Кирилл Үсэгтэй ЛаTEX

MonTEX
Mongolian and Manju
for
LaTEX 2ε

Implementation Level IVu
v. IVu.04.092

Oliver Corff[1]   Dorjpalam Dorj[2]

2002/07/01

[1]The author received a DFG (Deutsche Forschungsgemeinschaft) postdoctoral grant (Habilitationsstipendium) while developing the first version of MonTEX and expresses his sincere gratitude.

[2]Institute for Language and Literature, Mongolian Academy of Sciences, Ulaanbaatar.

Abstract

MonTEX is now released in Implementation Level IVu implying that there is not only Cyrillic Support for Modern Mongolian (Xalx dialect), Buryat and Russian; this version also includes comprehensive support for Mongolian Script (also known as Uighur Script) and Manju. All writings can be mixed freely within the same document, and within the same page.

The present release (IVu.04.092) is still very early; expect bugs and irregular behaviour. Especially the Mongolian full page mode still shows occasional quirks, depending on the document class loaded.

Contents

List of Tables

List of Figures

Part I

MonTeX: System Overview

Chapter 1

How to read and use the MonTEX documentation

According to your specific needs, you can read various parts of this documentation separately.

First-time users interested in generating *Cyrillic-only* documents can confine themselves to the sections beginning on part II, page 25.

First-time users primarily interested in creating *Mongol Bicig* and *Manju* documents or text insertions in these languages can directly jump to section III, page 39.

Advanced users interested in improved *Mongolian* and *Manju* display functions should directly go to section 11, page 59.

Advanced users already familiar with the majority of MonTEX functions can refer to the Alphabetic Command Reference beginning on page 84.

Nota Bene: Please note that MonTEX includes the complete functionality of ManjuTEX which is hereby declared obsolete. ManjuTEX will not longer be supported.

Chapter 2

System Features

2.1 Scope of MonTEX

MonTEX is a package which offers support for writing documents in Mongolian, Manju, Buryat and Russian. Mongolian can be represented in traditional Uighur script (also known as Classical or Traditional Script) and Cyrillic. Manju resembles the Traditional Mongolian script (from which it is derived) but uses a rich choice of diacritics in order to eliminate numerous ambiguities of the Mongolian script ancestor. Modern Buryat, like Mongolian in its present form, is written with a Cyrillic alphabet, but both Mongolian (35 letters) and Buryat (36 letters) use more letters than Russian (33 letters).

2.1.1 Mongolian

The word *Mongolian* is actually an umbrella term for several languages rather than the precise name of a single language. Things become more complicated when names of ethnic groups, languages and writing systems are mixed.

Xalx or Khalkha is the name of the Mongolian nationality residing in Mongolia proper. Their dialect forms the basis of Mongolian written with Cyrillic letters. Throughout this text, *Modern Mongolian* is used as a synonym.

Buryat is the name of the Mongolian nationality residing in Buryatia, north of Mongolia, east of Lake Baikal, being a part of the Russian Federation. The Buryat call themselves *Buryaad* while Xalx Mongolians call them *Buriad*. The English name follows the Russian orthography. Linguistically, Xalx and Buryat Mongol are fairly close languages; Buryat has a slightly different sound system in which the phoneme /s/ partially shifted to /h/; the modern Buryat Cyrillic alphabet (virtually identical with the Cyrillic alphabet used for writing Modern Mongolian) has one additional letter (H/h, һ/h) for marking the difference to /s/.

Bicig (literally *script* in Mongolian) denotes text written in the traditional Mongolian script which is also referred to as Uighur. Throughout this document, the term *Bicig* will be used on an equal footing with *Classical* and *Traditional* Mongolian. The latter term is used in the name of the Unicode/ISO10646 character plane U1800 which contains Mongolian, Manju, Sibe and sets of special characters called Ali Gali or Galig. In order to identify Mongolian script related commands distinct for Mongolian and Manju, the Mongolian commands have the name root `bicig` whereas the Manju commands have the name root `bithe`.

Xalx Mongolian, or Modern Colloquial Mongolian, is about as different from the form written in Classical script as modern English in phonetical spelling (assume it be written in Shavian letters) from the highly historical orthography of Standard English. Beyond these differences, Mongolian written in Classical Script usually preserves a substantial amount of historical grammatical features which make it look a bit like Elizabethan English.

2.1.2 Manju

Manju is a Tungusic language closely related to Mongolian. Though Manju is virtually not spoken anymore, it has been the official language during 300 years of Manju government in Qing Dynasty China. Vast amounts of official documents survive, as well as some of the finest multilingual dictionaries ever compiled, e. g. the Pentaglot, or Mirror in Five Languages, a dictionary with 18671 entries in five languages (Manju, Tibetan, Mongolian, Uighur and Chinese). Manju writing is derived from Uighur Mongolian by adding diacritics in the form of dots and circles.

2.2 MonTEX Implementation Levels

During several years of developing MonTEX, the desired capabilities of a software package serving the described scope were classified and implemented along four Implementation Levels which have the following, well-defined properties:

Implementation Level	Features
I	Modern Mongolian in Cyrillic Script and Buryat
II	I and Mongolian script in LR mode horizontal
IIa	II and Mongolian script portions in LR mode vertical
III	II and Mongolian script text in horizontal RL mode
IV	III and Mongolian script text printed vertically

Implementation Level I is good for producing documents in Modern Mongolian; Implementation Level II adds support for words and lines of Classical

Mongolian embedded in other (Cyrillic and Latin) scripts which is essentially useful for dictionaries etc.; Implementation Level IIa allows single words to be placed in vertical capsules; Implementation Level III allows the composition of purely Classical Mongolian documents while Implementation Level IV finally allows the combination of both scripts in freely assignable quantities and locations within the document.

Mongolian linguistic culture provides a perplexing richness of writing systems of varying regional, historical and socio-political importance. Developing a Mongolian system which covers *all* documented writing systems is tantamount to writing a Mongolian Babel system and cannot be done in a truly elegant manner with respect to the current LaTeX 2_ε limitations.

2.3 Requirements and Limitations

In order to run MonTeX a recent version of LaTeX 2_ε is necessary. MonTeX relies on the NFSS font selection mechanism and the ligature capacities of Metafont. This package has not been tested under LaTeX2.09 and will most certainly not function satisfactorily under that environment. Depending on the implementation level, further software support becomes necessary since not all features can be realized smoothly in LaTeX 2_ε alone.

Implementation Level	Requirements
I	A working LaTeX 2_ε system
II	A working LaTeX 2_ε system
IIa	like II, plus PostScript support
III	like II, plus functional `TeX-XeT` system
IV	like III, plus PostScript support

The eTeX (available for DOS and UNIX based computers alike) system provides full Right-to-Left writing support; eTeX and eLaTeX 2_ε are part of all modern TeX implementations for the majority of operating systems[1].

PostScript support is a standard feature of most UNIX installations and is also supplied with most of the available TeX for Windows distributions[2].

A word of warning is necessary here. MonTeX is not a small, convenient system which can be used without any effort. Much like its very foundation LaTeX it requires some willingness to study a few (and indeed simple) rules; occasionally one or the other old habit has to be overcome. The reward is text typeset in excellent quality so that scholarly achievement no longer disappears in badly typeset documents.

[1]It is also possible to build an eTeX system from scratch using the `web2c` (or teTeX) sources, replacing `tex.web` with `tex-xet.web` and `tex.ch` with `tex-xet.ch`.

[2]The authors used Linux (Red Hat 4.2 through 7.1) systems for the developing work; on the same hardware, PostScript under Windows*xx* is significantly slower than under Linux; this holds true for document and font compilation as well.

```
$ elatex montex.tex     # Compile document
$ dvips montex.dvi      # Create PostScript out of DVI
$ gv montex.ps          # Preview document
```

Figure 2.1: PostScript Compilation and Preview Cycle

2.4 PostScript Support

PostScript is used for creating vertical capsules of text within horizontal text for MonTEX implementation levels IIa as well as for complete pages with implementation level IV. This requires the presence of the `rotating` package for LATEX which itself relies on the `graphics` package. The `rotating` and `graphics` packages come with teTeX but do not come with emtex. They can be found at CTAN.

Besides the above-mentioned packages it is necessary that the generated `.dvi` files can be processed further, e. g. by `dvips` which generates a PostScript file out of a `.dvi` file. If there is no PostScript printer at your site, PostScript emulation is necessary which is usually provided by GhostScript and GhostView. Implementations of these systems are available for a large number of operating systems and can also be found at CTAN. Linux, a free UNIX system, comes which GhostScript, and the `winemtex` distribution of LATEX includes GhostScript as well. A sample command sequence to produce and preview a document with these utilities can be found in illustration 2.1.

Without PostScript support, only implementation level II can be realized (instead of IIa and IV): Mongolian script can be printed horizontally but not vertically. It must be noted here that most DVI viewers are *not* capable of presenting vertical text correctly; the conversion step from DVI to PostScript is virtually always necessary.

2.5 PDF Support

With the arrival of PDFTEX it is possible to generate a PDF (Adobe's Portable Document Format, that is) directly from the `.tex` sources without going through the `.dvi` stage. All systems offering `pdfelatex` can be used to compile a MonTEX Implementation Level IV document. provided the necessary PostScript is installed. PDF is the recommended form of output on systems without PostScript views and printers (like, unfortunately, most of the Windows*xx* world). PDF documents reproduce everything MonTEX generates as is, and with Type1 fonts for Classical Mongolian, the on-screen display of Classical Mongolian and Manju material is fast and pleasant.

A sample command sequence to produce and preview a document with PDF output can be found in illustration 2.2.

It is recommended to users with menu-driven environments (WinEdt, TeXnicCenter, TeXshell etc.) to set the compilation commands to the effect

```
$ pdfelatex montex.tex      # Compile document
$ acroread montex.pdf       # View PDF Document with Acrobat Reader
```

Figure 2.2: PDF Compilation and Preview Cycle

that `pdfelatex` is invoked as the default compilation engine, and the Acrobat Reader is invoked as the default viewer. Please consult the software documentation of these products for the necessary steps and procedures.

Chapter 3

Acknowledgements

The authors wish to thank the creators of TEX and LATEX as well as the designers of the existing fonts for their generosity of providing the world with such inspiring pieces of software. The packages from which pieces of code originated by inspiration or blunt copy are far too numerous; the Russian captions were taken from the file `russian.sty` (as were the English captions), most of the Cyrillic letters were produced with fonts by Nana Glonti and Alexander Samarin; additional letters were taken from J. Knappen's font files. Special thanks go to David Carlisle who offered the solution for a serious problem with the ligature mechanism in TEX. During the development of Implementation Level IV, important suggestions came from David Kastrup, Robin Fairbairns, Dan Luecking, e. a. Intensive communications about Cyrillic fonts and integrating MonTEX with the LH fonts took place with Vladimir Volovich, and other problems were discussed with Werner Lemberg.

Among the friends and colleagues in Mongolia and Germany who offered information, support and encouragement the authors wish to name B. Nerguy, Urgamal, M. Balk, Q. Änxzayaa and K. Maezono (without implication of any particular order or precedence). They contributed test data as well as their ideas for encoding, font shapes, user interfaces, and, last but not least, were patient alpha testers who helped the authors with numerous problem reports.

Many of the improvements between version 0.1 and the present version are not actually improvements; they are simply eliminations of partially awful bugs as well as ugly hacks (rather than *code*) and aim to make this package simply usable (if not useful).

3.1 Sources of Code and Inspiration

Some Cyrillic packages have been available for a few years. All Cyrillic packages available for TEX/LATEX stem from one of two lines of ancestry:

- Fonts developed at the University of Washington
- Fonts by Nana Glonti and Alexander Samarin

The two lines differ substantially in scope of characters and printing quality. The University of Washington series in OT2 encoding has a broad support for East European languages, but the praise for printing quality is given to the characters designed by Nana Glonti and Alexander Samarin. The Glonti/Samarin line of characters has undergone numerous minor modifications, not so much in glyph shapes but basically in determining encoding slots. Fortunately, the fonts are set up in a way that allows for convenient redefining of individual code positions.

Only in 1999, a comprehensive set of Cyrillic glyphs in various encodings called LH was finally implemented as standard Cyrillic support for LaTeX 2_ε, but at that time it was decided that for the time being MonTeX will continue to offer its own Cyrillic font set, for which there are mainly three reasons:

1. The MonTeX set has a seven-bit basis and allows for the all-Latin, all-ASCII communication of Mongolian texts, while Mongolian hyphenation is active—a matching LH encoding is not yet established;
2. MonTeX offers a Mongolian currency sign ₮ which is not yet included in the LH fonts;
3. Mongolia, on of the prime markets for MonTeX, continues to use partially outdated LaTeX 2_ε installations.

Neither original line of Cyrillic characters offers the additional characters necessary for writing non-Slavic languages like Mongolian; already in the beginning of the 1990s, Jörg Knappen filled the gap and designed additional letters which were intended to be used with Bashkir, a Turkic language. In fact, most of the letter forms employed there can also be used in other non-Slavic languages used throughout Central Asia since these letters are not specific to Bashkir. Some of J. Knappen's letter forms (accidentally mostly those which are not necessary for writing Modern Mongolian) do need some refinement, and are then immediately suitable for a range of languages including Kasakh, Tuvinian etc. In the present stage of the system, only those letters used in Mongolian and Buryat are incorporated from J. Knappen's files.

After discussing the typeface issues with Mongolian specialists, the Glonti/Samarin letter forms were chosen for their superior appearance in volume text. The fonts had to be renamed; failing to do so would have resulted in unpermissible ambiguity.

One feature of the traditional Cyrillic font packages for TeX (besides their lacking support for non-Slavic languages) is the intimate relationship between input encoding and output encoding. The first step in building Mongolian support was to separate these two spheres as numerous Mongolian encodings exist which should all be supported by the Mongolian package. A new encoding was then defined (LMC — Local Mongolian Cyrillic) which is a close approximation of a transliteration based on Latin1 encoding, notably with front vowels ä, ö, ü (э, ө, ү) and ï (й) in matching positions.

The encoding is completely detached from the existing Cyrillic codepages of which there are too many; in addition it should be possible to produce Mongolian documents in 7-bit environments so as to assure maximum document portability.

An additional ligature table for Metafont was then supplied which takes care of most of the two-letter combinations necessary for entering Cyrillic since the Cyrillic alphabet has more letters (36 in the present version) than the Latin alphabet which prohibited any 1:1-mapping scheme.

The used transliteration is very closely modelled after the MLS system yet provides enough transparency for accepting alternative spellings in some cases.

Chapter 4

Input and Output Encodings

4.1 The Need for Encodings

Any Mongolian text system has to deal with the issues of how to store, transmit, process and represent the following entities:

- Normal Latin letters, numbers and punctuation marks: a, b, c, etc.;
- Cyrillic letters, including those not present in basic Cyrillic but needed by Mongolian: а, б, в, ө, ү;
- Special symbols like the Mongolian Currency sign: ₮;
- Classical Mongolian script;
- Special symbols used in Latin scripts for purposes of transliterating Mongolian scripts: ä, ö, ü, γ e. a.

All these sets of symbols, alphabets and characters have their own unique properties, especially when it comes to non-Latin writings like Mongolian or Tibetan.

Unfortunately, prior to the arrival of Unicode, all computer systems based on 8-bit encodings (with a maximum character set of 256 characters) can only represent subsets of the above-mentioned entities as basic characters. All computer systems with 8-bit character encodings must either switch between several character sets (or code pages) or use non-standard commands to invoke individual character entities.

It is important to understand that the issue of how to enter all these characters is more or less completely detached from the issue of how to represent these characters on screen or in a document. Misleadingly, the usage of Latin characters in modern computers seems to suggest that there is a simple, 1:1 relationship (or mapping) between input and output, but for a number of languages, including Mongolian, this is simply not the case. While, due to the origin and history of computers, simple-minded systems do not make any difference between the two realms, TEX separates these two domains clearly, allowing for the amazing flexibility TEX shows when treating languages and writing systems.

It must also be understood that even though Unicode allows for the unambiguous representation of the characters and symbols of the world's major languages, it does not define any output conventions, and thus, input and output domains should still be treated as separate areas.

4.2 Input Encodings

MonTEX is flexible enough to deal with several kinds of input encodings including code pages with Cyrillic letters and Unicode. Input encodings are declared as an option to the main package in the document preample. E. g., a user working on an IBM compatible DOS platform is likely to specify the option `mls`: `mls`

```
\usepackage[mls]{mls}
```

4.2.1 7-bit ASCII and Mongolian

Basically it is possible to use MonTEX without anything else but the plain 7-bit ASCII Latin character set since internal and external mechanisms are available which can render transliterated texts (both Cyrillic and Traditional Mongolian) into their appropriate script presentations.

4.2.2 The MLS Codepage

The MLS codepage was the ancestor of all comprehensive, IBM-compatible Mongolian systems which intended to cover both Cyrillic and Classical Mongolian. Developed in the early 1990s, the MLS system tried to offer full Mongolian support for existing hardware and software as it was available then. While modern technological developments have confined the original approach to history, it is preserved here for preserving backward compatibility.

The MLS codepage is compatible with the IBM 437 codepage as far as the front vowels are concerned but features additional Cyrillic letters and Classical Mongolian.

4.2.3 8-bit Encodings

Most available 8-bit input encodings support either front vowels or Cyrillic letters or Classical Mongolian but usually not several of them at the same time.

If a local environment supports Cyrillic and Script codepages then texts can be composed using these codepages. Table 4.1 shows which codepages are supported. Those codepage names which are followed by a '(+)' are supplied by MonTEX whereas the other codepage declarations are recognized and passed through to the system assuming that the appropriate table exists. The column "Front Vowels" indicates whether the vowels ö and ü (and their Mongolian counterparts ө, ү) are available in that particular codepage. MonTEX recognizes both numbers and numbers preceded by `cp`, like `1250` and `cp1250` as names of codepages which are known by their number.

4.2.4 utf-8 Unicode

In summer of 2002, a new input encoding was made available for existing LaTeX 2_ε installations which allows the processing of utf8-encoded Unicode material. This package can be invoked with the option `utf8`: `utf8`

```
\usepackage[utf8]{mls}
```

There are some caveats, however. The relevant code is still under development, and at present, MonTEX only deals with the Mongolian, Manju and Sibe subsets of the Traditional Mongolian Character Plane beginning at U1800; a Todo character set remains to be implemented, and some of the more arcane special characters present in Unicode are as yet unavailable in MonTEX. The resulting constraints do not affect the work with contemporary text material and are only felt when dealing with frequently bilingual, mostly Tibetan and Sanskrit, religious texts of earlier centuries. Consult chapter 12.4 on page 70 and table 12.1 for details.

4.3 Output Encodings

Several output encodings are defined for MonTEX:

LMC Local Mongolian Cyrillic: This encoding was defined in order to avoid collisions with existing Cyrillic encodings for TEX and LaTEX. LMC is a 7-bit encoding which implies that most of its characters are addressed in the range of ordinary ASCII characters; when this encoding is active, all text typed in ASCII Latin characters will automatically appear in Cyrillic. Unlike some other available 7-bit encodings (like WN Cyrillic) it provides characters used in Mongolian.

LMA Local Manju: Manju in Ligature Mode. Any text typed in romanized Manju is automatically converted into Manju characters. LMA acts thus like a typical 7-bit encoding.

LMO Local Mongolian: Similar to Manju in Ligature Mode, Mongolian in Ligature Mode is typed in a special romanized form and is then automatically converted into Uighur Mongolian characters. LMO, too, acts thus like a typical 7-bit encoding.

LMS Local Mongolian Script: The system's original encoding for the Mongolian script. Mongolian is represented by a Latin transliteration the

Enc. Option	Latin Transliteration Symbols							Cyrillic	Front Vowels	
	Ä/ä	Ö/ö	Ü/ü	Ë/ë	Ï/ï	Č/č	Š/š		Ө/ө	Y/Y
`mls` (+)	+	+	+	ë	ï	-	-	+	+	+
`ncc` (+)	-	-	-	-	-	-	-	+	+	+
`mos` (+)	-	-	-	-	-	-	-	+	+	+
`mnk` (+)	-	-	-	-	-	-	-	+	+	+
`dbk` (+)	-	-	-	-	-	-	-	+	+	+
`ctt` (+)	-	-	-	-	-	-	-	+	+	+
`ibmrus`(+)	-	-	-	-	-	-	-	+	-	-
`koi` (+)	-	-	-	-	-	-	-	+	-	-
`437`	+	+	+	ë	ï	-	-	-	-	-
`437de`	+	+	+	ë	ï	-	-	-	-	-
`850`	+	+	+	+	+	-	-	-	-	-
`852`	+	+	+	+	-	-	-	-	-	-
`865`	+	+	+	ë	ï	-	-	-	-	-
`1250`	Ä	+	ü	+	-	-	-	-	-	-
`1252`	+	+	+	+	+	-	-	-	-	-
`applemac`	+	+	+	+	+	-	-	-	-	-
`mac`	+	+	+	+	+	-	-	-	-	-
`ansinew`	+	+	+	+	+	-	+	-	-	-
`ascii`	-	-	-	-	-	-	-	-	-	-
`atari`	+	+	+	+	Ï	-	-	-	-	-
`decmulti`	+	+	+	+	+	-	-	-	-	-
`isolatin`	+	+	+	+	Ï	-	-	-	-	-
`latin1`	+	+	+	+	+	-	-	-	-	-
`latin2`	+	+	+	+	-	+	+	-	-	-
`latin3`	+	+	+	+	+	-	-	-	-	-
`latin5`	+	+	+	+	+	-	-	-	-	-
`next`	+	+	+	+	+	-	-	-	-	-
`pc850`	+	+	+	+	Ï	-	-	-	-	-
`roman8`	+	+	+	+	Ï	-	-	-	-	-

Table 4.1: MonTEX Input Encodings

letters of which are essentially treated as future Mongolian canonical code positions. Once Mongolian Unicode will be available, the Latin transliteration can be seamlessly replaced by Mongolian canonical characters. The arrangement of code positions in this encoding does not reflect Unicode but follows the MLS system's keyboard support.

LMT Local Mongolian-Tibetan: This encoding is reserved to ensure access to the characters in the future ZANABAZAR package: SOYOMBO[1] and XÄWTÄÄ DÖRWÖLJIN. It is designed to comprise Tibetan as well, and Sirlin's Tibetan fonts can be directly used with this encoding.

LMX Local Mongolian XÄWTÄÄ DÖRWÖLJIN: This encoding is used for the XÄWTÄÄ DÖRWÖLJIN Script (available on CTAN) but is not frozen yet. Individual code positions are still subject to change.

LMU Local Mongolian Superset (U stands for 'Umbrella", "Unknown", or whatever you like to pick): This encoding is used to access all glyphs of the `bxg` glyph container, but is not frozen yet. Individual code positions are still subject to change.

4.4 MonTEX and Recent TEX Trends

As soon as the LH Cyrillic fonts support the Mongolian currency sign, MonTEX will switch to this font set. At the moment the private encoding LMC is favoured over LH; future implementations of MonTEX will provide a smooth transition for the user: documents developed with older versions of MonTEX will be upward compatible.

The `babel` package will, perhaps, also be supported in due course; at the moment, `babel` support is lacking mainly due to font encoding questions and a private RL setup. At present, MonTEX is *not* built with `babel` compatibility in mind. It must be seen as a stand-alone extension similar to `german.sty` or the CJK package.

The future belongs to 16-bit character sets; the first TEX development supporting larger character sets is Ωmega of which experimental versions exist. One of the great features of Ωmega is the capability to process canonical input encodings in order to generate glyph variants for document presentation. These so-called translation processes are far more powerful than anything Metafont can offer via ligatures, and they are the only feasible way to avoid external preprocessors or internal retransliteration engines coded in TEX needed to process Mongolian script.[2] Prof. Lagally's ArabTEX is the

[1]It is possible to use the Soyombo package available since 1996 as long as ZANABAZAR is not available.

[2]The retransliteration engine provided with the LMS encoding of MonTEX has a rather 'combined' approach; basic letter forms are selected in the retransliteration section while typical ligatures are composed with the ligature tables of Metafont. The authors express their sincerest gratitude to David Carlisle who contributed the missing link between

only LaTeX package known to the authors where an extensive retransliteration engine is realized as pure TeX code; it is an impressive piece of work defying any simple-minded imitation. So far, Ωmega translation processes exist for Tibetan and Arabic (paragons of complex relations between original script and any attempted romanization).

characters in the output list and TeX/Metafont's ligature mechanism.

Chapter 5

Installation

Before this latest version of MonTEX is installed please make sure that old installations of MonTEX *and* ManjuTEX are purged from disk as there are file name conflicts between earlier and recent versions of this software. In addition, ManjuTEX *is not required* any more as its functionality is now completely covered by MonTEX.

5.1 Hyphenation Patterns

MonTEX provides hyphenation rules for Modern Mongolian (Xalx). Hyphenation patterns for English are activated with English as selected language; hyphenation patterns for Russian exist at CTAN but they are unfortunately not suited for MonTEX withour prior work. Hyphenation patterns for Buryat have not been developed yet.

Due to the very nature of TEX, hyphenation patterns for a given language cannot easily be loaded at run-time but must be compiled into a so-called format file which gets loaded by TEX whenever the command `latex` is executed. A format file is usually created when a new TEX or LATEX 2_ε system is installed, but creating a new format can be done at any later time again. A special variant of TEX called `initex` is used for this purpose. The procedure sounds more intimidating than it actually is. Since there are many different types of TEX installations, the procedure is somewhat system-dependent. There is detailed on-line documentation available for performing this task, either in form of a text file for emtex, or in form of a FAQ file which can be displayed using the command `texconfig faq` on teTeX systems.

Part II

User Commands I
General Settings
Cyrillic Input

Chapter 6

Introduction

With regard to the substantial differences between Latin-like scripts (including Cyrillic) and Mongolian scripts, the user documentation of MonTeX is divided into two parts. This part deals with general settings, like language choices and input encoding definitions, whereas the commands specific to Mongolian and Manju are dealt with in part III, "Mongol Bicig and Manju Bithe". An alphabetic command reference covering *all* commands is presented in part IV.

6.1 General Settings

In order to access the commands of MonTeX the package must be loaded in the document preamble by saying

`\usepackage[<language options>,<encoding options>]{mls}`

The options include choices for the basic document language and input encodings.

6.1.1 Document Language

The document language can be set with one of `bicig`, `bithe`, `buryat`, `english`, `russian` or `xalx` like in

`\usepackage[xalx]{mls}`

which issues all captions and the date in Modern Mongolian.

The options `bicig` (see section 8.2, page 41) and `bithe` (see section 8.2, page 41) are discussed extensively in part III, "Mongol Bicig and Manju Bithe".

The options `buryat` (see table 6.1), `russian` (see table 6.2) and `xalx` (see table 6.3) produce captions in Buryat, Russian and Modern Mongolian. `buryat` `russian`

The option `english`, at least as a `\usepackage` option, is essentially a do-nothing: it sets captions to English (which is the default of this package anyway). `xalx` `english`

Command	English	Buryat
\prefacename	Preface	Оршол
\refname	References	Ашаглаһан ном
\abstractname	Abstract	Тобшолол
\bibname	Bibliography	Ном зүй
\chaptername	Chapter	Бүлэг
\appendixname	Appendix	Хабсаралта
\contentsname	Contents	Гаршаг
\listfigurename	List of Figures	Зурагай жагсаалт
\listtablename	List of Tables	Хүснэгтэй жагсаалт
\indexname	Index	Бүгэд хэлхээс
\figurename	Figure	Зураг
\tablename	Table	Хүсэнэгт
\partname	Part	Бүлэг
\enclname	encl	Оруулаха
\ccname	cc	
\headtoname	To	
\pagename	Page	Хуудаһан
\seename	see	Үз
\alsoname	see also	Мүн үз

Table 6.1: Captions in Buryat

The date form follows TeX conventions and is thus a mixture of numbers and words. Thus for `\today` (July 2, 2002) we get[1] what is shown in table 6.1. `\today`
The Uighur Mongolian and Manju dates are presented in section 8.2, page 41.

Buryat	2002 оной июлиин 2-ой үдэр
Xalx	2002 оны долоодугаар сарын 2
Russian	2 июля 2002

Figure 6.1: Dates in Buryat, Xalx and Russian

The language specifiers `buryat`, `english`, `russian` and `xalx` can also be used anywhere in the document as arguments to the `\selectlanguage` command. Instead of stating an argument to `\usepackage[...]{mls}` it is possible to say in your document

```
\selectlanguange{xalx}
```

which would set captions to Xalx Mongolian.

[1]The actual date at compilation time is used for the examples.

Command	English	Russian
`\prefacename`	Preface	Предисловие
`\refname`	References	Литература
`\abstractname`	Abstract	Аннотация
`\bibname`	Bibliography	Литература
`\chaptername`	Chapter	Глава
`\appendixname`	Appendix	Приложение
`\contentsname`	Contents	Оглавление
`\listfigurename`	List of Figures	Список рисунков
`\listtablename`	List of Tables	Список таблиц
`\indexname`	Index	Предметный указатель
`\figurename`	Figure	Рис.
`\tablename`	Table	Таблица
`\partname`	Part	Часть
`\enclname`	encl	вкл.
`\ccname`	cc	из
`\headtoname`	To	в
`\pagename`	Page	стр.
`\seename`	see	см.
`\alsoname`	see also	см. также

Table 6.2: Captions in Russian

Command	English	Xalx
`\prefacename`	Preface	Оршил
`\refname`	References	Ашигласан ном
`\abstractname`	Abstract	Товчлол
`\bibname`	Bibliography	Ном зүй
`\chaptername`	Chapter	Бүлэг
`\appendixname`	Appendix	Хавсралт
`\contentsname`	Contents	Гарчиг
`\listfigurename`	List of Figures	Зургийн жагсаалт
`\listtablename`	List of Tables	Хүснэгтийн жагсаалт
`\indexname`	Index	Бүгд хэлхээс
`\figurename`	Figure	Зураг
`\tablename`	Table	Хүснэгт
`\partname`	Part	Хэсэг
`\enclname`	encl	Оруулах
`\ccname`	cc	
`\headtoname`	To	
`\pagename`	Page	Хуудас
`\seename`	see	үзнэ үү
`\alsoname`	see also	мөн үзнэ үү

Table 6.3: Captions in Xalx

Chapter 7

Cyrillic Text – **Кирилл үсэг**

7.1 Cyrillic Text in Transliteration (LMC) Mode

MonTEX provides two basic modes of operation: in

- Transliteration Mode (intimately linked to the LMC encoding) all incoming text is regarded as transliterated Cyrillic. This allows users to compose Cyrillic documents on pure ASCII machines. In contrast, the
- Immediate Mode does nothing and waits for explicit Cyrillic characters in the input in order to generate Cyrillic output.

Two commands are used to switch between these modes:

```
\SetDocumentEncodingLMC
\SetDocumentEncodingNeutral
```

The first command switches to Transliteration Mode, the second command deactivates the transliteration and thus, by definition, activates Immediate Mode.

In the LMC encoding, most Cyrillic characters are mapped directly to a single Latin character but for some characters there is a text command which became necessary since there are more Cyrillic than Latin characters. For convenience, a few ligatures were defined, too. Details are given in table 7.1.

Front vowels can be entered directly using the encoding slot of a valid and active input encoding, or they can be expressed via an abbreviated `"`v notation where v stands for any desired vowel. In the LMC encoding used by MonTEX, `"` is not an active character; selecting the proper letter is done by ligature statements in the Metafont sources.

Some letters can be entered with or without a preceding `\`, like ю and я. Both `\yu` and `yu` will produce a ю. While `yu` is interpreted as a ligature, `\yu` allows for the character ю to be combined with accents. Accents are not commonly used in Mongolian since there are precise rules for word stress.

	Cyrillic Letter		LMC Input		Generic Command	
1	А	а	A	a	\CYRA	\cyra
2	Б	б	B	b	\CYRB	\cyrb
3	В	в	W	w	\CYRV	\cyrw
4	Г	г	G	g	\CYRG	\cyrg
5	Д	д	D	d	\CYRD	\cyrd
6	Е	е	E	e	\CYRE	\cyre
7	Ё	ё	Ë/"E {\}YO	ë/"e {\}yo	\CYRYO	\cyryo
8	Ж	ж	J	j	\CYRZH	\cyrzh
9	З	з	Z	z	\CYRZ	\cyrz
10	И	и	I	i	\CYRI	\cyri
11	Й	й	Ï/"I {\}YI	ï/"i {\}yi	\CYRISHRT	\cyrishrt
12	К	к	K	k	\CYRK	\cyrk
13	Л	л	L	l	\CYRL	\cyrl
14	М	м	M	m	\CYRM	\cyrm
15	Н	н	N	n	\CYRN	\cyrn
16	О	о	O	o	\CYRO	\cyro
17	Ө	ө	Ö/"O	ö/"o	\CYROTLD	\cyrotld
18	П	п	P	p	\CYRP	\cyrp
19	Р	р	R	r	\CYRR	\cyrr
20	С	с	S	s	\CYRS	\cyrs
21	Т	т	T	t	\CYRT	\cyrt
22	У	у	U	u	\CYRU	\cyru
23	Ү	ү	Ü/"U	ü/"u	\CYRY	\cyry
24	Ф	ф	F	f	\CYRF	\cyrf
25	Х	х	X	x	\CYRH	\cyrh
26	Һ	һ	H	h	\CYRHSHA	cyrhsha
27	Ц	ц	C	c	\CYRC	\cyrc
28	Ч	ч	Q \Ch	q \ch	\CYRCH	\cyrch
29	Ш	ш	\Sh	\sh sh	\CYRSH	\cyrsh
30	Щ	щ	\Sc \Qh	\sc \qh	\CYRSHCH	\cyrshch
31	Ъ	ь	\Y	\y	\CYRHRDSN	\cyrhrdsn
32	Ы	ы	Y	y	\CYRERY	\cyrery
33	Ь	ь	\I	\i	\CYRSFTSN	\cyrsftsn
34	Э	э	Ä/"A	ä/"a	\CYREREV	\cyrerev
35	Ю	ю	{\}YU	{\}yu	\CYRYU	\cyryu
36	Я	я	{\}YA	{\}ya	\CYRYA	\cyrya

Table 7.1: Cyrillic Alphabet Input Methods

This feature is taken from the OT2 encoding and is included mainly for the sake of completeness, convenience and compatibility[1].

Here now a sample of Mongolian text:

«Халхын гурван өндөр» хэмээн алдаршсан, Зүүн хязгаарыг тохинуулах сайд Н. Дугаржав ардын хувьсгалын бүүр эхэн үеэс хамгийн эгзэгтэй амь дүйсэн албанд томилогдох цэрэг дайны олон чухал даалгаврыг хичээнгүйлэн биелүүлж явсан түүхтэй хүн.

```
{\mnr<<Xalxyn gurwan "ond"or>>
x"am"a"an aldarshsan, Z"u"un xyazgaaryg
toxinuulax sa"id N.~Dugarjaw ardyn
xuw\i sgalyn b"u"ur "ax"an "ue"as
xamgi"in "agz"agt"a"i am\i\ d"u"is"an
alband tomilogdox c"ar"ag da"iny olon
quxal daalgawryg xiq"a"ang"u"il"an
biel"u"ulj yawsan t"u"uxt"a"i x"un.}%
```

Figure 7.1: Romanized Cyrillic Input Example

In order to make the document you are reading at the moment truly portable, the somewhat more clumsy "*v* notation was used in this example; if your environments supports an 8-bit codepage (what it usually does), all front vowels can be entered as ä, ö and ü etc. using the slots of those vowels in the particular active codepage.[2]

7.2 Entering Cyrillic Text in Immediate Mode

For freely combining Latin and Cyrillic characters without using any explicite commands it is necessary that the codepage in use supports some Cyrillic encoding. It should be noted, however that these documents are not easily portable between different platforms anymore since they need recoding; some of the Cyrillic codepages are defective in one or the other way thus individual characters can get lost.

The user simply specifies the desired input encoding as a `\usepackage[<encoding>]{mls}` option, and MonTEX takes care of the rest. It is a feature and not a bug that input encoding and document language are chosen independently. It is well possible that a user working on a computer with default Mongolian codepage wants to create a document in Russian, English or any other language yet wants to include Mongolian fragments in her text without explicitely issuing any command.

In case a need arises for switching from Transliteration Mode to Immediate Mode the command can be issued anywhere in the preamble or

[1] The magic triple-C!

[2] Looking at the source code of this document the astute reader will discover that all front vowels are indeed produced using the `\"a` (etc.) notation; thus the document source can be viewn and manipulated on any 7-bit ASCII platform; it can also safely be transmitted via e-mail.

the document itself; like \SetDocumentEncodingLMC it affects the Cyrillic transliteration only and leaves the document language in its chosen state.

7.3 Entering Cyrillic Characters by Name

Outside the Cyrillic environments, individual Cyrillic characters can be entered by using the commands beginning with \cyr*x* from the two right columns of table 7.1 where x stands for the letter name. This command works in any encoding.

7.4 Entering Special Cyrillic Characters

A few special characters are available, notably the guillemots frequently used for quoting text, the currency symbol, the ordinal number symbol and the currency sign. See table 7.2.

Symbol	Command	Alternative
«	\lgu	<<
»	\rgu	>>
№	\No	
₮	\Togrog	
₮	\togrog	

Table 7.2: MonTEX Special Cyrillic Characters

The command producing the guillemots (\lgu, \rgu) *only* works in a Cyrillic environment — it is not a generic command.

There are actually two versions of the \togrog command. While \Togrog produces a sans serif ₮ (considered standard) with any font selected it is also possible to print serif (₮), italic (*₮*) and typewriter (₮) versions of this symbol.[3] For achieving this result the commands \MyTogrog and \mytogrog are available. Unlike the standard command they simply pick the current font style of the surrounding letters for the currency symbol. \togrog \Togrog \MyTogrog \mytogrog

7.5 Running Text with Embedded Words in Different Encodings

Independently of the document language it is possible to produce portions of Cyrillic text within Latin text and vice versa. The two commands \mnr and \rnm switch from ordinary Latin text to transliterated Cyrillic text and \mnr \rnm

[3]The currency symbol is not limited to these three typefaces; all typefaces can be selected.

`\fontfamily{...}` Parameter	Family Description
cmr	Computer Modern Serif
cmss	Computer Modern Sans Serif
cmtt	computer Modern Typewriter
cmvtt	Computer Modern Variable Width Typewriter
cmfr	Computer Modern Funny
cmfib	Computer Modern Fibonacci
cmdh	Computer Modern Dunhill
cmssq	Computer Modern Sans Serif Quotation Style 8pt

Table 7.3: Font Families Supported by MonTEX

back to Latin text. The command stands for *m*ongolian *n*ew *r*omanization and its reversal (which can, by accident, also be read as *r*eturn to *n*or*m*al). They can be used as stream commands or for initializing groups:

монгол хэл ба english text with a монгол word inserted	`\mnr mongol x"al ba \rnm english text with a {\mnr mongol} word inserted`

For enhanced convenience, portions of text can also be encapsulated into the commands `\xalx{...}` for Cyrillic text and `\lat{...}` for neutral (i. e. `\xalx`
Latin) texts. `\lat`

The commands `\mnr`, `\rnm`, `\xalx{...}` and `\lat{...}` do *not* switch the default encoding; this shows up when a construct like `\lat{\verb|article|}` is placed in Transliteration Mode; the result will be `артицле` rather than `article`; in order to generate the desired form, the mode switching commands must be used.

7.6 Font Selection Commands

The Cyrillic fonts are set up in a manner which allows for seamless switching between Roman and Cyrillic typefaces. The font switching commands used for modifying typefaces (by `\text..`) are completely transparent to the encoding; no precaution whatsoever has to be taken. Most of the typefaces supplied with the traditional OT1 encoding are also available for MonTEX; Dunhill and Funny Roman are included.[4] MonTEX offers the following font families as shown in table 7.3:

[4] A complete overview of the NFSS classification of the Computer Modern fonts can be found in The LATEX Companion, by Michel Goossens, Frank Mittelbach and Alexander Samarin, Addison-Wesley 1994, p. 181.

ABАБ

Figure 7.2: `inch` Font Examples

The word "Roman" was avoided since in MonTEX these families also cover matching typefaces in Cyrillic script. The first three families have support for combinations of different weights and shapes (e. g. bold and italic) whereas the other series usually only offer an italic variant. The Sans Serif Quotation Style 8pt typeface is not by default installed in standard LATEX distributions hence it cannot be guaranteed that switching to and from Cyrillic letters maintains the typeface. The fonts (upright and slanted) can be accessed via the `\fontfamily{cmssq}` command but are not shown in table 7.4. See table 7.4 for a therefore incomplete list of available typeface examples.

Besides these transparent commands for scalable fonts MonTEX also offers two inch-high variants of bold Computer Modern Sans typefaces for Latin and Cyrillic: `\cminch` and `\kminch`. These commands bypass the NFSS font setup and should only be used for book titles etc. The command sequence `{\cminch AB} {\kminch AB}` produces the output shown in figure 7.2. `\cminch` `\kminch`

7.7 Shorthands for Embedding Words in a Different Typeface

Sometimes it may be necessary to give short portions of text not only in a different encoding (for which the `\lat` (see section 7.5, page 34) `{...}` and `\mnr` (see section 7.5, page 33) `{...}` commands are useful) but it may also be necessary to switch the typeface temporarily. Usually capsules using `\text`*xx* do the work if only the typeface is concerned, and building nested commands like `\textsf{\lat{...}}` is cumbersome if these changes have to be applied very often. MonTEX provides an abbreviated style following the rule

`[k|l]`*two letter font style code*`{...}`

where the font style code is one of `rm`, `bf`, `it`, `sl`, `sf`, `sc` and `tt`, like `\ksl{...}`, `\lsc{...}`, etc.

Family and Command Example		Typeface Examples	
Computer Modern Serif			
cmr	(default)	Сүхбат	Süxbat Šaγdur
	`\textbf{...}`	**Сүхбат**	**Süxbat** **Šaγdur**
	`\textsl{...}`	*Сүхбат*	*Süxbat* *Šaγdur*
	`\textsc{...}`	СҮХБАТ	SÜXBAT ŠAΓDUR
	`\textit{...}`	*Сүхбат*	*Süxbat* *Šaγdur*
	`\fontseries{bx}\textit{...}`	***Сүхбат***	***Süxbat*** ***Šaγdur***
Computer Modern Typewriter			
cmtt	`\texttt{...}`	Сүхбат	Süxbat Šaγdur
	`\texttt{\textit{...}}`	*Сүхбат*	*Süxbat* *Šaγdur*
	`\texttt{\textsl{...}}`	*Сүхбат*	*Süxbat* *Šaγdur*
	`\texttt{\textsc{...}}`	СҮХБАТ	SÜXBAT ŠAΓDUR
Computer Modern Variable Width Typewriter			
cmvtt	`\fontfamily{cmvtt}...`	Сүхбат	Süxbat
	`\fontfamily{cmvtt}\textit{...}`	*Сүхбат*	*Süxbat*
Computer Modern Sans			
cmss	`\textsf{...}`	Сүхбат	Süxbat Šaγdur
	`\textsf{\textsl{...}}`	*Сүхбат*	*Süxbat* *Šaγdur*
	`\textsf{\fontseries{bx}...}`	**Сүхбат**	**Süxbat** **Šaγdur**
	`\textsf{\fontseries{sbc}...}`	**Сүхбат**	**Süxbat** **Šaγdur**
Computer Modern Funny			
cmfr	`\fontfamily{cmfr}...`	Сүхбат	Süxbat
Computer Modern Dunhill			
cmdh	`\fontfamily{cmdh}...`	СҮХБАТ	SÜXBAT

Table 7.4: Typeface Consistency for Cyrillic and Latin

Letter	Input	Letter	Input
č	\ch	Č	\Ch
ǰ	\jh	J̌	\Jh
š	\sh	Š	\Sh
ž	\zh	Ž	\Zh
ŋ	\ng	Ŋ	\Ng
γ	\g	Γ	\G

Table 7.5: Shortcuts for Mongolian Transliteration Symbols

7.8 Shorthands for Writing Transliterated Texts

MonTEX provides shortcuts for writing certain accented symbols used in conventional transliterating of Mongolian by haceks, the nasal and the gamma. These shortcuts are essentially mnemonics replacing the somewhat more tedious accent notation (see table 7.5).

It must be observed that these commands are by default dependent on the environment they are used in. `\Sh` yields a Š when used in a Latin environment but results in a Ш when used in a Cyrillic context[5]:

Šagdar and *Čadraa* are transliterations for Шагдар and Чадраа.

```
\emph{\Sh agdar} and \emph{\Ch adraa}
are transliterations for
{\mnr\Sh agdar} and {\mnr\Ch adraa}.
```

7.9 Gamma Typeface

If modern Greek is supported by your LATEX 2ε installation then the shape of the gamma will match the neighbouring typeface as closely as possible as can be seen from table 7.4; \g otherwise, the selection of gamma shapes and styles is limited to the gamma math typeface supplied by standard TEX installations. \g

7.10 Oirat Double Accents

All accented characters which are contained in the T1 encoding or can be generated out of these via accents can be produced. This comes in conveniently for transliterating Oirat texts which need vowels with double diacritics, like

[5]The authors wish to thank J. Knappen for resolving one instability in the original code for these letters.

ǟ which can be entered as any combination of two nested accent commands (like `\={\"a}`) or one accent command and a vowel with diacritics (provided an 8-bit input codepage is available).

7.11 Numbering by Cyrillic Letters

Analogous to the `\Alpha` command which provides an alphabetical counter in English, MonTEX features counters for Buryat, Modern Mongolian, and Russian.

Buryat The counter for Buryat is invoked with `\Uzeg{`n`}` or `\uzeg{`n`}` and is valid for $1 \leq n \leq 32$. `\Uzeg` `\uzeg`

^{1}А/а	^{2}Б/б	^{3}В/в	^{4}Г/г	^{5}Д/д	^{6}Е/е
^{7}Ё/ё	^{8}Ж/ж	^{9}З/з	^{10}И/и	^{11}К/к	^{12}Л/л
^{13}М/м	^{14}Н/н	^{15}О/о	^{16}Ө/ө	^{17}П/п	^{18}Р/р
^{19}С/с	^{20}Т/т	^{21}У/у	^{22}Ү/ү	^{23}Ф/ф	^{24}Х/х
^{25}Һ/һ	^{26}Ц/ц	^{27}Ч/ч	^{28}Ш/ш	^{29}Щ/щ	^{30}Э/э
^{31}Ю/ю	^{32}Я/я				

Modern, or Xalx Mongolian The counter for Modern Mongolian is invoked with `\Useg{`n`}` or `\useg{`n`}` and is valid for $1 \leq n \leq 31$. `\Useg` `\useg`

^{1}А/а	^{2}Б/б	^{3}В/в	^{4}Г/г	^{5}Д/д	^{6}Е/е
^{7}Ё/ё	^{8}Ж/ж	^{9}З/з	^{10}И/и	^{11}К/к	^{12}Л/л
^{13}М/м	^{14}Н/н	^{15}О/о	^{16}Ө/ө	^{17}П/п	^{18}Р/р
^{19}С/с	^{20}Т/т	^{21}У/у	^{22}Ү/ү	^{23}Ф/ф	^{24}Х/х
^{25}Ц/ц	^{26}Ч/ч	^{27}Ш/ш	^{28}Щ/щ	^{29}Э/э	^{30}Ю/ю
^{31}Я/я					

Russian The counter for Russian is invoked with `\Asbuk{`n`}` or `\asbuk{`n`}` and is valid for $1 \leq n \leq 28$. `\Asbuk` `\asbuk`

^{1}А/а	^{2}Б/б	^{3}В/в	^{4}Г/г	^{5}Д/д	^{6}Е/е
^{7}Ж/ж	^{8}З/з	^{9}И/и	^{10}К/к	^{11}Л/л	^{12}М/м
^{13}Н/н	^{14}О/о	^{15}П/п	^{16}Р/р	^{17}С/с	^{18}Т/т
^{19}У/у	^{20}Ф/ф	^{21}Х/х	^{22}Ц/ц	^{23}Ч/ч	^{24}Ш/ш
^{25}Щ/щ	^{26}Э/э	^{27}Ю/ю	^{28}Я/я		

Part III

Mongol Bicig and Manju Bithe

ᠮᠣᠩᠭᠣᠯ ᠪᠢᠴᠢᠭ ᠮᠠᠨᠵᡠ ᠪᡳᡨᡥᡝ

Chapter 8

Introduction

This part describes in detail all aspects of typesetting Mongolian and Manju with MonTEX. The following sections cover the various input methods for these languages, the commands for presenting small snippets, big portions and whole documents composed in Mongolian and Manju, as well as the relationship between input notations and script-related commands.

8.1 Mongolian and Manju Script Fundamentals

Mongolian Script, or *bicig*, is a writing with an intriguing and complex relationship between the canonical letters of the alphabet and their presentations in context. Virtually any canonical letter can assume several shapes. As a rule of thumb, there are three or four basic shapes: the letter in isolated form, the letter in initial, medial and final position of a word. Only a few letters stay the same, and in rare cases there are up to ten possibilities for representing a single letter.

On the other hand, some letters share the same shape in different contexts; one so-called *glyph* can represent more than one letter, sometimes three or four different letters.

The Manju writing, or *bithe* system is a close relative of the Mongolian system; the basical letter shapes are the same. Yet for Manju, a set of diacritics (*dots und circles*) was designed to the effect that all the ambiguities of Mongolian are eliminated.

Decomposing the writing system and using glyphs as the atoms of writing is one of several conceivable methods of writing Mongolian script.

In MonTEX, Mongolian script can be entered in three ways, either by writing transliterated Mongolian in one of two different romanization systems, by an approximated symbol for every glyph or by generic name. There are certain constraints concerning the possible combinations of Mongolian input methods and Mongolian writing display commands. Since Manju has only one input method, these constraints do not apply to Manju. The possible combinations are listed in table 9.2.

A complete guide to the principles of glyph analysis can be found in the MLS Report by one of the authors.[1]

Due to technical constraints of MonTEX, there is an intimate relationship between various script-related commands and Mongolian input methods.

8.2 General Settings

As for Modern (Xalx) Mongolian, Buryat and Russian documents, it is possible to set the document language to Uighur Mongolian or Manju with a language option:

```
\usepackage[<language options>,<encoding options>]{mls}
```

The two language options are `bicig` for Uighur Mongolian and `bithe` for Manju documents. Among other things, they set the document encoding, the captions and the date in either Uighur Mongolian or Manju. `bicig` `bithe`

The date form follows TEX conventions and is thus a mixture of numbers and words. Thus for `\today` (July 2, 2002) we get[2] what is shown in table 8.1. `\today`

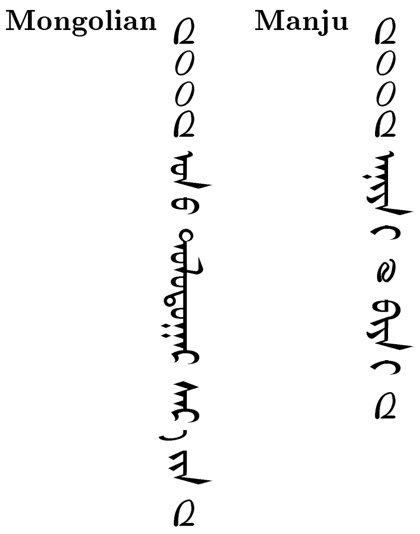


Figure 8.1: Dates in Uighur Mongolian and Manju

The document language option `bicig` can only be used with the Mongolian input method named "Simplified Transliteration" (see the following chapter and table 9.2).

[1] Oliver Corff: MLS Report. UNU/IIST Report No. 8, Macau 1993

[2] The actual date at compilation time is used for the examples.

Chapter 9

Mongolian Input

It has been mentioned before that each Mongolian input method provided by MonTEX has a slightly different application scope. The next sections cover

- *Simplified Transliteration Mode* which is the mode of choice for bulk text due to its availability as document encoding; the associated font encoding is labelled LMO and internally activated by the command `\SetDocumentEncodingBicig` (see section 9.1, page 49) .
- *MLS Transliteration Mode* which is most suitable for short portions of text, like dictionary entries, quotations, etc.
- *Immediate Mode* is the mode accepting Mongolian characters encoded in the MLS codepage. Together with this mode, the input encoding `mls` (see section 4.2, page 19) should be specified, ideally combined with the command `\SetDocumentEncodingNeutral` (see section 7.2, page 32) .
- *Glyph Input* which is useful mainly for rendering individual words in unorthodox or incorrect spellings, e. g. for reproducing idiosyncrasies found in old books.

A comprehensive table of the Mongolian alphabet and its MLS transliteration, the input conventions of the MLS transliteration in MonTEX and the Simplified Transliteration is given in table 9.1.

The possible combinations of Mongolian writing input methods and display commands are listed in table 9.2. The columns stand for each possible input encoding, the rows contain the display command types. Each table cell at the contains the command that is available for a given combination of input method and command.

9.1 Simplified Transliteration Mode

The broad romanization of the Mongolian script as realized in the MLS system focuses on lexical properties (the *information layer*) rather than graph-

Uighur Script	MLS Transl.	MLS Input	Simplified Input	Uighur Script	MLS Transl.	MLS Input	Simplified Input
	a	a	a		s	s	s
	ä	ä, E	e		sh	S	sh
	e	e	v		t	t	t
	i	i	i		d	d	d, t
	o	o	u		l	l	l
	u	u	u		m	m	m
	ö	ö, O	ui, u		c	c	c
	ü	ü, U	ui, u		z	z	z
	n	n	n		y	y	y
	*ng	ng	ng		r	r	r
	x	x	x		v	v	v
	γ	G	g		h	h	h
	k	k	k		j	j	j
	g	g	g, k		K	K	K
	b	b	b		[-]	Q	q
	p	p	p		C	C	C
	f	f	f		Z	Z	Z

Table 9.1: Mongolian Script Transliterations

ical properties (the *presentation layer*). The obvious advantage of such a method is the possibility to store and transmit Mongolian language information in and between systems without devices for displaying Mongolian writing.

With the ambiguities of the Mongolian script (the consonants *t/d*, the vowels *a/e*, and many other ambiguous shapes give vivid evidence hereof) it is however possible to enter misleading or wrong romanizations which lead to a desired yet semantically misleading *display* of Mongolian in which case the underlying information is not suitable for further processing.

Another aspect is the retrieval of information from e. g. library catalogues when only the display of potentially unknown words like in book titles is available. It must be possible to enter Mongolian script into an information processing system without knowing at every moment which underlying letter generates a given shape. This implies that glyph analysis does not decompose complex glyph shapes into atoms if the shape transformation is purely dictated by graphical rather than linguistical context.

Command	Mongolian		Manju
Type	MLS	Simplified	
Document Encoding	only available as font encoding LMS, not as document encoding	`LMO` (see section 9, page 42)	`LMA` (see section 10, page 56)
Horizontal Capsules	`\bcg` (see section 9.2, page 49)	`\bicig` (see section 11.1, page 59)	`\bithe` (see section 11.1, page 59)
Horizontal Paragraphs	not available	`bicigtext` (see section 11.2, page 59)	`bithetext` (see section 11.2, page 59)
Vertical Capsules	`\mbosoo` (see section 9.2, page 50)	`\mobosoo` (see section 16, page 60)	`\mabosoo` (see section 16, page 60)
Vertical Paragraph Boxes	not available	`\mobox` (see section 11.4, page 61)	`\mabox` (see section 17, page 61)
Vertical Pages	not available	`bicigpage` (see section 11.5, page 61)	`bithepage` (see section 11.5, page 61)

Table 9.2: Mongolian Input and Display Commands

Notwithstanding this fact, an obvious *n* appearing as ᠨ following a vowel should be entered as `n` while an *a* following a consonant which also appears as ᠨ should certainly entered as vowel, not as consonant. Anything going deeper in glyph analysis can only be considered as atomic coding which may be highly useful in special cases but renders the input process more than cumbersome in general cases.

The Mongolian Simplified Transliteration proposed here is based on principles laid out by Dr. Michael Balk of the Deutsche Staatsbibliothek, Stiftung Preussischer Kulturbesitz, Berlin. During its development, various proposals were discussed at DIN, MNISM and ISO standardization meetings during 1994 to 1997.

The most important principle of this simplified input method is the consequent elimination of ambiguities in the relation between romanized input (as performed on an ordinary computer keyboard) and its Mongolian script target. If the Mongolian language provides several readings for certain vowels, then only one vowel is available in the simplified method; if alternating consonants (like *k/g* can swap shapes, then each input letter is associated with one and only one output shape. Furthermore, the *input alphabet* (speaking in terms of computer theory) is limited to the basic Latin alphabet. It uses only `a b c d e f g h i j k l m n p q r s t u v x y z C K Z` and the following characters with special meanings: `- = ' "`. The first symbol is used to separate grammatical endings from preceding words, the second separates floating vowels from word stems, and the third and fourth character act as Variant Selectors. At present, the second Variant Selector is not yet assigned.

Unlike a purely atomic rendering, the resulting romanization as shown in table 9.1 is easy to learn, much easier to read than atomic code, and yet acceptably close to conventional Mongolian transliterations, as can be seen from table 9.3. Each row contains one or more instances of every letter listed in the first column. From left to right, these are the beginning of a word, the middle of a word and the end of a word. Every single cell features three elements: the *example* in MLS romanization appears in italics; the correct Simplified Transliteration `input` appears in typewriter style, and the word in Uighur Mongolian letters appears in the right half of the cell.

Letter	Beginning	Middle	End
a	*arad* `arad`	*baγatur* `bagadur`	*la* `la` *sanaγ=a* `sanag=a`
ä	*ärkä* `erke`	*cäcäg* `cecek`	*sükä* `suike`
e	*eKs* `evKs`	*geologi* `kvuluki`	
i	*izaγur* `izagur`	*minu* `minu`	*bandi* `bandi`
o	*olan* `ulan`	*aγul=a* `agul=a`	
u	*ulus* `ulus`		*γarxu* `garxu`
ö	*öndür* `uindur`	*cöm=ä* `cuim=e`	

ü	*üsüg* `uisuk`		
n + {V}	*nam* `nam`	*onol* `unul`	*bayin=a* `baiin=a`
n + {C}		*bandi* `bandi`	
n' + {V}	*n'am* `n'am`	*on'ol* `un'ul`	*bayin'=a* `baiin'=a`
n' + {C}		*KoNTor* `Kun'tur`	*ban'di* `ban'di`
ng		*mongγol* `munggul`	*vang* `vang`
x	*xota* `xuda`	*abxu* `abxu`	*mix=a* `mix=a`
γ	*γazar* `gazar`	*baγatur* `bagadur`	*tuγ* `tug`
γ'	*γ'azar* `g'azar`	*baγ'atur* `bag'adur`	
k	*käräg* `kerek`	*ärkiläkü* `erkileku`	
g	*gär* `ger`	*ügäi* `uigei`	*bicig* `bicik`
b	*baγ=a* `bag=a`	*däbtär* `tebder`	*äb* `eb`
p	*pangsa* `pangsa`		

f	*feodal* `fvudal`	*Cifr* `Cifr`	
s	*saxal* `saxal`	*basa* `basa`	*nas* `nas`
sh	*shaγdur* `shagdur`		
t	*tomu* `tumu`	*baγatur* `bagadur`	
d	*dumdadu* `dumdadu`	*odu* `udu` *sädgil* `sedkil`	*arad* `arad` *äD* `ed’`
l	*la* `la`	*aldar* `aldar` *blam=a* `blam=a`	*onul* `unul`
m	*mongγol* `munggul`	*nomin* `numin`	*nom* `num`
c	*caγan* `cagan`	*äcän* `ecen`	
z	*zam* `zam`	*γazar* `gazar`	
y	*yondan* `yundan`	*bayar* `bayar`	*xoriy=a* `xuriy=a`
r	*rashan* `rashan`	*oros* `urus`	*bolor* `bulur`

v	*vang* `vang`		
h	*heze* `hvzv`	*lhas* `lhas`	
j	*j* `j`		
K	*KoNTor* `Kun'tur`		
gh	*ghombo* `qumbu`		
C	*Cifr* `Cifr`		*sTan'C* `stan'C`
Z	*Zambu* `Zambu`	*aZi* `aZi`	

Table 9.3: Mongolian Simplified Transliteration by Example

While the input method for the majority of characters matches the transliteration conventions, some letters require a slightly different treatment:

1. Although the diphtong is usually rendered as *ayi*, it must be entered as `aii` in order to produce the desired effect.
2. The back vowels *o* and *u* are both rendered as `u`.
3. The front vowels *ö* and *ü* are both rendered as `ui` in first syllables and as `u` in later syllables.
4. Since means both *t* and *d*, it is necessary to spell this letter as `t` in the beginning of words, and `d` in the middle of words, regardless of the actual meaning.
5. The four consonants γ, *g*, *x* and *k* are constrained with regard to the following vowels. The Simplified Transliteration renders these as `g` (before *a* and *u* only), `g` (before *a* and *u* only), `x` and `k`.

As it was demonstrated in section 7.1, it is technically possible to choose between an automatic document encoding and the neutral mode. In the case of Uighur Mongolian, the mode of choice activates the Simplified Transliteration Mode and is called with

```
\SetDocumentEncodingBicig
```

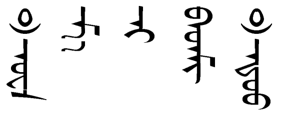
With `\SetDocumentEncodingBicig` set, it is possible to switch to the Simplified Transliteration Mode anywhere in the document, not only in the preamble.

Caveat: Since switching to Uighur Mongolian text requires a lot of settings to be effected at the same time, there are high-level commands available (see below, chapter 11: Mongolian and Manju Display Commands) which do all the work, including the definition of the document encoding. Thus, while `\SetDocumentEncodingBicig` is indeed classified as a user-level command, it is certainly not necessary for everyday work.

9.1.1 Character Variants

With the assistance of special, non-printing characters like the Form Variant Selectors, the appearance of certain characters can be modified in order to display typographical and orthographical variants. Notably, the *n* will loose its dot before vowels, as will *γ*. Let's assume the word "place" is written in an old book as ᠭᠠᠵᠠᠷ. It should be understood that this is a variant of ᠭᠠᠵᠠᠷ and should be spelled *γ'azar*, not *xazar*. With vowels, the Form Variant Selectors can change the shape that is usually required by graphical context. At present, only the first of two Form Variant Selectors actually does something, the exact behaviour of the second Form Variant Selector waits to be implemented.

The following short example shows a concrete application of this method. It renders the six syllable mantra *om ma ni padme hum* (tib. ཨོཾ་ མ་ ཎི་ པདྨེ་ ཧཱུྃ།) also featuring the special syllable `\om` as it is displayed on a huge `\om` bronze incense burner in front of the Gandan Monastery in Ulaanbaatar:



```
\mobox{3cm}{\noindent\sffamily
\om uva\\
\  ma'=a\\
\  n'i\\
\  badmi'\\
\om huu}
```

Figure 9.1: Mongolian Character Variants Example

9.2 MLS Transliteration Mode

In Transliteration Mode (activated with the commands `\bcg{...}` or `\bcg`

\mbosoo{...}) Mongolian text portions can be entered using a transliteration which is a rough approximation to the MLS system. The major difference is that only pure Latin alphabetical symbols can be used for virtually all letters. Front vowels are either entered via the traditional vowels with diacritics (*ä, ö, ü*) or can be entered with capitalized versions of the normal vowels. Capitalized letters have to be used for entering γ and š which are entered as *G* and *S*. Special variants for certain letters can be selected with Form Variant Selectors. [1] \mbosoo

The available Mongolian characters (*caγan toluγai*) are shown in table 9.1.[2]

9.3 Immediate Mode

For freely combining Mongolian Script with other characters without using any explicite commands it is necessary that the codepage in use supports Mongolian Script glyphs; currently this is the MLS codepage. The MLS input encoding is specified like `\usepackage[mls]{mls}`. As with Cyrillic codepages it should be noted that these documents are not easily portable between different platforms anymore since they cannot be recoded at ease. See table 9.4 for a list of available symbols.

9.4 Glyphs by Symbol

Without MLS codepage support, Mongolian words can also be entered using the `\glyphbcg{...}` command in running text. Within these groups, Mongolian Script glyphs are entered in the form of approximated symbols; sometimes these symbols reflect the underlying canonical letter, sometimes functional equivalents (for punctuation marks etc.) are chosen; sometimes there is no evident relation between glyph and input symbol simply because a free slot within the ASCII range $c \geq 32 \leq 127$ was chosen. Please consult table 9.5 of available glyphs and their input equivalents. \glyphbcg

9.5 Glyphs by Name

Without any preparations on the side of the text environment it is possible to enter individual Mongolian glyphs by name in a way similar for that of entering Cyrillic characters; the Mongolian glyph names can be found in

[1] The suggested solution has the advantage that it can be used on computers featuring codepages without umlaut symbols as most of the Cyrillic code pages are 'defective' in this point.

[2] The alphabetical arrangement follows large that given on p. 17 of N. Poppe's *Grammar of Written Mongolian*, Wiesbaden 1954, 1964, 1974 (third printing). Letters not given there are appended to Poppe's list.

table 9.4. Thus, `\shilbe` produces a . A number in the MLS column indicates the encoding position of the MLS codepage; a missing number in this column indicates that the glyph is part of extended MonTEX glyph set without being part of the original MLS.

Glyph	Generic Name	Input Char.	Glyph	Generic Name	Input Char.
	`\titem`	`@`		`\matgarshilbe`	`v`
	`\shud`	`a`		`\bituushilbe`	`h`
	`\secondaryshud`	`A`		`\secondaryqagt`	`K`
	`\shilbe`	`i`		`\qagt`	`k`
	`\gedes`	`o`		`\secnumtdelbenqix`	`P`
	`\secondarygedes`	`O`		`\numtdelbenqix`	`p`
	`\cegteishud`	`n`		`\secsertenqixtnum`	`F`
	`\lewer`	`l`		`\sertenqixtnum`	`f`
	`\suuliinlewer`	`L`		`\zadgaizardigt`	`Z`
	`\tertiarylewer`	`Q`		`\bituuzardigt`	`C`
	`\mewer`	`m`		`\malgaitaititem`	`j`
	`\suuliinmewer`	`M`		`\suul`	`e`
	`\xewteeqix`	`x`		`\orxic`	`E`
	`\dawxarcegtxewteeqix`	`X`		`\biodoisuul`	`Y`
	`\halfnum`	`g`		`\bagodoisuul`	`G`
	`\num`	`I`		`\nceg`	`-`
	`\halfnumtgedes`	`B`		`\gceg`	`=`
	`\numtaigedes`	`b`		`\ceg`	`,`
	`\buruuxarsangedes`	`t`		`\dorwoljin`	`;`
	`\gedesteishilbe`	`d`		-	`V`
	`\erweeljinshilbe`	`r`		-	`u`
	`\secerweeljin`	`R`		-	`T`
	`\bosooshilbe`	`z`		-	`U`
	`\etgershilbe`	`y`		-	`W`
	`\zawj`	`s`		-	`w`
	`\suuliinzawj`	`S`		-	`ml`
	`\dawxarcegtzawj`	`q`		-	`ll`
	`\sereeewer`	`c`			

Table 9.4: MLS Named Basic Glyphs

Glyph	Generic Name	MLS Code	Glyph	Generic Name	MLS Code
	`\titem`	`"C2`		`\matgarshilbe`	`"EB`
	`\shud`	`"C3`		`\bituushilbe`	`"EC`
	`\secondaryshud`	`"C5`		`\secondaryqagt`	`"ED`
	`\shilbe`	`"C6`		`\qagt`	`"EE`
	`\gedes`	`"C7`		`\secnumtdelbenqix`	`"EF`
	`\secondarygedes`	`"CF`		`\numtdelbenqix`	`"F0`
	`\cegteishud`	`"D0`		`\secsertenqixtnum`	`"F1`
	`\lewer`	`"D1`		`\sertenqixtnum`	`"F2`
	`\suuliinlewer`	`"D2`		`\zadgaizardigt`	`"F3`
	`\tertiarylewer`	`"D3`		`\bituuzardigt`	`"F4`
	`\mewer`	`"D4`		`\malgaitaititem`	`"F5`
	`\suuliinmewer`	`"D5`		`\suul`	`"F6`
	`\xewteeqix`	`"D6`		`\orxic`	`"F7`
	`\dawxarcegtxewteeqix`	`"D7`		`\biodoisuul`	`"F8`
	`\halfnum`	`"D8`		`\bagodoisuul`	`"F9`
	`\num`	`"DB`		`\nceg`	`"FA`
	`\halfnumtgedes`	`"DC`		`\gceg`	`"FB`
	`\numtaigedes`	`"DD`		`\ceg`	`"FC`
	`\buruuxarsangedes`	`"DE`		`\dorwoljin`	`"FD`
	`\gedesteishilbe`	`"DF`		-	
	`\erweeljinshilbe`	`"E0`		-	
	`\secerweeljin`	`"E3`		-	
	`\bosooshilbe`	`"E4`		-	
	`\etgershilbe`	`"E5`		-	
	`\zawj`	`"E6`		-	
	`\suuliinzawj`	`"E8`		-	
	`\dawxarcegtzawj`	`"E9`		-	
	`\sereeewer`	`"EA`			

Table 9.5: MLS Basic Glyph Positions

9.6 Special Characters

For the correct operation of retransliterating systems processing Mongolian script additional symbols are needed. These include Form Variant Selectors (FVS), the Vowel Separator, and other symbols like the Mongolian Positional Indicator. As can be seen from its usage in table 9.1, entering `*ng` tells the system to consider this *ng* to be in non-initial position.[3]

Besides these symbols, table 9.6 includes also some useful punctuation marks etc. as they are used in Mongolian Script.

9.7 Displaying Transliterations

For huge word lists and similar material it is convenient to enter the transliteration only once and use it as input both for the Mongolian retransliteration engine and the presentation of the transliteration. A construct like

```
\newcommand{\Keyword}[1]{#1 \bcg{#1}}
\Keyword{anda} / \emph{Looks nice.}
\Keyword{SaGdur} / \emph{Not as nice.}
```

anda / *Looks nice.*

SaGdur / *Not as nice.*

is helpful as long as no capitalized single-letter entity is used. Capitalized entities look less pleasing in conventional texts; for these purposes, the command `\PrettyMLS` is provided which takes input with single-letters entities and converts it to a more traditional representation.

```
\newcommand{\Keyword}[1]{%
\PrettyMLS{#1} \bcg{#1}}
\Keyword{anda} / \emph{Good.}\par
\Keyword{SaGdur} / \emph{Good again.}
```

anda / *Good.*

šaγdur / *Good again.*

Two additional flags, `\ShowSpecialMLStrue` and `\ShowSpecialMLSfalse`, can be used to activate canonical identifiers instead of the conventional notation for the special characters of table 9.6.

```
\PrettyMLS{SaGdur blam=a}
\ShowSpecialMLStrue
\PrettyMLS{SaGdur blam=a}
\ShowSpecialMLSfalse
\PrettyMLS{SaGdur blam=a}
```

šaγdur blam=a

šaγdur blam MVS a

šaγdur blam=a

The complete set of characters covered by `\PrettyMLS` is shown in table 9.7.

[3]Unfortunately, though it is now commonly agreed in the scientific community that these symbols are needed, their definition is still in a state of flux, and thus the symbols given here are presented on a preliminary basis.

Symbol	Name	Input
!	Exclamation Mark	!
?	Question Mark	?
!?	Exclamation Question Mark	!?
?!	Question Exclamation Mark	?!
$\boxed{*}$	Mong. Positional Indicator	*
$\boxed{\text{MSP}}$	Mongolian Space	-
︵	Opening Bracket	(
︶	Closing Bracket	)
︿	Opening Angle Bracket	<
﹀	Closing Angle Bracket	>
︽	Opening Guillemot	<<
︾	Closing Guillemot	>>
$\boxed{\text{FVS1}}$	Form Variant Selector 1	'
$\boxed{\text{FVS2}}$	Form Variant Selector 2	"
$\boxed{\text{MVS}}$	Mong. Vowel Separator	=
᠊	Mongolian Nuruu	|
᠃	Period	.
᠂	Comma	,
᠄	Colon	:
᠅	Dörwöljin	;
᠁	Ellipsis	..
᠐	Digit zero	0
᠑	Digit one	1
᠒	Digit two	2
᠓	Digit three	3
᠔	Digit four	4
᠕	Digit five	5
᠖	Digit six	6
᠗	Digit seven	7
᠘	Digit eight	8
᠙	Digit nine	9

Table 9.6: Mongolian Script Special Symbols and Punctuation Marks

MonTeX	\ShowSpecialMLS	
Input	true	false
E	ä	ä
O	ö	ö
U	ü	ü
G	γ	γ
S	š	š
-	$\boxed{\text{MSP}}$	-
=	$\boxed{\text{MVS}}$	=
'	$\boxed{\text{FVS1}}$	’
"	$\boxed{\text{FVS2}}$	”
*	$\boxed{*}$	∗

Table 9.7: MLS transliteration restauration

Chapter 10

Manju Input

Manju documents can be compiled with the `bithe` (see section 8.2, page 41) option to the `\usepackage` command, which will create complete documents in Manju. Anywhere in the document, it is possible to switch to Manju input (transliteration mode) with `\SetDocumentEncodingBithe` which internally activates the LMA encoding.

Caveat: Since switching to Manju text requires a lot of settings to be effected at the same time, there are high-level commands available (see below, chapter 11) which do all the work, including the definition of the document encoding. Thus, while `\SetDocumentEncodingBithe` is indeed classified as a user-level command, it is certainly not necessary for everyday work.

10.1 Basic Character Set and Romanization

Given by dictionary order, the system provides a basic character set as shown in table 10.1.

While the input method for the majority of characters matches the

Manju	Input	Latin	Manju	Input	Latin	Manju	Input	Latin
ᠠ	a	a	ᡥ	h	h	ᠴ	c	c
ᡝ	e	e	ᠪ	b	b	ᠵ	j	j
ᡳ	i	i	ᡦ	p	p	ᠶ	y	y
ᠣ	o	o	ᠰ	s	s	ᠺ	k'	k'
ᡠ	u	u	ᡧ	s'	š	ᡬ	g'	g'
ᡡ	v	ū	ᡨ	t	t	ᡭ	h'	h'
ᠨ	n	n	ᡩ	d	d	ᡵ	r	r
ᡴ	k	k	ᠯ	l	l	ᡶ	f	f
ᡤ	g	g	ᠮ	m	m	ᠸ	w	w

Table 10.1: Manju Basic Character Set

transliteration conventions, some letters require a slightly different treatment:

1. Although the diphtong is usually rendered as *ai*, it must be entered as `aii` in order to produce the desired effect.
2. The vowel which is conventionally rendered as *û* or *ū* can be entered as `v` or as `\={u}` due to the fact that a character *û* is not readily available on most systems.
3. The consonant *š* can be entered as `s'` or as `\v{s}`, but not as *`sh` as to avoid undesired mergers of *s* and *h* like in *ishun* which should not be **išun* !

10.2 Extended Character Set

The following special characters listed in major dictionaries are provided:

Manju	Input	Latin
	sy	sy
	cy	cy
	j'	jy
	dz	dz
	tsh	tsh
	tshy	tshy
	zr	zr

Please note that due to internal limitations of the retransliteration engine, *jy* has to be entered as `j'`.

10.3 Tibetan Transliteration Character Set

Besides these characters, an additional small set of special characters is provided for rendering Tibetan and Uighur transliterations:

Manju	Input	Latin
	z	z
	zh	zh
	ts	ts
	ng'	ng'
	l'	l'
	p'	p'
	t'	t'

ཀ་ ka	g'a	ཁ་ kha	k'a	ག་ ga	ga	ང་ nga	ng'a
ཅ་ ca	jiya	ཆ་ cha	cia	ཇ་ ja	ja	ཉ་ nya	niya
ཏ་ ta	t'a	ཐ་ tha	ta	ད་ da	da	ན་ na	na
པ་ pa	ba	ཕ་ pha	pa	བ་ ba	wa	མ་ ma	ma
ཙ་ tsa	tsa	ཚ་ tsha	tsha	ཛ་ dza	dza	ཝ་ wa	wa
ཞ་ zha	zha	ཟ་ za	za	འ་ '	ea	ཡ་ ya	ya
ར་ ra	ra	ལ་ la	la	ཤ་ sha	s'a	ས་ sa	sa
ཧ་ ha	h|a	ཨ་ a	a				

Table 10.2: Tibetan Transliteration Character Set

This allows to spell out the Tibetan alphabet in Manju writing, as used in the Pentaglot dictionary for Tibetan (see table 10.2) and Uighur transliterations. The following rules apply:

1. ང་ *nga* (ma. *ng'a*) is used for Tibetan initials and subscripts while finals are expressed as (ma. **ng*);
2. While ཧ་ *ha* is used for Tibetan initial ཧ་, a different form is taken for subscripted *ha*, as in ལྷ་ *lha* (ma. *l'a*).

10.3.1 Special Characters

Manju shares with Mongolian the complete set of numbers and punctuation marks as well as a few special characters used for influencing the presentation of the writing. See also section 9.6.

Provided a word should end with a non-final glyph shape then the Environment Marker $\boxed{*}$ is used which is entered as an asterisque `*`. This is helpful for writing abbreviated words or marking non-final vowels, like which is entered as `o*`.

Whenever the plethora of diacritics used in Manju writing causes ugly clashes between adjacent letters, then the 'backbone' (mong. *nirugu*), entered as `|`, can be used to stretch the distance between clashing letter elements, like in which should be entered `h|a` rather than `ha` resulting in .

Chapter 11

Display Commands

Depending on the size of the Mongolian or Manju material to be displayed, the user can choose between various commands and environments which have a similar structure for both Mongolian and Manju.

11.1 Small Portions of Mongolian and Manju in Running Text

For displaying short Mongolian snippets in running text use `\bicig{...}`. `\bicig`

For displaying short Manju snippets in running text use `\bithe{...}`. `\bithe`

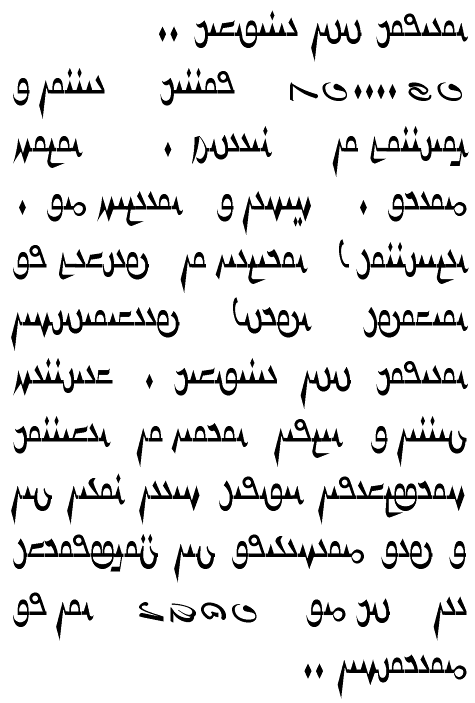
This is ᠮᠣᠩᠭᠣᠯ ᠪᠢᠴᠢᠭ.

That is ᠮᠠᠨᠵᡠ ᠪᡳᡨᡥᡝ.

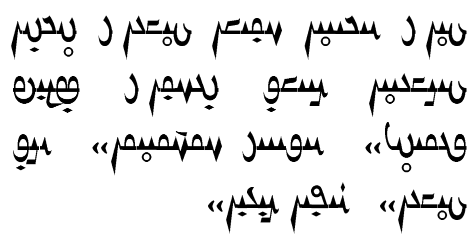
```
This is \bicig{munggul bicik}.
That is \bithe{manju bithe}.
```

11.2 Horizontal Paragraphs of Mongolian or Manju Text

If one needs more than a few words of Mongolian or Manju but does not want to change the line orientation, then the environments `bicigtext` for Mongolian (which should be entered in Mongolian Simplified Transliteration) and `bithetext` for Manju are useful. `bicigtext` `bithetext`



```
\begin{bicigtext}
uindur gegen zanabazar.
17..18 d'ugar zagun-u munggul-un
neiigem, ulus tuiru, shasin-u
uiiles-tu, ilangguy=a uralig-un
kuikzil-du uncukui ekurge
kuiicedgeksen uindur gegen
zanabazar, cinggis xagan-u
aldan urug-un izagur surbulzidan
abadai saiin nuyan xan-u kuiu
tuisiyedu xan gumbudurzi-yin
ger-tu 1635 un-du tuiruksen.
\end{bicigtext}
```



```
\begin{bithetext}
han-i araha sunja
hacin-i hergen kamciha
manju gisun-i buleku
bithe. abkai so\v{s}ohon.
emu hacin. nadan meyen.%
\end{bithetext}
```

11.3 Vertical Capsules

Individual Mongolian and Manju words can be placed vertically anywhere in otherwise horizontal text like in the keyword entry of dictionaries.[1]. The capsule containing the Mongolian or Manju word will automatically request sufficient space so that ugly overlaps with neighbouring lines will not happen.

For presenting text given in broad (or MLS) transliteration, use the command `\mbosoo{...}`; when writing in Mongolian Simplified Transliteration, use `\mobosoo{...}`; likewise for Manju, use `\mabosoo{...}`. All these commands are derived from a command `\bosoo{...}` which places text in ver-

`\mbosoo`
`\mobosoo`
`\mabosoo`
`\bosoo`

[1]Famous dictionaries with a mixture of vertical and horizontal printing are I. J. Schmidt's Mongolian-Russian-German dictionary (1835) and F. Lessing's Mongolian-English dictionary (1960).

This is vertical text. This is ᠮᠣᠩᠭᠣᠯ ᠪᠢᠴᠢᠭ, this is ᠮᠣᠩᠭᠣᠯ ᠪᠢᠴᠢᠭ, that is ᠮᠠᠨᠵᡠ ᠪᡳᡨᡥᡝ.

```
This is \bosoo{vertical}
\bosoo{text}.
This is \mbosoo{mongGol}
\mbosoo{bicig},
this is \mobosoo{munggul}
\mobosoo{bicik},
that is \mabosoo{manju}
\mabosoo{bithe}.
```

Figure 11.1: Vertical Text Capsules

tical capsules but leaves the contents untouched as far as the encoding is concerned.

ᠮᠣᠩᠭᠣᠯ ᠪᠢᠴᠢᠭ without PostScript support Mongolian text enclosed in vertical capsules will be printed *horizontally*!

11.4 Vertical Text Boxes

`\mobox` `\mabox`

For presenting individual paragraphs of Mongolian or Manju text in vertical manner in an otherwise horizontal text, there are the box commands `\mobox{...}{...}` for Mongolian[2] and `\mabox{...}{...}` for Manju. These boxes take two arguments. The first argument indicates the *vertical depth* of the box, or its line length. The second argument contains the desired text. An example is shown in figure 11.2 for Mongolian, and below for Manju.

ᡥᠠᠨ ᡳ ᠠᡵᠠᡥᠠ ᠰᡠᠨᠵᠠ ᡥᠠᠴᡳᠨ ᡳ ᡥᡝᡵᡤᡝᠨ ᡴᠠᠮᠴᡳᡥᠠ ᠮᠠᠨᠵᡠ ᡤᡳᠰᡠᠨ ᡳ ᠪᡠᠯᡝᡴᡠ ᠪᡳᡨᡥᡝ᠉ ᠠᠪᡴᠠᡳ ᡧᠣᡧᠣᡥᠣᠨ᠉ ᡝᠮᡠ ᡥᠠᠴᡳᠨ᠉ ᠨᠠᡩᠠᠨ ᠮᡝᠶᡝᠨ᠉

```
\mabox{3.75cm}{%
\noindent\raggedleft
han-i araha sunja
hacin-i hergen kamciha
manju gisun-i buleku
bithe. abkai so\v{s}ohon.
emu hacin. nadan meyen.%
}
```

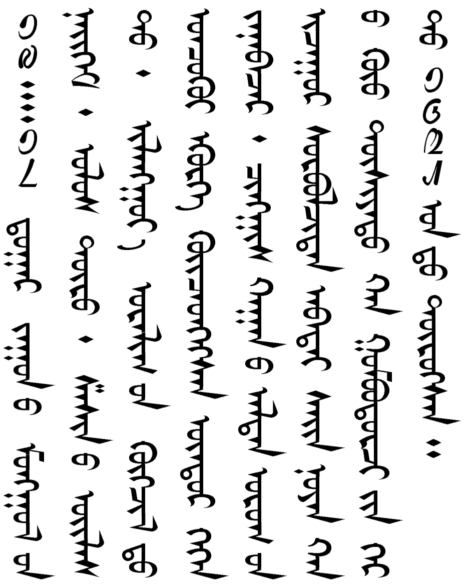
11.5 Full Vertical Text Pages

`bicigpage` `bithepage`

If you need several pages of Mongolian output, enclose your text in an environment `bicigpage`, and use `bithepage` likewise for Manju texts. Note that Mongolian must be entered in Simplified Transliteration.

Finally, if you want the whole document and its basic language to be Classical, or Uighur Mongolian, say `\usepackage[bicig,...]{mls}`. Likewise, complete Manju documents are produced with `\usepackage[bithe,...]{mls}`.

[2] Mongolian input *must* be coded in Mongolian Simplified Transliteration; MLS input won't work.



```
\mobox{7.5cm}{%
17..18 d'ugar zagun-u munggul-un
neiigem, ulus tuiru, shasin-u
uiiles-tu, ilangguy=a uralig-un
kuikzil-du uncukui ekurge
kuiicedgeksen uindur gegen
zanabazar, cinggis xagan-u
aldan urug-un izagur surbulzidan
abadai saiin nuyan xan-u kuiu
tuisiyedu xan gumbudurzi-yin
ger-tu 1635 un-du tuiruksen.%
}
```

Figure 11.2: A Vertical Text Box

If you start a document with a `\usepackage[bicig]{mls}` declaration you can still switch back to Latin by issuing an `\end{bicigpage}` command.

Likewise, if you start a document with a `\usepackage[bithe]{mls}` declaration you can still switch back to Latin by issuing an `\end{bithepage}` command.

The following snippet of Mongolian text is presented in full page mode on the next pages, first in Simplified Transliteration form, then in Uighur form; in order to achieve this result the text had to be included in the environment `bicigpage`.

```
\begin{bicigpage}
uindur gegen zanabazar.
17||18 d'ugar zagun-u munggul-un neiigem, ulus tuiru, shasin-u
uiiles-tu, ilangguy=a uralig-un kuikzil-du uncugui ekurge
kuiicedgeksen uindur gegen zanabazar, cinggis xagan-u aldan
urug-un izagur surbulzidan abadai saiin nuyan xan-u kuiu
tuisiyedu xan gumbudurzi-yin ger-tu 1635 un-du tuiruksen.
badumÖngke daiyan xagan-u 6-d'aki uiy=e-yin kuimun. gurban
nasudai-d'agan num ungsizu enedkek gazar tuibed kele-yi xar=a
ayandagan surcu, keuked axui cag-aca erdem num-un duiri-tei
bulugsan zanabazar 15 nasu-tai-dagan baragun zuu (lhasa)
uruzu tabudugar dalai lam=a-d'u shabilan saguzu, ulamar
zebCundamba-yin xubilgan tudurazei. uran barimalci, zirugaci,
kele sinzigeci, uran barilgaci, kuin uxagandan zanabazar ulan
zagun zil-un daiin tululdugan-d'u nerbekden suliduzu, zugsunggi
baiidal-d'u urugsan dumdadu zagun-u munggul-un suyul uralig-i
serkun manduxu-d'u yeke xubi nemekuri urugulugsan yum. tekun-u
abiyas bilig nuiri yeke kuidelmuri-ber munggul-un uralig nigen
uiy=e tanigdasi uigei uindurlik-tu kuiruksen azei. xarin 1654
un-d'u neiislel kuiriyen-u tulg=a-yin cilagu-yi tabilcagsan
zanabazar-un uran barilg=a-yin buidugel-ece uinudur-i uizeksen
zuiil barug uigei ni xaramsaldai. zanabazar uindesun-u bicig
uisuk-i kuikzikulku-d'u beyecilen urulcazu, suyungbu uisuk-i
zukiyazu ene uiy=e suyungbu ni man-u tusagar tugdanil-un belge
temdek bulugsagar baiin=a.  tere-ber <<cag-i tukinagulugci>>
gedek silukleksen zukiyal-d'agan arad tuimen-u-ben engke
amugulang, saiin saiixan-i imagda kuisen muirugedezu yabudag
sedkil-un-iien uige-i ilerkeiileksen baiidag. uindur gegen
duirsuleku uralig-un xubi-d'u uirun=e-yin sunggudag-ud-tai
eng zergeceku buidugel-tei kuimun abacu basa xari ulus-un
buzar bacir arg=a-d'u abdagdan yabugsan nigen.
...
... more text ...
              ...
\end{bicigpage}
```

Figure 11.3: Input Example of a Mongolian Text

ᠠᠷᠠᠳᠴᠢᠯᠠᠯ ᠤᠨ ᠬᠤᠪᠢᠰᠬᠠᠯ ᠃

[illegible]

[illegible]

11.6 Pure Uighur Mongolian and Manju Documents

Writing a complete document in Mongolian or Manju is as simple and straightforward as writing a document in English or Xalx Mongolian.

The example file, `zanabazr.tex` (shipped together with this documentation and located in the directory `../examples/`) demonstrates how a pure Mongolian Bicig document can be created.

```
\documentclass{article}
\usepackage[bicig]{mls}
\begin{document}
uindur gegen zanabazar.
17||18 d'ugar zagun-u munggul-un neiigem, ulus tuiru,
shasin-u uiiles-tu, ilangguy=a uralig-un kuikzil-du
...
... more text ...
              ...
\end{document}
```

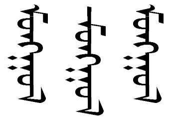
The concept is the same for Manju documents: instead of `bicig` one would use the `\usepackage[...]{mls}` option `bithe` and enter Manju text.

11.7 Font Selection Commands

There are two distinct styles of Mongolian script: one style is typically used for modern print, whereas the other style appears in old block prints and stone inscriptions.

Since there is no proper correspondance between Latin and Mongolian typographical features, a somewhat arbitrary assignment was made to the effect that the block print style can be activated by setting the font family sans serif with `\sffamily`. In contrast, setting the roman default family with `\rmfamily` switches back to the modern style. `\sffamily` `\rmfamily`

The same two commands can be applied to Manju, too. In this context it makes sense to assign, e. g., `\sffamily` to Mongolian and `\rmfamily` (which is the default anyway) to Manju. At one glance one can tell which writing represents which language.



```
\mobox{2cm}{\noindent
munggul\\
\sffamily munggul\\
\rmfamily munggul}
```

Figure 11.4: Mongolian Font Styles

Nota Bene: The MLS-related Mongolian display commands are internally limited to the sans serif, or block print style, so that there is always a clear visual distinction possible which input mode was chosen.

Chapter 12

MonTEX Software Internals

12.1 MonTEX System Layout

MonTEX consists of many files each performing dedicated functions. These files are listed here in systematical order.

12.1.1 Main Package

The main package is `mls.sty`. The RL capabilities are provided by `rlbicig.sty`.

12.1.2 Hyphenation Patterns

```
texinput/mnhyphen.tex   # Modern Mongolian
texinput/mnhyphex.tex   # Modern Mongolian Exceptions
```

12.1.3 Transliteration Engines

```
texinput/mlstrans.tex   # Main Transliteration Engine
texinput/mlsgalig.tex   # Latin Presentation Engine
```

12.1.4 Input Encodings

```
texinput/cpctt.def
texinput/cpdbk.def
texinput/cpibmrus.def
texinput/cpkoi.def
texinput/cpmls.def
texinput/cpmnk.def
texinput/cpmos.def
texinput/cpncc.def
```

12.1.5 Output or Font Encodings

```
texinput/lmaenc.def     # Manju
texinput/lmcenc.def     # Cyrillic
texinput/lmoenc.def     # Mongolian (Simplified Input)
texinput/lmsenc.def     # Mongolian Script (deprecated)
texinput/lmuenc.def     # Traditional Mongolian Glyph Container
```

12.1.6 Caption Translations

```
texinput/bicig.def      # Mongolian
texinput/bithe.def      # Manju
texinput/buryat.def     # Buryat
texinput/english.def    # English
texinput/kazakh.def     # Kazakh, implementation pending
texinput/russian.def    # Russian
texinput/xalx.def       # Modern Mongolian
```

12.1.7 Font Definitions

```
texinput/lmabthhs.fd    # Manju horizontal 'steel'
texinput/lmabthhw.fd    # Manju horizontal 'wood'
texinput/lmabthvs.fd    # Manju vertical 'steel'
texinput/lmabthvw.fd    # Manju vertical 'wood'
texinput/lmccmdh.fd     # Cyrillic Dunhill
texinput/lmccmfib.fd    # Cyrillic Fibonacci
texinput/lmccmfr.fd     #
texinput/lmccmiss.fd    #
texinput/lmccmr.fd      # Cyrillic CM Roman
texinput/lmccmss.fd     # Cyrillic CM Sans Serif
texinput/lmccmssq.fd    # Cyrillic CM Sans Serif Quotes
texinput/lmccmtt.fd     # Cyrillic CM TeleType
texinput/lmccmvtt.fd    # Cyrillic CM Variable TeleType
texinput/lmclcmss.fd    #
texinput/lmobcghs.fd    # Mongolian horizontal 'steel'
texinput/lmobcghw.fd    # Mongolian horizontal 'wood'
texinput/lmobcgvs.fd    # Mongolian vertical 'steel'
texinput/lmobcgvw.fd    # Mongolian vertical 'wood'
texinput/lmsbcgh.fd     #
texinput/lmsbcgv.fd     #
texinput/lmubxghs.fd    # Glyph container horizontal 'steel'
texinput/lmubxghw.fd    # Glyph container horizontal 'wood'
texinput/lmubxgvs.fd    # Glyph container vertical 'steel'
texinput/lmubxgvw.fd    # Glyph container vertical 'wood'
```

12.1.8 Miscellae

```
texinput/mtdocmac.tex   # Macro collection for this document
texinput/TODO           # The Author's To Do List
```

12.2 MonTEX Mongolian Font Layout

Mongolian and Manju fonts are generated from common sources in `mfinput/bcgbase`. Mongolian-specific material is kept in `mfinput/bicig`, Manju-specific material is kept in `mfinput/bithe`. All Mongolian fonts can be used for RL *and* for LR typesetting. Individual font names are best described by the following regular expression:

$$(bcg|bth|bxg)[hv][sw][mb]$$

Here, *bcg* stands for Mongolian, *bth* for Manju and *bxg* for the generic Mongolian glyph container. The next letter indicates whether the material is to be typeset *h*orizontally or *v*ertically. The next letter indicates the typeface: *s*teel or *w*. The last letter indicates a *m*edium or *b*old font.

12.3 bxg: A Generic Mongolian Glyph Container

Besides dedicated fonts for Mongolian and Manju, the MonTEX font system offers a generic glyph container which is accessible through the LMU encoding. The name of this glyph container is `bxg`, and all glyphs (the superset of Mongolian and Manju) are available in both font families (block print and modern print styles) of the LMO and LMA encodings. Please note that at present there is no working ligature mechanism associated with `bxg`; hence it cannot be used for general-purpose text at the moment.

In the future, the `bxg` generic glyph container will manage the Unicode interface.

12.4 Unicode Mongolian and MonTEX

In the present version, a first attempt was made to provide Unicode compatibility. Please note that at this stage the Unicode of MonTEX is purely experimental!

Unicode-encoded Traditional Mongolian is located at plane `U+1800` and contains canonical characters for Mongolian, Sibe, Manju and Todo. There is also a rich collection of Ali Gali (or Galig) characters used for transliterating Sanskrit, Tibetan and other languages into Mongolian, Manju etc.

MonTEX covers a subset of Unicode Traditional Mongolian which is sufficient to typeset modern Mongolian and Manju texts as well as a choice of Tibetan words transliterated in Manju (as in the Pentaglot dictionary, e. g.).

At the moment, language-specific groups of Unicode characters are mirrored into the related encodings. It is therefore necessary to tag the desired language with the `SetDocumentEncoding<...>` command in order to achieve the appropriate ligature behaviour.

The availability of individual Unicode Mongolian characters and their canonical names are shown in table 12.1.

The astute observer will note several discrepancies between the official Unicode standard documentation and this particular, *experimental* implementation:

1. Canonical letter shapes differ from those shown in the standard documentation. In context, however, the characters behave as they should.
2. There is not yet a third MVS in MonTEX.
3. The Mongolian front vowels are not yet treated properly.
4. The complete Todo range of characters is missing.
5. Most of the Mongolian Ali Gali (Galig) characters are missing; there are, however some Manju Galig characters.
6. Unicode decided to choose a special space to separate morpheme boundaries; this character is *not* part of the Traditional Mongolian plane (sic!). Also, this character is defined as a non-breaking space, which contrasts with the understanding of the MonTEX authors.

This list of differences between Unicode Traditional Mongolian and MonTEX Mongolian and Manju is incomplete.

Code	Character	Unicode Name MonTEX Name	MonTEX Encoding
U+1800		(MONGOLIAN BIRGA) \textmongolianbirga	(n. a.)
U+1801	᠁	MONGOLIAN ELLIPSIS \textmongolianellipsis	LMO, LMA
U+1802	᠂	MONGOLIAN COMMA \textmongoliancomma	LMO
U+1803	᠃	MONGOLIAN FULL STOP \textmongolianfullstop	LMO
U+1804	᠄	MONGOLIAN COLON \textmongoliancolon	LMO, LMA
U+1805	᠅	MONGOLIAN FOUR DOTS \textmongolianfourdots	LMO, LMA
U+1806		(MONGOLIAN TODO SOFT HYPHEN) \textmongoliantodosofthyphen	(n. a.)
U+1807		(MONGOLIAN SIBE SYLLABLE BOUNDARY MARKER) \textmongoliansibesyllableboundarymarker	(n. a.)

U+1808	᠈	MONGOLIAN MANCHU COMMA \textmongolianmanchucomma	LMA
U+1809	᠉	MONGOLIAN MANCHU FULL STOP \textmongolianmanchufullstop	LMA
U+180A	᠊	MONGOLIAN NIRUGU \textmongoliannirugu	LMO, LMA
U+180B	FVS1	MONGOLIAN FREE VARIATION SELECTOR ONE \textmongolianfreevariationselectorone	LMO, LMA
U+180C	FVS2	MONGOLIAN FREE VARIATION SELECTOR TWO \textmongolianfreevariationselectortwo	LMO, LMA
U+180D		(MONGOLIAN FREE VARIATION SELECTOR THREE) \textmongolianfreevariationselectorthree	(n. a.)
U+180E	MVS	MONGOLIAN VOWEL SEPARATOR \textmongolianvowelseparator	LMO, LMA
U+1810	᠐	MONGOLIAN DIGIT ZERO \textmongolianzero	LMO, LMA
U+1811	᠑	MONGOLIAN DIGIT ONE \textmongolianone	LMO, LMA
U+1812	᠒	MONGOLIAN DIGIT TWO \textmongoliantwo	LMO, LMA
U+1813	᠓	MONGOLIAN DIGIT THREE \textmongolianthree	LMO, LMA
U+1814	᠔	MONGOLIAN DIGIT FOUR \textmongolianfour	LMO, LMA
U+1815	᠕	MONGOLIAN DIGIT FIVE \textmongolianfive	LMO, LMA
U+1816	᠖	MONGOLIAN DIGIT SIX \textmongoliansix	LMO, LMA
U+1817	᠗	MONGOLIAN DIGIT SEVEN \textmongolianseven	LMO, LMA
U+1818	᠘	MONGOLIAN DIGIT EIGHT \textmongolianeight	LMO, LMA
U+1819	᠙	MONGOLIAN DIGIT NINE \textmongoliannine	LMO, LMA
U+1820	ᠠ	MONGOLIAN LETTER A \textmongoliana	LMO, LMA
U+1821	ᠡ	MONGOLIAN LETTER E \textmongoliane	LMO
U+1822	ᠢ	MONGOLIAN LETTER I \textmongoliani	LMO, LMA
U+1823	ᠣ	MONGOLIAN LETTER O \textmongoliano	LMO, LMA
U+1824	ᠤ	MONGOLIAN LETTER U \textmongolianu	LMO

U+1825		MONGOLIAN LETTER OE \textmongolianoe	LMO
U+1826		MONGOLIAN LETTER UE \textmongolianue	LMO
U+1827		MONGOLIAN LETTER EE \textmongolianee	LMO
U+1828	ᠨ	MONGOLIAN LETTER NA \textmongolianna	LMO, LMA
U+1829	ᠩ	MONGOLIAN LETTER ANG \textmongolianang	LMO, LMA
U+182A	ᠪ	MONGOLIAN LETTER BA \textmongolianba	LMO, LMA
U+182B	ᠫ	MONGOLIAN LETTER PA \textmongolianpa	LMO
U+182C	ᠬ	MONGOLIAN LETTER QA \textmongolianqa	LMO
U+182D	ᠭ	MONGOLIAN LETTER GA \textmongolianga	LMO
U+182E	ᠮ	MONGOLIAN LETTER MA \textmongolianma	LMO, LMA
U+182F	ᠯ	MONGOLIAN LETTER LA \textmongolianla	LMO, LMA
U+1830	ᠰ	MONGOLIAN LETTER SA \textmongoliansa	LMO, LMA
U+1831	ᠱ	MONGOLIAN LETTER SHA \textmongoliansha	LMO
U+1832	ᠲ	MONGOLIAN LETTER TA \textmongolianta	LMO
U+1833	ᠳ	MONGOLIAN LETTER DA \textmongolianda	LMO
U+1834	ᠴ	MONGOLIAN LETTER CHA \textmongoliancha	LMO, LMA
U+1835	ᠵ	MONGOLIAN LETTER JA \textmongolianja	LMO, LMA
U+1836	ᠶ	MONGOLIAN LETTER YA \textmongolianya	LMO, LMA
U+1837	ᠷ	MONGOLIAN LETTER RA \textmongolianra	LMO
U+1838	ᠸ	MONGOLIAN LETTER WA \textmongolianwa	LMO
U+1839	ᠹ	MONGOLIAN LETTER FA \textmongolianfa	LMO
U+183A	ᠺ	MONGOLIAN LETTER KA \textmongolianka	LMO

U+183B	ᠻ	MONGOLIAN LETTER KHA	LMO
		\textmongoliankha	
U+183C	ᠼ	MONGOLIAN LETTER TSA	LMO
		\textmongoliantsa	
U+183D	ᠽ	MONGOLIAN LETTER ZA	LMO
		\textmongolianza	
U+183E	ᠾ	MONGOLIAN LETTER HAA	LMO
		\textmongolianhaa	
U+183F	ᠿ	MONGOLIAN LETTER ZRA	LMO
		\textmongolianzra	
U+1840	ᡀ	MONGOLIAN LETTER LHA	LMO
		\textmongolianlha	
U+1841	ᡁ	MONGOLIAN LETTER ZHI	LMO
		\textmongolianzhi	
U+1842	ᡂ	MONGOLIAN LETTER CHI	LMO
		\textmongolianchi	
U+1843		(MONGOLIAN LETTER TODO LONG VOWEL SIGN)	(n. a.)
		\textmongoliantodolongvowelsign	
U+1844		(MONGOLIAN LETTER TODO E)	(n. a.)
		\textmongoliantodoe	
U+1845		(MONGOLIAN LETTER TODO I)	(n. a.)
		\textmongoliantodoi	
U+1846		(MONGOLIAN LETTER TODO O)	(n. a.)
		\textmongoliantodoo	
U+1847		(MONGOLIAN LETTER TODO U)	(n. a.)
		\textmongoliantodou	
U+1848		(MONGOLIAN LETTER TODO OE)	(n. a.)
		\textmongoliantodooe	
U+1849		(MONGOLIAN LETTER TODO UE)	(n. a.)
		\textmongoliantodoue	
U+184A		(MONGOLIAN LETTER TODO ANG)	(n. a.)
		\textmongoliantodoang	
U+184B		(MONGOLIAN LETTER TODO BA)	(n. a.)
		\textmongoliantodoba	
U+184C		(MONGOLIAN LETTER TODO PA)	(n. a.)
		\textmongoliantodopa	
U+184D		(MONGOLIAN LETTER TODO QA)	(n. a.)
		\textmongoliantodoqa	
U+184E		(MONGOLIAN LETTER TODO GA)	(n. a.)
		\textmongoliantodoga	
U+184F		(MONGOLIAN LETTER TODO MA)	(n. a.)
		\textmongoliantodoma	
U+1850		(MONGOLIAN LETTER TODO TA)	(n. a.)
		\textmongoliantodota	

U+1851		(MONGOLIAN LETTER TODO DA) \textmongoliantododa	(n. a.)
U+1852		(MONGOLIAN LETTER TODO CHA) \textmongoliantodocha	(n. a.)
U+1853		(MONGOLIAN LETTER TODO JA) \textmongoliantodoja	(n. a.)
U+1854		(MONGOLIAN LETTER TODO TSA) \textmongoliantodotsa	(n. a.)
U+1855		(MONGOLIAN LETTER TODO YA) \textmongoliantodoya	(n. a.)
U+1856		(MONGOLIAN LETTER TODO WA) \textmongoliantodowa	(n. a.)
U+1857		(MONGOLIAN LETTER TODO KA) \textmongoliantodoka	(n. a.)
U+1858		(MONGOLIAN LETTER TODO GAA) \textmongoliantodogaa	(n. a.)
U+1859		(MONGOLIAN LETTER TODO HAA) \textmongoliantodohaa	(n. a.)
U+185A		(MONGOLIAN LETTER TODO JIA) \textmongoliantodojia	(n. a.)
U+185B		(MONGOLIAN LETTER TODO NIA) \textmongoliantodonia	(n. a.)
U+185C		(MONGOLIAN LETTER TODO DZA) \textmongoliantododza	(n. a.)
U+185D	ᡝ	MONGOLIAN LETTER SIBE E \textmongoliansibee	LMA
U+185E		(MONGOLIAN LETTER SIBE I) \textmongoliansibei	(n. a.)
U+185F		(MONGOLIAN LETTER SIBE IY) \textmongoliansibeiy	(n. a.)
U+1860	ᡠ	MONGOLIAN LETTER SIBE UE \textmongoliansibeue	LMA
U+1861	ᡡ	MONGOLIAN LETTER SIBE U \textmongoliansibeu	LMA
U+1862	ᡢ	MONGOLIAN LETTER SIBE ANG \textmongoliansibeang	LMA
U+1863		(MONGOLIAN LETTER SIBE KA) \textmongoliansibeka	(n. a.)
U+1864	ᡤ	MONGOLIAN LETTER SIBE GA \textmongoliansibega	LMA
U+1865	ᡥ	MONGOLIAN LETTER SIBE HA \textmongoliansibeha	LMA
U+1866	ᡦ	MONGOLIAN LETTER SIBE PA \textmongoliansibepa	LMA

U+1867	ᡧ	MONGOLIAN LETTER SIBE SHA \textmongoliansibesha	LMA
U+1868	ᡨ	MONGOLIAN LETTER SIBE TA \textmongoliansibeta	LMA
U+1869	ᡩ	MONGOLIAN LETTER SIBE DA \textmongoliansibeda	LMA
U+186A		(MONGOLIAN LETTER SIBE JA) \textmongoliansibeja	(n. a.)
U+186B		MONGOLIAN LETTER SIBE FA \textmongoliansibefa	LMA
U+186C	ᡬ	MONGOLIAN LETTER SIBE GAA \textmongoliansibegaa	LMA
U+186D	ᡭ	MONGOLIAN LETTER SIBE HAA \textmongoliansibehaa	LMA
U+186E	ᡮ	MONGOLIAN LETTER SIBE TSA \textmongoliansibetsa	LMA
U+186F	ᡯ	MONGOLIAN LETTER SIBE ZA \textmongoliansibeza	LMA
U+1870	ᡰ	MONGOLIAN LETTER SIBE RAA \textmongoliansiberaa	LMA
U+1871	ᡱ	MONGOLIAN LETTER SIBE CHA \textmongoliansibecha	LMA
U+1872		(MONGOLIAN LETTER SIBE ZHA) \textmongoliansibezha	(n. a.)
U+1873		(MONGOLIAN LETTER MANCHU I) \textmongolianmanchui	(n. a.)
U+1874	ᡴ	MONGOLIAN LETTER MANCHU KA \textmongolianmanchuka	LMA
U+1875	ᡵ	MONGOLIAN LETTER MANCHU RA \textmongolianmanchura	LMA
U+1876	ᡶ	MONGOLIAN LETTER MANCHU FA \textmongolianmanchufa	LMA
U+1877	ᡷ	MONGOLIAN LETTER MANCHU ZHA \textmongolianmanchuzha	LMA
U+1880	ᢀ	MONGOLIAN LETTER ALI GALI ANUSVARA ONE \textmongolianaligalianusvaraone	LMO
U+1881		(MONGOLIAN LETTER ALI GALI VISARGA ONE) \textmongolianaligalivisargaone	(n. a.)
U+1882		(MONGOLIAN LETTER ALI GALI DAMARU) \textmongolianaligalidamaru	(n. a.)
U+1883		(MONGOLIAN LETTER ALI GALI UBADAMA) \textmongolianaligaliubadama	(n. a.)

U+1884	(MONGOLIAN LETTER ALI GALI INVERTED UBADAMA) \textmongolianaligaliinvertedubadama	(n. a.)
U+1885	(MONGOLIAN LETTER ALI GALI BALUDA) \textmongolianaligalibaluda	(n. a.)
U+1886	(MONGOLIAN LETTER ALI GALI THREE BALUDA) \textmongolianaligalithreebaluda	(n. a.)
U+1887	(MONGOLIAN LETTER ALI GALI A) \textmongolianaligalia	(n. a.)
U+1888	(MONGOLIAN LETTER ALI GALI I) \textmongolianaligalii	(n. a.)
U+1889	(MONGOLIAN LETTER ALI GALI KA) \textmongolianaligalika	(n. a.)
U+188A	(MONGOLIAN LETTER ALI GALI NGA) \textmongolianaligalinga	(n. a.)
U+188B	(MONGOLIAN LETTER ALI GALI CA) \textmongolianaligalica	(n. a.)
U+188C	(MONGOLIAN LETTER ALI GALI TTA) \textmongolianaligalitta	(n. a.)
U+188D	(MONGOLIAN LETTER ALI GALI TTHA) \textmongolianaligalittha	(n. a.)
U+188E	(MONGOLIAN LETTER ALI GALI DDA) \textmongolianaligalidda	(n. a.)
U+188F	(MONGOLIAN LETTER ALI GALI NNA) \textmongolianaligalinna	(n. a.)
U+1890	(MONGOLIAN LETTER ALI GALI TA) \textmongolianaligalita	(n. a.)
U+1891	(MONGOLIAN LETTER ALI GALI DA) \textmongolianaligalida	(n. a.)
U+1892	(MONGOLIAN LETTER ALI GALI PA) \textmongolianaligalipa	(n. a.)
U+1893	(MONGOLIAN LETTER ALI GALI PHA) \textmongolianaligalipha	(n. a.)
U+1894	(MONGOLIAN LETTER ALI GALI SSA) \textmongolianaligalissa	(n. a.)
U+1895	(MONGOLIAN LETTER ALI GALI ZHA) \textmongolianaligalizha	(n. a.)
U+1896	(MONGOLIAN LETTER ALI GALI ZA) \textmongolianaligaliza	(n. a.)
U+1897	(MONGOLIAN LETTER ALI GALI AH) \textmongolianaligaliah	(n. a.)
U+1898	(MONGOLIAN LETTER TODO ALI GALI TA) \textmongoliantodoaligalita	(n. a.)

U+1899		(MONGOLIAN LETTER TODO ALI GALI ZHA) \textmongoliantodoaligalizha	(n. a.)
U+189A		(MONGOLIAN LETTER MANCHU ALI GALI GHA) \textmongolianmanchualigaligha	(n. a.)
U+189B	ᢛ	MONGOLIAN LETTER MANCHU ALI GALI NGA \textmongolianmanchualigalinga	LMA
U+189C	ᢜ	MONGOLIAN LETTER MANCHU ALI GALI CA \textmongolianmanchualigalica	LMA
U+189D		(MONGOLIAN LETTER MANCHU ALI GALI JHA) \textmongolianmanchualigalijha	(n. a.)
U+189E		(MONGOLIAN LETTER MANCHU ALI GALI TTA) \textmongolianmanchualigalitta	(n. a.)
U+189F		(MONGOLIAN LETTER MANCHU ALI GALI DDHA) \textmongolianmanchualigaliddha	(n. a.)
U+18A0		(MONGOLIAN LETTER MANCHU ALI GALI TA) \textmongolianmanchualigalita	(n. a.)
U+18A1		(MONGOLIAN LETTER MANCHU ALI GALI DHA) \textmongolianmanchualigalidha	(n. a.)
U+18A2		(MONGOLIAN LETTER MANCHU ALI GALI SSA) \textmongolianmanchualigalissa	(n. a.)
U+18A3		(MONGOLIAN LETTER MANCHU ALI GALI CYA) \textmongolianmanchualigalicya	(n. a.)
U+18A4	ᢤ	MONGOLIAN LETTER MANCHU ALI GALI ZHA \textmongolianmanchualigalizha	LMA
U+18A5	ᢥ	MONGOLIAN LETTER MANCHU ALI GALI ZA \textmongolianmanchualigaliza	LMA
U+18A6		(MONGOLIAN LETTER ALI GALI HALF U) \textmongolianaligalihalfu	(n. a.)
U+18A7		(MONGOLIAN LETTER ALI GALI HALF YA) \textmongolianaligalihalfya	(n. a.)
U+18A8		(MONGOLIAN LETTER MANCHU ALI GALI BHA) \textmongolianmanchualigalibha	(n. a.)
U+18A9		(MONGOLIAN LETTER ALI GALI DAGALGA) \textmongolianaligalidagalga	(n. a.)

Table 12.1: Unicode Mongolian Code Positions and Associated Commands

Chapter 13

External Support Software

13.1 MLS Software

With MonTEX, it is still possible to process documents generated with the MLS software package. The MLS converter produces Cyrillic and Mongolian Script texts out of transliterations using the MLS codepage. Documents encoded in MLS can be directly processed, no further conversion is necessary.

13.2 Simplified Transliteration Converter

The directory `../source/` contains a small MLS to Simplified Transliteration converter written in Perl. This file can be used directly if Perl exists on your system. Perl is available under a Public Licence for a huge variety of platforms. Consult CPAN (`www.cpan.org`) for information and downloads.

Chapter 14

Shortcomings, Bugs and Desiderata

14.1 Hyphenation Patterns

The Mongolian hyphenation patterns delivered with MonTEX are still under development, so please expect occasional hyphenation errors. It must be also noted that for proper hyphenation of critical words š should be entered as `\sh`, not as `sh` since the first is a character command processed by LATEX 2ε while the latter is a ligature processed by Metafont. If a wrong hyphenation is spotted please check first whether the word in question contains ligature statements (`sh`, `ya`, `yu` etc.) which should then replaced by the proper character commands. It is usually sufficient to add a leading back slash and include the entity in braces: `{\ya}` is as good as `\ya␣`.

Hyphenation patterns for Russian exist but are still to be re-encoded in LMC encoding; Buryat hyphenation rules are still to be defined.

14.2 Retransliteration Problems

Apart from being incomplete as far as some arcane writing variants are concerned, the MLS (Broad Romanization) retransliteration engine provided with MonTEX has two serious shortcomings. Firstly, the input can only consist of letters, punctuation marks and numbers. Any TEX or LATEX command (including `\"a` for *ä* and friends) makes the retransliteration engine fail. Secondly, for large quantities of text, the retransliteration system is agonizingly slow.

The Simplified Transliteration is incorporated into a fontencoding, LMO, which can be selected as default encoding. This allows for complete freedom of all LATEX commands but requires an initial amount of practise.

14.3 Missing Caption Definitions

The translated captions provided with MonTEX are not completely translated at the moment. Notably `\ccname` and `\headtoname` are missing in Mongolian and Buryat, mainly due to grammatical reasons. This will be fixed in later versions.

14.4 Page Headers and Text Encodings

In rare cases it is possible that a `\section` text appearing in a header or footer which is supposed to be typeset in Cyrillic letters is output in Latin letters. This happens if the text on that very page contains encoding selection commands which happen to fall near the page boundary. The only remedy is to enclose the argument text in an additional `{\mnr ...}` command (or vice versa for Latin).

14.5 The `kminch` Font

The Cyrillic typefaces of MonTEX are completed by inch-high sans serif capital letters good for book titles etc. Unfortunately, they cannot be used orthogonally with the other fonts in T1 (Latin characters) and LMC (Cyrillic characters) encoding since their definition is based on TEX primitives rather than LATEX 2ε's NFSS font selection scheme.

14.6 Problems with PostScript Fonts

Any attempt to compile this document with `pslatex` or declare `\usepackage{pslatex}` in the preamble works for the bulk of the document but reduces the Cyrillic typefaces to Roman only (see tables 7.2 and 7.4) and eliminates some of the transliteration symbols (see table 7.5). A solution has not yet been defined.

14.7 Error Message regarding `\selectlanguage`

There seem to be differences in the nature of installed LATEX 2ε platforms; emtex shows a behaviour different from teTeX with regard to pre-loaded language options. On teTeX systems, no error message concerning the redefinition of the `selectlanguage` command occurs, on emtex systems such a message may occur if no other language support packages are loaded. This error message can be safely ignored but the author hopes to find a solution later.

14.8 Printer Memory Overflow

Depending on the printing system it may happen that a Printer Memory Overflow message is generated when attempting to print this text. So far, this happened only on emtex systems running on plain DOS. This is an exceptional situation caused by the very high number of fonts used for this document. The error message never occurred on systems with PostScript postprocessing of the DVI file.

It is very simple and straightforward to reduce the number of typeface examples of this document. Near the beginning of the source file of this very text, the lines

```
% If emtex goofs with (printer) memory overflow
% when attempting to print this document then
% set the following number to "1", recompile and
% increase the number step by step until all
% examples are printed. The maximum value is 6.
\newcounter{FontSamples}
\setcounter{FontSamples}{6} % <--- Modify this number!
```

can be found. It is now possible to increase the number of printed typeface samples step by step until either memory saturation is reached or the system manages to print all fonts. In addition, it should be noted that printing this documentation for the first time may take some time until all fonts are computed.

14.9 Error Reports

Time is a most precious resource and one of the main reasons why the authors decided *not* to support other environments than LaTeX 2_ε. If MonTeX does not work for you because you use a LaTeX2.09 installation, or expect to find a working system for plain TeX support, then the author cannot assist you beyond the advice to install the most recent version of LaTeX 2_ε.

If you find a bug or think a feature is missing which you'd like to see included then your comments are most welcome. One of the authors can be reached by e-mail (`corff@zedat.fu-berlin.de`), and available updates will appear in Infosystem Mongolei (`http://userpage.fu-berlin.de/~corff`). Please check the MLS directory for available releases and patches.

14.10 Outlook and Desiderata

Unfortunately, some code positions in the Metafont sources of MonTeX haven't been frozen yet. In addition, the authors are not happy yet with some of the interaction performed by certain glyph combinations. This will have to be refined definitely! Last but not least, some of the font metrics will undergo further tuning which all implies that documents containing Mongolian or Manju text should be recompiled once a new version of this software is issued.

With Ωmega lurking around, MonTEX should actually be obsolete work. A unified encoding comprising all Mongolian writings has been integrated into Unicode 3.0 and ISO 10646. The author needed a quick solution for ongoing lexicographical work (the Pentaglot database, that is) and will merge Unicode support with the existing MonTEX system later. At a future point, there will also be full-featured Ωmega support.

Anyway, whatever the mistakes and the shortcomings are that have crept into this system, I can only kindly ask you to blame me.

Мигжэд Жанрайсиг Бурхны мэлмий нээсэн ойн өлзийтэй өдөр бичив.

Now go forth and create beautiful Manju text!
Oliver Corff, Shenyang, April 1st, 2001

Part IV

Commands in Alphabetical Order

Chapter 15

Alphabetical Command Reference

All user level commands available in MonTEX are given here in alphabetical order. Every entry in the following list has up to seven sections which are only present if necessary:

Synopsis shows the usage of the command;

Function states its purpose and function;

Limitations in functional range, allowed input etc. are stated here;

Comments and additional information about purpose and nature of the command;

Related commands in the command reference;

See page of the main text;

Example shows a typical application. If several related commands have the same usage and command syntax, then only one example is given which is typically found at the first place a command is mentioned. One example is the command for numbering by letters: The commands `\Asbuk`, `\Useg` and `\Uzeg` are similar, and an example is only given under `\Asbuk`.

15.1 \Asbuk

Synopsis `\Asbuk`{*<number>*}

Function *(Command)* Provides counting by upper case Cyrillic letters, Russian style.

Limitations *<number>* must be between 1 and 28.

Related commands `\asbuk \Useg \useg \Uzeg \uzeg`

See page 38

Example

Position 25 is һ) in Buryat, Ц) in Xalx Mongolian and Ш) in Russian.	`Position 25 is` `\Uzeg{25}) in Buryat,` `\Useg{25}) in Xalx Mongolian and` `\Asbuk{25}) in Russian.`

15.2 \asbuk

Synopsis `\asbuk{`*<number>*`}`

Function *(Command)* Provides counting by lower case Cyrillic letters, Russian style.

Limitations *<number>* must be between 1 and 28.

Related commands `\Asbuk \Useg \useg \Uzeg \uzeg`

See page 38

15.3 \bcg

Synopsis `\bcg{`*<text>*`}`

Function *(Command)* Generates Classical Mongolian out of MonTEX-ified MLS transliteration.

Limitations *<text>* can only consist of unexpandable characters; any TEX or LATEX 2_ε command sequence (even those for dotted vowels like `\"a`) make the system derail.

Related commands `\glyphbcg \PrettyMLS`

See page 49

Example

mongγol bicig is ᠮᠣᠩᠭᠣᠯ ᠪᠢᠴᠢᠭ.	`\emph{mong\g ol bicig}` `is \bcg{mongGol bicig}.`

15.4 \bicig

Synopsis `\bicig{`*<text>*`}`

Function *(Command)* Generates Classical Mongolian out of Simplified Transliteration.

Related commands `\bcg \bithe`

See page 59

Example

munggul bicik is ᠮᠣᠩᠭᠣᠯ ᠪᠢᠴᠢᠭ.	`\emph{munggul bicik}` `is \bicig{munggul bicik}.`

15.5 `bicig`

Function *(Environment* or *Option)* Sets document language to Uighur, or Bicig Mongolian.

Limitations Cooperates well only with Simplified Transliteration as its underlying encoding is LMO.

Related commands `bithe buryat english russian xalx`

See page 41

15.6 `bicigpage`

Function *(Environment* or *Option)* Similar to `bithepage`, it provides full pages of vertical Mongolian text.

Limitations Like all commands of the vertical output family, this command requires PostScript support for proper vertical display. In addition, a functional e-LaTeX environment is mandatory.

Mongolian must be entered in Simplified Transliteration.

Related commands `bithepage bicigtext bithetext`

See page 61

15.7 `bicigtext`

Function *(Environment* or *Option)* Similar to `bicigpage`, it provides full paragraphs of Uighur Mongolian text, but in horizontal line orientation.

Limitations Mongolian must be entered in Simplified Transliteration, and a functional e-LaTeX environment is mandatory.

Related commands `bicigpage bithepage bithetext`

See page 59

15.8 `\BicigToday`

Function *(Command)* Provides the date in Uighur Mongolian.

Comments Internal command. Authors should use `\today` which is redefined automatically by the `bicig` option when calling the `mls` package.

Related commands `\BitheToday \BuryatToday \RussianToday \XalxToday`

See page 27

15.9 `\bithe`

Synopsis `\bithe{`*<text>*`}`

Function *(Command)* Generates Manju out of transliterated material.

Related commands `\bicig`

See page 59

Example

manju bithe is ᠮᠠᠨᠵᡠ ᠪᡳᡨᡥᡝ.

```
\emph{manju bithe}
is \bithe{manju bithe}.
```

15.10 `bithe`

Function *(Environment* or *Option)* Sets document language to Manju.

Related commands `bicig buryat english russian xalx`

See page 41

15.11 `bithepage`

Function *(Environment* or *Option)* Similar to `bicigpage`, it provides full pages of vertical Manju text.

Limitations Like all commands of the vertical output family, this command requires PostScript support for proper vertical display. In addition, a functional e-LaTeX environment is mandatory.

Related commands `bicigpage bicigtext bithetext`

See page 61

15.12 `bithetext`

Function *(Environment* or *Option)* Similar to `bithepage`, it provides full pages of Manju text, but in horizontal line orientation.

Limitations A functional e-LaTeX environment is mandatory.

Related commands `bicigpage bithepage bicigtext`

See page 59

15.13 \BitheToday

Function *(Command)* Provides the date in Manju.

Comments Internal command. Authors should use `\today` which is redefined automatically by the `bithe` option when calling the `mls` package.

Related commands `\BicigToday \RussianToday \XalxToday`

See page 27

Example

```
\marginpar{\mabosoo{\BitheToday}}
```

15.14 \bosoo

Synopsis `\bosoo{`*<text>*`}`

Function *(Command)* Prints text in vertical capsules.

Limitations PostScript support is required for presenting the output. The `rotating` package must be installed. If MonTEX cannot find `rotating` encapsulated material will be printed horizontally.

Comments Line spacing etc. adjust automatically. Useful for dictionaries etc.

Related commands `\mabosoo \mbosoo \mobosoo`

See page 60

Example

A vertical word.

```
A \bosoo{vertical} word.
```

15.15 buryat

Synopsis `buryat`

Function *(Environment* or *Option)* Sets document language to Buryat.

Related commands `bicig bithe english russian xalx`

See page 26

15.16 \BuryatToday

Function *(Command)* Provides the date in Buryat.

Comments Internal command. Authors should use `\today` which is redefined automatically by the `buryat` option when calling the `mls` package.

Related commands `\BicigToday \BitheToday \RussianToday \XalxToday`

See page 27

Example

2002 оной июлиин 2-ой үдэр 2 июля 2002 2002 оны долоодугаар сарын 2	`\BuryatToday\par` `\RussianToday\par` `\XalxToday\par`

15.17 \ch

Function *(Command)* Creates a *ch* which is used for Mongolian transliterations.

Related commands `\g \sh`

See page 37

Example

Šagdar and *Čadraa* are transliterations for Шагдар and Чадраа.	`\emph{\Sh agdar} and \emph{\Ch adraa}` `are transliterations for` `{\mnr\Sh agdar} and {\mnr\Ch adraa}.`

15.18 \cminch

Function *(Command)* Produces inch-high bold sans serif latin letters for book titles etc.

Limitations Only capital letters and numbers available.

Comments This command bypasses the NFSS font setup, hence deprecated since the font provided by this command does not orthogonally follow with the font changes of the main document.

Related commands `\kminch`

See page 35

15.19 \CYR

Synopsis `\CYR`{*<letter name>*}

Function *(Command)* Allows writing of Cyrillic letters in non-Cyrillic environments without changing the document language.

Limitations *letter name* must be one of A, B, V, G, D, E, YO, ZH, Z, I, ISHRT, K, L, M, N, O, OTLD, P, R, S, T, U, Y, F, H, HSHA, C, CH, SH, SHCH, HRDSN, ERY, SFTSN, EREV, YU or YA.

Comments This set of letter names provides compatibility with the forthcoming T2 Cyrillic encoding designed to be the future LaTeX 2_ε standard encoding for the extended Cyrillic alphabets.

Related commands `\cyr`

See page 31

15.20 \cyr

Synopsis `\cyr`{*<letter name>*}

Function *(Command)* Allows writing of Cyrillic letters in non-Cyrillic environments without changing the document language.

Limitations *letter name* must be one of a, b, v, G, D, e, yo, zh, z, i, ishrt, k, l, m, n, o, otld, p, r, s, t, u, y, f, h, hsha, c, ch, sh, shch, hrdsn, erevy, hrdsn, erev, yu or ya.

Related commands `\CYR`

See page 31

Example

Монгол `\CYRM\cyro\cyrn\cyrg\cyro\cyrl`

15.21 \g

Function *(Command)* Creates a *gamma* which is used for Mongolian transliterations.

Limitations Only a limited number of typefaces is available in standard MonTeX.

Comments More *gamma* shapes are provided by the Modern Greek package which is loaded automatically by MonTeX if available.

Related commands `\ch \sh`

See page 37

Example

mongγol-un γazar nutuγ `mong\g ol-un \g azar nutu\g`

15.22 \glyphbcg

Synopsis `\glyphbcg{`*<text>*`}`

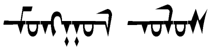
Function *(Command)* Accepts MLS glyph transliteration as input for Mongolian.

Comments Inconvenient for anything longer than five glyphs.

Related commands `\bcg`

See page 50

Example



```
\glyphbcg{@moaNnnoL @aoloS}
```

15.23 \ImplementationLevel

Function *(Command)* Shows the Implentation Level of MonTEX.

Comments Only for administrative purposes.

Related commands `\Version(Date|Kirill|Mongol|Release)`

Example

This is MonTEX IVu

```
This is \MonTeX\
\ImplementationLevel
```

15.24 \kbf

Synopsis `\kbf{`*<text>*`}`

Function *(Command)* Cyrillic boldface capsule.

Related commands `\k(it|rm|sc|sf|sl|tt) \l(bf|it|rm|sc|sf|sl|tt)`

See page 35

Example

This is **кирилл бүдүүн** writing.

```
This is
\kbf{kirill b\"ud\"u\"un}
writing.
```

15.25 `\kit`

Synopsis `\kit{`*<text>*`}`

Function *(Command)* Cyrillic italic capsule.

Related commands `\k(bf|rm|sc|sf|sl|tt)` `\l(bf|it|rm|sc|sf|sl|tt)`

See page 35

Example

This is *кирилл бичмэл* writing.

```
This is
\kit{kirill biqm\"al}
writing.
```

15.26 `\kminch`

Function *(Command)* Produces inch-high bold sans serif cyrillic letters for book titles etc.

Limitations Only capital letters and numbers available.

Comments This command bypasses the NFSS font setup, hence deprecated since the font provided by this command does not orthogonally follow with the font changes of the main document.

Related commands `\cminch`

See page 35

15.27 `\krm`

Synopsis `\krm{`*<text>*`}`

Function *(Command)* Cyrillic «roman» capsule.

Related commands `\k(bf|it|sc|sf|sl|tt)` `\l(bf|it|rm|sc|sf|sl|tt)`

See page 35

Example

This is кирилл шулуун writing.

```
This is
\krm{kirill shuluun}
writing.
```

15.28 `\ksc`

Synopsis `\ksc{`*<text>*`}`

Function *(Command)* Cyrillic small caps capsule.

Related commands `\k(bf|it|rm|sf|sl|tt) \l(bf|it|rm|sc|sf|sl|tt)`

See page 35

Example

This is КИРИЛЛ ЖИЖИГ ТОМ ҮСГИЙН writing.	`This is \ksc{kirill jijig tom \"usgi\"in} writing.`

15.29 `\ksf`

Synopsis `\ksf{`*<text>*`}`

Function *(Command)* Cyrillic sans serif capsule.

Related commands `\k(bf|it|rm|sc|sl|tt) \l(bf|it|rm|sc|sf|sl|tt)`

See page 35

Example

This is кирилл огтолсон writing.	`This is \ksf{kirill ogtolson} writing.`

15.30 `\ksl`

Synopsis `\ksl{`*<text>*`}`

Function *(Command)* Cyrillic slanted capsule.

Related commands `\k(bf|it|rm|sc|sf|tt) \l(bf|it|rm|sc|sf|sl|tt)`

See page 35

Example

This is *кирилл налуу* writing.	`This is \ksl{kirill naluu} writing.`

15.31 `\ktt`

Synopsis `\ktt{`<*text*>`}`

Function *(Command)* Cyrillic typewriter capsule.

Related commands `\k(bf|it|rm|sc|sf|sl) \l(bf|it|rm|sc|sf|sl|tt)`

See page 35

Example

This is `кирилл бичгийн машины` writing.	`This is \ktt{kirill biqgi\"in mashiny} writing.`

15.32 `\lat`

Synopsis `\lat{`<*text*>`}`

Function *(Command)* Latin capsule.

Related commands `\xalx`

See page 34

Монгол ба English	`{\mnr Mongol ba \lat{English}}`

15.33 `\lbf`

Synopsis `\lbf{`<*text*>`}`

Function *(Command)* Latin boldface capsule.

Related commands `\k(bf|it|rm|sc|sf|sl|tt) \l(it|rm|sc|sf|sl|tt)`

See page 35

Example

Энэ бол **latin boldface** шрифт.	`{\mnr \"An\"a bol \lbf{latin boldface} shrift.}`

15.34 `\lit`

Synopsis `\lit{`*<text>*`}`

Function *(Command)* Latin italic capsule.

Related commands `\k(bf|it|rm|sc|sf|sl|tt) \l(bf|rm|sc|sf|sl|tt)`

See page 35

Example

Энэ бол *latin italic* шрифт.

```
{\mnr \"An\"a bol
\lit{latin italic}
shrift.}
```

15.35 `\lrm`

Synopsis `\lrm{`*<text>*`}`

Function *(Command)* Latin roman capsule.

Related commands `\k(bf|it|rm|sc|sf|sl|tt) \l(bf|it|sc|sf|sl|tt)`

See page 35

Example

Энэ бол latin roman шрифт.

```
{\mnr \"An\"a bol
\lrm{latin roman}
shrift.}
```

15.36 `\lsc`

Synopsis `\lsc{`*<text>*`}`

Function *(Command)* Latin small caps capsule.

Related commands `\k(bf|it|rm|sc|sf|sl|tt) \l(bf|it|rm|sf|sl|tt)`

See page 35

Example

Энэ бол LATIN SMALL CAPS шрифт.

```
{\mnr \"An\"a bol
\lsc{latin small caps}
shrift.}
```

15.37 `\lsf`

Synopsis `\lsf{`*<text>*`}`

Function *(Command)* Latin sans serif capsule.

Related commands `\k(bf|it|rm|sc|sf|sl|tt) \l(bf|it|rm|sc|sl|tt)`

See page 35

Example

Энэ бол latin sans serif шрифт.	`{\mnr \"An\"a bol` `\lsf{latin sans serif}` `shrift.}`

15.38 `\lsl`

Synopsis `\lsl{`*<text>*`}`

Function *(Command)* Latin slanted capsule.

Related commands `\k(bf|it|rm|sc|sf|sl|tt) \l(bf|it|rm|sc|sf|tt)`

See page 35

Example

Энэ бол *latin slanted* шрифт.	`{\mnr \"An\"a bol` `\lsl{latin slanted}` `shrift.}`

15.39 `\ltt`

Synopsis `\ltt{`*<text>*`}`

Function *(Command)* Latin typewriter capsule.

Related commands `\k(bf|it|rm|sc|sf|sl|tt) \l(bf|it|rm|sc|sf|sl)`

See page 35

Example

Энэ бол `latin typewriter` шрифт.	`{\mnr \"An\"a bol` `\ltt{latin typewriter}` `shrift.}`

15.40 `\mabosoo`

Synopsis `\mabosoo`{*<text>*}

Function *(Command)* Similar to \mobosoo, it provides vertical capsules of text, but *<text>* is treated as Manju.

Limitations Like all commands of the `\bosoo` family, this command requires PostScript support for proper vertical display.

Related commands `\bosoo` `\mbosoo` `\mobosoo`

See page 60

Example

manju ᠮᠠᠨᠵᡠ᠈ writing ᠪᡳᡨᡥᡝ᠉ looks beautiful indeed.

```
\emph{manju} \mabosoo{manju}
writing \mabosoo{bithe}
looks beautiful indeed.
```

15.41 `\mabox`

Synopsis `\mabox`{*<vertical length>*}{*<text>*}

Function *(Command)* Similar to \mobox, it provides boxes of vertical text, but *<text>* is treated as Manju.

Limitations Like all commands of the `\box` family, this command requires PostScript support for proper vertical display.

Related commands `\mobox`

See page 61

Example

ᠮᠠᠨᠵᡠ᠈ ᠪᡳᡨᡥᡝ᠉

```
\mabox{1.5cm}{%
\noindent manju\\bithe.
}
```

15.42 `\mbosoo`

Synopsis `\mbosoo`{*<text>*}

Function *(Command)* Similar to \bosoo, it provides vertical capsules of text, but *<text>* is converted to Mongolian.

Limitations Like `\bosoo`, this command requires PostScript support. Like `\bcg`, the input text may only contain letters, transliteration symbols and numbers but no TEX commands. The command is internally defined as `\bosoo\bcg{...}`.

Related commands `\bosoo \mabosoo \mobosoo`

See page 50

Example

mongγol ᠮᠣᠩᠭᠣᠯ writing ᠪᠢᠴᠢᠭ looks beautiful indeed.

```
\emph{mong\g ol}
\mbosoo{mongGol}
writing \mbosoo{bicig}
looks beautiful indeed.
```

15.43 `\mobosoo`

Synopsis `\mobosoo`{*<text>*}

Function *(Command)* Similar to \mbosoo, it provides vertical capsules of text, but *<text>* is converted to Mongolian using the Simplified Transliteration.

Limitations Like all commands of the `\bosoo` family, this command requires PostScript support for proper vertical output.

Related commands `\bosoo\mabosoo \mbosoo`

See page 60

Example

mongγol ᠮᠣᠩᠭᠣᠯ writing ᠪᠢᠴᠢᠭ looks beautiful indeed.

```
\emph{mong\g ol}
\mobosoo{munggul}
writing \mobosoo{bicik}
looks beautiful indeed.
```

15.44 `\mobox`

Synopsis `\mobox`{*<vertical length>*}{*<text>*}

Function *(Command)* Similar to \mabox, it provides boxes of vertical text, but *<text>* is treated as Mongolian.

Limitations Mongolian must be entered in Simplified Transliteration. Currently, LMS input is not accepted.

Like all commands of the `\box` family, this command requires PostScript support for proper vertical display.

Related commands `\mobox`

See page 61

Example

ᠮᠣᠩᠭᠣᠯ ᠪᠢᠴᠢᠭ

```
\mabox{2cm}{munggul\\bicik}
```

15.45 \mnr

Function *(Command)* Switches the current stream to Xalx transliteration of Latin characters.

Comments `\mnr` can be understood as Mongolian New Romanization.

Related commands `\rnm`

See page 34

Example

First кирилл, then latin.

```
First \mnr kirill,
\rnm then latin.
```

15.46 \MonTeX

Function *(Command)* Produces the MonTEX- logo.

Example

You are using MonTEX, a LATEX 2$_\varepsilon$ package providing Mongolian.

```
You are using \MonTeX,
a \LaTeXe\ package
providing Mongolian.
```

15.47 \MyTogrog

Function *(Command)* Provides the Mongolian currency denominator.

Comments Matches the typeface of the environment.

Related commands `\mytogrog` `\Togrog` `\togrog`

See page 33

Example

Үнэ 200 ₮.

```
\kit{\"Un\"a 200 \MyTogrog}.
```

15.48 \mytogrog

Function *(Command)* Provides the Mongolian currency denominator, lower case variant (not considered standard).

Comments Matches the typeface of the environment.

Related commands `\MyTogrog \Togrog \togrog`

See page 33

15.49 \PrettyMLS

Synopsis `\PrettyMLS{`*<text>*`}`

Function *(Command)* Replaces some of the Mongolian transliteration shorthands with nicer output.

Related commands `\glyphbcg \ShowSpecialMLS`

See page 55

Example

SaGdur  is *šaγdur* .

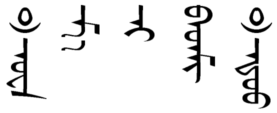
```
\emph{SaGdur} \cbg{SaGdur}
\emph{\PrettyMLS{SaGdur}}.
```

15.50 \om

Function *(Command)* Used for Tibetan transliterations.

See page 34

Example



```
\mobox{3cm}{\noindent\sffamily
\om uva\\
\  ma'=a\\
\  n'i\\
\  badmi'\\
\om huu}
```

15.51 `\rmfamily`

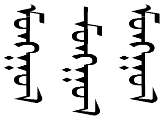
Function *(Command)* Sets normal Mongolian or Manju font family.

Limitations Works only for LMA and LMO encodings.

Comments There is no good equivalent between *Roman* and Mongolian typographical styles.

See page 66

Example



```
\mobox{2cm}{\noindent
munggul\\
\sffamily munggul\\
\rmfamily munggul}
```

15.52 `\rnm`

Function *(Command)* Disables transliteration of Latin characters to Xalx in the current stream.

Comments `\rnm` can be understood as Return to NorMal.

Related commands `\mnr`

See page 34

Example

Одоо кирилл дараа latin

```
\mnr Odoo kirill
daraa \rnm latin
```

15.53 `russian`

Synopsis `russian`

Function *(Environment* or *Option)* Sets document language to Russian.

Related commands `bicig bithe buryat english xalx`

See page 26

15.54 `\RussianToday`

Function *(Command)* Provides the date in Russian.

Comments Internal command. Authors should use `\today` which is redefined automatically by the `russian` option when calling the `mls` package.

Related commands `\BuryatToday \XalxToday`

See page 27

15.55 \SetDocumentEncodingBicig

Function *(Command)* Sets the document encoding to Classical Mongolian, also known as Uighur.

Comments The romanization used for this encoding is a simplified system with an emphasis on graphical, not phonetical properties of the Uighur writing system.

Related commands `\SetDocumentEncodingBithe`

See page 49

15.56 \SetDocumentEncodingBithe

Function *(Command)* Sets the document encoding to Classical Manju.

Comments The romanization used for this encoding is, with a few simple exceptions, a close match of Hauer's system which is the *de facto* standard.

Related commands `\SetDocumentEncodingBicig`

See page 56

15.57 \SetDocumentEncodingLMC

Function *(Command)* Sets the document encoding to Modern Mongolian (Xalx in Cyrillic writing).

Comments Used for writing Mongolian texts on Latin-only platforms.

Related commands `\SetDocumentEncodingNeutral`

See page 33

Example

Кирилл үсэг, монгол хэл
Latin üsäg, mongol xäl

```
\SetDocumentEncodingLMC
Kirill \"us\"ag, mongol x\"al\\
\SetDocumentEncodingNeutral
Latin \"us\"ag, mongol x\"al
```

15.58 \SetDocumentEncodingNeutral

Function *(Command)* Resets the document encoding so that Latin appears as Latin again and is not anymore converted to Cyrillic automatically.

Related commands `\SetDocumentEncodingLMC`

See page 32

15.59 `\sffamily`

Function *(Command)* Sets Block Print Style Mongolian or Manju font family.

Limitations Works only for LMA and LMO encodings.

Comments There is no good equivalent between *Roman* and Mongolian typographical styles.

See page 66

15.60 `\sh`

Function *(Command)* Creates a *š* which is used for Mongolian transliterations.

Related commands `\ch` `\g`

See page 37

15.61 `\ShowSpecialMLS`

Synopsis `\ShowSpecialMLS(true|false)`

Function *(Command)* Controls the behaviour of `\PrettyMLS` and either reveals or hides FVS and other codes for input of `-'*` etc.

Limitations This function accepts only character tokens as input, no TEX commands.

Related commands `\bcg` `\glyphbcg` `\PrettyMLS`

See page 53

15.62 `\Togrog`

Function *(Command)* Provides the Mongolian currency denominator.

Comments Never changes the typeface. If you want to match `\Togrog` with the environment use `\MyTogrog` instead.

Related commands `\togrog` `\MyTogrog` `\mytogrog`

See page 33

Example

Yнэ 200 ₮.	`\xalx{\"Un\"a 200 \Togrog}.`

15.63 `\togrog`

Function *(Command)* Provides the Mongolian currency denominator, lower case variant (not considered standard).

Comments Never changes the typeface. If you want to match \Togrog with the environment use \MyTogrog instead.

Related commands `\Togrog \MyTogrog \mytogrog`

See page 33

15.64 `\Useg`

Synopsis `\Useg`{*<number>*}

Function *(Command)* Provides counting by upper case Cyrillic letters, Xalx Mongolian style.

Limitations *<number>* must be between 1 and 31.

Related commands `\Asbuk \asbuk \useg \Uzeg \uzeg`

See page 38

15.65 `\useg`

Synopsis `\useg`{*<number>*}

Function *(Command)* Provides counting by lower case Cyrillic letters, Xalx Mongolian style.

Limitations *<number>* must be between 1 and 31.

Related commands `\Asbuk \asbuk \Useg \useg \Uzeg`

See page 38

15.66 `\Uzeg`

Synopsis `\Uzeg`{*<number>*}

Function *(Command)* Provides counting by upper case Cyrillic letters, Buryat style.

Limitations *<number>* must be between 1 and 32.

Related commands `\Asbuk \asbuk \Useg \useg \Uzeg`

See page 38

15.67 `\uzeg`

Synopsis `\uzeg{`*<number>*`}`

Function *(Command)* Provides counting by lower case Cyrillic letters, Buryat style.

Limitations *<number>* must be between 1 and 32.

Related commands `\Asbuk \asbuk \Useg \useg \Uzeg`

See page 38

15.68 `\VersionDate`

Function *(Command)* Provides the release date of of the current version

Comments Only for administrative purposes.

Related commands `\Version(Kirill|Mongol|Release) \ImplementationLevel`

Example

This version was officially released 2002/07/01.

```
This version was officially
released \VersionDate.
```

15.69 `\VersionKirill`

Function *(Command)* Provides the version number of the MonTEX code related to Cyrillic.

Comments Only for administrative purposes.

Related commands `\Version(Date|Mongol|Release) \ImplementationLevel`

Example

Cyrillic version: 04

```
Cyrillic version: \VersionKirill
```

15.70 `\VersionMongol`

Function *(Command)* Provides the version number of the MonTEX code related to Mongolian.

Comments Only for administrative purposes.

Related commands `\Version(Date|Kirill|Release) \ImplementationLevel`

Example

Mongolian version: 092

```
Mongolian version: \VersionMongol
```

15.71 `\VersionRelease`

Function *(Command)* Comprehensive version information.

Comments Only for administrative purposes.

Related commands `\Version(Date|Kirill|Mongol)` `\ImplementationLevel`

Example

This is MonTEX IVu.04.092	`This is \MonTeX\ \VersionRelease`

15.72 `xalx`

Synopsis `xalx`

Function *(Environment* or *Option)* Sets document language to Xalx, or Modern Mongolian.

Related commands `bicig bithe buryat english russian`

See page 26

15.73 `\xalx`

Synopsis `\xalx{`*<text>*`}`

Function *(Command)* Creates capsules with Modern Mongolian transliteration for including Xalx words in other languages.

Related commands `\lat`

See page 34

Example

English and монгол	`English and \xalx{mongol}`

15.74 `\XalxToday`

Function *(Command)* Provides the date in Xalx Mongolian.

Comments Internal command. Authors should use `\today` which is redefined automatically by the `xalx` option when calling the `mls` package.

Related commands `\BuryatToday` `\RussianToday`

See page 27