MonTEX
A Quick Guide
(*Draft*)

Oliver Corff

July 1, 2002

Contents

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.

1.0.1 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 3.10, page 17) and `bithe` (see section 3.10, page 17) are introduced in part 3.9, "Full Vertical Text Pages".

The options `buryat`, `russian` and `xalx` produce captions in Buryat, Russian and Modern Mongolian. `buryat` `russian` `xalx`

Buryat	2002 оной июлиин 1-нэй үдэр
Xalx	2002 оны долоодугаар сарын 1
Russian	1 июля 2002

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). `english`

2 Cyrillic Text – Кирилл үсэг

2.1 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 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

	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 1: Cyrillic Alphabet Input Methods

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. 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.}
```

2.2 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{...}` and `\mnr{...}` 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

`\lat`
`\mnr`

> `[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.

2.3 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 2).

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

[1] The magic triple-C!

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 2: Shortcuts for Mongolian Transliteration Symbols

environment but results in a Ш when used in a Cyrillic context[2]:

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}.
```

3 Uighur Mongolian and Manju Input

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 3.

The possible combinations of Mongolian writing input methods and display commands are listed in table 4. 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.

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.

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

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 3: Mongolian Script Transliterations

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 subsection 2.1, 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 all kinds of 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,

Command Type	Mongolian		Manju
	MLS	Simplified	
Document Encoding	only available as font encoding `LMS`, not as document encoding	`LMO`	`LMA`
Horizontal Capsules	`\bcg` (see section 3.5, page 10)	`\bicig` (see section 3.5, page 10)	`\bithe` (see section 3.5, page 10)
Horizontal Paragraphs	not available	`bicigtext` (see section 3.6, page 10)	`bithetext` (see section 3.6, page 10)
Vertical Capsules	`\mbosoo` (see section 4, page 11)	`\mobosoo` (see section 4, page 11)	`\mabosoo` (see section 4, page 11)
Vertical Paragraph Boxes	not available	`\mobox` (see section 3.8, page 12)	`\mabox` (see section 5, page 12)
Vertical Pages	not available	`bicigpage` (see section 3.9, page 13)	`bithepage` (see section 3.9, page 13)

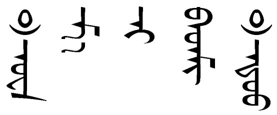
Table 4: Mongolian Input and Display Commands

it is certainly not necessary for everyday work.

3.1 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) as it is displayed on a huge bronze incense burner in front of the Gandan Monastery in Ulaanbaatar:



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

3.2 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 3, entering `*ng` tells the system to consider this *ng* to be in non-initial position.[3]

Besides these symbols, table 5 includes also some useful punctuation marks etc. as they are used in Mongolian Script.

3.3 Manju Input

Manju documents can be compiled with the `bithe` (see section 3.10, page 17) 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) 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.

3.4 Basic Character Set and Romanization

Given by dictionary order, the system provides a basic character set as shown in table 6.

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 *ai*, it must be entered as `aii` in order to produce the desired effect.

[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 5: Mongolian Script Special Symbols and Punctuation Marks

Manju	Input	Latin	Manju	Input	Latin	Manju	Input	Latin
	a	a		h	h		c	c
	e	e		b	b		j	j
	i	i		p	p		y	y
	o	o		s	s		k’	k’
	u	u		s’	š		g’	g’
	v	ū		t	t		h’	h’
	n	n		d	d		r	r
	k	k		l	l		f	f
	g	g		m	m		w	w

Table 6: Manju Basic Character Set

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* !

3.5 Small Portions of Mongolian and Manju in Running Text

For displaying short Mongolian snippets in running text use

MLS Romanization `\bcg{...}`. `\bcg`

Simplified Transliteration `\bicig{...}`. `\bicig`

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

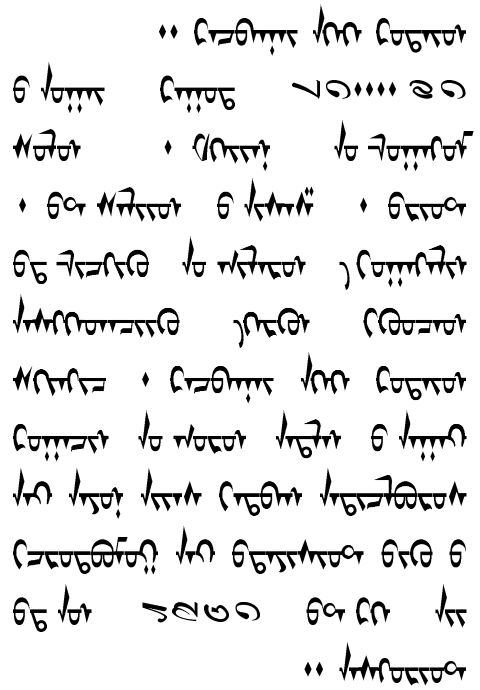
This is .

That is .

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

3.6 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, cingkis 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}
```

3.7 Vertical Capsules

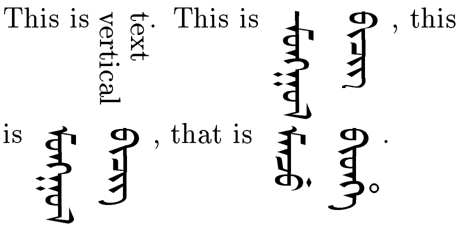
Individual Mongolian and Manju words can be placed vertically anywhere in otherwise horizontal text like in the keyword entry of dictionaries.[4]. The capsule containing the Mongolian or Manju word will automatically request sufficient space so that ugly overlaps with neighbouring lines will not happen.

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

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

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

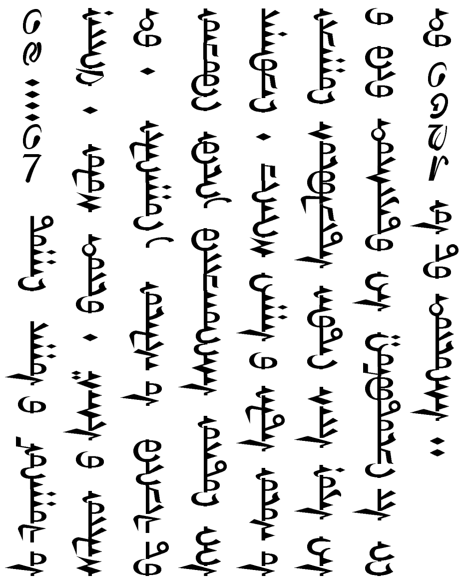
[4]Famous dictionaries with a mixture of vertical and horizontal printing are I. J. Schmidt's Mongolian-Russian-German dictionary (1835) and F. Lessing's Mongolian-English dictionary (1960).




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

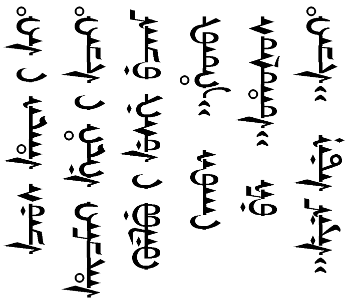
3.8 Vertical Text Boxes

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[5] 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. `\mobox` `\mabox`



```
\mobox{7.5cm}{%
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, cingkis xagan-u
aldan urug-un izagur surbulzidan
abadai saiin nuyan xan-u kuiu
tuisiyedu xan gumbudurzi-yin
ger-tu 1635 un-du tuiruksen.%
}
```

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



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

3.9 Full Vertical Text Pages

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.

`bicigpage`
`bithepage`

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}`.

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, cingkis xagan-u aldan
urug-un izagur surbulzidan abadai saiin nuyan xan-u kuiu
tuisiyedu xan gumbudurzi-yin ger-tu 1635 un-du tuiruksen.
badu muingke dayan 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 1: Input Example of a Mongolian text

3.10 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.

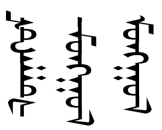
```
\documentclass{article}
\usepackage[bicig]{mls}
\begin{document}
uindur gegen zanabazar.
17||18 d'ugar zagun-u munggul-un neiigem, ulus tuiru,
shasin-u uiiles-tu, ilangguy=a uralig-un kuikzil-du
...
... more text ...
              ...
\end{document}
```

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

3.11 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 equivalent 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 `\sffamily`
`\rmfamily` switches back to the modern style. `\rmfamily`



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