





consciously from a wide range of possibilities and will be valuable for applications.

3. In addition to its original intention of allowing phonetically differentiating words, the great flexibility also allows intended alternative character sequences of a single word according to different phonetic letter identifications.
4. It was also underestimated that how unintended mistakes would hide in confusables and make the phonetic information in text unreliable from the moment text is input.
5. Significant security issues arise when the large number of possible character sequences for a single grapheme sequence can be exploited maliciously.
6. Many confusables can not be even easily detected manually by users (eg,  U+1823 0 and  U+1824 U).

且对具体应用会有价值。

- 在允许区分单词语音的原本意图之外，巨大的灵活性还允许根据不同的语音字母识别为一个单词故意提供不同的字符序列。
- 无意的错误会隐藏在混淆结构里使文本中的语音信息从输入的那一刻起就不可靠，这一问题被低估了。
- 单个字位序列有大量可能的字符序列可受到恶意利用是显著的安全问题。
- 许多混淆结构甚至无法由用户简单地手动检测（比如  U+1823 0 和  U+1824 U）。

1.1 Controversial phonetic letter sequences 有争议的的语音字母序列


See figure 1 for how different phonetic letter sequences can be intendedly identified according to different considerations that are etymological (word 1), orthographical (word 2), graphetic (word 3 and 4), morphological (word 5), or phonetic (word 6), as well how certain words (7, 8, and 9) are simply opaque.

如图1所示，根据在语源（单词1）、正词法（单词2）、字形（单词3和4）、形态（单词5）、语音（单词6）上的不同考量，以及一些单纯就是难解的单词（7、8、9），会有意识地识别出不同的语音字母序列。

This complication requires applications to have knowledge of the exact user intention in advance in order to utilize the phonetic information, or to ignore many phonetic details and only to fuzzily match encoded words to a limited built-in dictionary.

这一复杂情况要求在具体应用中为了利用语音信息，对具体的用户意图要有预知，或者就要忽略掉许多语音细节然后仅仅模糊地把编码的单词映射到有限的内置词典中。

1.2 Unintended or malicious character sequences 无意或恶意的字符序列

Partly due to the difficulty of mastering the complicated contextual rules of the Mongolian encoding, users often freely assemble unexpected characters together to form the desired grapheme sequence, without caring about if the phonetic information is correct. This is especially common for spellings that require special format characters  U+202F NNBS and FVSes, which are hard to ma-


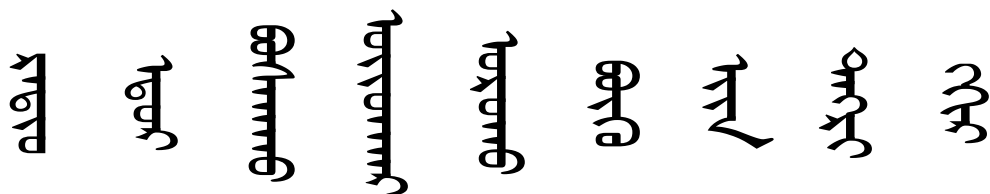
在一定程度上正因为蒙古文编码的复杂上下文规则难以掌握，用户经常自由地把意料之外的字符拼凑为想要的字位序列，而不管语音信息是否正确。对于难以操控的特殊格式字符  U+202F NNBS 和 FVS 这个问题尤为常见。

Figure 图 1: Intended differences in phonetic letter identification. 语音字母识别中有意的区别。



$$1) \begin{matrix} \text{𑌵} & 1837 & \text{RA} & \text{𑌶} & 1820 & \text{A} & \text{𑌷} & 1833 & \text{DA} & \text{𑌸} & 1822 & \text{I} \end{matrix} \left\{ \begin{matrix} \text{𑌹} & 1823 & \text{O} & \text{FVS} \\ \text{𑌺} & 1824 & \text{U} & \text{FVS} \end{matrix} \right\}$$

$$2) \begin{matrix} \text{𑌻} & 1821 & \text{E} & \text{𑌷} & 1833 & \text{DA} \end{matrix} \left\{ \begin{matrix} \text{𑌼} & 1826 & \text{UE} \\ \text{𑌽} & 1825 & \text{OE} \end{matrix} \right\} \text{𑌵} & 1837 & \text{RA}$$

$$3) \begin{matrix} \text{𑌾} & 182\text{A} & \text{BA} & \text{𑌹} & 1823 & \text{O} & \text{𑌿} & 1830 & \text{SA} \end{matrix} \left\{ \begin{matrix} \text{𑍀} & 182\text{C} & \text{QA} \\ \text{𑍁} & 182\text{D} & \text{GA} \end{matrix} \right\} \text{𑌶} & 1820 & \text{A} & \text{𑍂} & 182\text{C} & \text{QA} & \text{𑌺} & 1824 & \text{U}$$

$$4) \begin{matrix} \text{𑌹} & 1824 & \text{U} & \text{𑌸} & 1822 & \text{I} \end{matrix} \left\{ \begin{matrix} \text{𑌷} & 1833 & \text{DA} & \text{𑍃} & 182\text{C} & \text{QA} \\ \text{𑍄} & 1832 & \text{TA} & \text{𑍁} & 182\text{D} & \text{GA} \end{matrix} \right\} \text{𑌶} & 1820 & \text{A} & \text{𑌵} & 1837 & \text{RA}$$

$$5) \begin{matrix} \text{𑌹} & 1823 & \text{O} & \text{𑌵} & 1837 & \text{RA} & \text{𑌹} & 1823 & \text{O} & \text{𑍃} & 182\text{C} & \text{QA} \end{matrix} \left\{ \begin{matrix} \text{𑌹} & 1824 & \text{U} \\ \text{𑌹} & 1823 & \text{O} \end{matrix} \right\}$$

$$6) \begin{matrix} \text{𑌾} & 182\text{A} & \text{BA} \end{matrix} \left\{ \begin{matrix} \text{𑌼} & 1826 & \text{UE} \\ \text{𑌽} & 1825 & \text{OE} \end{matrix} \right\} \text{𑍃} & 182\text{C} & \text{QA} \left\{ \begin{matrix} \text{𑌼} & 1826 & \text{UE} \\ \text{𑌽} & 1825 & \text{OE} \end{matrix} \right\}$$

$$7) \left\{ \begin{matrix} \text{𑌻} & 1821 & \text{E} & \text{𑌸} & 1822 & \text{I} \\ \text{𑌸} & 1822 & \text{I} & \text{𑌸} & 1822 & \text{I} \end{matrix} \right\} \text{𑍅} & 1828 & \text{NA}$$

$$8) \begin{matrix} \text{𑍄} & 1832 & \text{TA} \end{matrix} \left\{ \begin{matrix} \text{𑍆} & 1829 & \text{ANG} \\ \text{𑌻} & 1821 & \text{E} & \text{𑍁} & 182\text{D} & \text{GA} \end{matrix} \right\} \text{𑌵} & 1837 & \text{RA} & \text{𑌸} & 1822 & \text{I}$$

$$9) \left\{ \begin{matrix} \text{𑍁} & 182\text{D} & \text{GA} & (\text{FVS})? \\ \text{𑍃} & 182\text{C} & \text{QA} & (\text{FVS})? \end{matrix} \right\} \text{𑍃} & 182\text{C} & \text{QA} & \text{𑌸} & 1822 & \text{I} & \text{𑌵} & 1837 & \text{RA}$$

nipulate.

The number of resulted possibilities is much higher than that of intended, which is the reason of security concerns and why the Mongolian encoding will not be supported in crucial identifiers like internationalized domain names and email addresses.

See also [UTS #39 Unicode Security Mechanisms](#). However it should be note that, the mechansim provided by UTS #39 is insufficient to handle the complication of the Mongolian encoding.

See chart *Variants required by the modern Hudum* for a full list of variants in the Mongolian encoding needed for the modern Hudum, from which confusables between characters and character sequences can be summarized. Note the Mongolian encoding basically allow any variant of a character to be requested manually.

2 Positional mismatches 位置失配

See [WG2 4884 / L2/17-332 Positional mismatches in Mongolian encoding](#) for a complete discussion.

1. The Mongolian encoding is expected to be implemented with the Unicode–OpenType Arabic cursive joining model, and its standardized variation sequence names also suggest so. But the positional variants do not actually conform the model.
2. The variants do not form a self-consistent system either, therefore can not even be losslessly mapped to the Arabic cursive joining model.
3. The positional variants in the Mongolian encoding were the result of inconsistently analyzing the shaping behavior with controversial grammatical and orthographical concepts of the Mongolian language (such as word and enclitic, which were not well defined in the specification).
4. Certain variants are bound to special contexts. Eg, $\! \not\! \text{NA.1.fina}$ is considered “medial” because *in word* it usually preceds a word-fina disjointed tail $\! \not\! \text{ᠰ}$, but it also has usage unrelated to the disjointed tail which is a typical “final” no matter “in word” or “cursive joining” is considered.

这样导致的可能性数量比有意的还高许多，这就是安全隐患的原因，并且是为什么蒙古文编码不会在国际化域名和邮箱地址这样关键的标识符中受到支持。

参见 [UTS #39 Unicode Security Mechanisms](#) (Unicode 安全机制)。不过应当注意，UTS #39 提供的机制对于处理蒙古文编码的复杂情况来说是不足的。

图表《现代胡都木文要求的变体》展示了蒙古文编码中现代胡都木文所需变体的总表，从中可以总结中字符及字符序列间的混淆结构。注意，蒙古文编码基本上允许手动请求一个字符的任何变体。

参见 [WG2 4884 / L2/17-332 《蒙系文字编码中的位置失配》](#) 中的完整讨论。

- 蒙古文编码是预期由 Unicode–OpenType 阿拉伯连写模型实现的，它的标准化变体序列名称也表明这一点。但其位置变体并不遵守该模型。
- 其变体也不形成自洽的系统，所以甚至都不能无损映射到阿拉伯连写模型。
- 蒙古文编码的位置变体来自用有争议的蒙古语语法与正词法概念（比如没有在规范中良好定义的词和附加成分）对成形行为的不一致分析。
- 某些变体限制于特殊的上下文。例如， $\! \not\! \text{NA.1.fina}$ 认为是“中形”因为它一般“在词里”出现在词末的分尾前，但它也有和分尾无关的不论考虑“在词里”还是“连写”都是典型“下形”的

- Usage of nirugu is explicitly mentioned in the GB but actually breaks contextual rules.
- ZWNJ and ZWJ's effect to contextual rules is not defined.
- An open set of additions of a grapheme used for different letters. (Missing tailless U.1.fina , now OE alt.)

FVS0。

- Nirugu 的用法在国标中明确提及了但实际上会打破上下文规则。
- ZWNJ 和 ZWJ 对上下文规则的效用未定义。
- 会不断为不同字母增加字位。(之前缺失的 U.1.fina ，现在要加入 OE 的变体。)

3.1 The disjointed tail 分尾

The disjointed tail (U+1820 A or U+1821 E triggered by a format character U+180E MVS) is a special case of letters U+1820 A or U+1821 E triggered by a format character U+180E MVS .

- The special graphemes of certain letters preceding the disjointed tail are contextually selected and have inconsistent cursive joining positions.
- The syllable structure limitation forces U+1836 YA and U+1838 WA to have confusable and debatable (vowels are sometimes identified, especially for U+1838 WA) variants only used here.
- The interaction between each other of the preceding letter, U+180E MVS , and the vowel (U+1820 A or U+1821 E) is not clearly defined and has usability issues.
- It is undefined what should happen when a suffix is appended.

分尾 (U+1820 A 或 U+1821 E 由格式字符 U+180E MVS 触发的特例。

- 分尾前某些字母的特殊字位是上下文选择的，有不一致的连写位置。
- 音节结构的限制强迫 U+1836 YA 和 U+1838 WA 有只在这里使用的易混淆且有争议 (有时会识别为元音，尤其对于 U+1838 WA) 的变体。
- 前侧字母、 U+180E MVS 、元音 (U+1820 A 或 U+1821 E) 之间的交互没有清楚定义，而且有可用性问题。
- 未定义当附加后缀时应当发生什么。

3.2 Syllables 音节

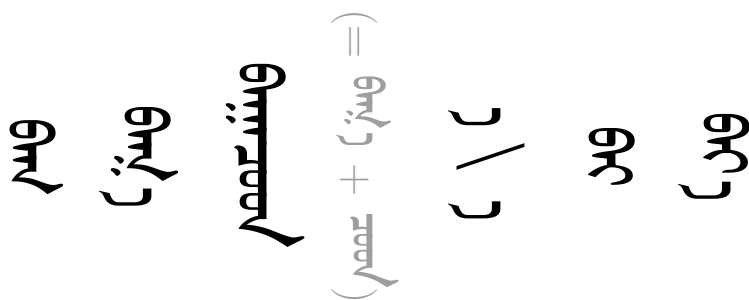
Syllabification algorithm is not defined.

- The identification of underlying phonetic letters of diphthongs is heavily debated.
- Expected behavior of stray consonants is not fully specified.
- The behavior of coda consonants in syllables is relatively well specified, but special syllable boundaries often lead to the need of FVSes for overriding syllabification.

音节划分算法没有定义。

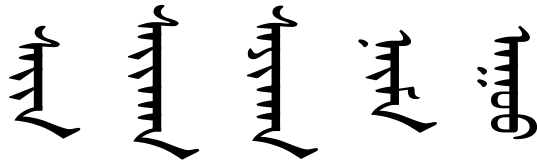
- 对底层语音字母的识别有激烈的争论。
- 游离辅音的预期行为没有充分定义。
- 音节尾辅音的行为相对来说是良好定义的，但特殊的音节界限常常导致需要 FVS 来手动控制音节划分。

Figure 图 2: The disjointed tail. 分尾。



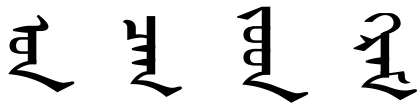
- 1) מִשְׁ 182A BA מִשְׁ 1820 A מִשְׁ 182D GA
- 2) מִשְׁ 182A BA מִשְׁ 1820 A מִשְׁ 182D GA MVS 180E MVS מִשְׁ 1820 A
- 3) מִשְׁ 182A BA מִשְׁ 1820 A מִשְׁ 182D GA (MVS 180E MVS)? מִשְׁ 1820 A מִשְׁ 1834 CHA מִשְׁ 1824 U מִשְׁ 1833 DA
- 4) MVS 202F NNBSP $\left\{ \begin{array}{l} מִשְׁ 1820 A \\ מִשְׁ 1821 E \end{array} \right\}$
- 5) מִשְׁ 182A BA מִשְׁ 1820 A $\left\{ \begin{array}{l} מִשְׁ 1822 I \\ מִשְׁ 1836 YA \end{array} \right\}$
- 6) מִשְׁ 182A BA מִשְׁ 1821 E מִשְׁ 1836 YA MVS 180E MVS מִשְׁ 1821 E

Figure 图 3: Postvocalic /i/ and related cases. 元音后 /i/ 以及相关事例。



- 1) $\text{ᠰᠠ} 1830 \text{ SA } \text{ᠠ} 1820 \text{ A } \left\{ \begin{array}{l} \text{ᠢ} 1822 \text{ I (FVS)?} \\ \text{ᠶᠠ} 1836 \text{ YA (FVS)? } \text{ᠢ} 1822 \text{ I} \\ \dots \end{array} \right\} \text{ᠨᠠ} 1828 \text{ NA}$
- 2) $\text{ᠰᠠ} 1830 \text{ SA } \text{ᠠ} 1820 \text{ A } \left\{ \begin{array}{l} \text{ᠢ} 1822 \text{ I (FVS)?} \\ \text{ᠶᠠ} 1836 \text{ YA (FVS)? } \text{ᠢ} 1822 \text{ I} \\ \dots \end{array} \right\} \text{ᠠ} 1820 \text{ A } \text{ᠶᠠ} 1820 \text{ YA } \text{ᠨᠠ} 1828 \text{ NA}$
- 3) $\text{ᠰᠠ} 1830 \text{ SA } \text{ᠠ} 1820 \text{ A } \text{ᠶᠠ} 1836 \text{ YA (FVS)? } \text{ᠢ} 1822 \text{ I } \text{ᠠ} 1820 \text{ A } \text{ᠶᠠ} 1820 \text{ YA } \text{ᠨᠠ} 1828 \text{ NA}$
- 4) $\text{ᠨᠠ} 1828 \text{ NA } \text{ᠠ} 1820 \text{ A } \left\{ \begin{array}{l} \text{ᠢ} 1822 \text{ I (FVS)?} \\ \text{ᠶᠠ} 1836 \text{ YA (FVS)?} \end{array} \right\} \text{ᠮᠠ} 1820 \text{ MA } \text{ᠠ} 1820 \text{ A}$
- 5) $\text{ᠠ} 1820 \text{ A } \text{ᠠ} 1820 \text{ GA } \text{ᠤ} 1824 \text{ U } \left\{ \begin{array}{l} \text{ᠤ} 1824 \text{ U} \\ \text{ᠠ} 1838 \text{ WA (FVS)?} \end{array} \right\}$

Figure 图 4: Coda and stray consonants. 音节尾与游离辅音。



- 1) $\text{ᠣ} 1823 \text{ O } \text{ᠨᠠ} 1828 \text{ NA}$
- 2) $\text{ᠴᠬᠠ} 1834 \text{ CHA } \text{ᠠ} 1820 \text{ A } \text{ᠠ} 1820 \text{ GA}$
- 3) $\text{ᠵᠠ} 1835 \text{ JA } \text{ᠤ} 1824 \text{ U } \text{ᠳᠠ} 1833 \text{ DA}$
- 4) $\text{ᠠ} 1820 \text{ GA } \text{ᠷᠠ} 1837 \text{ RA } \text{ᠠ} 1820 \text{ A } \text{ᠮᠠ} 1820 \text{ MA}$

3.3 Word stems 词干

There are three mechanisms directly related to word stems, but word stems are not marked in encoding:

1. Stem-beginning vowels: aleph-ed *init* or *medi*, and aleph-less enclitics.
2. First-in-stem vowels: the first or second stem, and enclitics.
3. Vowel harmony.

These word stem related mechanisms are bound to the cursive joining mechanism which is not directly related, leading the cases not shown in the cursive joining mechanism being overlooked.

4. Vowels have two major trans-graphemic shaping mechanisms: *stem-beginning* vowels (the prepended aleph, see figure 5) and *first-in-stem* vowels (see figure 6).
5. These two mechanisms are bound to the cursive joining mechanism. Such a strongly coupled logic leads to complication in related mechanisms.
6. The first-in-stem vowel behavior is not specified on *final*. Second-stem (see figure 7) and enclitic (see figure 9) situations are complicated.¹
7. Since stem boundaries inside compound words are also syllable boundaries, coda consonants of the first stems have to be manually requested with FVSes when the second stems start with vowels.
8. The major source of long-distance effect is vowel-harmony-affected consonants. The long-distance effect might seem convenient in simple words but it quickly gets complicated in non-harmonious compound words and loanwords, then require manual overriding. Actually only one sequence “ig” actually needs vowel-harmony across syllables. Parallel character-based rules are complicated and still inadequate.
9. Enclitics are highly related to word stem mechanisms but have additional features.

有 3 个机制与词干直接相关，但词干在编码中没有标记：

- 词干起始元音：有 aleph 的 *init*（上形）或 *medi*（中形）、没有 aleph 的附加成分。
- 词干第一元音：第一或第二词干、附加成分。
- 元音和谐。

这些词干相关的机制被限制在并不直接相关的连写机制中，导致那些没有在连写机制中体现的情况被忽视。

- 元音有两大跨字位的成形机制：“词干起始”元音与“词干第一”元音。
- 这两个机制被限制在连写机制中。这样强耦合的逻辑导致相关机制的复杂情况。
- 词干第一元音行为没有在 *final*（下形）指定。第二词干和附加成分情况很复杂。¹
- 因为复合词内的词干边界也是音节边界，当第二词干以元音开头时第一个词干的音节尾辅音必须用 FVS 手动请求。
- 长程作用的主要来源是受元音和谐影响的辅音。这一长程作用作用在简单的词中可能看起来方便，但在不和谐的复合词和外来词中就迅速变得复杂了，然后就需要手动控制。其实只有 ig 这一个序列需要跨音节的元音和谐。并行的基于字符的规则很复杂而且依旧不完备。
- 附加成分与词干机制高度相关但有更多特性。

¹Some sample words are from Siqinbilige's discussion:
<https://lists.w3.org/Archives/Public/public-i18n-mongolian/2015JulSep/0355.html>

¹一些例词来自 Siqinbilige 的讨论：
<https://lists.w3.org/Archives/Public/public-i18n-mongolian/2015JulSep/0355.html>

Figure 图 5: Stem-beginning vowel alephs. 词干起始元音的 aleph。

𐤀 𐤁 𐤂 𐤃 𐤄 𐤅 𐤆

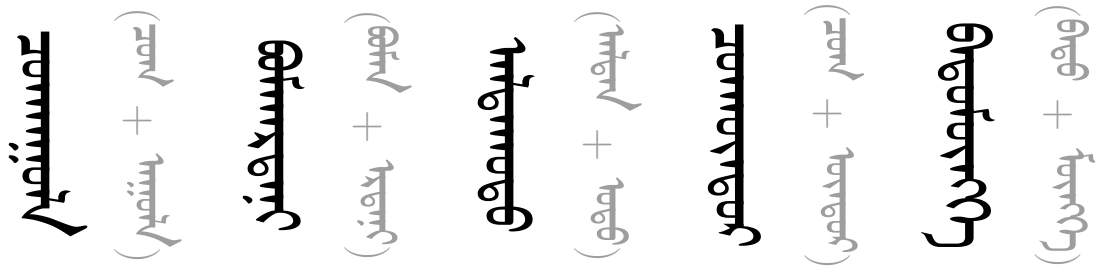
1) 𐤀 1820 A 𐤁 1821 E 𐤂 1822 I 𐤃 1823 O 𐤄 1824 U 𐤅 1825 OE 𐤆 1826 UE

Figure 图 6: First-in-stem vowels. 词干第一元音。

𐤀 𐤁 𐤂
 𐤃 𐤄 𐤅
 𐤆 𐤇 𐤈
 𐤉 𐤊 𐤋

- 1) 𐤀 1828 NA 𐤁 1823 O (FVS)? 𐤂 1828 NA 𐤃 1824 U (FVS)? 𐤄 1828 NA 𐤅 1825 OE FVS 𐤆 1828 NA 𐤇 1826 UE FVS
- 2) 𐤈 1830 SA 𐤉 1823 O (FVS)? 𐤊 1830 SA 𐤋 1824 U (FVS)? 𐤌 1830 SA 𐤍 1825 OE FVS 𐤎 1830 SA 𐤏 1826 UE (FVS)?
- 3) 𐤐 1834 CHA 𐤑 1823 O (FVS)? 𐤒 1834 CHA 𐤓 1824 U FVS 𐤔 1834 CHA 𐤕 1825 OE FVS 𐤖 1834 CHA 𐤗 1826 UE FVS

Figure 图 7: Compared to compound words. 与复合词比较。



- 1) ገ 1834 CHA ገ 1824 U ብ 182D GA FVS ገ 1820 A FVS ብ 182D GA ገ 1824 U ብ 182F LA ገ 1820 A
- 2) ገ 182A BA ገ 1824 U ብ 182E MA ገ 1820 A ብ 1828 NA FVS ገ 1821 E ገ 1837 RA ገ 1833 DA ገ 1821 E ብ 1828 NA ገ 1822 I
- 3) ገ 1820 A ብ 182F LA ገ 1832 TA ገ 1820 A ብ 1828 NA FVS ገ 1823 O FVS ገ 1833 DA ገ 1823 O
- 4) ገ 1834 CHA ገ 1824 U ብ 182D GA FVS ገ 1826 UE FVS ብ 1828 NA ገ 1833 DA ገ 1826 UE ገ 1837 RA
- 5) ገ 182A BA ገ 1820 A ገ 1832 TA ገ 1824 U ብ 182E MA ገ 1825 OE FVS ገ 1829 ANG ብ 182C QA ገ 1821 E

Figure 图 8: Vowel harmony affected consonants. 受元音和谐影响的辅音。



- 1) ገ 1834 CHA ገ 1821 E ገ 1837 RA ገ 1822 I ብ 182D GA
- 2) ገ 1835 JA ገ 1820 A ገ 1837 RA ብ 182F LA ገ 1822 I ብ 182D GA
- 3) ገ 1830 SA ገ 1822 I ብ 182D GA
- 4) ገ 1830 SA ገ 1822 I ብ 182D GA FVS
- 5) ገ 182A BA ገ 1822 I ብ 182F LA ገ 1822 I ብ 182D GA FVS ገ 182A BA ገ 1820 A ገ 1832 TA ገ 1824 U

3.4 Enclitics 附加成分

The complicated situation for enclitics is the result of analyzing all enclitics (while the categorization work is incomplete in the first place) as a dictionary of special cases triggered by NNBS instead of productive cases.

1. Vowels not stem-beginning.
2. Vowels not first-in-stem.
3. D takes its disambiguating form.
4. Historical forms of Y in yi... and iy... carried over when the Manchu Y form was introduced.

Issues:

Not based on graphematics/orthography but grammar categorization.

Not always have visible effect. (ügei)

Cases not grammatically considered enclitics are handled as pure special cases in dictionary.

Relying on the NNBS imposes a typography preference not widely accepted to shaping logic:

5. Narrower than normal word space: Manchu and Sibe have different convention. Definition of width is wrong possibly due to lost in translation.
6. Forbidding line break. Not a convention in the publishing industry.
7. Extending word boundary: Not widely accepted. Different languages, Manchu and Sibe have different convention.

3.5 Dictionary-based special cases 基于词典的特例

The Mongolian encoding has a set of special cases that should be handled by automatical rules or manual overriding.

3.6 Other manually requested variants 其他手动请求的变体

In addition to the aforementioned mechanisms that are expected to work automatically, a lot of alternative forms

附加成分的复杂情况是因为把所有附加成分分析为（尽管分类工作一开始就不完整）一个由 NNBS 触发的特例词典而非能产的情况。

- 元音不是词干起始。
- 元音不是词干第一。
- D 取其消歧义形态。
- 在引入满文 Y 时在 yi... 和 iy... 里沿用了历史上的 Y 形态。

问题:

没有基于书写法而是基于语法分类。

并不总有可见的视觉效用。(ügei)

语法上不认为是附加成分的当做词典纯特例处理。

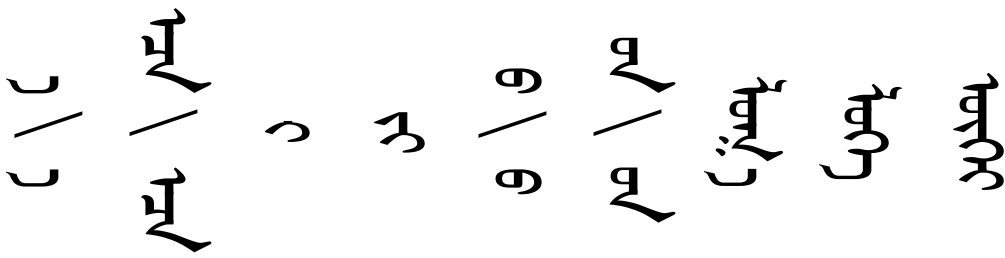
对 NNBS 的依赖向成形逻辑强加了并没有广泛接受的排版偏好:

- 比普通词间空格窄: 满语和锡伯语有不同的习惯。宽度的定义错误, 可能是翻译问题。
- 禁止断行。在出版业不是习惯。
- 扩展词界: 没有广泛接受。满语和锡伯语有不同的习惯

蒙古文编码有一组特例应当用自动规则或手动控制解决。

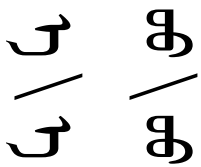
在上述要求自动工作的机制之外, 很多其他的形式需要手动用 FVS 请求, 比如

Figure 图 9: Enclitics? 附加成分?



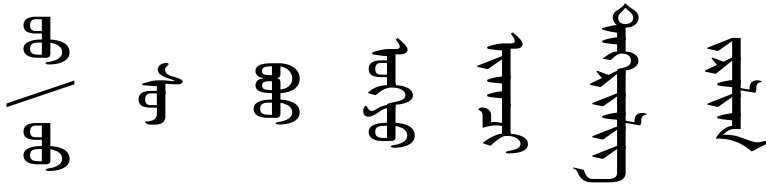
- 1) $\left\{ \begin{array}{l} \text{𠃉} \text{ 1820 A} \\ \text{𠃊} \text{ 1821 E} \end{array} \right\}$
- 2) $\left\{ \begin{array}{l} \text{𠃉} \text{ 1820 A} \\ \text{𠃊} \text{ 1821 E} \end{array} \right\} = \text{𠃋} \text{ 1834 CHA} \left\{ \begin{array}{l} \text{𠃉} \text{ 1820 A} \\ \text{𠃊} \text{ 1821 E} \end{array} \right\}$
- 3) $\left\{ \begin{array}{l} \text{𠃉} \text{ 1822 I} \end{array} \right\}$
- 4) $\left\{ \begin{array}{l} \text{𠃌} \text{ 1836 YA} \\ \text{𠃍} \text{ 1822 I} \end{array} \right\}$
- 5) $\left\{ \begin{array}{l} \text{𠃎} \text{ 1824 U} \\ \text{𠃏} \text{ 1826 UE} \end{array} \right\}$
- 6) $\left\{ \begin{array}{l} \text{𠃎} \text{ 1824 U} \\ \text{𠃏} \text{ 1826 UE} \end{array} \right\} \text{ 𠃐} \text{ 1828 NA}$
- 7) $\left\{ \begin{array}{l} \text{𠃑} \text{ 182F LA} \\ \text{𠃒} \text{ 1824 U} \\ \text{𠃓} \text{ 182D GA} \end{array} \right\} (\text{𠃔} \text{ 180E MVS})? \text{ 𠃕} \text{ 1820 A}$
- 8) $\left\{ \begin{array}{l} \text{𠃑} \text{ 182F LA} \\ \text{𠃒} \text{ 1826 UE} \\ \text{𠃓} \text{ 182D GA} \\ \text{𠃔} \text{ 1821 E} \end{array} \right\}$
- 9) $\left\{ \begin{array}{l} \text{𠃒} \text{ 1826 UE} \\ \text{𠃓} \text{ 182D GA} \\ \text{𠃔} \text{ 1821 E} \\ \text{𠃕} \text{ 1822 I} \end{array} \right\}$

Figure 图 10: Non-enclitics? 非附加成分?



- 1) $\left\{ \begin{array}{l} \text{𠃉} \text{ 1820 A FVS} \\ \text{𠃊} \text{ 1821 E} \end{array} \right\}$
- 2) $\left\{ \begin{array}{l} \text{𠃎} \text{ 1824 U} \\ \text{𠃏} \text{ 1826 UE} \end{array} \right\}$

Figure 图 11: Dictionary-based special cases. 基于词典的特例。

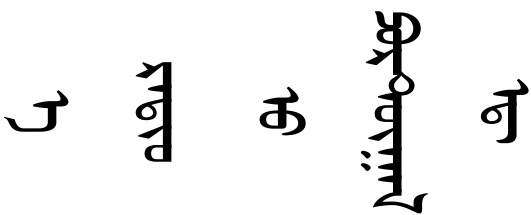


- 1) $\left\{ \begin{array}{l} \text{འ 1824 U (FVS)? འ 1824 U} \\ \text{འ 1826 UE (FVS)? འ 1826 UE} \end{array} \right\}$
- 2) འ 1830 SA འ 1826 UE (FVS)?
- 3) འ 182A BA འ 1826 UE (FVS)? འ 1826 UE
- 4) $\left\{ \begin{array}{l} \text{འ 1823 O} \\ \text{འ 1824 U} \end{array} \right\}$ འ 182D GA (FVS)? འ 1836 YA འ 1824 U
- 5) འ 1822 I འ 182D GA (FVS)? འ 1834 CHA འ 1822 I
- 6) འ 1832 TA འ 1829 ANG འ 1837 RA འ 1822 I འ 182F LA འ 1822 I འ 182D GA (FVS)?
- 7) འ 1835 JA འ 1837 RA འ 182F LA འ 1822 I འ 182D GA (FVS)?

have to be manually requested with FVSes, such as loanword O, Chinese U, loanword and Chinese UE, loanword TA, native or loanword DA.

借词 O、汉语 U、借词与汉语 UE、借词 TA、原始或借词 DA。







Figure 图 12: Other manually requested variants. 其他手动请求的变体




- 1) འ 1820 A FVS
- 2) འ 1837 RA འ 1820 A འ 1833 DA འ 1822 I འ 1823 O FVS
- 3) འ 1826 UE FVS
- 4) འ 182B PA འ 1823 O འ 1837 RA འ 1832 TA FVS འ 1826 UE FVS འ 182D GA འ 1820 A འ 182F LA
- 5) འ 1821 E འ 1833 DA FVS

4 Stylistic variants 风格变体







The Mongolian encoding contains a couple of stylistic variants that are not significant in graphematics or orthography of the modern Hudum.


1. The variant  is a historically preferred and temporarily alternative stylistic form of  BA.1.fina. Similar to the case of  vs  MA.1.fina.
2. And  was historically possibly distinguished from  SA.1.fina but now is an alternative stylistic form.

A large number of historically used variants is proposed in [L2/16-309 Proposed additions for Mongolian in 5th edition of UCS](#), in which all additions for  U+1820 A are historically stylistic.

Such attempts of encoding all manuscript visual details in plain text with format characters will further reduce the usability of the Mongolian encoding. Fonts and markup languages should be employed to deal with these.

蒙古文编码包含一些在现代胡都木文正写法中无意义的风格变体。

-  是  BA.1.fina 在历史上受偏好的变体，在现代也只是风格变体。类似于  与  MA.1.fina 的情况。
-  在历史上有可能区别于  SA.1.fina 但在现代只是风格变体。

大量历史上使用的变体在 [L2/16-309](#) 《Proposed additions for Mongolian in 5th edition of UCS》中提交。其中所有为  U+1820 A 新增的都是历史上的风格变体。

这样尝试把所有古籍中的视觉细节都编码在纯文本中会进一步降低蒙古文编码的可用性。字体和标记语言应当用于处理这些。