Filling the gap between orthographic patterns and shaping specs: a case study of Hudum *g*

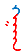
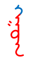
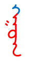
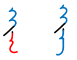
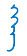
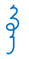
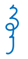
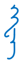
Author: SHEN Yilei

Date: 2019-04-03

0 Setting the stage

0.1 Cheat sheet

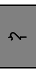
Table 1 A cheat sheet for Hudum *g*'s shaping

Stray	F							
Onset	M	M	M	F	F	F	F	F
Coda	M	M	M	**M/F**	F	F	F	F
Form								
Xlit.	*ggag*	*ggog*	*ggug*	*ggig*	*ggeg*	*ggög*	*ggüg*	*ggëg*

0.2 Notation

- ⟨ɢ̤⟩, ⟨ɢ⟩, and ⟨g⟩:

Table 2 Forms and graphemic transliterations of Hudum *g*

		Init.	Medi.	Fina.
Dotted masculine	⟨ɢ̤⟩			
Dotless masculine	⟨ɢ⟩			
Feminine	⟨g⟩			

- °, ', ", `: "FVS0" (an unencoded character that requests no-FVS forms), FVS1, FVS2, FVS3.
- $^{\mathrm{m}}$ and $^{\mathrm{f}}$: Appropriate FVSes to request dotless masculine and feminine forms of *g* respectively.
- Interpunct (·): The internal boundary of a compound.
- ʔ: U+1807 MONGOLIAN SIBE SYLLABLE BOUNDARY MARKER (SSBM).

0.3 Definition

A consonant is

- An onset, if immediately preceding a vowel (***C****V*, ***C****VC*); or otherwise
- A coda, if immediately following a vowel (*CV****C***); or otherwise
- A stray, which does not belong to any orthographic syllable (***C****CVC*, *CVC****C***).

A Mongolian vowel is

- A masculine vowel if it is *a*, *o*, or *u*;
- A feminine vowel if it is *e*, *ö*, *ü*, or *ë*;
- A neuter vowel if it is *i*.

A Mongolian word is

- A masculine word, if it contains masculine vowels (*a*, *o*, *u*) and NO feminine vowels (*e*, *ë*, *ö*, *ü*); or otherwise
- A feminine word, if it contains feminine (*e*, *ë*, *ö*, *ü*) vowels and NO masculine vowels (*a*, *o*, *u*); or otherwise
- A neuter word, if it contains no vowel other than *i*; or otherwise
- A bigender word (rare, mostly loans or compounds).

1 Orthographic patterns and shaping specs of *g* in non-compounds

1.1 Onset logic

Orthographic patterns:

- An onset *g* in *ga*, *go*, *gu* takes the masculine onset form ⟨ɢ̤⟩.
  - Examples: ⟨ɢ̤⟩*ar* | ᠭᠠᠷ, *xu*⟨ɢ̤⟩*ur* | ᠬᠤᠭᠤᠷ, *so*⟨ɢ̤⟩*o* | ᠰᠣᠭᠣ, *ba*⟨ɢ̤⟩*_a* | ᠪᠠᠭ᠎ᠠ, ….
  - Exceptions: (none)
- An onset *g* in *ge*, *gi*, *gö*, *gü*, *gë* takes the feminine form ⟨g⟩.
  - Examples: ⟨g⟩*er* | ᠭᠡᠷ, ⟨g⟩*ërman* | ᠭᠧᠷᠮᠠᠨ, *ǰe*⟨g⟩*ün* | ᠵᠡᠭᠦᠨ, *ǰaŋ*⟨g⟩*i* | ᠵᠠᠩᠭᠢ, *möŋ*⟨g⟩*ö* | ᠮᠥᠩᠭᠥ….
  - Exceptions: (none)

Shaping specs = orthographic patterns.

- Required overrides = exceptions.

These patterns/specs also apply to *x*, which appears in a Mongolian word always as an onset.

1.2 Stray logic

Orthographic patterns:

- A stray *g* takes the feminine form ⟨g⟩.
  - Examples: (14 of 26k)[1] ⟨g⟩*xib* | ᠭᠬᠢᠪ, ⟨g⟩*xir* | ᠭᠬᠢᠷ, ⟨g⟩*šaŋ* | ᠭᠱᠠᠩ, ⟨g⟩*ranat* | ᠭᠷᠠᠨᠠᠲ, ⟨g⟩*ram* | ᠭᠷᠠᠮ, ⟨g⟩*radü's* | ᠭᠷᠠᠳᠦᠰ, *aŋ*⟨g⟩*li* | ᠠᠩᠭᠯᠢ, *sam*⟨g⟩*rida* | ᠰᠠᠮᠭᠷᠢᠳᠠ…. (More loans like *bël*⟨g⟩*rad'* | ᠪᠧᠯᠭᠷᠠᠳ, ⟨g⟩*ladiyat'or* | ᠭᠯᠠᠳᠢᠶᠠᠲᠣᠷ, ⟨g⟩*reg* | ᠭᠷᠡᠭ)
  - Exceptions: (none)

Shaping specs = orthographic patterns.

- Required overrides = exceptions.

To my knowledge, this generalization has never been stated before. It eluded all practitioners because no one tried to seriously apply syllable structure analysis to Mongolian. However, any attempt to build Mongolian shaping logic coherent inevitably leads to syllable structure analysis.

1.3 Gendered coda logic (*CVg* except *Cig*)

Orthographic patterns:

- A coda *g* in *ag*, *og*, *ug* takes the masculine form ⟨ɢ⟩.
  - Examples: *ča*⟨ɢ⟩ | ᠴᠠᠭ, *bo*⟨ɢ⟩*da* | ᠪᠣᠭᠳᠠ….
  - Exceptions: (5 of 26k) *o*⟨g⟩*yu~u*⟨g⟩*yu* | ᠣᠭᠶᠤ, *pro*⟨g⟩*ram* | ᠫᠷᠣᠭᠷᠠᠮ (*pro*⟨g⟩*ramči* | ᠫᠷᠣᠭᠷᠠᠮᠴᠢ), *da*⟨g⟩*yig*[f] | ᠳᠠᠭᠶᠢᠭ, *kilo*⟨g⟩*ram* | ᠺᠢᠯᠣᠭᠷᠠᠮ. (More loans like *za*⟨g⟩*rëb* | ᠽᠠᠭᠷᠧᠪ)
- A coda *g* in *eg*, *ög*, *üg*, *ëg* takes the feminine form ⟨g⟩.
  - Examples: *xere*⟨g⟩ | ᠬᠡᠷᠡᠭ, *ö*⟨g⟩*xü* | ᠥᠭᠬᠦ….
  - Exceptions: (none)

Shaping specs = orthographic patterns.

- Required overrides = exceptions.

[1] Number of word types in the 26k-word dictionary *Mongolian–Chinese Dictionary* (Inner Mongolian University Press, 1999).

Alternatively, one may wish to capture these exceptions in fonts so as to reduce the use of FVSes. This may work when exception is clearly definable (e.g., the offglide *i* in only *naima* | and its derivatives is exceptionally written as a single shin), but fails here because of logical incoherence as new loan exceptions join in.

1.4 Neuter coda logic (*Cig*)

Orthographic patterns:

- A neuter coda *g* in a masculine word takes the masculine form ⟨ɢ⟩.
  - Masculine examples: *ǰarli*⟨ɢ⟩ | , …; *ni*⟨ɢ⟩*ta* | , ….
  - Feminine exceptions: (3 of 26k) *abisi*⟨g⟩ | , *taximli*⟨g⟩ | , *dag*ᶠ*yi*⟨g⟩ | ; (4 of 26k) *mi*⟨g⟩*man* | , *milli*⟨g⟩*ram* | ~ *mili*⟨g⟩*ram* | , *d'i*⟨g⟩*da* | , *di*⟨g⟩*baranǰa* | .
- A neuter coda *g* in a feminine word takes the feminine form ⟨g⟩.
  - Feminine examples: *čeri*⟨g⟩ | , …; *ǰi*⟨g⟩*de* | , ….
  - Masculine exceptions: (attested to in premodern dictionaries: *xereli*⟨ɢ⟩ | , *xersli*⟨ɢ⟩ | , *sinesi*⟨ɢ⟩ |)
- A neuter coda *g* in a neuter word takes either the masculine form ⟨ɢ⟩ or the feminine form ⟨g⟩.
  - Masculine examples: (11 of 26k) *ti*⟨ɢ⟩ | , *i*⟨ɢ⟩*či* | , *iǰili*⟨ɢ⟩ | , …;
  - Feminine examples: (51 of 26k) *ti*⟨g⟩ | , *ǰi*⟨g⟩*sil* | , *bili*⟨g⟩ | , ….
- A neuter coda *g* in a bigender word takes the masculine form ⟨ɢ⟩ if it is closer to a masculine vowel, or the feminine form ⟨g⟩ if closer to a feminine vowel.
  - Masculine examples: (5 of 26k) *mëxani*⟨ɢ⟩ | (*mëxani*⟨ɢ⟩*ǰixu* | , …);
  - Feminine examples: (1 of 26k) *bolšëwi*⟨g⟩ | .

Shaping specs = orth…

—*Uh… hold on; how can we formulate these patterns in the language of shaping specs?*

2 Stipulating the neuter coda g's (*Cig*) shaping specs

2.1 Gaps between orthographic patterns and shaping specs

There are, apparently, wide gaps between orthographic patterns and shaping specs. To name a few, …

Firstly, shaping results must be unique and determinate. In neuter words we see both masculine *Ci*⟨ɢ⟩ and feminine *Ci*⟨g⟩ as regular patterns, which means we need to choose one possibility as default. Not surprisingly, the feminine form is chosen as default for *Cig* in neuter words, as there are more BILI⟨g⟩ words than I⟨ɢ⟩ČI words.

Secondly, shaping context must cover all possible input sequences. Although no *Cig* is attested to appear between a masculine and a feminine vowel in non-compounds, shaping specification cannot underspecify these marginal situations. A typist-friendly specification should stipulate that progressive propagation takes precedence in these cases. Another example of marginal situation would be incomplete context, i.e., input sequences with ZWJs or *nirugu*'s. Syllable structure analysis needs to address how these characters would contribute to the context.

Thirdly, shaping specs must be implementable. It seems that the standard setters simply assumed global gender propagation without any articulated or even concrete technical solutions at that time. As a result, implementers have struggled to fill this wide gap between the specification and implementable shaping of *Cig*…

2.2 The pseudo-global approach

The pseudo-global approach to shaping *Cig* is the approach adopted by all fonts in this industry as far as I know, where enumerated local contextual rules are listed in the font. Popular as it is, partial enumeration of infinite rules is not a favorable approach:

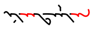
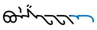
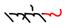
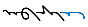
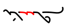
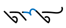
- It a pitfall in logic. It does work for the majority of words if plenty of rules (say, several score) are enumerated, but will certainly fail for propagations beyond a specific limit. (For instance, the font Orhon can handle *alllllllllig* (9 *l*'s) but fails for *allllllllllig* (10 *l*'s).)
- It is a burden to implementation. Building, verifying, and maintaining these rules would be particularly painful.

This popular but clumsy workaround resulted from an underestimation of the OpenType's shaping ability. We can achieve unbounded gender propagation by exploiting chaining contextual substitutions and dummy intermediate glyphs. (See Appendix A for details.)

2.3 New approaches

Now that the device of unbounded gender propagation is available, we have the following approaches to shaping *Cig*:

Table 3 Representations in various approaches to *Cig*

			V…Cig		*Cig…V*	
			Regular	Exceptional	Regular	Exceptional
			341/26k	3/26k	33/26k	4/26k
						
ni⟨G⟩tali⟨G⟩	*bolšëwi⟨g⟩*		*ǰarli⟨G⟩*	*abisi⟨g⟩*	*ni⟨G⟩ta*	*d'i⟨g⟩da*
*nigt**a**lig*	***b**o**l**š**ë**wig*	Bidirectional prop.	*ǰ**a**rlig*	***abisig***$^{\text{f}}$	*nigt**a***	*d'ig*$^{\text{f}}$*d**a***
nig$^{\text{m}}$*t**a**lig*	***b**o**l**š**ë**wig*	Progressive-only prop.	*ǰ**a**rlig*	***abisig***$^{\text{f}}$	*nig*$^{\text{m}}$*t**a***	*d'igd**a***
nig$^{\text{m}}$*talig*$^{\text{m}}$	*bolšëwig*	No prop.	*ǰarlig*$^{\text{m}}$	*abisig*	*nig*$^{\text{m}}$*ta*	*d'igda*

2.4 Comparison of various approaches

2.4.1 Typists' perspective

- 😊 Bidirectional propagation: The most sparing FVS usage.
- 😐 Progressive-only propagation: Additional FVSes in 33 words like *ni⟨G⟩ta* (out of 26k words).
- 😟😟😟 No propagation: Much more FVS occurrences.

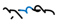
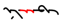
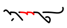
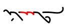
2.4.2 Implementers' perspective

- 😟 Bidirectional propagation & Progressive-only propagation: Some vendors (especially individual font developers) may fail to produce conformant fonts due to the circuitous nature of global gender propagation, and text typed under these fonts may contaminate the whole Mongolian script community, wasting the efforts we and the conformant font vendors make.
- 😊 No propagation: Good.

2.4.3 NLP's perspective

- 😟 Bidirectional propagation: Regressive propagation encourages vacuous FVSes. Some typist may type additional FVSes along with *g* prematurely without anticipating the subsequent gendered vowel that is going to propagate the gender context back.

Table 4 How vacuous FVSes come about

Typist A		Typist B		
nig			*nig*	
nig			*nig*$^{\text{m}}$	B fails to anticipate the subsequent *a* and adds a vacuous FVS.
nigt			*nig*$^{\text{m}}$*t*	
*nigt**a***			*nig*$^{\text{m}}$*t**a***	The vacuous FVS typed by B results in a duplicate as the eventual forms look the same.
ni⟨G⟩ta				

- 😊 Progressive-only propagation & No propagation: Good.

2.5 Local summary

Table 5 Comparison of various approaches to *Cig*

Progressive propagation	Regressive propagation		More FVSes to type	Implementation difficulty	Encouraging vacuous FVSes
✓	✓	Bidirectional prop.		😟	😟
✓		Progressive-only prop.	😐	😟	
		No prop.	😟😟😟		

3 Merging compounds into non-compound patterns

Generally, *g* in a compound word retains the original graphemic forms in the components that build the compound. As the components in a compound do not necessarily harmonize with each other, the definition of word gender introduced above do not apply to heterogender compounds anymore; original syllables may be "restructured" as well. However, we are interested in what exceptions may emerge if we mechanically merge these compounds into the pattern classification discussed above.

3.1 *Cig* and heterogender compounds

3.1.1 New orthographic patterns of *Cig* in heterogender compounds

- Feminine *Ci*⟨g⟩ in "masculine" compounds:
  - Noun: *batu·bili*⟨g⟩, *arigun·bili*⟨g⟩, *uyun·bili*⟨g⟩, *xas·bili*⟨g⟩, *xous·bili*⟨g⟩, *xurča·bili*⟨g⟩, *masi·bili*⟨g⟩, *saran·bili*⟨g⟩, *sain·bili*⟨g⟩, *sodo·bili*⟨g⟩, *todo·bili*⟨g⟩, …; *bili*⟨g⟩*·batu*, *bili*⟨g⟩*·xurča*, *bili*⟨g⟩*·sodon*, …;
  - Verb: *alab·xi*⟨g⟩, ….
- Masculine *Ci*⟨ɢ⟩ in "feminine" compounds:
  - Noun: (Theoretically possible through combining a feminine word and an I⟨ɢ⟩ČI word, but not readily attested to in common names);
  - Verb: (No construction attested to)
- Coexisting masculine *Ci*⟨ɢ⟩ and feminine *Ci*⟨g⟩ in "neuter" compounds:
  - Noun: (Theoretically possible through combining an I⟨ɢ⟩ČI word and a BILI⟨g⟩ word, but not readily attested to in common names)
  - Verb: (Theoretically possible for -·*xig* forms derived from masculine roots, yet not attested to)
- *Cig* between a masculine vowel and a feminine vowel in bigender compounds:
  - Noun: *xesi*⟨g⟩*·dalai*, *xesi*⟨g⟩*·manda*, *xesi*⟨g⟩*·buyan*, *xesi*⟨g⟩*·batu*, …; *bayali*⟨ɢ⟩*·önir*, …;
  - Verb: *šab·xi*⟨g⟩*sen*, ….

3.1.2 Handling heterogender compounds with Root Delimiter: a sensible move at all?

Root Delimiter, a new character that would block gender propagation to reduce the use of FVSes in heterogender compounds, has been proposed:

Table 6 Representations of heterogender compounds in various approaches

	Compound verbform		Compound proper name			
	M+N⟨g⟩	M+F⟨g⟩	N⟨g⟩+M	M+N⟨g⟩	F⟨g⟩+M⟨ɢ⟩	M⟨ɢ⟩+F⟨g⟩
Form	ᠠᠯᠠᠪᠬᠢᠭ *alab·xi*⟨g⟩	ᠱᠠᠪᠬᠢᠭᠰᠡᠨ *šab·xi*⟨g⟩*sen*	ᠪᠢᠯᠢᠭᠪᠠᠲᠤ *bili*⟨g⟩*·batu*	ᠪᠠᠲᠤᠪᠢᠯᠢᠭ *batu·bili*⟨g⟩	ᠬᠡᠰᠢᠭᠪᠠᠶᠠᠯᠢᠭ *xesi*⟨g⟩*·bayali*⟨ɢ⟩	ᠪᠠᠶᠠᠯᠢᠭᠬᠡᠰᠢᠭ *bayali*⟨ɢ⟩*·xesi*⟨g⟩
RtDel + Bidi. prop.	*alab|xig*	*šab|xigsen*	*bilig|batu*	*batu|bilig*	*xesig|bayalig*	*bayalig|xesig*
Bidi. prop.	*alabxig*^{f}	*šabxig*^{f}*sen*	*bilig*^{f}*batu*	*batubilig*^{f}	*xesigbayalig*	*bayaligxesig*
Prog.-only prop.	*alabxig*^{f}	*šabxig*^{f}*sen*	*biligbatu*	*batubilig*^{f}	*xesigbayalig*	*bayaligxesig*
No prop.	*alabxig*	*šabxigsen*	*biligbatu*	*batubilig*	*xesigbayalig*^{m}	*bayalig*^{m}*xesig*

Using Root Delimiter along with gender propagation does reduce FVS occurences in compound. However, it is a downright pitfall:

- It is an invisible character. It encourages representation inconsistency. It perplexes and daunts typists.
- It is a new character. It needs time to be accepted by the standards and to gain support.

In a word, Root Delimiter will not be a viable option.

3.1.3 Cutting the knot: splitting Mongolian *g*

Badral et al. (L2/18-294) have proposed splitting Mongolian *g* into two characters, one for both ⟨ɢ̣⟩ (dotted masculine) and ⟨ɢ⟩ (dotless masculine), and one for ⟨g⟩ (feminine). This is a favorable move that can fundamentally put out the fire.

3.2 *gV* and resyllabified compounds

3.2.1 Orthographic pattern of resyllabified compounds

A resyllabified compound is formed when a stem beginning with *alep* (zero consonant; written as a tooth) is joined to a preceding stem. In an alternative analysis, upon which the present Hudum and Todo encoding is based, *alep* is an integral part of the following vowel. Thus, the compound is "resyllabified" in the sense that the …*C·V*… or …*V·V*… character sequence at the juncture will be identified by the shaping process as tautosyllabic ("belonging to the same syllable") if medial *alep* is not represented by a dedicated character. Here are some examples of resyllabified compounds:

Table 7 Examples of resyllabified compounds

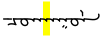
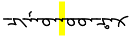
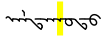
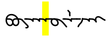
xöxe·agula	*čiŋ·ünen*	*nasun·urtu*	*čog·agula*
ᠬᠥᠬᠡᠠᠭᠤᠯᠠ	ᠴᠢᠩᠦᠨᠡᠨ	ᠨᠠᠰᠤᠨᠤᠷᠲᠤ	ᠴᠣᠭᠠᠭᠤᠯᠠ

3.2.2 Handling resyllabified compounds with SSBM

SSBM is originally devised for Manchu and Sibe to mark their medial *alep* (also written as a tooth)[2]. The same approach could have been adopted for Hudum and Todo, but the standard setters for some reasons decided to use FVSes throughout the two scripts. A comparison of the two approaches to Hudum medial *alep* are shown below:

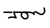
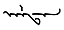
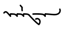
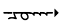
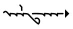
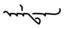
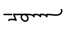
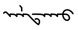
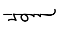
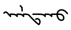
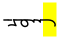
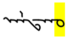
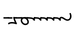
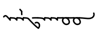
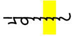
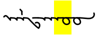
[2] It may appear in Chinese names of Manchus (*kiʔiŋ* | ᡴᡳ᠇ᡳᠩ < Chinese *Qíyīng* | 耆英; *peiʔioi* | ᡦᡝᡳ᠇ᡳᠣᡳ < Chinese *Péiyù* | 培豫). Other examples include Manchu *kuiʔi* | ᡴᡠᡳ᠇ᡳ 'spoon' (simplex native word), *guʔioi* | ᡤᡠ᠇ᡳᠣᡳ 'a kind of jade' (< native *gu* 'jade' + Chinese *yǔ* | 瑀 'a kind of jade'), and Sibe *juʔi* | ᠵᡠ᠇ᡳ '-ism, doctrine' (< Chinese *zhǔyì* ｜ 主义), *tuŋʔi* | ᡨᡠᠩ᠇ᡳ 'consent' (< Chinese *tóngyì* | 同意).

Table 8 Representation of resyllabified compounds

Form	 *čog·agula*	 *čimed·odcar*	 *altan·odo*	 *buyan·ölǰei*
The FVS approach	*čog°a'gula*	*čimed°o'dcar*	*altan°o'do*	*buyan°ö"lǰei*
The SSBM approach	*čogʔagula*	*čimedʔodcar*	*altanʔodo*	*buyanʔö'lǰei*

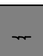
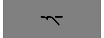
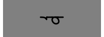
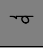
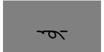
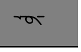
Typist-friendliness is an important consideration in evaluating an encoding scheme, and is vital especially for compounds, because Mongolian compounds are usually proper names that cannot count on a smart input method. Let's see what happens in typing with a bare keyboard:

Table 9 Comparison of the SSBM approach and the FVS approach in typing

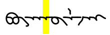
The SSBM approach		The FVS approach		Comment on the FVS approach
čog 	*altan* 	*čog*	*altan* 	
čogʔ 	*altanʔ* 	*čog°*	*altan°* 	Opaque FVS without immediate effect
čogʔa 	*altanʔo* 	*čog°a* 	*altan°o* 	
		čog°a' 	*altan°o'* 	Opaque FVS, whose immediate effect gets overridden
čogʔag 	*altanʔod* 	*čog°a'g* 	*altan°o'd* 	

The SSBM approach disuses those FVS usages intended only for resyllabified compounds, whereby opaque FVS effects during typing are eliminated. Compared with irrational FVS usages in resyllabified compounds that demands additional memorization, "always type an SSBM before the leading vowel of the second stem" is much easier to learn.

Table 10 Variants with medial *alep* (shaded; to be deprecated) are handled with SSBM instead

	No FVS	FVS1	FVS2			
…*a*…				…*a'*…	>	…*ʔa*…
…*i*…				…*i'*…	>	…*ʔi*…
…*o*…				…*o'*…	>	…*ʔo*…
…*u*…				…*u'*…	>	…*ʔu*…
…*ö*…				…*ö"*…	>	…*ʔö'*…
…*ü*…				…*ü"*…	>	…*ʔü'*…

Unlike Root Delimiter, SSBM will not be designed as a one-of-a-kind special character. Instead, in terms of shaping logic, SSBM will behave like nothing more than an ordinary consonant with no extra global or local complexity, leaving all other required adjustments to FVSes. In particular,

- It does not break gender propagation;
- It does not supply the following vowel *ö* or *ü* with the "first-within-stem" context. (*buyan·ölǰei* |  should be represented as *buyanʔö'lǰei*, with a FVS1 override to the leading vowel of the second stem, rather than simply allowing *buyanʔölǰei*.)

Were these complexities built into SSBM's logic, logical incoherence between SSBM and normal consonants outweighs the gain of saving a few FVSes.

4 Revisiting FVS assignment

Let’s turn to the issue of typist-friendliness regarding FVS assignment, which has been touched upon in the previous sections. As to Hudum *g*, we know FVS override is needed in various situations, at least for OGYU words (exceptional gendered coda) and IGČI words (masculine *Ci*⟨ɢ⟩ in neuter words) whichever scheme is adopted. Below is the typing process we expect for *o*⟨g⟩*yu* | ᠣᠭᠶᠤ:

Table 11 Expected typing process of *o*⟨g⟩*yu*

og	ᠣᠭ
og^{f}	ᠣᠭ᠋
og^{f}*yu*	ᠣᠭᠶᠤ

But if we adopt the current FVS assignment paradigm where *g*’s medial and final feminine forms are unaligned,

Table 12 Current FVS paradigm of medial and fina *g* (unaligned feminine forms)

	No FVS	FVS1	FVS2	FVS3
medi	‍ᠭ‍	‍ᠭ᠋‍	▨	‍ᠭ᠍‍
fina	‍ᠭ	‍ᠭ᠋	▨	▨

this is what a novice may experience in typing:

Table 13 How a novice may type *o*⟨g⟩*yu*

og	ᠣᠭ	Not in the correct gender form
og<FVS1>	ᠣᠭ᠋	Fixed with FVS
og<FVS1>*y*	ᠣᠭ᠋ᠶ	Fooled you! (immediate effect gets overridden)
og<FVS3>*y*	ᠣᠭ᠍ᠶ	Going back to modify the FVS

and they might end up getting familiar with the opaque FVS usages after being fooled a thousand times:

Table 14 How a veteran may type *o*⟨g⟩*yu*

og	ᠣᠭ	Not in the right gender form
og<FVS3>	ᠣᠭ᠍	Anticipating the FVS assignment opaque for the moment
og<FVS3>*y*	ᠣᠭ᠍ᠶ	Thank goodness!

Therefore, a typist-friendly shaping spec should not allow graphetically equivalent variants unaligned in terms of FVS assignment, at least for final & medial, and isolate & initial, which regularly alternate in typing.

5 Summary

- Apply syllable structure analysis in shaping.
- Use SSBM (U+1807) for Hudum medial zero consonant (*alep*) and shape it just like an ordinary consonant; deprecate all vowel variants that incorporate a medial zero consonant (*alep*).
- Split Hudum *g* (U+182D). If not possible, adopt one of the following approaches instead:
  - Perform bidirectional gender propagation;
  - Perform progressive-only gender propagation; or
  - Perform no propagation and use FVS override for every masculine *Ci*⟨ɢ⟩.
- Align FVS assignment.
- Don’t introduce invisible magic characters to Mongolian encoding anymore, as no ordinary user would ever learn.

A Implementing global gender propagation

Having no devices like infinite wildcards or quantifiers doesn't mean that OpenType cannot manage genuine global gender propagation. Instead, this can be achieved with chaining contextual substitutions and intermediate glyphs. Here is one possible way of implementing unbounded gender propagation:

- Step 1: Build intermediate glyphs, and apply cursive joining rules.
  - Define masculine vowels and neutral Mongolian glyphs as below:

```
@VOWEL_MASC    = [@a @o @u];
@NEUT_MINUS_G = [   @i @CONSONANT_MINUS_G @FVS \MVS \nirugu \ZWJ \ZWNJ];
@NEUT         = [@g @i @CONSONANT_MINUS_G @FVS \MVS \nirugu \ZWJ \ZWNJ];
```

  - Build an intermediate glyph named `\XXX.MASC` for every neutral Mongolian glyph `\XXX` that may occur after cursive joining rules are applied. The shapes of these intermediate glyphs are arbitrary because they will not appear in the eventual shaping results.

Shaping results:

mëxani⟨ɢ⟩ = `\m.init \ee.medi \x.medi \a.medi \n.medi \i.medi \g.fina`

bolšëwi⟨g⟩ = `\b.init \o.medi \l.medi \sh.medi \ee.medi \w.medi \i.medi \g.fina`

- Step 2: Propagate masculine context.
  - Insert the following pseudocode (involving chaining contextual substitution) after cursive joining rules:

```
lookup propagate_gender {
      sub [@VOWEL_MASC @NEUT.MASC] @NEUT' by @NEUT.MASC;
} propagate_gender;
```

Shaping results:

mëxani⟨ɢ⟩ = `\m.init \ee.medi \x.medi \a.medi` **`\n.medi.MASC \i.medi.MASC \g.fina.MASC`**

bolšëwi⟨g⟩ = `\b.init \o.medi` **`\l.medi.MASC \sh.medi.MASC`** `\ee.medi \w.medi \i.medi \g.fina`

- Step 3: Absorb masculine context.
  - Insert the following pseudocode after Step 2:

```
lookup absorb_gender {
      sub @NEUT_MINUS_G.MASC by @NEUT_MINUS_G;
} absorb_gender;
```

Shaping results:

mëxani⟨ɢ⟩ = `\m.init \ee.medi \x.medi \a.medi` **`\n.medi \i.medi`** `\g.fina.MASC`

bolšëwi⟨g⟩ = `\b.init \o.medi` **`\l.medi \sh.medi`** `\ee.medi \w.medi \i.medi \g.fina`

- Step 4: Perform local shaping.
  - Modify remaining rules so that:

    (1) `@g.MASC` are treated as if they were `@g` except in neuter coda (*Cig*) context;

    (2) In neuter coda (*Cig*) context, `@g.MASC` shape into dotless masculine ⟨ɢ⟩, and `@g` into feminine ⟨g⟩.

Shaping results:

mëxani⟨ɢ⟩ = `\m.init \ee.medi \x.medi \a.medi \n.medi \i.medi` **`\g.masc_coda.fina`**

bolšëwi⟨g⟩ = `\b.init \o.medi \l.medi \sh.medi \ee.medi \w.medi \i.medi` **`\g.femi.fina`**

Regressive gender propagation is similar in principle but involves reverse chaining contextual substitutions, whose implementation method is left to the reader.

(End of document)