The United Nations University

UNU/IIST

International Institute for Software Technology

Traditional Mongolian Script in the ISO/IEC 10646 and Unicode Standards

Myatav Erdenechimeg, Richard Moore and Yumbayar Namsrai

August 1999

UNU/IIST Report No. 170

UNU/IIST and UNU/IIST Reports

UNU/IIST is a Research and Training Center of the United Nations University. It was founded in 1992, and is located in Macau. UNU/IIST is jointly funded by the Governor of Macau and the Governments of China and Portugal through contribution to the UNU Endowment Fund.

The mission of UNU/IIST is to assist developing countries in the application and development of software technology.

UNU/IIST contributes through its programmatic activities:

1. advanced development projects in which software techniques supported by tools are applied,
2. research projects in which new techniques for software development are investigated,
3. curriculum development projects in which courses of software technology for universities in developing countries are developed,
4. courses which typically teach advanced software development techniques,
5. events in which conferences and workshops are organised or supported by UNU/IIST, and
6. dissemination, in which UNU/IIST regularly distributes to developing countries information on international progress of software technology.

Fellows, who are young scientists and engineers from developing countries, are invited to actively participate in all these projects. By doing the projects they are trained.

At present, the technical focus of UNU/IIST is on formal methods for software development. UNU/IIST is an internationally recognised center in the area of formal methods. However, no software technique is universally applicable. We are prepared to choose complementary techniques for our projects, if necessary.

UNU/IIST produces a report series. Reports are either Research [R], Technical [T], Compendia [C] or Administrative [A]. They are records of UNU/IIST activities and research and development achievements. Many of the reports are also published in conference proceedings and journals.

Please write to UNU/IIST or visit UNU/IIST home page: http://www.iist.unu.edu, if you would like to know more about UNU/IIST and its report series.

Zhou Chaochen, Director — 01.8.1997 – 31.7.2001

The United Nations University

UNU/IIST

International Institute for Software Technology

P.O. Box 3058
Macau

Traditional Mongolian Script in the ISO/IEC 10646 and Unicode Standards

Myatav Erdenechimeg, Richard Moore and Yumbayar Namsrai

Abstract

Traditional Mongolian script has recently been the subject of an international standardisation process within the scope of ISO/IEC 10646, which specifies an encoding scheme covering the set of characters occurring in the written forms of all the world's languages together with more general symbols (punctuation marks, mathematical symbols, and so on). This paper gives an overview of this encoding and the principles on which it is based and explains how the full range of positional variants of characters and of ligatures are obtained from it.

Myatav Erdenechimeg worked as a UN Fellow at UNU/IIST from September 1995 until August 1996 and has continued to work at UNU/IIST as a Visiting Scientist since September 1996. She graduated from the National University of Mongolia with a degree in Mathematical Calculation and then became a lecturer in computer science at the same university. Her more recent work has concentrated on the problem of the computerisation of the Mongolian language. As part of this she has developed coding systems for both traditional Mongolian script and Mongolian Cyrillic letters and software for Mongolian word processing. She has also worked on software for translation between Mongolian script and Cyrillic. She has published several papers in these fields.

Richard Moore is a Research Fellow on the staff of UNU/IIST, a position he took up on October 1st 1995. He has an M.A. in mathematics from the University of Cambridge and a Ph.D. in physics from the University of Manchester. He has been engaged in computing science research in the field of formal methods since 1985, a large part of which was carried out in the formal methods group at Manchester University. He has written several papers on formal methods and is co-author of two books on formal methods – mural: a Formal Development Support System; and Proof in VDM: A Practitioner's Guide. He has also worked for the Defence Research Agency in Malvern, UK, on various formal methods projects, both as a consultant and as a full-time member of staff.

Yumbayar Namsrai was a UN Fellow at UNU/IIST from September 1996 to June 1997 and from May 1999 to August 1999. He studied mathematics at The National University of Mongolia in Ulaanbaatar, Mongolia, from 1967 to 1972, and worked at the Joint Institute for Nuclear Research (JINR), Dubna, USSR from 1972 to 1981 where he was awarded a Candidate of Science degree in Physics and Mathematics (Software Systems of Computers and Computing Systems) in 1982. After returning to Mongolia he became the Head of the Programming Department in the Mathematics Faculty at The National University of Mongolia, and in January 1998 moved to the Computer Science and Management School of the Mongolian Technical University, also in Ulaanbaatar, where he is now Head of the Computer Science Department. His research interests are in the computerization of the traditional Mongolian script.

Contents

1 Introduction

Although most countries in the world have had national standard encoding schemes for the characters of their own language or languages for some time, these could differ wildly even between countries sharing the same written language. As a result, an electronic document written using a piece of software based on a particular encoding scheme could only be read by someone possessing either software based on the same encoding scheme or software for translating between the two different encoding schemes.

As the volume of international communications increased, especially the international exchange of electronic data, not least via the internet, it became clear that this situation was completely impractical and that some internationally accepted universal encoding scheme, which could form the basis for multi-lingual software, was needed. A joint technical committee (ISO/IEC JTC1) was therefore set up by the International Organization for Standardisation (ISO) and the International Electrotechnical Committee (IEC) to work on this, and, initially independently though later in collaboration with ISO/IEC, the Unicode Consortium embarked on a similar project.

The resulting ISO/IEC international standard 10646 and the Unicode standard, which uses the identical encoding but which additionally includes information which is important for people wishing to implement computer software based on the standard, offer a universal international standard encoding scheme covering not only all the characters used in the written forms of the languages of the world but also more general symbols.

The current versions of the standards [8,2] cover the majority of the European scripts (Latin, with extensions including characters with various accents; Greek; Cyrillic; etc.), various Indian scripts, including Devanagari, Bengali and Gujarati, and several other Asian scripts, including Thai and Tibetan as well as the Chinese, Japanese and Korean ideographic scripts. New versions, which are expected to be published shortly, will extend these with the scripts that have been undergoing standardization since 1993. These include, among many others, the traditional Mongolian script which is the subject of this paper.

Traditional Mongolian script, which is properly written vertically in columns ordered from left to right, was derived around the 12th century from the Uighur script, which was in turn developed from the Sogdian Aramaic script in the 8th or 9th century. It has been in continuous use since that date, even though in 1946 in Mongolia proper it was supplanted as the official written form of the Mongolian language by a Cyrillic script, written horizontally from left to right: the traditional script continued to be used in preference to the Cyrillic in certain disciplines, including history, literature and linguistics, and beginning in 1994 it has started to be used more widely and is now being taught in schools once again.

Mongolian is a cursive script, so individual letters in a word are joined together as illustrated in Figure 1. In addition, the actual written form of each individual letter in a word generally depends on the position of the letter within a word – specifically on whether it appears at the beginning (initial form), in the middle (medial form), or at the

end (final form) of the word (see Figure 2) – and may also depend on the preceding letter with which it can form a ligature (see Figure 3). In abstract terms, therefore, each letter has what might be called a *basic form* together with various *variant forms*, while certain combinations of letters combine to form ligatures.

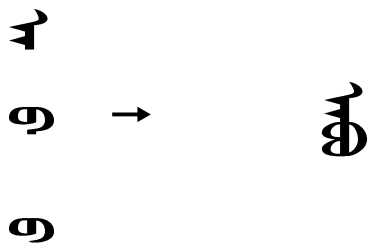


Figure 1: Joining of Characters in Mongolian Script

Mongolian letter (transliteration)	Isolate	Initial	Medial	Final
(A)				
(OE)				
(L)				

Figure 2: Initial, Medial and Final Forms of Mongolian Script Letters

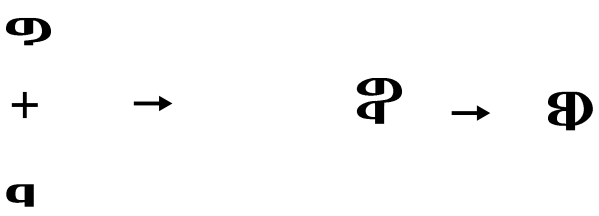


Figure 3: Mongolian Script Ligatures

The standard encoding which is to appear in the ISO/IEC and Unicode standards in fact codes only the basic character set, together with special punctuation symbols and numerals, but does not explicitly encode the variant forms or the ligatures since the correct variant form or ligature can, at least in most cases, be determined from context. Instead, control symbols are encoded which can be used to resolve ambiguities in those few cases where the context rules are inadequate and which can also be used to override the default contextual forms if so required.

In this paper, we present the basic Mongolian character set and explain the principles behind the selection of the characters which comprise it in Section 2. We also explain the use of the special characters in the character set. Section 3 then describes all the variant forms of each basic character and indicates how they are generated, and Section 4 contains a description of all the ligatures.

The final section of the paper gives some information about implementing software based on this encoding and also discusses how traditional Mongolian text can be intermixed with other scripts.

2 The Basic Character Set

The standard encoding covers not only traditional Mongolian script but also related scripts: Todo and Manchu, which are derivatives of Mongolian; Sibe, which is derived from Manchu; and Ali Gali, which was used for transcriptions of Tibetan and Sanskrit texts. Todo, Manchu and Sibe all share Mongolian characters.

The characters in the encoding are named according to the scripts they are used in as follows: letters used only in traditional Mongolian and letters shared between traditional Mongolian and other scripts are named MONGOLIAN LETTER; letters used exclusively in Todo are named MONGOLIAN LETTER TODO; letters used exclusively in Sibe and those shared between Sibe and Manchu are named MONGOLIAN LETTER SIBE; and letters used exclusively in Manchu are named MONGOLIAN LETTER MANCHU. Similarly, the Ali Gali letters are named after the script with which they are associated: MONGOLIAN LETTER ALI GALI, MONGOLIAN LETTER TODO ALI GALI, and MONGOLIAN LETTER MANCHU ALI GALI for Mongolian, Todo and Manchu respectively.

The basic character set encodes the Mongolian numerals together with precisely one form of each different letter. Generally this is the isolated form for the vowels and the initial form for the consonants, with the particular variant form occurring when the consonant is followed by the letter "A" being chosen in cases where this initial form has several alternative variants.

However, the various forms that the characters can take are not all unique: in some cases one character can have the same form in different positions (e.g. the initial and medial forms of the Mongolian letter "B" look the same (ᠪ)), while in other cases two different characters can look the same, either in the same position (e.g. the initial form

of the Mongolian letter "O" (ᠣ) looks the same as the initial form of the Mongolian letter "U" (ᠤ)) or in different positions (e.g. the first medial form of the Mongolian letter "TA" (ᠲ) looks the same as the second initial form of the Mongolian letter "DA" (ᠳ)). And the same can also be true across the different scripts (e.g. the medial form of the Mongolian letter "ANG" (ᠩ) is the same as that of the Todo letter "ANG" (ᡢ), though their final forms are different (ᠩ, ᡢ)).

This duplication can in fact lead to complete words being visually indistinguishable: for example, the Mongolian words "bodo" (to think) and "budu" (to dye) both have the same printed form ᠪᠣᠳᠣ because the positional forms of the Mongolian letters "O" (ᠣ) and "U" (ᠤ) occurring in the words are identical. However, the coding must be able to distinguish between the letters in order to be able to distinguish between the words.

We therefore choose the particular variant forms of the characters of the basic character set in such a way that no two different characters in the set have the same glyph, thus allowing us to distinguish the characters of the basic character set by sight. Thus, for example, the isolated form of the Mongolian letter "O" (ᠣ) is used, but the initial form of the Mongolian letter "U" (ᠤ) is used instead of the isolated form because the isolated form looks the same as the isolated form of "O" (ᠣ). Similarly, the medial form of the Mongolian letter "ANG" (ᠩ) is used but the final form of the Todo letter "ANG" (ᡢ).

The basic character set also encodes various punctuation symbols which are specific to Mongolian, including character 1800, MONGOLIAN BIRGA (᠀), character 1805, MONGOLIAN FOUR DOTS (᠅), and character 180A, MONGOLIAN NIRUGU (᠊). The Mongolian Birga is a symbol which is used to mark the beginning of a piece of text, such as the beginning of a section, a paragraph, or a line, and the Mongolian Four Dots is similarly used to mark the end of a piece of text. The Mongolian Nirugu is basically used simply to lengthen the cursive connection between letters, as in the following examples:

(Normal characters)

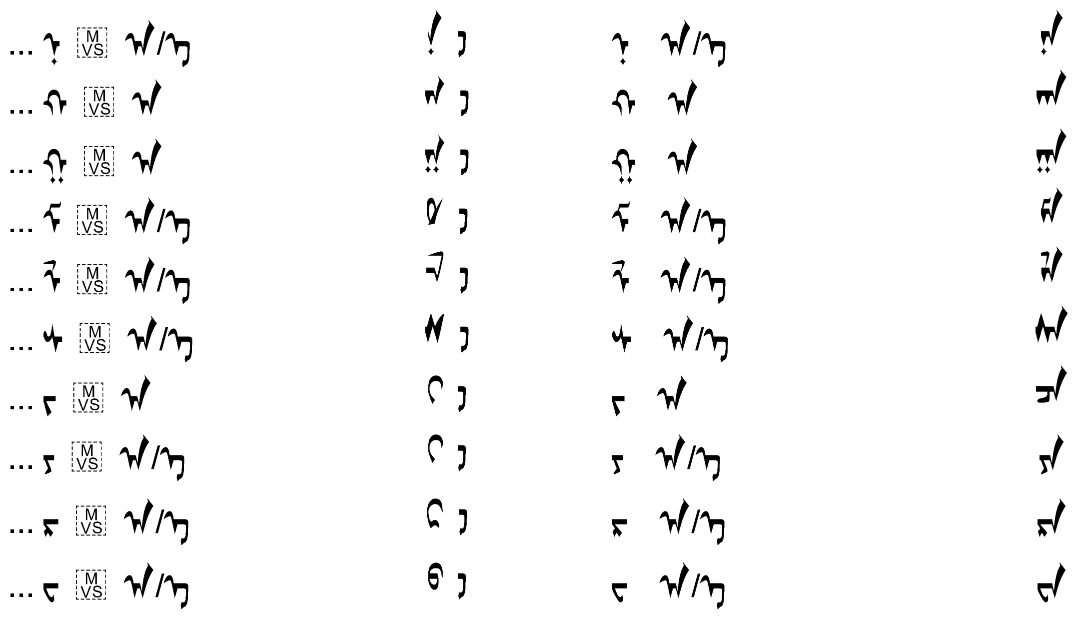
(Lengthened characters)

Finally, the basic character set includes four control characters: three for selecting alternative variants of a given positional form (the Mongolian free variant selectors, characters 180B (FVS1), 180C (FVS2) and 180D (FVS3)), and the Mongolian vowel separator (character 180E (MVS)).

The Mongolian vowel separator serves to separate the vowels "A" and "E", when they occur as the final letter in a word, from the consonant preceding them. A given sequence of characters with a connected final "A" or "E" (ᠠ) and the same sequence of characters with a separated final "A" or "E" (᠎ᠠ) can both correspond to Mongolian words, though these have different meanings and are thus different words. For example, the word "xana" with a separated final "a" ᠬᠠᠨ᠎ᠠ means " the wall of a tent", while the word "xana" with the final "a" connected ᠬᠠᠨᠠ means "the outer casing of a vein".

The following examples illustrate the use of the Mongolian vowel separator MVS :

Character sequence | Display | Character sequence | Display

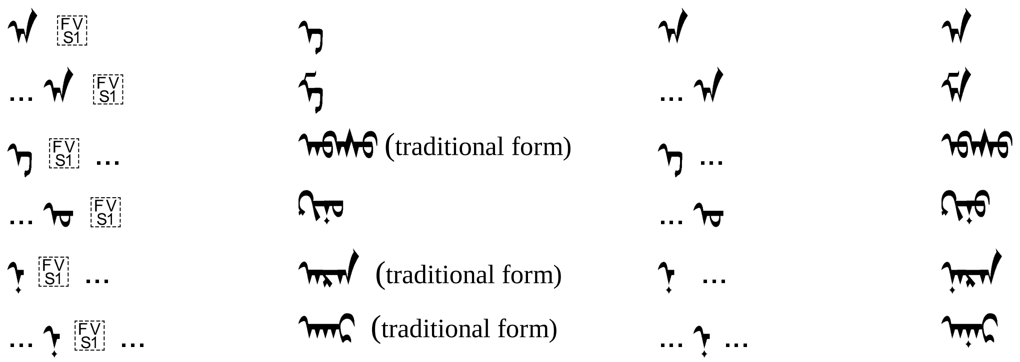


The Mongolian free variant selectors are used to distinguish different variants of the same positional form of a character. They modify only the character immediately preceding them and have no effect on the character following. Basically, the three variant selectors indicate the second, third and fourth variant form of a particular positional variant respectively the default (first) variant being oftained if no variant selector is included. The order of the different variants follows that given in the Mongolian Reference Table in Section 3.2.

Note that a free variant selector applied to a character for which no corresponding variant exists is assumed to have no effect.

The following examples illustrate some uses of the free variant selectors:

Character sequence | Display | Character sequence | Display



The full basic character set is shown in the tables on pages 7 and 8 and the names of these characters are given in the tables on pages 9 and 10.

Basic Character Set

	180	181	182	183	184	185	186	187
0	0	16	32	48	64	80	96	112
1	1	17	33	49	65	81	97	113
2	2	18	34	50	66	82	98	114
3	3	19	35	51	67	83	99	115
4	4	20	36	52	68	84	100	116
5	5	21	37	53	69	85	101	117
6	6	22	38	54	70	86	102	118
7	7	23	39	55	71	87	103	119
8	8	24	40	56	72	88	104	120
9	9	25	41	57	73	89	105	121
A	10	26	42	58	74	90	106	122
B	FVS1 11	27	43	59	75	91	107	123
C	FVS2 12	28	44	60	76	92	108	124
D	FVS3 13	29	45	61	77	93	109	125
E	MVS 14	30	46	62	78	94	110	126
F	15	31	47	63	79	95	111	127

Basic Character Set (continued)

	188	189	18A
0	 128	 144	 160
1	 129	 145	 161
2	 130	 146	 162
3	 131	 147	 163
4	 132	 148	 164
5	 133	 149	 165
6	 134	 150	 166
7	 135	 151	 167
8	 136	 152	 168
9	 137	 153	 169
A	 138	 154	170
B	 139	 155	171
C	 140	 156	172
D	 141	 157	173
E	 142	 158	174
F	143	159	175

Names of Basic Characters

dec	hex	Name
000	00	MONGOLIAN BIRGA
001	01	MONGOLIAN ELLIPSIS
002	02	MONGOLIAN COMMA
003	03	MONGOLIAN FULL STOP
004	04	MONGOLIAN COLON
005	05	MONGOLIAN FOUR DOTS
006	06	MONGOLIAN TODO SOFT HYPHEN
007	07	MONGOLIAN SIBE SYLLABLE BOUNDARY MARKER
008	08	MONGOLIAN MANCHU COMMA
009	09	MONGOLIAN MANCHU FULL STOP
010	0A	MONGOLIAN NIRUGU
011	0B	MONGOLIAN FREE VARIATION SELECTOR ONE
012	0C	MONGOLIAN FREE VARIATION SELECTOR TWO
013	0D	MONGOLIAN FREE VARIATION SELECTOR THREE
014	0E	MONGOLIAN VOWEL SEPARATOR
015	0F	(THIS POSITION SHALL NOT BE USED)
016	10	MONGOLIAN DIGIT ZERO
017	11	MONGOLIAN DIGIT ONE
018	12	MONGOLIAN DIGIT TWO
019	13	MONGOLIAN DIGIT THREE
020	14	MONGOLIAN DIGIT FOUR
021	15	MONGOLIAN DIGIT FIVE
022	16	MONGOLIAN DIGIT SIX
023	17	MONGOLIAN DIGIT SEVEN
024	18	MONGOLIAN DIGIT EIGHT
025	19	MONGOLIAN DIGIT NINE
026	1A	(THIS POSITION SHALL NOT BE USED)
027	1B	(THIS POSITION SHALL NOT BE USED)
028	1C	(THIS POSITION SHALL NOT BE USED)
029	1D	(THIS POSITION SHALL NOT BE USED)
030	1E	(THIS POSITION SHALL NOT BE USED)
031	1F	(THIS POSITION SHALL NOT BE USED)
032	20	MONGOLIAN LETTER A
033	21	MONGOLIAN LETTER E
034	22	MONGOLIAN LETTER I
035	23	MONGOLIAN LETTER O
036	24	MONGOLIAN LETTER U
037	25	MONGOLIAN LETTER OE
038	26	MONGOLIAN LETTER UE
039	27	MONGOLIAN LETTER EE
040	28	MONGOLIAN LETTER NA
041	29	MONGOLIAN LETTER ANG
042	2A	MONGOLIAN LETTER BA
043	2B	MONGOLIAN LETTER PA
044	2C	MONGOLIAN LETTER QA
045	2D	MONGOLIAN LETTER GA
046	2E	MONGOLIAN LETTER MA
047	2F	MONGOLIAN LETTER LA
048	30	MONGOLIAN LETTER SA
049	31	MONGOLIAN LETTER SHA
050	32	MONGOLIAN LETTER TA
051	33	MONGOLIAN LETTER DA
052	34	MONGOLIAN LETTER CHA
053	35	MONGOLIAN LETTER JA
054	36	MONGOLIAN LETTER YA
055	37	MONGOLIAN LETTER RA
056	38	MONGOLIAN LETTER WA
057	39	MONGOLIAN LETTER FA
058	3A	MONGOLIAN LETTER KA
059	3B	MONGOLIAN LETTER KHA
060	3C	MONGOLIAN LETTER TSA
061	3D	MONGOLIAN LETTER ZA
062	3E	MONGOLIAN LETTER HAA
063	3F	MONGOLIAN LETTER ZRA

dec	hex	Name
064	40	MONGOLIAN LETTER LHA
065	41	MONGOLIAN LETTER ZHI
066	42	MONGOLIAN LETTER CHI
067	43	MONGOLIAN LETTER TODO LONG VOWEL SIGN
068	44	MONGOLIAN LETTER TODO E
069	45	MONGOLIAN LETTER TODO I
070	46	MONGOLIAN LETTER TODO O
071	47	MONGOLIAN LETTER TODO U
072	48	MONGOLIAN LETTER TODO OE
073	49	MONGOLIAN LETTER TODO UE
074	4A	MONGOLIAN LETTER TODO ANG
075	4B	MONGOLIAN LETTER TODO BA
076	4C	MONGOLIAN LETTER TODO PA
077	4D	MONGOLIAN LETTER TODO QA
078	4E	MONGOLIAN LETTER TODO GA
079	4F	MONGOLIAN LETTER TODO MA
080	50	MONGOLIAN LETTER TODO TA
081	51	MONGOLIAN LETTER TODO DA
082	52	MONGOLIAN LETTER TODO CHA
083	53	MONGOLIAN LETTER TODO JA
084	54	MONGOLIAN LETTER TODO TSA
085	55	MONGOLIAN LETTER TODO YA
086	56	MONGOLIAN LETTER TODO WA
087	57	MONGOLIAN LETTER TODO KA
088	58	MONGOLIAN LETTER TODO GAA
089	59	MONGOLIAN LETTER TODO HAA
090	5A	MONGOLIAN LETTER TODO JIA
091	5B	MONGOLIAN LETTER TODO NIA
092	5C	MONGOLIAN LETTER TODO DZA
093	5D	MONGOLIAN LETTER SIBE E
094	5E	MONGOLIAN LETTER SIBE I
095	5F	MONGOLIAN LETTER SIBE IY
096	60	MONGOLIAN LETTER SIBE UE
097	61	MONGOLIAN LETTER SIBE U
098	62	MONGOLIAN LETTER SIBE ANG
099	63	MONGOLIAN LETTER SIBE KA
100	64	MONGOLIAN LETTER SIBE GA
101	65	MONGOLIAN LETTER SIBE HA
102	66	MONGOLIAN LETTER SIBE PA
103	67	MONGOLIAN LETTER SIBE SHA
104	68	MONGOLIAN LETTER SIBE TA
105	69	MONGOLIAN LETTER SIBE DA
106	6A	MONGOLIAN LETTER SIBE JA
107	6B	MONGOLIAN LETTER SIBE FA
108	6C	MONGOLIAN LETTER SIBE GAA
109	6D	MONGOLIAN LETTER SIBE HAA
110	6E	MONGOLIAN LETTER SIBE TSA
111	6F	MONGOLIAN LETTER SIBE ZA
112	70	MONGOLIAN LETTER SIBE RAA
113	71	MONGOLIAN LETTER SIBE CHA
114	72	MONGOLIAN LETTER SIBE ZHA
115	73	MONGOLIAN LETTER MANCHU I
116	74	MONGOLIAN LETTER MANCHU KA
117	75	MONGOLIAN LETTER MANCHU RA
118	76	MONGOLIAN LETTER MANCHU FA
119	77	MONGOLIAN LETTER MANCHU ZHA
120	78	(THIS POSITION SHALL NOT BE USED)
121	79	(THIS POSITION SHALL NOT BE USED)
122	7A	(THIS POSITION SHALL NOT BE USED)
123	7B	(THIS POSITION SHALL NOT BE USED)
124	7C	(THIS POSITION SHALL NOT BE USED)
125	7D	(THIS POSITION SHALL NOT BE USED)
126	7E	(THIS POSITION SHALL NOT BE USED)
127	7F	(THIS POSITION SHALL NOT BE USED)

Names of Basic Characters (continued)

dec	hex	Name
128	80	MONGOLIAN LETTER ALI GALI ANUSVARA ONE
129	81	MONGOLIAN LETTER ALI GALI VISARGA ONE
130	82	MONGOLIAN LETTER ALI GALI DAMARU
131	83	MONGOLIAN LETTER ALI GALI UBADAMA
132	84	MONGOLIAN LETTER ALI GALI INVERTED UBADAMA
133	85	MONGOLIAN LETTER ALI GALI BALUDA
134	86	MONGOLIAN LETTER ALI GALI THREE BALUDA
135	87	MONGOLIAN LETTER ALI GALI A
136	88	MONGOLIAN LETTER ALI GALI I
137	89	MONGOLIAN LETTER ALI GALI KA
138	8A	MONGOLIAN LETTER ALI GALI NGA
139	8B	MONGOLIAN LETTER ALI GALI CA
140	8C	MONGOLIAN LETTER ALI GALI TTA
141	8D	MONGOLIAN LETTER ALI GALI TTHA
142	8E	MONGOLIAN LETTER ALI GALI DDA
143	8F	MONGOLIAN LETTER ALI GALI NNA
144	90	MONGOLIAN LETTER ALI GALI TA
145	91	MONGOLIAN LETTER ALI GALI DA
146	92	MONGOLIAN LETTER ALI GALI PA
147	93	MONGOLIAN LETTER ALI GALI PHA
148	94	MONGOLIAN LETTER ALI GALI SSA
149	95	MONGOLIAN LETTER ALI GALI ZHA
150	96	MONGOLIAN LETTER ALI GALI ZA
151	97	MONGOLIAN LETTER ALI GALI AH
152	98	MONGOLIAN LETTER TODO ALI GALI TA
153	99	MONGOLIAN LETTER TODO ALI GALI ZHA
154	9A	MONGOLIAN LETTER MANCHU ALI GALI GHA
155	9B	MONGOLIAN LETTER MANCHU ALI GALI NGA
156	9C	MONGOLIAN LETTER MANCHU ALI GALI CA
157	9D	MONGOLIAN LETTER MANCHU ALI GALI JHA
158	9E	MONGOLIAN LETTER MANCHU ALI GALI TTA
159	9F	MONGOLIAN LETTER MANCHU ALI GALI DDHA
160	A0	MONGOLIAN LETTER MANCHU ALI GALI TA
161	A1	MONGOLIAN LETTER MANCHU ALI GALI DHA
162	A2	MONGOLIAN LETTER MANCHU ALI GALI SSA
163	A3	MONGOLIAN LETTER MANCHU ALI GALI CYA
164	A4	MONGOLIAN LETTER MANCHU ALI GALI ZHA
165	A5	MONGOLIAN LETTER MANCHU ALI GALI ZA
166	A6	MONGOLIAN LETTER ALI GALI HALF U
167	A7	MONGOLIAN LETTER ALI GALI HALF YA
168	A8	MONGOLIAN LETTER MANCHU ALI GALIBHA
169	A9	MONGOLIAN LETTER ALI GALI DAGALGA
170	AA	(THIS POSITION SHALL NOT BE USED)
171	AB	(THIS POSITION SHALL NOT BE USED)
172	AC	(THIS POSITION SHALL NOT BE USED)
173	AD	(THIS POSITION SHALL NOT BE USED)
174	AE	(THIS POSITION SHALL NOT BE USED)
175	AF	(THIS POSITION SHALL NOT BE USED)

2.1 Other basic Mongolian characters

The basic Mongolian character set described above only includes characters which are peculiar to Mongolian and its related scripts. Other symbols which are used not only in the Mongolian scripts but also in other scripts are encoded as general punctuation symbols in the General Punctuation block of the standards. These include the two combination symbols "?!" and "!?" and the Mongolian space.

The combination symbols "?!" and "!?" are represented by characters 2048, QUESTION EXCLAMATION MARK, and 2049, EXCLAMATION QUESTION MARK, respectively.

The Mongolian space is not coded explicitly in the standards, but its functionality is provided by character 202F, NARROW NO-BREAK SPACE (NNBSP). The Mongolian space occurs frequently in Mongolian: many words are formed by the addition of one or more suffixes (which indicate for example different case endings of nouns and pronouns, ownership, and negation) to a basic stem word, and each individual suffix is separated from the stem or from the preceding suffix by the Mongolian Space. Visually, it appears as a small white space. It also affects the forms of the letters preceding and following it, the preceding character adopting its final form. However, it does not mark

a break between words, the stem word together with all its suffixes being considered to form a single word.

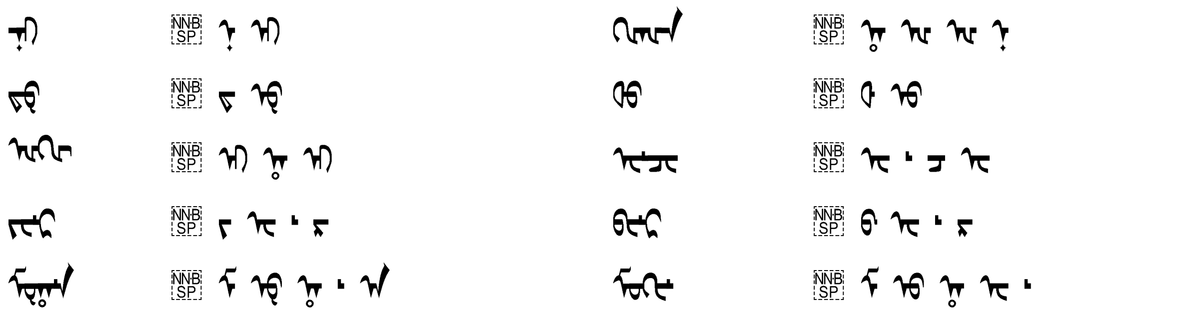
Note that the functionality of character 202F, NARROW NO-BREAK SPACE, is different from that of character 00A0, NO-BREAK SPACE, which does mark a division between words but which forbids a line of text to be broken at that division.

The following examples illustrate how narrow no-break space NNBSP is used to generate the most commonly occurring case endings in Mongolian:

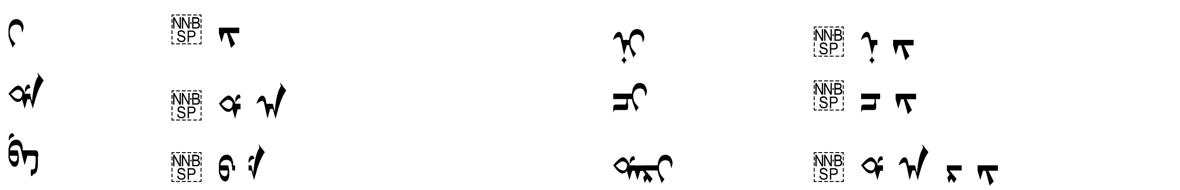
Case - ending	Character sequence	Case - ending	Character sequence

TRADITIONAL MONGOLIAN :

TODO :



SIBE and MANCHU :



3 The Variant Forms

As indicated in Section 1, the actual written form of any given letter in Mongolian generally depends on its position within a word: the letter assumes its initial form when it is the first letter in the word, its final form when it is the last letter in the word, and its medial form when it occurs somewhere in the middle of the word. However, there can also be a number of possible variations of a given positional form, and these variations can depend on a number of factors including the preceding and following letters, which syllable in the word contains the letter, and the gender of the word. In fact, a given positional form may have as many as four different variants, while a letter may have as many as nine different variant forms altogether. The complete set of variants of each letter are shown in the Mongolian Reference Table in Section 3.2.

Taking account of the fact that different variants may look the same as discussed in Section 2, a set of presentation forms is defined, all of which are visually distinct not only from each other but also from all the characters in the basic character set. These are shown in the "Presentation Character Set" tables on pages 13 and 14, and their names are given in the following tables on pages 15 and 16. The set of all possible character shapes in Mongolian is therefore represented by the basic character set together with the set of presentation forms.

These presentation forms are not encoded explicitly in the ISO/IEC10646 standard. Instead, the propriate positional form of a character is determined directly from its position in a word and the first (which is the most commonly occurring) variant of this is taken as the default form. Then a different (non-default) variant of that positional form is obtained by appending one of the Mongolian free variant selectors (characters 180B, 180C and 180D (FVS1, FVS2, FVS3) in the basic character set) to the code for the basic character. Thus, for example, the word "DUG" means "deep sleep" (ᠳᠤᠭ) when spelt with the first (default) variant of the initial form of the letter "D" (ᠳ) (when the actual sequence of characters would be ᠳ ᠤ ᠭ) but means "to put in check using the bishop in the game of chess" (ᡩᠤᠭ) when spelt using the second variant (when the actual sequence of characters would include the first free variant selector: ᠳ FVS1 ᠤ ᠭ).

Presentation Character Set

	F30	F31	F32	F33	F34	F35	F36	F37
0	 0	 16	 32	 48	 64	 80	96	 112
1	 1	 17	 33	 49	65	 81	 97	 113
2	2	 18	 34	 50	66	 82	 98	 114
3	 3	 19	 35	 51	 67	 83	 99	 115
4	 4	20	 36	 52	68	 84	100	 116
5	5	21	37	53	69	85	101	117
6	6	22	38	54	70	86	102	118
7	7	23	39	55	71	87	103	119
8	8	24	40	56	72	88	104	120
9	9	25	41	57	73	89	105	121
A	10	26	42	58	74	90	106	122
B	11	27	43	59	75	91	107	123
C	12	28	44	60	76	92	108	124
D	13	29	45	61	77	93	109	125
E	14	30	46	62	78	94	110	126
F	15	31	47	63	79	95	111	127

Presentation Character Set (continued)

	F38	F39
0	 128	144
1	129	145
2	 130	146
3	 131	147
4	132	148
5	 133	149
6	 134	150
7	135	151
8	136	152
9	 137	153
A	 138	154
B	 139	155
C	 140	156
D	 141	157
E	 142	158
F	143	159

Names of Presentation forms

dec	hex	Name
000	00	MONGOLIAN BIRGA FIRST FORM
001	01	MONGOLIAN BIRGA SECOND FORM
002	02	MONGOLIAN BIRGA THIRD FORM
003	03	MONGOLIAN BIRGA FOURTH FORM
004	04	MONGOLIAN LETTER A INITIAL FORM
005	05	MONGOLIAN LETTER A FIRST MEDIAL FORM
006	06	MONGOLIAN LETTER A SECOND MEDIAL FORM
007	07	MONGOLIAN LETTER A THIRD MEDIAL FORM
008	08	MONGOLIAN LETTER A FIRST FINAL FORM
009	09	MONGOLIAN LETTER A SECOND FINAL FORM
010	0A	MONGOLIAN LETTER I INITIAL FORM
011	0B	MONGOLIAN LETTER I FINAL FORM
012	0C	MONGOLIAN LETTER O FIRST MEDIAL FORM
013	0D	MONGOLIAN LETTER O SECOND MEDIAL FORM
014	0E	MONGOLIAN LETTER O FIRST FINAL FORM
015	0F	MONGOLIAN LETTER O SECOND FINAL FORM
016	10	MONGOLIAN LETTER OE THIRD MEDIAL FORM
017	11	MONGOLIAN LETTER OE SECOND FINAL FORM
018	12	MONGOLIAN LETTER EE INITIAL FORM
019	13	MONGOLIAN LETTER EE FINAL FORM
020	14	MONGOLIAN LETTER NA FIRST MEDIAL FORM
021	15	MONGOLIAN LETTER NA THIRD MEDIAL FORM
022	16	MONGOLIAN LETTER NA MEDIAL SEPARATE FORM
023	17	MONGOLIAN LETTER ANG FINAL FORM
024	18	MONGOLIAN LETTER BA FINAL FORM
025	19	MONGOLIAN LETTER PA FINAL FORM
026	1A	MONGOLIAN LETTER QA SECOND MEDIAL FORM
027	1B	MONGOLIAN LETTER QA THIRD MEDIAL FORM
028	1C	MONGOLIAN LETTER QA FOURTH MEDIAL FORM
029	1D	MONGOLIAN LETTER QA FEMININE SECOND ISOLATE FORM
030	1E	MONGOLIAN LETTER GA FEMININE MEDIAL FORM
031	1F	MONGOLIAN LETTER GA FEMININE FINAL FORM
032	20	MONGOLIAN LETTER MA MEDIAL FORM
033	21	MONGOLIAN LETTER MA FINAL FORM
034	22	MONGOLIAN LETTER LA MEDIAL FORM
035	23	MONGOLIAN LETTER LA FINAL FORM
036	24	MONGOLIAN LETTER SA MEDIAL FORM
037	25	MONGOLIAN LETTER SA FIRST FINAL FORM
038	26	MONGOLIAN LETTER SA SECOND FINAL FORM
039	27	MONGOLIAN LETTER SA THIRD FINAL FORM
040	28	MONGOLIAN LETTER SHA MEDIAL FORM
041	29	MONGOLIAN LETTER SHA FINAL FORM
042	2A	MONGOLIAN LETTER TA SECOND MEDIAL FORM
043	2B	MONGOLIAN LETTER TA FINAL FORM
044	2C	MONGOLIAN LETTER DA SECOND MEDIAL FORM
045	2D	MONGOLIAN LETTER DA FIRST FINAL FORM
046	2E	MONGOLIAN LETTER DA SECOND FINAL FORM
047	2F	MONGOLIAN LETTER CHA MEDIAL FORM
048	30	MONGOLIAN LETTER CHA FINAL FORM
049	31	MONGOLIAN LETTER JA FIRST MEDIAL FORM
050	32	MONGOLIAN LETTER JA SECOND FINAL FORM
051	33	MONGOLIAN LETTER RA FINAL FORM
052	34	MONGOLIAN LETTER FA FINAL FORM
053	35	MONGOLIAN LETTER KA FINAL FORM
054	36	MONGOLIAN LETTER KHA FINAL FORM
055	37	MONGOLIAN LETTER TSA MEDIAL FORM
056	38	MONGOLIAN LETTER TSA FINAL FORM
057	39	MONGOLIAN LETTER ZA MEDIAL FORM
058	3A	MONGOLIAN LETTER ZA FINAL FORM
059	3B	MONGOLIAN LETTER HAA FINAL FORM
060	3C	MONGOLIAN LETTER ZRA FINAL FORM
061	3D	MONGOLIAN LETTER LHA MEDIAL FORM
062	3E	MONGOLIAN LETTER TODO LONG VOWEL SIGN FINAL FORM
063	3F	MONGOLIAN LETTER TODO E FIRST MEDIAL FORM

dec	hex	Name
064	40	MONGOLIAN LETTER TODO E SECOND MEDIAL FORM
065	41	MONGOLIAN LETTER TODO I INITIAL FORM
066	42	MONGOLIAN LETTER TODO I FIRST MEDIAL FORM
067	43	MONGOLIAN LETTER TODO I SECOND MEDIAL FORM
068	44	MONGOLIAN LETTER TODO I FINAL FORM
069	45	MONGOLIAN LETTER TODO O INITIAL FORM
070	46	MONGOLIAN LETTER TODO O FIRST MEDIAL FORM
071	47	MONGOLIAN LETTER TODO O SECOND MEDIAL FORM
072	48	MONGOLIAN LETTER TODO O FINAL FORM
073	49	MONGOLIAN LETTER TODO U SECOND ISOLATE FORM
074	4A	MONGOLIAN LETTER TODO U INITIAL FORM
075	4B	MONGOLIAN LETTER TODO U SECOND MEDIAL FORM
076	4C	MONGOLIAN LETTER TODO U THIRD MEDIAL FORM
077	4D	MONGOLIAN LETTER TODO U FIRST FINAL FORM
078	4E	MONGOLIAN LETTER TODO OE INITIAL FORM
079	4F	MONGOLIAN LETTER TODO OE FIRST MEDIAL FORM
080	50	MONGOLIAN LETTER TODO OE SECOND MEDIAL FORM
081	51	MONGOLIAN LETTER TODO OE FINAL FORM
082	52	MONGOLIAN LETTER TODO PA FINAL FORM
083	53	MONGOLIAN LETTER TODO GA FIRST MEDIAL FORM
084	54	MONGOLIAN LETTER TODO GA SECOND MEDIAL FORM
085	55	MONGOLIAN LETTER TODO GA FINAL FORM
086	56	MONGOLIAN LETTER TODO TA FINAL FORM
087	57	MONGOLIAN LETTER TODO CHA MEDIAL FORM
088	58	MONGOLIAN LETTER TODO CHA FINAL FORM
089	59	MONGOLIAN LETTER TODO JA MEDIAL FORM
090	5A	MONGOLIAN LETTER TODO JA FINAL FORM
091	5B	MONGOLIAN LETTER TODO WA FINAL FORM
092	5C	MONGOLIAN LETTER TODO KA FINAL FORM
093	5D	MONGOLIAN LETTER TODO HAA MEDIAL FORM
094	5E	MONGOLIAN LETTER TODO DZA MEDIAL FORM
095	5F	MONGOLIAN LETTER TODO DZA FINAL FORM
096	60	MONGOLIAN LETTER SIBE E FIRST MEDIAL FORM
097	61	MONGOLIAN LETTER SIBE I THIRD MEDIAL FORM
098	62	MONGOLIAN LETTER SIBE I SECOND FINAL FORM
099	63	MONGOLIAN LETTER SIBE I THIRD FINAL FORM
100	64	MONGOLIAN LETTER SIBE IY FINAL FORM
101	65	MONGOLIAN LETTER SIBE UE INITIAL FORM
102	66	MONGOLIAN LETTER SIBE UE FIRST MEDIAL FORM
103	67	MONGOLIAN LETTER SIBE UE FIRST FINAL FORM
104	68	MONGOLIAN LETTER SIBE KA SECOND MEDIAL FORM
105	69	MONGOLIAN LETTER SIBE GA MEDIAL FORM
106	6A	MONGOLIAN LETTER SIBE GA FEMININE ISOLATE FORM
107	6B	MONGOLIAN LETTER SIBE HA MEDIAL FORM
108	6C	MONGOLIAN LETTER SIBE HA FEMININE ISOLATE FORM
109	6D	MONGOLIAN LETTER SIBE SHA MEDIAL FORM
110	6E	MONGOLIAN LETTER SIBE SHA FINAL FORM
111	6F	MONGOLIAN LETTER SIBE TA SECOND MEDIAL FORM
112	70	MONGOLIAN LETTER SIBE DA SECOND INITIAL FORM
113	71	MONGOLIAN LETTER SIBE DA FIRST MEDIAL FORM
114	72	MONGOLIAN LETTER SIBE DA SECOND MEDIAL FORM
115	73	MONGOLIAN LETTER SIBE TSA MEDIAL FORM
116	74	MONGOLIAN LETTER SIBE ZA SECOND INITIAL FORM
117	75	MONGOLIAN LETTER SIBE ZA FIRST MEDIAL FORM
118	76	MONGOLIAN LETTER SIBE ZA SECOND MEDIAL FORM
119	77	MONGOLIAN LETTER SIBE CHA MEDIAL FORM
120	78	MONGOLIAN LETTER MANCHU KA FEMININE SECOND MEDIAL FORM
121	79	MONGOLIAN LETTER MANCHU KA FEMININE FIRST FINAL FORM
122	7A	MONGOLIAN LETTER MANCHU KA FEMININE SECOND FINAL FORM
123	7B	MONGOLIAN LETTER MANCHU ZHA MEDIAL FORM
124	7C	MONGOLIAN LETTER ALI GALI ANUSVARA ONE SECOND FORM
125	7D	MONGOLIAN LETTER ALI GALI VISARGA ONE SECOND FORM
126	7E	MONGOLIAN LETTER ALI GALI A SECOND ISOLATE FORM
127	7F	MONGOLIAN LETTER ALI GALI A FIRST FINAL FORM

Names of Presentation forms (continued)

dec	hex	Name
128	80	MONGOLIAN LETTER ALI GALI A SECOND FINAL FORM
129	81	MONGOLIAN LETTER ALI GALI A THIRD FINAL FORM
130	82	MONGOLIAN LETTER ALI GALI A FOURTH FINAL FORM
131	83	MONGOLIAN LETTER ALI GALI I FIRST FINAL FORM
132	84	MONGOLIAN LETTER ALI GALI KA INITIAL FORM
133	85	MONGOLIAN LETTER ALI GALI NGA SECOND INITIAL FORM
134	86	MONGOLIAN LETTER ALI GALI NGA FIRST MEDIAL FORM
135	87	MONGOLIAN LETTER ALI GALI NGA SECOND MEDIAL FORM
136	88	MONGOLIAN LETTER ALI GALI CA MEDIAL FORM
137	89	MONGOLIAN LETTER ALI GALI SSA MEDIAL FORM
138	8A	MONGOLIAN LETTER ALI GALI ZA MEDIAL FORM
139	8B	MONGOLIAN LETTER MANCHU ALI GALI GHA MEDIAL FORM
140	8C	MONGOLIAN LETTER MANCHU ALI GALI NGA MEDIAL FORM
141	8D	MONGOLIAN LETTER MANCHU ALI GALI CA MEDIAL FORM
142	8E	MONGOLIAN LETTER MANCHU ALI GALI JHA MEDIAL FORM
143	8F	MONGOLIAN LETTER MANCHU ALI GALI TTA MEDIAL FORM
144	90	MONGOLIAN LETTER MANCHU ALI GALI DDHA MEDIAL FORM
145	91	MONGOLIAN LETTER MANCHU ALI GALI DHA MEDIAL FORM
146	92	MONGOLIAN LETTER MANCHU ALI GALI CYA MEDIAL FORM
147	93	MONGOLIAN LETTER MANCHU ALI GALI ZHA MEDIAL FORM
148	94	MONGOLIAN LETTER MANCHU ALI GALI ZA MEDIAL FORM

In normal Mongolian text, the correct variant of any given positional form of a letter can in most cases be determined unambiguously from the context using a set of rules involving the preceding and following letters, the syllable in the word, and the gender of the word. In these cases, software supporting Mongolian could generate the appropriate variant form of each letter automatically on input.

In a few situations, however, the rules are not sufficient to determine the correct variant form uniquely, and there can be an essentially arbitrary choice between two or more possible alternatives. Then a software system could at best generate one of the possible alternatives automatically as a default, the other possible alternatives being obtained by manually overriding this default as described in Section 3.1.

3.1 Overriding the Defaults

The default positional form of a Mongolian letter can be overridden using the zero width joiner (Z-W Join) and non-joiner (Z-W N-J) (characters 200D and 200C in the General Puntuation block respectively): in the rules for determining the correct positional form the non-joiner effectively acts as an invisible space while the joiner acts as an invisible letter.

Thus, for example, the initial, medial and final forms of any character can be printed as a single character surrounded by white space as follows:

initial form: space + character + zero-width joiner + space
medial form: space + zero-width joiner + character + zero-width joiner + space
final form: space + zero-width joiner + character + space

More generally, appending a zero-width joiner to the beginning of a sequence of two or more letters converts the first letter in the sequence from initial form to medial form, while appending it to the end of such a sequence converts the last letter in the sequence

from final form to medial form. Inserting a zero-width joiner into the middle of such a sequence has no effect. Thus, for example, the Mongolian word ᠰᠤᠷᠭᠠᠭᠤᠯᠢ (school) can be split into its separate syllables ᠰᠤᠷ ᠭᠠ ᠭᠤᠯᠢ using the zero-width joiner as follows:

ᠰ ᠤ ᠷ [ZWJ] [ZWJ]ᠭ ᠠ[ZWJ] [ZWJ]ᠭ ᠤ ᠯ ᠢ

and the Todo word [illegible] (school) can similarly be split into its syllables [illegible] thus:

[illegible] .

The zero-width non-joiner only produces a visible effect when it is inserted between two letters. In such a situation, it has the effect of breaking the cursive connection between the two letters, thus effectively splitting the sequence into two at that position. The letter immediately preceding the non-joiner is thus reated as if it were the end of one sequence, and hence would default to final form (assuming there were some other letters preceding it), while the letter immediately following the non-joiner is treated as if it were the beginning of another sequence, and hence would default to initial form (assuming there were some other letters following it).

A combination of one zero-width joiner and one zero-width non-joiner, in either order, also has a visible effect when inserted into the middle of a sequence of letters. If the joiner precedes the non-joiner, the cursive connection is only broken on the right, so the letter to the left of the break retains its original default positional form while the one on the right becomes initial form. If, on the other hand, the joiner follows the non-joiner, the cursive connection is only broken on the left, so the letter on the left of the break becomes final form while the letter on the right retains its original default positional form.

Two adjacent joiners have the same effect as a single joiner, and similarly two adjacent non-joiners have the same effect as a single non-joiner.

Finally, two joiners separated by a non-joiner and two non-joiners separated by a joiner have the same effect as a single joiner or a single non-joiner respectively. Any sequence consisting of three or more joiners and non-joiners in any order can therefore be reduced to either a single joiner, a single non-joiner or a joiner/non-joiner pair.

The default or correct variant forms can simply be overridden by inserting the appropriate Mongolian free variant selector after the letter to be changed.

The following examples illustrate the use of the zero-width joiner and zero-width non-joiner**:**

Display | Character sequence | Display | Character sequence

3.2 The Mongolian Reference Table

The Mongolian Reference Table on the following pages shows all the different variant forms of each of the basic Mongolian characters.

The basic characters, together with their (decimal) codes and glyphs, are listed in the first column of the table (headed "Basic Characters"). The next column (headed "Variant Forms") shows the glyphs and the names of all the variant forms of each character.

The particular variant form which occurs on the same horizontal line as the name of the basic character in the first column is the variant which belongs to the basic character set. All other variant forms are numbered as follows: if the glyph of the variant has the same shape as that of one of the basic characters, the (decimal) number of that basic character is shown in the left-hand column under the heading "No."; if, on the other hand, the glyph of the variant corresponds to one of the presentation forms, the (decimal) number of that presentation form is shown in the right-hand column under the heading "No.".

The final column (headed "Rule") of this section of the table shows the sequence of basic characters (including zero-width joiners and non-joiners where necessary) which can be used to generate the particular variant in isolation (that is, as a single character surrounded by white space).

Note that not all positional variants are defined for all characters: for example, only medial and final forms are given for character 1829, MONGOLIAN LETTER ANG (ᠸ). This means that the particular character is not found at all positions in words in normal Mongolian text: in the case of the letter "ANG" it is never found as the first character of a Mongolian word. However, this does not mean that such a character can never occur in one of these "impossible" positions – it is, of course, quite possible to use the zero-width joiner to build an arbitrary string of characters with the letter "ANG" at the beginning even though this would not correspond to a real Mongolian word.

The last column (headed "Usage") in the table shows, for each of the four scripts, the letter in that script to which each particular variant corresponds. A blank space in this column indicates that the particular letter is not used in that script.

4 The Ligatures

In Mongolian script, a pair of letters consisting of a "bowed" consonant (that is a consonant without a trailing vertical stem, for example characters 182A MONGOLIAN LETTER BA (ᠪ), 183A MONGOLIAN LETTER KA (ᠺ), and 183B MONGOLIAN LETTER KHA (ᠻ)) followed by a vowel generally combine to form a ligature. The set of all different ligatures in Mongolian is shown in table "Mongolian Ligatures".

As for the basic characters and the presentation forms, all ligatures in the table have visually distinct forms, though one such form may in fact correspond to more than one different combination of letters.

Each ligature is assigned a (decimal) identification number, which appears in the first column of the table, and this is followed by the ligature's glyph and its (unique) name. The remainder of the table shows, for each of the different scripts, to which combinations of letters at which positions the ligature corresponds: isolate (column headed "ISO"); initial (column headed "INI"); medial (column headed "MED); and final column headed "FIN"). Finally, the column headed "RULE" shows sequences of characters which generate the versions of each ligature corresponding to its different possible spellings as a stand-alone symbol.

Ligature Set

	F40	F41	F42	F43	F44	F45	F46	F47
0	 0	 16	 32	 48	 64	 80	 96	 112
1	 1	 17	 33	 49	 65	 81	 97	 113
2	 2	 18	 34	 50	 66	 82	 98	114
3	 3	 19	 35	 51	 67	 83	 99	115
4	 4	 20	 36	 52	 68	 84	 100	116
5	5	21	37	53	69	85	101	117
6	6	22	38	54	70	86	102	118
7	7	23	39	55	71	87	103	119
8	8	24	40	56	72	88	104	120
9	9	25	41	57	73	89	105	121
A	10	26	42	58	74	90	106	122
B	11	27	43	59	75	91	107	123
C	12	28	44	60	76	92	108	124
D	13	29	45	61	77	93	109	125
E	14	30	46	62	78	94	110	126
F	15	31	47	63	79	95	111	127

Ligature Set (continued)

	F48	F49	F4A	F4B	F4C
0	128	144	160	176	192
1	129	145	161	177	193
2	130	146	162	178	194
3	131	147	163	179	195
4	132	148	164	180	196
5	133	149	165	181	197
6	134	150	166	182	198
7	135	151	167	183	199
8	136	152	168	184	200
9	137	153	169	185	201
A	138	154	170	186	202
B	139	155	171	187	203
C	140	156	172	188	204
D	141	157	173	189	205
E	142	158	174	190	206
F	143	159	175	191	207

5 Implementing Software for Mongolian

A text processing system supporting Mongolian requires a font containing all the characters of the basic Mongolian character set as well as their variant presentation forms and the ligatures. To conform to the standards, the characters in the basic character sets must be situated at the coding positions given in this paper. However, the presentation forms and the ligatures are not explicitly part of the standards so they have no fixed coding positions; they should instead be coded at some point within what is known as the "private use area". Interchange of documents which include characters outside the basic character set is then only guaranteed to respect the sense of the document if the various parties have all agreed on the coding positions of the presentation forms and the ligatures within the private use area.

The tables given in this paper make a specific choice for the coding positions within the private use area and in fact code the presentation forms at positions F300 to F395 and the ligatures at positions F400 to F4C1.

The mechanism of inputting characters is not specified by the standard, so any keyboard driver capable of generating the appropriate 16-bit character encodings can be used. However, the input mechanism should ideally generate the correct positional forms, variants and ligatures on input by analysis of the context of each letter.

The standard also does not specify how traditional Mongolian should be intermixed with other scripts. This is an important question because the traditional Mongolian script is correctly written vertically in columns progressing from left to right while most other scripts in the world are written in a different orientation: for example, the Cyrillic script, which frequently appears together with traditional Mongolian script on official documents in Mongolia, is properly written in horizontal lines which are read from left to right.

To be absolutely correct, when Mongolian script is intermixed with a script having horizontal, left-to-right orientation like the Cyrillic script each script should retain in its own individual orientation. However, in cases where this correct bidirectionality cannot be achieved, one of the scripts can lose its natural orientation and instead adopt the orientation of the other. In such cases, it is often written with its characters rotated through 90 degrees. Thus, for example, if the Mongolian script adopts the horizontal, left-to-right orientation of the Cyrillic script, its characters are rotated by 90 degrees anticlockwise, and the columns are transcribed to the equivalent lines (first column becomes first line, etc.), while if the Cyrillic script adopts the vertical, left-to-right orientation of the Mongolian script (though this is much less common) its characters are rotated by 90 degrees clockwise and the lines are transcribed to columns in the opposite order (last line becomes first column, etc.). Examples of the rotation of traditional Mongolian script to bring it into alignment with English text can in fact be found throughout this paper.

Mixing traditional Mongolian with script which have other orientations is also possible in a similar way: if the two scripts cannot both retain their correct orientation one can adopt that of the other, usually rotating its characters when one script is horizontally oriented and the other vertical. The basic rule to follow is that the text of the modified

script should be readable normally if the whole "page" is rotated in such a way as to return it to its original orientation.

Since the standardisation of traditional Mongolian is comparatively recent, most of the software supporting traditional Mongolian does not conform fully to the standard. UNU/IIST's Multiscript project has in fact designed and is building a prototype software system which not only supports traditional Mongolian in its correct orientation, but which also supports more general multi-directional multi-lingual documents. This Multiscript system is also compatible with the ISO/IEC 10646 and Unicode standards, and it supports traditional Mongolian script basically as described here, including supporting all the presentation forms and ligatures: in fact the traditional Mongolian font which has been used in the preparation of this paper (and which is available from UNU/IIST) has been created as part of the implementation of the Multiscript prototype. More information about the Multiscript system can be found in the range of reports and papers [1,3,4,5,6,7,9] or can be obtained direct from the authors.

Acknowledgements

The encoding scheme for traditional Mongolian script described here forms part of the international standard encoding system ISO 10646 and was developed collaboratively by a group of members of ISO/IEC JTC1 SC2 WG2. The authors are grateful for numerous discussions with other members, especially with Ken Whistler and Asmus Freytag of the Unicode Consortium and Choijinjaw of the Inner Mongolian University, Huhhot. Myatav Erdenechimeg thanks UNU/IIST for its hospitality during the course of this work.

References

[1] Avirmed Amar, Myatav Erdenechimeg, and Richard Moore. Implementation of the MultiScript Multi-lingual Document Processing System. Technical Report 160, UNU/IIST, P.O.Box 3058, Macau, March 1999.

[2] The Unicode Consortium. *The Unicode Standard, Version 2.0*. Addison Wesley, 1996.

[3] Myatav Erdenechimeg and Richard Moore. Multi-directional Multi-lingual Script Processing. Technical Report 75, UNU/IIST, P.O.Box 3058, Macau, June 1996. Published in Proceedings of the Seventeenth International Conference on the Computer Processing of Oriental Languages, Hong Kong, April 2 - 4, 1997, under the title *Multi-directional Multi-lingual Script Processing*.

[4] Myatav Erdenechimeg and Richard Moore. Multi-directional Multi-lingual Script Processing. In *Proceedings of the Seventeenth International Conference on the Computer Processing of Oriental Languages, Vol. 1*, pages 29 -- 34. Oriental Languages Computer Society, Inc., 1997.

[5] Myatav Erdenechimeg and Richard Moore. MultiScript III: Creating and Editing Multi-lingual Documents. Technical Report 113, UNU/IIST, P.O.Box 3058, Macau, September 1997. Revised June 1998.

[6] Myatav Erdenechimeg, Richard Moore, and Yumbayar Namsrai. MultiScript I: The Basic Model of Multi-lingual Documents. Technical Report 105, UNU/IIST, P.O.Box 3058, Macau, June 1997. A part of the work has been presented at and published in the proceedings of the Workshop on the *Principles of Digital Document Processing*, March 1998, St. Malo, France, Ethan V. Munson, Charles Nicholas and Derick Wood (Eds), Lecture Notes in Computer Science 1481, Springer Verlag, 1998, pages 70 - 81.

[7] Myatav Erdenechimeg, Richard Moore, and Yumbayar Namsrai. On the Specification of the Display of Documents in Multi-lingual Computing. In Ethan V. Munson, Charles Nicholas, and Derick Wood, editors, *Principles of Digital Document Processing*, volume 1481 of *Lecture Notes in Computer Science*, pages 70--81. Springer Verlag, 1998.

[8] International Organization for Standardization. *ISO 10646-1: Information Technology -- Universal Multiple-Octet Coded Character Set (UCS) -- Part 1: Architecture and Basic Multilingual Plane.*

[9] Yumbayar Namsrai and Richard Moore. MultiScript II: Displaying and Printing Multi-lingual Documents. Technical Report 112, UNU/IIST, P.O.Box 3058, Macau, June 1997. A part of the work has been presented at and published in the proceedings of the Workshop on the *Principles of Digital Document Processing*, March 1998, St. Malo, France, Ethan V. Munson, Charles Nicholas and Derick Wood (Eds), Lecture Notes in Computer Science 1481, Springer Verlag, 1998, pages 70 - 81.