neighbors in prehistoric and early historic times. *In* David N. Keightley, ed., *The origins of Chinese civilization.* Berkeley: University of California Press, 411-466.

—— 1984. *Middle Chinese: a study in historical phonology.* Vancouver: University of British Columbia Press.

—— 1991. *Lexicon of reconstructed pronunciation in Early Middle Chinese, Late Middle Chinese, and Early Mandarin.* Vancouver: UBC Press.

—— 1995. The historical and prehistorical relationships of Chinese. *In* Wang 1995, 145-194.

—— 1996. Zou 鄒 and Lu 魯 and the Sinification of Shandong. *In* P.J. Ivanhoe, ed., *Chinese Language, Thought, and Culture.* Chicago: Open Court, 39-57.

Sagart, Laurent 1995. Some remarks on the ancestry of Chinese. *In* Wang 1995, 195-223.

—— 1999. *The roots of Old Chinese.* Amsterdam: John Benjamins.

Sakakura, Atsuyoshi 1993. *Nihongo hyôgen no nagare* [= Iwanami seminâ bukkusu 45]. Tokyo: Iwanami Shoten.

Schuessler, Axel 1987. *A dictionary of early Zhou Chinese.* Honolulu: University of Hawaii Press.

Starostin, Sergei A. 1989. *Rekonstrukciya drevnekitayskoy fonologičeskoy sistemï.* Moscow: Nauka.

Takata, Tokio 1988. *Tonkô shiryô ni yoru chûgokugoshi no kenkyû* [= A historical study of the Chinese language based on Dunhuang materials]. Tokyo: Sobunsha.

Vovin, A. 1997. The comparative method and ventures beyond Sino-Tibetan. *Journal of Chinese Linguistics* 25.2: 308-336.

Wang, William S.Y., ed. 1995. *Languages and dialects of China.* n.p.: Journal of Chinese Linguistics Monograph Series, No. 8.

Yakhontov. S.E. 1965. *Drevnekitayskiy yazïk.* Moscow: Nauka.

This page intentionally left blank

A GLOSSARY OF PYU

CHRISTOPHER I. BECKWITH (INDIANA UNIVERSITY)

There are two main published sources for the Pyu language: the urn inscriptions dated to the seventh and eighth centuries A.D. and the Myazedi inscription of 1112 A.D. The lexical material in these sources has been collected and published by the late Robert Shafer in his study of the Pyu language (1943) based on the pioneering decipherment of Otto Blagden (1911, 1913-1914). Shafer's work (including the glossed text) is the basis for the present glossary, the compiler of which is responsible for any errors of transmission.

The pronunciation of Pyu is not always clear from the script, due largely to the scanty corpus. Because there are differences between the earlier language of the urn inscriptions, called by Shafer 'Old Pyu', and that of the later inscriptions, forms known only from the urn inscriptions are marked 'OPyu' in the glosses. The transcription follows Shafer's, with the following exceptions: his velar nasal symbol (*ṅ*) has been converted mechanically to the IPA symbol (*ŋ*), his use of an apostrophe (') for aspiration has been converted mechanically to superscript *h* (^{h}), all macron accents have been converted mechanically to circumflex accents, and syllable hyphens have been omitted. Note that many of the diacritics (except the circumflex, which indicates length, and the subscript dot, which in some cases indicates retroflexion) are tone markers. However, it is probable that the combination *ạ* (as in *tạ* 'one' and *pạ:* 'give') represents a vowel different from [a]. In view of the existence of minimal pairs such as *plạ* 'four' and *plị* 'grandson', and of the relationship of Pyu *hrạ* 'eight' to Old Tibetan *brgyad*, the value of *ạ* was perhaps closer to [ɛ] than to [i] as in Burmese.

Words which have unknown or uncertain meaning, or which are obvious loans from Pali or Mon, have been excluded; those which are not obvious are included, whether or not they could be loans. (However, the words for 'two' and 'seven' are included, though this compiler has reversed Shafer's glosses.) It should be emphasized that there are numerous additional words in the inscriptions which recur in similar syntactic positions and could perhaps yet be identified.

Pyu	Gloss
ḅà:	honorific prefix (also OPyu)
ḅí:	past tense
ḅû:	lord
ḅû:	optative: may
bo	form
dà	prioritive (OPyu)
ḍʰau	that
ḍị	sight
ḍí	passive
ḍọ:	benevolent, compassionate
gà:	first person pro.: I
gi	first person possessive pro.: my
hi	die
hlau	dedication formula
hni	seven (or two)
hŋu	persons
ho:	three (OPyu also *hau*)
hrà	eight
kle:	to repose (OPyu)
kni	two (or seven)
la	or
la	month (OPyu)
lo:	postp.: into, to, upon
ma	relative pronoun
ma	negative: not
maya	queen
mdau	remember, recall
mî	name
mtau	nourish
ŋa, piŋa, miŋa	five (OPyu)
ŋu	exclaim
o	village
pà:	give; permit
pʰwụ	day
plà	four (OPyu)
pli	grandson
pra	good
pri:	city
rí	to delight
ru	bone (OPyu)
sa:	son

sȁ:	pronounce (a dedication)
se	make
si	speak, say, call (by name)
sni	year (also OPyu)
sru:	kinsman
sû	ten (OPyu perh. also *sau*)
t(k)ûo	nine (OPyu)
ta	perfect (also OPyu)
tȁ	one (also OPyu)
tḅa:	again (also OPyu)
tc^ha	pour
tc^ha	likeness
tc^he	do (present)
$tc^hí$:	obtain, get, attain
tc^ho	aspiration
tc^ho	exclamatory particle
tdạ:	king (also OPyu)
tdị:	speak
tdû:	water
tí	preposition: in, for, on (also OPyu)
to:	terminal particle: topic shift marker
tpû	twenty (also OPyu; perh. *npû*)
tra	goods, ornaments
tra:	slave
tro:	favor
tru	six (OPyu)
tû	all, entirely, only of
tu:	bring
twạ:	elapse (of time)
u	third person pro. (genitive?) (also OPyu)
-wȁ	dative (of third person pro.)
yȁ	this (also OPyu: 'here')

References

Blagden, C. Otto. 1911: A Preliminary Study of the Fourth Text of the Myazedi Inscriptions. *JRAS*: 365-388.

—— 1913-1914: The 'Pyu' Inscriptions'. *Epigraphia Indica* 12: 127-132.

Shafer, Robert, 1943: Further analysis of the Pyu inscriptions. *HJAS* 6: 313-366 + 5 plates.

This page intentionally left blank

A LIST OF OLD BURMESE WORDS FROM 12TH CENTURY INSCRIPTIONS

RUDOLF A. YANSON (ST. PETERSBURG STATE UNIVERSITY)

The language of the inscriptions of the Pagan period (12-13th centuries) is called Old Burmese. This list is restricted to the inscriptions of the 12th century, otherwise it would increase considerably.

The material was collected from *Shehaun Myanmar kyauksa mya* ['Old Burmese Inscriptions'], Vol. 1, Rangoon, 1972. Only lexical items and those which are not already present (in Modern Burmese form) in Benedict's list of Tibeto-Burman roots (Benedict 1972), or which are present therein but need comments, are included here.

By the time the Burmese started to write, their language had been influenced by Sanskrit and Pali, by Mon, and by Pyu. Indo-Aryan words are in most cases easily identified as alien and thus pose no problem. For Mon, I have relied on Dr. Hla Pe's list of Mon loanwords in Burmese (Hla Pe 1967), although it has some drawbacks. Of Pyu we know too little to have grounds for suspecting any Burmese word to be of Pyu origin.

Words are given in transliteration in Burmese alphabetical order. My transliteration differs from that of Benedict's, which sometimes reflects pronounciation and not spelling. Instead of *ts* I use *c* which is more common in transliterations of Burmese words in different publications. Instead of the sequence *ui* I use *iu* for the following reason. In the inscriptions some rhymes have two spellings, e.g., *ai* ~ *ay, au* ~ *aw*; the spellings *ui* and *uy* also occur, but (unlike the former two pairs) they reflect different rhymes, so the spelling *ui* might be misleading. Besides, in the earliest inscriptions, the rhymes which Benedict spells with *ui* were spelled *iw*; only later did the spelling became *iuw*. Now final *w* is dropped, which justifies both ways of transliteration—*ui* and *iu*—but based on the historical evidence, transliterating the rhymes under discussion as *iu* is more justified.

Instead of Benedict's *au,* which reflects a modern phonetic transcription, I use *o*, the transliterated value.

Three words in the inscriptions, which are the commonest there, viz., *puhra* ~ *purha* 'Buddha, pagoda', *kyon* ~ *kywan* 'slave', *kloŋ* 'a

monastery', pose a problem. They are no doubt of late origin, which unfortunately I have failed to trace. They should be Mon loans, because the Burmese got acquainted with Buddhism via the Mons, but neither Hla Pe nor the dictionaries I have consulted shed light on the problem of their origin. In all probability these words are not of Tibeto-Burman stock and therefore they are not included in the list.

OB	MB	GLOSS
akliw	akyiu	'benefit, profit'
achaŋ	achaŋ	'statue, form, image'
achuy	chwe	'friend'. In inscriptions mostly used within a compound *achuy amliw* 'friends and relatives'. The word *amliw* is given in Benedict with the meaning 'race, lineage' (1972: 43, #150).
ati	ati	'only, nothing but'; in inscriptions 'pure (gold)'.
atiw	tiu	pro.'we'; in inscriptions mostly 'I', but very respectful, e.g., *atiw kywan* 'I, your slave' (lit., 'we, the slave').
ahmok	ahmok	'in front of, in presence of'
ahriy	ahre	'front, east'
ahraŋ	ahraŋ	'master, lord, owner'. Though Benedict considers this word to be of TB origin (1972: 48, #205), it is very doubtful. It derives from Pali *araham* 'Arhat'. Burmese resonants can not be postaspirated, but only preaspirated; this is why the Pali word was respelled according to Burmese phonological rules (*rh* › *hr*)
i	-	Feminine prefix(?). Many feminine names in inscriptions start with this word. Not preserved in MBur.
iy	i	'this'
uy	we	'to distribute, share'
kin	kyin	'to conjure, entreat'
koŋ	koŋ	'be good'
kra	kra	'to hear, witness'
kri	kri	'be big,important'
klok	kyok	'stone'. This word is among Benedict's reconstructions (1972: 32, #88), but all the cognates of it have final *-ŋ*, whereas the Burmese word has *-k*, which is very unusual.
khlyaŋ	khyaŋ	'to want, wish' (bound form)
khwai~ khway	khwai	'to divide into separate parts'
ŋa	-	masculine prefix(?). Many masculine names start with this word. Not preserved in MBur in this function.
ŋai~ŋay	ŋay	'small, young'
cac	cac	'war'. The old spellings *-ac* and *–añ*, which form a correlated pair, are very peculiar in the sense that they are very far from representing the phonetic value of corresponding rhymes. I reconstruct the spelling *ac*,

		which appears in inscriptions also as *–ec, ic* and even as *–iec,* as [iet], and the spelling *–añ* as [ieñ]. (For details see Yanson 1990: 90-101).
chip	chip	'poison'
chu	chu	'a reward'
chut	chut	'retreat, recede' (about the moon)
chok	chok	'to build'
chiw~ chiuw	chiu	'bad, evil'. Often used in combination with *ŋray* 'hell' (a LW, from Pali *nirayo*), *chiw ŋray* 'hell'.
toŋ	toŋ	'to ask for'
toŋ	toŋ	'south, mountain'
twak	twak	'to count'
than	than	'toddy palm'
thi	thi	'umbrella'
thiw	thiu	demonstrative pronoun 'that'
thiw	thiu	'to strike, pierce'
nok	nok	'back, west'
noŋ	noŋ	'future'
niw	niu	'be awake, wake up'
pay	-	'beloved, dear'. This word is suspicious. It is not preserved in MBur, and in the Myazedi inscription it is spelled with long *â*, which is typical for Pali loans, but I did not find a Pali cognate for this word.
prañ	prañ	'country, city'
plu	pru	'to do, make'
pham	pham	'to catch', cf. Benedict's TB **bam ~ pam* 'to be beaten, defeated' (1972: 125, #471). This Burmese word is not among the cognates for this root. Probably it should be.
phu	phu	'to pray, worship'
phlac	phrac	'to be'. Benedict attributes the meaning 'to be' to Burmese *hri* (1972: 62, #264). In fact *hri* means 'to have'.
maya	maya	'wife'
min	min	' to say, order'
miy ma	min ma	'woman'
muy	mwe	'to give birth, bring up'
mya	mya	'many, much'. All Benedict's cognates for this word have medial *-r-* (p.43,#148), but the Burmese word has *–y-*. In some cases Burmese medial *y* evolves from *l* or *r*, but in this particular case *y* is original.
mlac	mrac	'river'
mlac ok	mrok	'north' (*mlac* 'river', *ok* 'under')
mlat	mrat	'be excellent, noble'
mwat	mwat	'to starve'
yok	yok	classifier for people
yok kya	yok kya	'a man, male'
rac	rac	'remain behind, leave without care'
rañ	rañ	'liquid, juice'
riy	re	'to write, to paint'
riy	re	'water'. According to Benedict, the TB root for 'water'

		is *ti(y)* (1972: 26, #55). It is difficult to assume that the Burmese word is not of TB origin, but Benedict's reconstruction is well supported by many cognates. Not to make an important Burmese word the outcast of the family, maybe we could blame for the situation a quality of Burmese *r*. There are examples that Pali cerebral *d* is reflected in Burmese as *r*, e.g., Pali *nadi* 'period of time'—Burmese *nari* 'hour, clock, watch'. Some more examples could be found. So can't we suppose that when the Burmese started to write, for some reason they used the symbol *r* for TB initial *t*?
rok	rok	'to arrive'
rya	ya	'dry field'.Benedict remarks that OBur *ry-* has become simply *r-* in MBur (1972: 54, #229). This example shows he is not right. Hla Pe lists this word as a borrowing from Mon (1967: 75). He states that the word first occurs in Mon and Burmese writings in the form of *ya* in the 15th century. If in Mon it can be traced only from the 15th century and in the form *ya*, it means it was borrowed by the Mons from Burmese.
hraw~ hrow~ sraw	saro	'time, when'. The word is interesting because it is the only one in inscriptions containing initial *sr-*, though it interchanges with *hr-*. It is a well established fact that Burmese initial *h-* in complex initials descends from PTB **s-*, so why this particular word retains the PTB initial in the Pagan period remains unclear. It is noteworthy also that in MBur the word is preserved with an obsolete initial the components of which belong now to different syllables.
lay~lai	lay	'rice field'
lon	lwan	'pass over, be in past'
liw~liuw	liu	'to want, wish'
hlu	hlu	'to donate'
su	su	'he, she'
sim	sim	'secluded place in monasteries for monks'
hu	hu	In OBur 'to say'; in MBur, marks the end of quotations

Several lexical complexes were widely used in the Old Burmese inscriptions. In Modern Burmese the components of some of them can function independently, but some are bound forms with no independent meaning, and their origin cannot be traced.

ariy aram	are aram	in OBur: 'a protection, care'; in MBur not used as a compound. The component *are* means 'matter, case'; *aram* is given by Benedict in the form *ram* 'to surround' (1972: 68, #313).
ahnip acak	ahnip acak	'a torture' (from *hnip cak* 'to torture); *hnip* is among Benedict's roots with the meaning 'crush, put down'(1972: 83, #400), but *cak* is not considered by Benedict. It means 'to scratch'.

ok mi	ok me	'to remember, bear in mind'. The components cannot be used independently. Their meaning is unknown, yet the complex does not seem to be a loan.
khi paŋ	khyi paŋ	'to extoll, lift, raise'. Both components have the same meaning the complex has.
hŋan chay	hñañ chay	'to hurt, oppress'. The components are bound forms with the meaning of the whole complex.
taŋ tay	taŋ tay	'to appear well, be comely'. The component *taŋ* has the same meaning as the whole complex; the meaning of *tay* unknown.
taw hlan	to hlan	'to rebel, rise up'. The meaning of the components is unknown.
thaŋ hra	thaŋ hra	'be famous, glorious', *thaŋ* 'be visible', *hra* 'be rare'.
pyak ci	pyak ci	'be destroyed, ruined', *pyak* 'be destroyed'; *ci* is a bound form with no known meaning.
phyak chi	phyak chi	'to destroy, ruin'
miy lyaw	me lyo	'to forget, leave without care', *me* 'to forget', *lyo* 'be reduced'.
lu yak	lu yak	'take by force, snatch', *lu* 'take by force', *yak* 'be rough'.
lu ca	lu ca	*lu*—'take by force'; *ca* can mean 'to eat' or 'to pick a quarrel', depending on the tone. Since with this word the tone is never marked in inscriptions, the general meaning of the compound is either 'to take by force' or 'to take by force and eat (use)'.

References

Benedict, Paul K. 1972. *Sino-Tibetan: A Conspectus*. Cambridge University Press.

Hla Pe 1967. A Tentative List of Mon Loanwords in Burmese. *JBRS*, L.1: 71-94.

Yanson, Rudolf A. 1990. *Voprosï fonologii drevnebirmanskogo yazika* ['Phonological problems of Old Burmese']. Moscow: Nauka.

This page intentionally left blank

A GLOSSARY OF EARLY CLASSICAL TIBETO-BURMAN NEWARI

KASHINATH TAMOT (TRIBHUVAN UNIVERSITY)

ABBREVIATIONS OF SOURCES, CHRONOLOGICALLY ARRANGED

P235	Palmleaf deed of	NS 235 (AD 1114)
P405	Palmleaf of	NS 405 (AD 1285)
P420	Palmleaf of	NS 420 (AD 1300)
I454	Copperplate inscription of	NS 454 (AD 1334)
P483	Palmleaf of	NS 484 (AD 1363)
H	*Haramekhalâ* of	NS 494 (AD 1374)
N	*Nâradasmṛti* of	NS 500 (AD 1380)
A1	*Amarakośa* 1 of	NS 501 (AD 1381)
A2	*Amarakośa* 2 of	NS 506 (AD 1386)
A2a	*Amarakośa* 2a of *circa*	NS 520 (AD 1400)
Gv	*Gopâlarâjavaṃśâvalî* of	NS 508 (AD 1388)
P511	Palmleaf of	NS 511 (AD 1391)
I543	Copperplate inscription of	NS 543 (AD 1423)
P546	Palmleaf of	NS 550 (AD 1430)
A3	*Amarakośa* 3 of	NS 550 (AD 1430)
P565	Palmleaf of	NS 565 (AD 1445)
P575	Palmleaf of	NS 575 (AD 1455)
A4	*Amarakośa* 4 of	NS 591 (AD 1471)

ECNEWARI	CATEGORY	GLOSS	SOURCE
û	adj.	the same	N65a.4
ekâ	n.	black mustard seed	H2b.4; Gv30b.5
o	pro.	that	Gv37b.3
osa	pro.	his/her	N12a.1
kasatî	n.	honey	H6a.4
kâm-	rt.	tell, inform	-
kâṃ-ŋa	v.	hearing (*kaṃ-ŋa taya* = *śruta*)	N14a.4
kâṃ-ja	v.t.	to tell, inform	N66a.1
kâṃ-ña	v. inf.	to tell, inform	N66a.1
kâ-ñe	v.t.	to tell, inform	N56a.3
kâya	n.	son	Gv43a.4
kâ-	rt.	take, occupy	-

kâ-yâ	v.	taken, occupied	Gv34a.3
kâ-sa-na	-	taking	Gv46b.2
kâ-syaṃ	-	taking	Gv52a.2
kâ-yu	v.	will take	N64a.3
kâ-ye	v.t.	to take	N11b.5
kâ-va	adj.	one who takes	N20a.1
kiñja-	n.	younger brother	Gv61a.1
kejâ-	var. of	*kiñja*	Gv63b.3
kisi	n.	elephant	N57b.4
kâ-syaṃ-	v.	slitting	Gv41b.1
kuṭa	n.	piece	H6a.4
ke	n.	grain	Gv60a.1
ke-	rt.	throw	-
ke-kva	v.	thrown to	Gv56a.4
ko-	adv.	down, low	-
kotha	adv.	downside	Gv57b.2
kokhâyake	v.c.	to cause to hang	N54a.3
ko-	rt.	irrigate	-
ko-le	v.t.	to irrigate	N55b.1
ko-nhu	adv.	particular day	Gv52b.3
ku-nhu	var. of	*konhu*	Gv62b.2
ko-nho	var. of	*konhu*	H2a.4
kau-nho	var. of	*konhu*	H1a.4
kyaṃ-ŋâ	v.	shown	Gv30b.4
kyaṃ-ñe	v.t.	liable to be (fined)	N55a.4
kyaṃ-baṃ	n.	garden (lit., 'vegetable land')	P511
kyeṃ-ñe	v.t.	to show	N14a.3
kye	n.	pulse, bean	Gv58b.2
khaṃ	n.	matter, fact, talk	N15b.5
khaṃ-gva	v.	seen	N17a.3
khaṃ-gva	adj.	one who has seen	N15a.5
khaṃ-nâ-va	v.	seeing	Gv55b.1
khaṃ-ja	v.t.	to be seen	N19a.2
khava	adj.	left	H1b.4
khâ	n.	fowl	Gv48a.2
khâ	n.	door	Gv50b.5
khâ-kva	v.	trembled	N56a.4
khâ-ŋa	v.	observing, seeing	N65a.2
khâ-yâ	v.	shaved	N24b.3
khâ-syaṃ	v.	hanging	Gv56b.2
khi-câ	n.	dog	H1a.1
khi-ti	n.	dirt	H3a.2
khuṃ	n.	thief	N17a.1; Gv56b.5
khum-ñe	v.t.	to boil, cook	Gv30b.5

khu	num.	six	N11b.2; Gv59a.2
khu-nhu	num.	six days	Gv33b.2
khu-yâ	v.	stolen	Gv56b.4
khu-ra	v.inf.	to steal	N17a.1
khu-syaṃ	v.	stealing	N69b.2
kho	n.	stream, river	N53b.1; Gv61b.2
kho-ca	v.t.	to tear (paper)	I543.11
khobî	n.	tears	Gv44a.4
kho-ya-kaṃ	v.	causing to weep	Gv58a.2
kholâ	n.	bark	H72b.3
khau	var. of	*kho*	N55b.1
khyâ-ca	v.t.	to threaten	N22b.4
khyâ-ta	v.inf.	to chase	N57b.4
khva-pvâ	n.	water/rain and hailstone	Gv44a.4
khvâla	n.	face	N68a.4
gaṃ-gva	adj.	dried	N20a.4
gaṃ-ja-kâ-vo	v.	drying	H6a.4
gaṃ-ja-ke	v.c.	to make dry	N22a.1
ga-thyaṃ	adv.	how	N60a.4
gâ	n.	garment	N21b.5
gâ-kva	v.	sufficed	N63a.4
gâ-kva	adj.	darkened, dark fortnight	H2b.4; Gv29b.1
gâ-kva	v.	rained	Gv30b.2
gâ-ye	v.t.	to cross	N58a.4
guṃ	num.	nine	Gv50b.5
gu-nhu	-	nine days	Gv29b.3
guṃ	n.	forest, mountain	Gv36b.5
gai-va	adj.	ferocious	N20a.2
go-kâle	n.	while torn	N15a.4
go-sâla	n.	plan	P511
gva-tai	pro.	which	N59a.3
gva-na	pro.	anyone	N16a.5
ŋaṃ	num.	five	Gv40b.1
ŋaṃ-nhu	num.	five days	Gv46a.3
ŋaṃ-mha	-	five persons	Gv43a.5
ŋatârha	n.	cheek	N62b.1
ŋâ	n.	fish	H1b.4
ŋâ	num.	five	P235.2
ŋâ-ye	v.i.	to walk	N62b.1
ŋâ-syaṃ	v.	walking	50b.1
ŋeṃ-gva	adj.	what is heard	N15b.1
ŋe-gva	adj.	what has been heard	N15a.5

ŋe-	rt.	twist	-
ŋe-la	v.	twisting	H3b.4
ŋai-yu	num.	seventy (seven times ten)	P235.3
ŋoya	adj.	mad	H1a.1
ŋhâ	adv.	before	N11b.2
ŋhâ-tho	adv.	former, earlier	N63b.4
cacâ	adj.	small	H6a.4
calati	n.	perspiration, sweat	N22a.1; Gv44a.4
calasa	var. of	*cvalasa*	Gv48a.2
ca-saṃ	n.	hair	N18b.2
câ	adv.	night	Gv29b.2
câ-nasa	adv.	night	Gv55a.1
câ	n.	child, offspring	N58b.4; Gv48a.2
câ-	rt.	wash	-
câ-yâ	v.	washed	Gv56a.4
câ-syaṃ	-	washing	Gv56b.2
câ-	rt.	open, penetrate	-
câ-la-kaṃ	v.	opening, penetrating	Gv54b.5
câ-la-kâ	-	caused to open	Gv41b.4
câ-lya-kâ	-	caused to open	Gv33b.4
câ-lva	-	opened, penetrated	Gv33a.4
ci-	rt.	pave	-
ci-ŋâ	v.	paved	Gv63a.5
ci	n.	salt	N20a.4; Gv58b.2
cepa	n.	act of eating	Gv58b.5
ce-	rt.	roof, pave	-
ce-yâ	v.	roofed	Gv56b.1
ce-syaṃ	-	roofing	Gv56a.5
ce-ye	v.	to roof	Gv62b.5
ce-le	v.t.	to use	N15a.3
ce-/cai-	rt.	tie, bind, fetter	-
ce-syaṃ	v.	binding	Gv38a.5
cai-raṃ	v.	tied, fettered	N24b.4
coṃ-	rt.	stay	-
coṃ-ko	adj.	one who stayed	P235.3
coṃ-ŋâ	v.	stayed, occupied	Gv36b.2
co	n.	urine	N62b.2
co-	rt.	fire	-
co-kva	adj.	one who fires	N20a.1
co-yâ	v.	burnt	Gv34a.1
co-/cho-	rt.	send	-
co-ya	n.	to send	Gv61b.1

co-	rt.	write	-
co-ye	v.i.	to write	N13b.1
co-sa-naṃ	v.	writing	I543.8
cyâ	num.	eight	I454.2; N11a.5
cvaṃ	rt.	stay, live	-
cvañ-ja	v.i.	to live, to stay	N11b.4
cvalasa	n.	she-goat	N57b.4
cha-naṃ	adv.	once	N66b.1
cha-nhaṃ	adv.	once	N66b.1
châ-	rt.	dare	-
châ-la-ke	v.c.	to cause to dare, in: *machâlake*	N22b.4
châ-lya	v.i.	to dare, in: *machâlya*, to be bashful	N22a.1
chiṃ-ñu	v.	be well; in: *machiṃ-ñu*, not be well	N60a.2
chî	num.	one, entire	P235.4
che	pro.	you	N22a.4
che	n.	house	N14a.2
che-	rt.	lay (brick)	-
che-ca-kâ	v.	caused to lay brick	Gv63a.5
cho	pro.	any	P484; H4b.1
cho-	rt.	fire, burn	-
cho-yâ	v.	burnt	H1a.2
cho-	rt.	send	-
cho-ye	v.t.	to send	N13a.4
chyaṃ	n.	house	N54b.4
chyaṃ	af.	agentive suffix	I454.1
chyâ-	rt.	mix	-
chyâ-ya	v.t.	to mix	H4a.4
cheṃ	n.	house	A2 35b.2
ña	rt.	eat	-
ña-ke	v.c.	to make to eat	H5b.1
java	adj.	right (hand)	H1a.1
jâ-	rt.	being friend	-
jâ-yâ	n.	friend (lit., 'be friended')	N19a.2
jâ-va	adj.	full	N62b.2
jâ-syaṃ	v.	unitedly	N69b.1
jî	num.	ten	N57b.3
ju-ko	adv.	what happened, only	N14b.2
ju-	rt.	be	-
ju-ya	v.i.	to be	I454.7
ju-roṃ	v.	was	N11a.3
ju-rvaṃ	v.	was	N11a.5

ju-va	adj.	one who becomes	N20b.3
jo-/jau-	rt.	walk, go	-
jo-ye	v.i.	to walk	P565.5
jo-va	adj.	one who goes	N18b.2
jau-ye	v.t.	to go	N20a.5
jau-va	adj.	one who goes	N18b.3
jyâ-pe	v.t.	to work	N60a.2
jva-	rt.	hold, seize, capture	-
jva-ŋa	v.	holding, seizing,capturing	P235.4
jhâ-	rt.	tire, fatigue	-
jhâ-va	adj.	one who is tired or fatigued	N20a.2
jhâ-so	n.	sighs	N21b.5
ṭuṃthi	var. of	*tuṃthi*	N55a.2
ṭe-/te-	rt.	ought, permit	-
ṭe-vu	v.	ought, permitted	P235.4
ṭo	prt.	emphatic particle	P235.3
ṭvaṃ	var. of	*tvaṃ*	P483
ḍâ-kva	adj.	all, everything	N52b.3
ḍâ-	rt.	measure	-
ḍâ-ya	v.t.	to measure	N54b.2
ḍe-	rt.	cut	-
ḍe-ñu	v.	will cut	N56b.2
ḍeṃ	adj.	times	N11a.5
ḍyaṃ	adj.	times	N11a.5
ḍvaṃ	rt.	mistake	-
ḍvaṃ-ja-ka	v.	mistaken	N15a.4
taṃ-	rt.	add	-
taṃ-ŋa	rv.	adding	P235.2
tala-nhaṃ	adj.	many times	H3b.3
ta-va	adj.	great, big	I454.2
tava-mi	adj.	courtier (lit., 'great person')	I454.2
tâṃ-	rt.	be lost	-
tâṃ-ñu	v.	lost	N60a.2
tâ	cl.	types	H5b.4
tâ-va	adj.	what is heard	N59a.5
tuṃ-thi	n.	well	N54b.4
tu	prt.	emphatic particle	N54a.2
te	v.t.	to keep	N12a.4
te-va	v.	ought to be	P511
to	adj.	till, up to	N11a.5
toyu	adj.	white	H72b.2
tau	var. of	*to*	A2 70b.2
tyâ-	rt.	borrow, to take a loan	-
tyâ-ṇâ	v.	borrowed	N11a.4
tvaṃ	prt.	emphatic particle	N12a.1

thaṃ-tho	adv.	upper	P575
thaṃ-	rt.	raise	-
tha-khai-ra	v.	may it be	N13a.1
tha-ju-ra	v.	let it be	H2b.1
tha-ma	pro.	oneself	H2b.1
tha-ma-vo	pro.	oneself	N21a.1
tha-va	pro.	one's	H5b.1
tha-va-ke	pro.	with oneself	N67b.2
thu-te	prn.	these	N57b.4
theṃ	conj.	as, like	N11a.3
the	v.t.	to distribute, to divide	N71b.2
the-the	adv.	each other	N62a.1
thai-thai	var. of	*thethe*	N17a.4
thai-lva	adj.	owner	N72a.4
tho-sâ	n.	bull	N72a.4
thau-lva	adj.	owner, one who posseses	N21a.2
thau-sâ	var. of	*thosa*	N72a.4
thyaṃ	var. of	*theṃ*	N13a.5
thvaṃ	n.	local beer	N69b.2
thva	pro.	this	I454.11
thva-te	prno	these	I454.3; H1b.4; N11a.2
thva-tai	var. of	*thvate*	N14a.1
thva-na-li	adv.	after this	N62a.1
tho-/thva-	rt.	posess, own, endow	-
tho-lva	adj.	who possessed, endowed with	N16a.4
daṃ	adv.	year	P483; N57a.1
daṃ-ko	adv.	all, entire	N59b.4
daṃ-	rt.	fill	-
daṃ-gva	v.	filled with	N62b.1
daṃ-	rt.	be eradicated	-
daṃ-ja	v.i.	to be eradicated	I543.9
da-	rt.	have	-
da-ya-ke	v.c.	to make	H1b.4
dâ	n.	fat	H3b.4
dâ-	rt.	be mistaken	-
dâ-ca-ke	v.c.	to be mistaken	N12b.1
dâ-thau	adv.	middle	P420
dî-	rt.	stop	-
dî-syaṃ	v.	stopping	N21b.5
do	v.aux.	have, is	P483; N55b.1
dau	var. of	*do*	N19a.2

dyâ-	rt.	lose	-
dyâ-kva	adj./n.	lost/loss	N13a.5
dyâŋa	n.	embankment (of a field)	N52b.1
dva-/dvâ-	rt.	have	-
dva-ya-kaṃ	v.	having being	I543.7
dva-ya-ke	v.c.	to make	N12b.2
dvâ-kâ-le	v.	while having being	I543.11
dvâ-ko	adj.	whoever, all	N16b.1
dvâ-lyaṃ	v.	while having been	I454.6
dhara	n.	canal, water course	N53a.4
dhâ-	rt.	say, tell	-
dhâ-yâ	v.	told	N21a.5
dhuṃ-	rt.	complete, finish	-
dhuṃ-gva	v.	completed	N61a.1
dhuṃ-ŋâ-va	v.	having finished	N21a.5
na	af.	ergative suffix; adverbial suffix	P235.2; I454.2
naka	adv.	at first, initially, newly	N17a.5
nanâ	adv.	quickly, soon	N66a.3
nanâ-na	var. of	*nanâ*	N65b.1
napaṃ-	postp.	with, including	N11a.4
nâ-	rt.	smell	-
nâ-ja-ke	v.c.	to make smell	H5b.3
nâ	rt.	accept, recognize	-
nâ-le	v.t.	to accept	N67a.4
niṃ-	n.	sun	A1 7b.3
niṃ-bhâra	n.	sun-light	A2 12b.1
ne	num.	two	P235.5
ne-/nai-	rt.	grind, pound	-
ne-ca-kâ-vo	v.	grinding	H1a.4
ne-yâ-vo	v.	grinding	H1b.4
nai	var. of	*ne* 'two'	H6a.2
nai-mha	adj.	two (animate)	N14b.4
nai-sa-na	v.	grinding	H2b.4
no	conj.	also	I454.b; H4b.1
no	n.	mouth	A1 32a.5
no-ca	v.t.	to start, to speak	A1 11a.2
nvâ-	rt.	scold, blame	-
nvâ-kva	adj.	one who scolds or blames	N68a.1
nhaṃ	cl.	times, in: *tala-nhaṃ*, many times	H3b.3
nhasa	var. of	*nhâsa*	A2 99a.5
nhasa	num.	seven	N57a.1
nhasa	n.	ear	A2 17b.3

nhasa-pvaṃja	n.	ear-wax	H3a.2
nhâsa	n.	nose	A2 17b.3
nhiṃḍa	n.	sleeping	N69b.2
nhi	n.	snot, mucus	H3a.2
nhî	n.	mucus, phlegm	A1 30a.3
paṃ-	rt.	obstruct, block	-
paṃ-ñe	v.t.	to obstruct	N55a.2
palakâ	n.	yellow mustard seed	Gv30b.5
pâta-lai	n.	foot	A2 70b.2
pâ(ṃ)-	rt.	get warm	-
pâ-ñe	v.t.	to get warm	H1a.2
pâ-pha-ke	v.c.	to take an oath	N22b.1
pâ-phe	v.c.	to take an oath	N22b.4
pâ-phe-ke	v.c.	to take an oath	N22b.1
piva-laṃ	n.	crossroad (lit., 'four road')	N55a.2
pu	n.	seed	H1a.3
pû	var. of	*pu*	N72a.2
pecaḍa-va	n.	secretion from the eyes	H3a.2
pai-tâ	num.	four types	N11a.2
po-/pau-	rt.	pay	-
po-le	v.t.	to pay	P405
po-va	v.	covered	A2 70b.2
pau-	rt.	cross, jump	-
pau-la	v.inf.	to cross	N19b.5
pau-la-ke	v.c.	to make pay	N12b.2
pau-lṛ	n.	knee	N62b.1
pau-	rt.	pay, return (money)	-
pau-le	v.t.	to pay	N11b.5
pyaṃ-nu	conj.	than	N61b.2
pyaṃ (-nho)	num.	four (days)	H2b.2
pvaṃ-cirhi	adj.	naked	N24b.3
phâ	n.	boar, pig	H3b.3
phupa-kiṃja	n.	elder and younger brother	N64b.3
pheṃ-	rt.	release, unfasten	-
pheṃ-da	v.inf.	to release, unfasten	N18b.3
pho-/phvâ-	rt.	be able	-
pho	adj.	able (person)	N64b.1
pho-/phva-	rt.	beg	-
pho-ṇa-na	v.	begging	P235.4
phya-/phyâ-	rt.	retake, return	-
phya-ca	v.t.	to retake, return (land)	P405
phyâ-ca-ke	v.c.	to cause to retake	P553
phyâ-ya	rt.	to retake, in: *liphyâya*	P553
phvâ-taṃ	v.	could	N11b.4
phvâ-yu	v.	will be able	N60a.4

baṃ	n.	ground, land	N21b.5
ba-chi	num.	half, initial form of *bâ* (lit., 'half-one')	P235.2
barâ	n.	an arrow	N22b.2
bala-nhi	adv.	night	A3 12a.5
bâ	num.	half	N57a.2
bâ-câ	adv.	midnight	A2 13a.3
bâ-nhîṃ	adv.	midday	A4 4b.2
bâ-la-chi	adv.	fortnight (lit., 'half-month-one')	N63b.2
bi-	rt.	give	-
bi-mâ-lu	v.	needed to give	P235.4
bi-bola	n.	abuse, revile	N20b.3
bi-ya	v.t.	to give	P483
bi-ya-kaṃ	v.	causing to give	N68a.3
bi-yu	v.	will give	P235.2
bi-vu	v.	will give	P235.2
bi-ye	v.t.	to give	N13b.2
bi-sa-na	v.	giving	P235.3
bi-syaṃ	v.	giving	Gv40a.3
bî	n.	snake	N20a.1
bû	n.	land, field	N32b.3
be-kva-vuṃ	v.	deformed one	N68a.1
bo	n.	portion, part	N13a.1
bola	n.	abuse, revile	N20b.3
bo-	rt.	emerge	-
bo-/bva-	rt.	fly	-
bo-ya-kaṃ	v.	flying	A2 26a.6
bo-va	v.	emerged	H1a.4
byaṃ-	rt.	release	-
byaṃ-gva	v.	released	N13a.1
bvaṃ-	rt.	invite	-
bvaṃ-ŋa	v.	having invited	N69a.3
bhiṃ-	rt.	be good	-
bhiṃ-gva	adj.	good	N13b.1
bhiṃ-ñe	v.i.	to be good	N55b.3
maṃ	n.	chaff (of grain)	N53a.3
ma	af.	negative affix	I454.4
ma-khâ	prt.	is not so	N13b.2
ma-châ-la-ke	v.c.	to cause not to dare; to be ashamed	N22b.4
ma-châ-lya	v.i.	to be bashful	N22a.1
ma-do	v. aux.	is not	N20b.2
ma-dau	var. of	*mado*	N11b.1
ma-phau	v.	unable to do	N11b.5
ma-/mâ-	rt.	need	-

ma-la-ko	adv.	as much as needed, enough	P235.5; N67b.1
masa-mo	adj.	young girl	H1a.4
mâ-ma	n.	mother	N62a.1
mâ-lu	v.	needed	P235.2
mâ-sâ	n.	mother-cow	N57b.3
mi	n.	eye	H2b.1
mi-kâ-sâ	n.	woman (lit., 'male-take-nom. suffix')	A1 26b.2
mi-sâ	n.	female, woman	N11a.5
mî	n.	person, man	I454.2
mî-	rt.	sell	-
mî-ra	v.inf.	to sell	N20a.4
mî-va	adj.	seller	N20a.1
meṃ-va	pro.	others	N20b.2
me	n.	fire	H1a.2
me	n.	tongue	H1a.4
me-lâ	n.	moon-light	A2 10b.2
me-lyaṃ	adv.	elsewhere	N70a.4
me-sâ	var. of	*misâ*	H1a.2
mai	n.	fire	N14b.2
mai-sâ	var. of	*misâ*	A1 64b.5
moṃ-	rt.	speak	-
moṃ-ŋa	v.	speaking	N67b.4
moṃ-câ	n.	child	N65b.1
mo-	rt.	lose, destroy	-
mo-kva	v.	lost, destroyed	N14b.3
mo-câ	var. of	*moṃcâ*	P235.3
mau-pa	n.	maternal uncle	N64b.4
myaṃ-va	var. of	*meṃva*	N11b.4
mya-sâ	var. of	*misâ*	N16b.2
mvaṃ-câ	var. of	*moṃcâ*	H1a.4
mhaṃ	n.	body, person	H1b.2
mhaca-moṃ	n.	wife	P235.3
mhaca-mo	var. of	*mhacamoṃ*	N64a.3
mhâca	var. of	*mhyâca*	N67b.3
mhe-/mhai-	rt.	sow	-
mhe-yâ-va	v.	sowing	N72a.2
mhai-kvo	adj.	one who sows seeds	N72a.3
mhoya	n.	tenant	N72b.2
mhyâca	n.	daughter	A2 55b.1
mhyâca-moṃ -câ	n.	daughter	N68a.1
mho-	rt.	erase	-
mho-kâ-le	v.	while erased	N15a.4

yaṃ	adv.	north	A2 9a.2
yaṃ-kuli	adv.	north-west	A1 6b.7
yaṃ-/ye-	rt.	take	-
yaṃ-gva	v.	taken	I454.6; Gv38a.5
yaṃ-ja	v.t.	to take	P553
yaṃ-tâ	adv.	northside	A1 6b.7
ya-/yâ-	rt.	do, work	-
ya-ca	var. of	*yâca*	A1 36b.3
yâ	af.	genitive suffix	P235.1
yâ-ke	af.	comitative suffix	I454.4
yâ-kva	v.	done	I454.8; N15b.5
yâ-ŋa	v.	doing	I454.3; Gv46a.5
yâ-ca	v.t.	to do, work	N15a.1; I543.6
yâ-ca-kâ	v.	caused to do	Gv58b.5
yâ-ca-ke	v.c.	to cause to do	N13a.4
yâ-ye	v.t.	to do	N17a.5
yâ-va	adj.	doer	N62b.2
yi-kuli	adv.	south-east	A3 8b.2
yi-nu-ko	var. of	*yuniko*	A2 13a.2
yu-kuli	adv.	south-west	A3 8b.3
yu-ni	adv.	afternoon	A1 7b.6
yu-ni-ko	adv.	afternoon	A3 12b.5
ye-kuli	adv.	south-east	P483; A45a.1
ye-kva	adj.	a lot, very much	N15b.5
ye-ja	var. of	*yaṃja*	P484
ye-tâ	adv.	southside	P565
ye-ne	var. of	*yaṃja*	P585
ye-	rt.	like, tend toward	-
ye-ye	v.t.	to like	-
ye-va	adj.	one who tends/likes/desires to	N18b.4
yela	n.	spit	H4a.4
yo-kuli	adv.	south-west	A1 6b.7
yo-tâ	adv.	westside	A3 8b.3
yau	adv.	west	A2 35b.5
yau-thva	adv.	westside	P546
laṃ	n.	way, road	N58a.4
laṃ-khu	n.	street	N55a.2
laṃ-khva	n.	water	H2b.2
laṃ-ñe	v.t.	to weigh	A2 101b.4
laŋa	n.	upper garment	N61a.1

la-/lâ-	rt.	acquire, get, receive, hold	-
la-ca	v.t.	to acquire	A4 70b.4
la-cu	v.	acquiring	A1 65b.4
la-ni	n.	moon	A1 8a.4
la-paṃ-chi	n.	fathom	A2 67a.5
lâṃ-kha	var. of	*laṃ-khva*	N55b.1
lâṃ-khva	var. of	*laṃ-khva*	N55b.1
lâ	adv.	month	Gv57a.1
lâ	n.	hand	H1a.1
lâ-kva	v.	acquired, got	I454.7
lâ-ŋa	v.	holding	Gv44a.4
lâ-ŋa-na	v.	while waiting	N63b.3
lâ-ñe	v.t.	to wait	N63b.2
lâ-hâṭha	var. of	*lâhâtha*	A2 111a.2
lâ-hâtha	n.	hand (*-hâtha* < Skt *hasta*)	A2 105b.5
li	af.	later, after	-
li-	rt.	follow, pursue, drive out	-
li-ca	v.t.	to follow, pursue	A4 61a.3
li-co	adv.	later, after	N17a.5
li-to	adv.	later	P483
li-thu	adv.	later	A2 a2b.5
li-tho	var. of	*lithu*	A1 28a.2
li-thau	var. of	*lithu*	N61b.4
li-bî	adv.	back	P446
li-va	adv.	after	Gv33b.2
li-phyâ-va	v.t.	to retake	P553
lî	var. of	*li*	N11b.2
lî-kâ-ye	v.t.	to take back	N57b.1
lî-kva	adj.	one who follows	N71a.1
lî-ye	var. of	*lica*	N61b.1
luṃ	n.	gold	P235.2; N11a.4
luṃ	n.	heart, in: *luṃgvarha*	A2 26a.3
luṃ-pe	v.t.	to wish, to like, to satisfy	A4 63b.4
luṃ-siṃ	n.	finger-nail	N20a.3
lu-daṃ-ja	v.i.	to be contented	N54b.2
lṛ	n.	leg, foot	A4 11a.4
le	var. of	*lṛ*	H5b.2
lena	n.	radish	A2 5a.6
le-pâta	n.	foot	A4 30a.4
lo-/lva-/lvâ-	rt.	fight, quarrel	-
lo-ca	v.i.	to fight, quarrel	A1 44a.3
lo-ṭha-ya	v.i.	to make quarrel	A1 39a.5
lyâseṃ	adj.	young woman	N66b.3
lva-cu	v.	quarrelled, fought	A1 44a.3

lvâ-ca	v.i.	to fight, to quarrel	A4 42a.3
lvâ-ta	v.inf.	to quarrel	N18b.2
lvâ-mî	adj.	fighter, warrior	A2 85b.4
lvâ-ye	v.i.	to fight	A4 42a.4
lhaṃ-	rt.	repair	-
lhaṃ-ŋe	v.t.	to repair	N55a.1
lhâ-	rt.	to say, talk, speak	-
lhâ-ca	v.t.	to tell	N21a.4
lhâ-ca-kaṃ	v.	causing to tell	N21b.1
lhâ-yu	v.	will tell	N21a.1
lhâ-ye	v.t.	to tell, talk	N17a.2
lhâ-ra	v.inf.	to tell, to say	N17a.2
lhu-	rt.	pay	-
lhu-ye	v.t.	to pay	N59b.1
lhvaṃŋa	adv.	again	N70a.4
vaṃ	adv.	east	A2 9a.2
vaṃ-kuli	adv.	north-east	P405; A3 8b.3
vaṃ-gu	v.	gone	P235.4
vaṃ-gva	adj.	one who goes	N58a.4
vaṃcu	adj.	blue	N20a.3
vaṃ-	rt.	go	-
vaṃ-ja	v.i.	to go	N20a.1
vaṃ-ju	v.	will go, gone	N24b.3
vaṃ-ña-tvaṃ	v.	went	N65b.3
vaṃ-ñe	v.i.	developed form of *vaṃja*	N57b.4
vaṃ-tho	adv.	eastside	P557
va-ŋâ-va	v.	going	Gv61b.2
va-ñe	var. of	*vaṃñe*	N12a.2
va-bî	adj.	the same	N64b.4
va-va	adj.	one who comes	N70b.1
vâ	n.	tooth	H1a.3
vâ	n.	paddy	P235.4
vâ-	rt.	mix	-
vâ-la	v.	mixing	H2a.4
vâ-le	v.t.	to mix	H6a.4
vâ-	rt.	plough	-
vâ-syaṃ	v.	ploughing	N60a.2
vuṃ	prt.	emphatic particle	N11a.2
vo	af.	sociative case suffix	N68a.4
voho	n.	silver	N22b.2
śâ	var. of	*sâ*	N22b.2
śi	n.	lip	N22a.1
saṃ	n.	hair	N24b.3
saṃ-/sâṃ-	rt.	move, try	-

saṃ-ñu	v.	will move, will try	N21b.5
sa	af.	locative case suffix/genitive case suffix (*usa*-his/her)	P235.2
sa-ke	af.	comitative case suffix	P235.2
sa-câ	n.	calf (lit., 'cow-child')	-
saḍaṃ	var. of	*sarhaṃ*	H6a.3
sa-na	af.	agentive case suffix	P235
sa-ni	adv.	evening	A1 7b.6
sa-ni-vela	adv.	evening time (*vela* is derived from Skt. *velâ* 'time')	A2 12b.5
sara (TB?)	num.	hundred	P235.3
sarhaṃ	n.	horse	H58a.4
sa-	rt.	learn	-
sa-va	adj.	one who has learnt	N62b.2
sâṃ-gva	adj.	one who tries to do something	N68a.4
sâ	n.	cow	A1 48b.4; A166b.6
sâlu	adj.	thin	N62b.1
siṃ	n.	wood, tree	N53a.5
si-	rt.	rinse, wash	-
si-ca	v.t.	to rinse, wash	-
si-ye	-	developed form of *sica*	A2 77a.1
sîṃ	var. of	*siṃ*	N20a.4
suṃ-ḍaṃ	adv.	morning	A3 106b.2
su	n.	cloud	A1 7a.2
su	n.	hay	A3 33b.1
su	pro.	who, anyone	P235.4
su-na	pro.	anybody	P235.4
su-ni-tyaṃ	adv.	morning	A2 12b.4
seṃ-/syaṃ-	rt.	destroy, violate, spoil, damage	-
seṃ-ña-ke	-	developed form of *syaṃjake*	N61a.2
se-ŋa	v.	destroying, violating, spoiling	P235.4
se-/sya-/sai-	rt.	know, recognize	-
se-ye	v.t.	to know	N52b.2
se-va	adj.	what is known	N15a.5
se-hva	v.imp.	know	N21a.1
sai-ko	adj.	what is known	N22a.5
sai-ye	v.t.	to recognize	N53b.1
sai-raṃ	v.	knew	N17a.5
sai-hvana	v.imp.	be known (hon.)	N19a.1
so-/sva-	rt.	see, view, look	-
so-kva	adj.	viewer, one who has seen	N15a.5
so-ya	v.t.	to look	N22a.1
sau	n.	oil	N11a.5
stâ-kva	v.	fixed	I454.3

syaṃ-ja-ke	v.c.	to cause to destroy	N17a.1
syâ-	rt.	kill	-
syâ-ŋâ	v.	killed	Gv41b.1
syâ-ta	v.inf.	to kill	N20a.3
svaṃ-ḍyaṃ	num.	three times	N11a.5
haṃ-/ha-	rt.	bring	-
haṃ-ñe	v.t.	to bring	N64b.3
ha-kâ-le	v.	while brought	N61b.3
ha-yâ	v.	brought	Gv44a.4
hala	n.	leaf	H61.2
hâ	n.	root	H1a.4
hâ-	rt.	pour	-
hâ-ya-ke	v.c.	to pour water	P511
hi-puḍi	n.	bile	H1b.4
hî	n.	blood	H3b.3
he	v.t.	to bring	N12a.3
heŋâla	n.	charcoal	H1a.2
heŋau	adj.	red	H2b.3
hyâŋu	var. of	*heŋau*	H72b.3; N54a.3
hvaṃ-ŋâ	v.	wedded	N70a.4

A PRELIMINARY GLOSSARY OF TANGUT FROM THE TIBETAN TRANSCRIPTIONS

KSENIA B. KEPPING (RUSSIAN ACADEMY OF SCIENCES, ST. PETERSBURG)
AND
CHRISTOPHER I. BECKWITH (INDIANA UNIVERSITY)

The phonology of Tangut, a language written in a complex native script modeled on the Chinese, has been much studied but is still problematic. The existence of a small number of Tangut texts with Tibetan interlinear transcriptions, most of which are preserved in the library of the Institute of Oriental Studies of the Russian Academy of Sciences in St. Petersburg, has long been noted (Nevsky 1926) and, *contra* Nevsky, their importance stressed (Wolfenden 1931), but a definitive reading of these texts (most of which are part of a Tantric Buddhist work) remains to be done. The present glossary is the result of a joint effort of the authors, who were able to go through, rather quickly, only a few pages of the manuscripts. It is not, therefore, by any means a finished work, but it is hoped that it will be of use for Tibeto-Burman comparative-historical linguistics until a more complete study (cf. van Driem and Kepping 1991) can be done. No attempt has been made to determine the etymology of the Tangut words, but obvious loans from Sanskrit and Chinese (mostly Buddhist terms) have been excluded. It is to be noted that the 'subscribed' letter *y* of the Tibetan script is here transcribed as (^{y}) to distinguish it from the 'base-letter' *y*, and the aspiration inherent in the 'aspirated series' of consonants in the Tibetan script is transcribed as (^{h}) to distinguish it from the subscribed letter *h*, which is part of a digraph that apparently records velar fricatives—thus, for example, *gha* is undoubtedly a transcription of [ɣa].

TANGUT	GLOSS
bgu	'center'
bgu	'head'
bg^{y}i	'say; word'
bŋifi	'five'
bts^{h}afi	'to heat'

bu, bufi	'fire; to fire' (cf. *dmu*)
c^{h}ifi [ʧ^{h}ifi]	'six'
dgafi	'ten'
dkafi	'inside'
dk^{h}i	'to do, become'
dmefi	'girl, young woman'
dmi, dmifi	'eye'
dmi	negative
dmi	'room, house'
dmu	'fire'
dŋafi	'sky'
du	'by'
gha	'at, in, on' (postp.)
gha	'what'
gla	'false, mistake, lie; falsely'
gle	'great' (cf. *lefi* 'big')
gñafi	'seven'
gne, ne	'heart'
gnifi	'two'
gnofi	'spokes (of a wheel)'
gsofi	'three'
gye	'self'
gyi	'say'
gze	'when, if' (subord. conj.)
gzi	'water'
gzifi	'all'
fibhi	'low, below'
fidzi	'to teach, teacher'
fidzo	'person, man'
figifi, dgifi	'nine'
fig^{y}i	'one'
fijafi [fidʑafi]	'to start a fire'
jefi [dʑefi]	'wheel (cakra)'
k^{h}ha	'in' (postpos.)
k^{h}ha	'inside'
k^{h}o	'exist, have'
k^{h}o	'meaning'
ldi	'and'
ldi	'four' (cf. *zlafi*)
ldi	'to know, remember; to think'

lefi	'big' (cf. *gle* 'great')
lhe	'to get, receive'
lho	'to come out'
li, lifi	'wind'
lufi	'body'
mu	'move'
mu, dmu	'kind, variety'
ni	'arrive'
ni	plural marker
ni	'sun, day'
ŋu	'copula'
p^hi	causative (postp.)
p^ho	'above'
re	'pleasure'
rko	'throat'
rnu	'destroy'
śe	'to achieve, become'
śi	'previously'
śo	'to go'
ta	topic marker (postp.)
te	'center, navel'
t^ha	'he, that (dem. pro.)'
t^hi	'this (dem. pro.)'
ye	genitive marker
yi	'again'
yu	'to seek, look for'
ze	'knowledge'
zi	'trees'
zi, zifi	'all'
zlafi	'four'
zlifi	'arrow'

References

van Driem, George and Ksenija B. Kepping 1991. The Tibetan transcriptions of Tangut (Hsi-hsia) ideograms. *Linguistics of the Tibeto-Burman Area* 14.1: 117-128.

Nevsky, Nicolas (=Nevskiy, Nikolay Aleksandrovič) 1926. A brief manual of the Si-Hia characters with Tibetan transcriptions. *Research review of the Osaka Asiatic Society*, No. 4.

Wolfenden, Stuart N. 1931. On the Tibetan transcriptions of Si-Hia words. *JRAS*: 47-52.

This page intentionally left blank

A GLOSSARY OF 39 BASIC WORDS IN ARCHAIC AND MODERN MEITHEI

SHOBHANA CHELLIAH (UNIVERSITY OF NORTH TEXAS)
AND
SOHINI RAY (HARVARD UNIVERSITY)

We have compiled this comparative list of basic Archaic and Modern Meithei words from the publications of N.K. Singh (1964 and 1975) and from three manuscripts that we have partially translated. As can be seen there is little difference in these basic words between the Modern Meithei forms and the Meithei forms found in the manuscripts. As tones are not indicated in Archaic Meithei, they have been omitted throughout.

ABBREVIATIONS

ALG: N.K. Singh, *Ariba Manipuri longei.* Imphal, 1978.
MME: N.K. Singh, *Manipuri to Manipuri & English dictionary.* 1964
LP: the manuscript *Langpum* (a creation myth)
PB: the manuscript *Pombirol* ('The knowledge of birds')
CN: the manuscript *Chainaron* ('The art of war')

SOURCE	ARCHAIC MEITHEI	MODERN MEITHEI	GLOSS
LP	khoy	khoy	bee
PB	hak	sa	body[1]
LP	pok	pok	born
LP	ŋaŋ	əŋaŋ	child
ALG	cəŋ	hiŋ	cold
LP	si	si	die
MME	ləy	ləybak	earth
LP	mit	mit	eye

[1] *hak* occurs in Modern Meithei words to denote proximity, as in *məhak* 's/he (here).'

CR	pənthaw	məpa	father
CR	khambi	məy	fire
LP	məy	məy	fire
MME	məkiŋ	mətu	fur, thatch
LP	khut	khut	hand
LP	lu	kok	head[2]
PB	ciŋ	ciŋ	hill
LP	khəŋ	khəŋ	know
MME	ləm	ləybak	land[3]
LP	mi	mi	man, person
MME	oŋ	luhoŋ	marry
MME	unbi	puknıŋ	mind
CR	palen	ima	mother
PB	miŋ	miŋ	name
LP	khoy	khoy	navel
LP	siŋli	siŋli	nerve
LP	thap	hap	place
MME	məl	məl	price
LP	u	u	see
PB	til	til	saliva
MME	khum	cəphu məca	small pot[4]
CR	pari	ica nupa	son
LP	nuŋ	nuŋ	stone
PB	ya	ya	tooth
MME	pe	catin	umbrella[5]
LP	ceŋ	ceŋ	uncooked rice
LP	isiŋ	isiŋ	water
LP	nurabi	nupi	woman
MME	kum	kumsi	year[6]
MME	məki	məri	thread[7]

[2] *lu* is still used in traditional oral literature.

[3] *ləmbi* is used to mean 'path'.

[4] The stem *khum* occurs in MMei: e.g. *cakkhum* 'kitchen', where *cak* is 'cooked rice'.

[5] *pe* is recognized but not used; *catin* is borrowed from Bengali.

[6] *cəhi* is used more commonly.

[7] *ləŋ* is used as a general term. *məri* is most often used to refer to silk thread.