EGD Supplement: Symbol Taxonomy
<mark>*work in progress*</mark>
Introduction

Preliminary Remarks

– as indicated in EGD §4.2 (to which we refer for further details), the use of a `<g>` tag is mandatory on all symbol characters
  – when the function of a symbol is confidently established, this tag shall be used as a wrapper which qualifies a transliteration character (that in turn represents a symbol):
    – if it is a punctuation mark in the narrow sense, then `<g>` wraps the character . (period, full stop), representing an abstract punctuation mark
    – if it is a space filler, then `<g>` wraps the character § representing an abstract space filler
  – when the function of a (non-numeral) symbol is not clearly one of the above, then the same tag shall be used as an empty element which represents and qualifies a symbol
  – in all of the above cases, `<g>` mandatorily takes the attribute @type for encoding the physical appearance of the symbol character
– the **values** available **for @type**, hereafter referred to as "(symbol) tokens", **are at present not constrained**, except by the following practical considerations:
  – for the sake of XML validity, a token must not contain a space
  – for the sake of simplicity, a token should preferably not contain characters other than letters of the English alphabet and numerals, and should only include uppercase letters to mark the initials of second or subsequent words within a token (camelCase)
– our **long-term goal is** to gather the symbol tokens used by encoders and use it as a starting point for **the creation of a constrained list for symbol classification**
  – to facilitate this, it is strongly recommended that you consult this Supplement whenever you encode a symbol, and
    – whenever possible, use a token recommended herein for symbols resembling your symbol in shape
    – whenever you find no symbol listed here that is close enough in shape to yours, create a new "species" to an existing genus or a new "subspecies" to an existing species; see also "Hierarchical Taxonomy" below
    – if possible, add a clipping of symbols you encounter to enrich this Supplement (see below about how to contribute)

A Hierarchical Taxonomy of Symbols

– eventually, we intend to classify symbols by using a limited number of values for @type and a likewise limited number of permitted values of @subtype for each @type
– with this aim in mind, we shall attempt already at this stage to describe symbols in this hierarchical manner, but without constraints on permitted values and with the hierarchy compressed into a single token, using camelCase to segment the token into "genus" and "species" (and optional "subspecies") as a precursor to future @type and @subtype
– thus, when encoding a symbol, proceed in the following steps:
   1. see if it can be classified as a variety of one of the numbered species described in the following subsections
   – if it can, proceed to step 2
   – if it cannot, because it is indistinct, encode it as per "Unclear symbols" below
   – if it cannot, because it is a radically different shape, assign it a genus of your choice, comparing it to specimens already recorded under "Miscellaneous" below
   2. use the selected genus name as the token for your symbol
   – classifying symbols by genus is sufficient: there is no obligation to add a species name
     – so you may stop at this step and encode e.g. `<g type="floret"/>`
   – species names should generally be used if a particular symbol is not a typical representative of the genus as a whole, but they may be used at any time when you think encoding this level of detail is advantageous enough to be worth the trouble
   – if you deem that a species name should be used for more accurate classification, proceed to step 3
   3. have a look at the specimens below for the genus you have selected and see if any of them are a close enough match for your particular symbol
   – if yes, choose the species name listed as preferred for that specimen, and add it to your token with a capital initial
     – you may stop at this step and encode e.g. `<g type="floretComplex"/>`
   – if not, you may do either of the following:
     – create a new species name for your symbol and use that, e.g. `<g type="floretSixpetalled"/>`
     – choose a roughly similar species name from what is already listed below, and add a "subspecies", again with a capital initial, after the species name, e.g. `<g type="floretComplexDottedcircle"/>`
– **always keep in mind that the diversity of actual symbols can never be fully represented in a classification scheme and you should not attempt to create tokens that provide an accurate and detailed description of symbols**
   – thus, while it is possible and permitted to create tokens more complex than the above (adding sub-subspecies, etc.), it is recommended that you never go beyond the subspecies level (i.e. three components in your token) and generally stop at the species level (two components in your token)
   – an accurate and detailed human-readable description can, and for unusual symbols should, be recorded in the Hand Description

section of your TEI header (EGD §11.2/The Hand Description)

Using the Tables

– each of the numbered subsections concerns one "approved" symbol genus, containing
  – a description of the significant features or criteria on the basis of which a symbol can be allocated to that genus
  – some notes on encoding shorthand and display
  – a table of species (and potential subspecies) belonging to that genus
– each primary row of the tables contains one "approved" symbol species or subspecies, described using the following columns
    1. archetype: an image that we consider to be a typical form of the subspecies or species concerned
    – this image may be a clipping from a facsimile, a drawing, or it may (especially at this early stage) be missing
    2. description: a concise description of what we consider to be the significant features or criteria on the basis of which a symbol can be allocated to that species or subspecies
    3. preferred token: the "approved" token to use for that species or subspecies
    4. specimens: where available, this cell contains one or more specimens encoded as that species or subspecies
    – each specimen is recorded in a new row within the primary row, and numbered for reference
    5. alternative tokens: this cell will contain tokens that have previously been used for symbols to which we have assigned a new preferred token
    – if you have already encoded a symbol with one of the tokens listed here as "alternative", there is no obligation to change your encoding to the preferred token, but newly encoded symbols should receive the preferred token
    – if we have a specimen of the actual symbol encoded with one of the alternative tokens, then the alternative token will have a number indicating the number of the corresponding specimen in the previous column
    6. remarks, clipping source
    – this column may contain further notes about the actual species or subspecies
    – this is also the place where the source (DHARMA repository and filename) of specimens may be recorded, numbered to indicate the corresponding specimen

Contributing to the Tables

– when you encode a symbol using a genus, species or subspecies already featured in the tables below, it is generally not necessary to add anything to the tables
  – but if the tables do not already have a specimen for the species or subspecies that you use, or if your symbol looks quite different from the specimen(s) already featured, then it would be appreciated if you could create a clipping of your symbol and add it to the table
– when you choose to employ a species or subspecies not already listed in the tables below, then

- mandatorily create a new row in the table pertaining to the genus of your token, and record your new species or subspecies there
- optionally add a clipping
- to **add a new species or subspecies** to the table,
  - create a new row at the end of the table for a new species, or below the applicable species for a new subspecies
  - write a description for your (sub)species, making an effort to summarise its most salient features concisely
  - add the token you have used for that symbol as the preferred token for the new (sub)species
  - if possible, add a clipping of your symbol as a specimen
- to **add a specimen** to a pre-existing row or one that you have just created,
  - create a clipping of your specimen
  - add it to the column "specimens", giving it the number 1 or, if there are already specimens in the table for the applicable (sub)species, the next higher available number
  - under "remarks", record the repository and filename of the inscription from which your symbol has been taken, and give this record the same number as that given to the specimen
    - optionally, especially for long inscriptions, also record the line number from which the clipping was taken
    - there is no need to record the name of the image file from which your clipping comes
- to **create a clipping**, use any image processing software you are familiar with
  - as far as possible, choose an instance of your symbol that is clear and typical of its kind
  - clip the symbol from your digital surrogate, and if possible, include one or two alphabetic characters before and/or after the symbol so its relative size and position can be seen in the clipping
  - if your surrogate has a very high resolution, preferably reduce the size of your clipping so that it is no more than 200 pixels in height
    - but feel free to use larger images if you cannot manage reducing it
    - do not enlarge your clipping if it is smaller in size to begin with
  - if possible, save your clipping as a PNG file (your graphic editor will probably be able to do that) before inserting it into this document
    - but if saving as PNG is a problem, do not worry and use whatever image format you can handle (though if you must save as JPG, do try to avoid high compression and choose high quality instead)
- our plan is that the tables will be **reviewed** every now and then, in the course of which we may ask you for further information/clarification on your contributions; moreover,
  - descriptions provided by you may be rephrased (to make them clearer or improve their English), and
  - new (sub)species created by you may be merged into other (sub)species, or the preferred token suggested by you may be overruled by a different token
    - in this case, the token originally suggested by you will be relegated to the status of an alternative token

Symbol Taxonomy

1. Tall, single vertical bars (*daṇḍa*)

– for symbols that consist of, or are palaeographically derived from, a single vertical bar that is about as tall as an average character body, use the genus token "danda"
– the transliteration shorthand | is recommended for `<g type="danda">.</g>`
– the transliteration shorthand / is recommended for `<g type="dandaOrnate">.</g>`
– a `<g>` element (empty or containing a . character) with a @type starting with "danda" shall be provisionally displayed as |

archetype	description	preferred token	specimens	alternative token(s)	remarks, clipping source
|	plain vertical bar	danda		dandaPlain	
	vertical bar with a headmark or small horizontal line on top	dandaSerif	1.		1. tfb-vengicalukya-epigraphy/CalE41-Diggubarru-Bhima2
	vertical bar with a hook on top	dandaHooked			
	vertical bar crossed by a predominantly horizontal line	dandaCross			
	vertical bar with short horizontal appendix at middle on the left	dandaStrokeLeft			
ꝫ	vertical bar with more complex ornamentation	dandaOrnate			

2. Tall, double vertical bars (double *daṇḍa*)
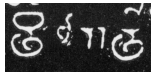
– for symbols that consist of, or are palaeographically derived from, a double vertical bar as typically used for punctuation in many Indic scripts, use the genus token "ddanda" (for "double *daṇḍa*")
– note that the reason we use a separate token for double *daṇḍa*s is that such marks usually comprise a single unit of punctuation, as demonstrated by the ornate variants of the double *daṇḍa*
– however, we see no need to dedicate tokens to triple or multiple vertical bars, which should be encoded as multiple instances of a
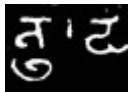
single (or, as the case may be, double) *daṇḍa*

−the transliteration shorthand || is recommended for `<g type="ddanda">.</g>`
−the transliteration shorthand // is recommended for `<g type="ddandaOrnate">.</g>`
−a `<g>` element (empty or containing a . character) with a @type starting with "ddanda" shall be provisionally displayed as ‖
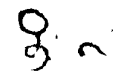
archetype	description	preferred token	specimens	alternative token(s)	remarks, clipping source
‖	plain double vertical bar	ddanda		ddandaPlain	
	double vertical bar with a headmark or small horizontal line on top	ddandaSerif	1.		1. tfb-vengicalukya-epigraphy/CalE41-Diggubarru-Bhima2
	double vertical bar with a hook on top of one or both bars	ddandaHooked	1. 2. 3.		1. Guntur, Java, very early 10th c. 2. Munduan, Java, very early 9th c. 3. DHARMA_INSBengalCharters00032
	double vertical bar crossed by by one or more further strokes that meet the vertical lines more or less at right angles	ddandaCross	1. 2. 3.		1. Pandaan inscription, Java, 11th c. 2. Luitan inscription, Java, 823 Śaka 3. Guntur inscription, Java, very early 10th c.

	double vertical bar with short horizontal appendix at middle of left bar	ddandaStrokeLeft	1.		1. Khalimpur plate, Dharmapāla, Bengal, 9th c.
	double vertical bar with short horizontal appendix at middle of right bar	ddandaStrokeRight	1.	ddandaStroke	1. Cambodia, 7th c. (K. 1455)
	double vertical bar enclosing three dots	ddandaDotTriple	1.	ddandaTripleDot	1. Cambodia, 7th c. (K. 1330)
౨	double vertical bar with more complex ornamentation	ddandaOrnate	1. 2. 3. lanaḥ // nṛpa 4. 5.		1. 2. 3. INSPallava00314 4. MpuMano (Java, 13th c.) 5. Marinchi (Java, 14th c.)

3. Short vertical bars (half *daṇḍa*)

- for symbols that consist of, or are palaeographically derived from, a short, predominantly vertical bar that may be straight or curved, use the genus token "comma"
- the transliteration shorthand , is recommended for `<g type="comma">.</g>`

−a `<g>` element (empty or containing a . character) with a @type starting with "comma" shall be provisionally displayed as ,

archetype	description	preferred token	specimens	alternative token(s)	remarks, clipping source
☐	a short, curved vertical stroke at median height	comma	1.		1. Kurangan, Java, 807 Śaka
	a very short vertical line at median height	commaSmall	1.	dotCommaMid	1. tfb-vengicalukya-epigraphy/CalE29-Ederu-Vijayaditya1-2
	a short vertical bar raised to head height	commaHigh	1.		1. tfb-vengicalukya-epigraphy/CalE36-Cevuru-Amma1
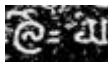
4. Short horizontal strokes, single and double

−for symbols that consist of, or seem to be essentially derived from, a single horizontal bar as used for punctuation in some regions and periods, use the genus token "dash"

−the transliteration shorthand ~ is recommended for `<g type="dash">~</g>`

−a `<g>` element (empty or containing a . character) with a @type starting with "dash" shall be provisionally displayed as ~

archetype	description	preferred token	specimens	alternative token(s)	remarks, clipping source
-	short and quite straight horizontal line	dash	1.		1. tfb-badamicalukya-epigraphy/CalE05-Aihole-Pulakesin2
	horizontal line with the middle bulging downward	dashConcave			
	horizontal line with the middle bulging upward	dashConvex	1.		1. tfb-badamicalukya-epigraphy/CalE01-Makarappi-Pulakesin2
	horizontal line with a hook at one end or the other	dashHooked	1.		1. tfb-vengicalukya-epigraphy/CalE12-Timmapuram-Visnuvardhana1

archetype	description	preferred token	specimens	alternative token(s)	remarks, clipping source
~	prominently sinuous horizontal line	dashWavy			
—	conspicuously long and quite straight horizontal line	dashLong			
=	two short and quite straight horizontal lines	dashDouble	1. 2.		1. tfb-badamicalukya-epigraphy/CalE05-Aihole-Pulakesin2 2. tfb-badamicalukya-epigraphy/CalE06-Chiplun-Pulakesin2
÷	a dash with a pair of dots above and below	dashDoubledot	1.		1. tfb-daksinakosala-epigraphy/Dk0005
	three predominantly horizontal lines one below the other, including cases where the middle line may be a slightly more complex figure	dashTriple	1.		1. C. 66

5. Dots

−for symbols that consist of one or more dots or very small circles and no other marks, use the genus token "dot"
−a `<g>` element (empty or containing a . character) with a @type starting with "dot" shall be provisionally displayed as °

archetype	description	preferred token	specimens	alternative token(s)	remarks, clipping source
ABC·	median dot	dotMid			
	two dots, one above the other, like a visarga	dotDouble			
	three dots arranged in a triangle	dotTriangle	1.	dotTriple	1. tfb-vengicalukya-epigraphy/CalE11-Cipurupalle-Visnuvardhana1

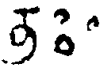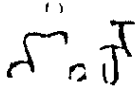@if we separate rings from dots, ◦ may be used as display for rings and • for dots
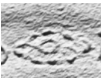
6. Circles

−for symbols that are circular (including ellipses) or have a circle as their most prominent element, use the genus token "circle"
−the transliteration shorthand @ is recommended for `<g type="circle">.</g>`
−a `<g>` element (empty or containing a . character) with a @type starting with "circle" shall be provisionally displayed as ◯

archetype	description	preferred token	specimens	alternative token(s)	remarks, clipping source
ABC○	small circle (noticeably smaller than a typical character body), placed at median height; Cambodian "Chicken egg"	circleMed	1.	circle circleSmall	1. Alasantan, Java, 861 Śaka
ABC◯	large circle (about the size of a typical character body or larger)	circleLarge			
ABC°	small circle above median line	circleHigh	1. (the two circles on the left are a visarga; the symbol is the single circle on the right)		1. tfb-badamicalukya-epigraphy/CalE 01-Makarappi-Pulakesin2
ABC。	small circle below median line	circleLow	1.		1. tfb-badamicalukya-epigraphy/CalE 01-Makarappi-Pulakesin2
	circle with cross inside	circleCross	1.		
	two circles, one above the other, like a visarga	circleDouble	1.		
⊛	circle with a floret inside	circleFloret		circleFinial	

ꙫ	circle topped by a double (or multiple) curved stroke	circleHorned	1. ꙫ 2. ꙫ 3. ꙫ	egg2apo	
◉	circle containing one or more concentric circles, Cambodian "Chicken eye"	circleConcentric	1. ◉ 2. ◎	eye circleTarget	
⊙	circle with a concentric dot, Cambodian "Chicken eye"	circleTarget	1.	eye	1. Alasantan, Java 861 Śaka
	oval containing one or more concentric circles or ovals	circleConcentricOval	1. 2.		
	three small circles arranged in a triangle	circleTriangle			
	a circular shape depicting a wheel with spokes	circleWheel	1.		1. DHARMA_INSBengalCharters00063.xml
	circle with one or more curved lines	circleCurve	1. 2.	tennisBall	1. Wurudu Kidul, Java, 929 CE 2. Alasantan, Java 861 Śaka

I am provisionally pasting here a clipping from DHARMA_INSCIC00171 of a symbol that daba proposes to classify as "circleConcentric" — for me, however, the ovoid aspect is important and hence I have tentatively labelled this "eye" in the xml edition. Once decision on classification is taken, I will modify xml file as needed.

7. Spirals

−for symbols based on a spiral line, use the genus token "spiral"
−a `<g>` element (empty or containing a . character) with a @type starting with "spiral" shall be provisionally displayed as @

archetype	description	preferred token	specimens	alternative token(s)	remarks, clipping source
◎ ◎	clockwise (right-turning) spiral (like a figure 6 or 9)	spiralR	1.		1. Vappaghosavata grant, Bengal, 7th c.
◉	counter-clockwise (left-turning) spiral (like a mirrored figure 6 or 9)	spiralL	1. 2.		2. Munduan, Java, very early 9th c.

8. Crosses

−for symbols that consist of, or are predominantly derived from, two straight lines crossing at a right angle, use the genus token "cross"
−a `<g>` element (empty or containing a . character) with a @type starting with "cross" shall be provisionally displayed as □
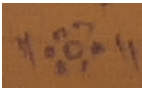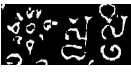
archetype	description	preferred token	specimens	alternative token(s)	remarks, clipping source
	cross in the shape of a + sign	crossPlus			

	cross in the shape of an ×	crossX	1.		

9. Florets

−for symbols resembling stylised flowers, generally consisting of several petals arranged radially around a central circle, use the genus token "floret"

−a `<g>` element (empty or containing a . character) with a @type starting with "floret" shall be provisionally displayed as ❧

archetype	description	preferred token	specimens	alternative token(s)	remarks, clipping source
	flower-shaped designs not matching any particular species of floret described below	floret	1. 2.		1. Odisha State Museum, ms. Dh/95 2. Odisha State Museum, ms. Dh/95
✤ ✤	floret with four petals, potentially with spikes or smaller petals interspersed between the four cardinal petals	floretQuatrefoil	1. 2. 3. 4.	simpleFinial	1. tfb-vengicalukya-epigraphy/CalE43-Akulamannandu-Bhima2 2. Mpu Mano's inscription, Java, 14th c. 3. tfb-vengicalukya-epigraphy/CalE44-Elavarru-Amma2 4. Sangguran, Java, 850 Śaka
⚙	floral design with more detail than a centre and petals (including flowers with eight identical petals)	floretComplex	1. 2.	complexFinial (3: spider) (4: simpleFinial)	

		3. ⊞ (image) 4. 𑀰		

10. *Gomūtra*

−for symbols that consist of, or are stylised renderings of, a wavy line of gradually increasing or decreasing amplitude (with or without a squiggle at the widest end), use the genus token "gomutra"

−display suggestion: U+00B6 ¶ Pilcrow Sign; possibly also U+204B ⁋ Reversed Pilcrow Sign so that one of these stands for initial, and the other for final gomūtras

archetype	description	preferred token	specimens	alternative token(s)	remarks, clipping source
ℯ〜	final *gomūtra*	gomutraFinal	1. @ ● ⑴⼃ 2. ⊖ 3. ⑰ 4. ◖ 5. ‖∿	finalGomutra	
〜⑨	initial *gomūtra*	gomutraInitial	1. 〜⑰ 2. ⩗⑨	initialGomutra	

ALTERNATIVE GOMŪTRA CLASSIFICATION, with a bit more detail

archetype	description	preferred token	specimens	alternative token(s)	remarks, clipping source
ℯ〜	generic final *gomūtra*	gomutraFinal		finalGomutra	
	final *gomūtra* rendered as a tapering series of vertical bars	gomutraFinalBars	1. ⼃⼇		

				2.		
	final *gomūtra* involving a more complex design than the archetypal rendering	gomutraFinalCompl ex	1. 2.		
	generic initial *gomūtra*	gomutraInitial	1. 2.	initialGomutra	
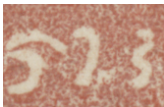
− in this alternative, the variants ⊖ and ⊕ should be reclassified into the genus "circle", and the variant .↶ should be reclassified into whatever spiral genus we create for these

11. Oṁ/Siddham

− siddham, for any variant of the signs discussed by
− Roth, Gustav. 1986. "Mangala-Symbols in Buddhist Manuscripts and Inscriptions." In Deyadharma: Studies in Memory of Dr. D.C. Sircar, edited by Gouriswar Bhattacharya, 239–49. Sri Garib Dass Oriental Series 33. Delhi: Sri Satguru Publications.
− Sander, Lore. 1986. "Om or Siddham — Remarks on Openings of Buddhist Manuscripts and Inscriptions from Gilgit and Central Asia." In Deyadharma: Studies in Memory of Dr. D.C. Sircar, edited by Gouriswar Bhattacharya, 239–49. Sri Garib Dass Oriental Series 33. Delhi: Sri Satguru Publications.

Miscellaneous/unclassified

− symbols that don't seem to fit any of the numbered species above may be added here
− a `<g>` element (empty or containing a . character) with a @type that does not start with one of the genus names listed above shall be provisionally displayed as NEED SUGGESTION

archetype	description	preferred token	specimens	alternative token(s)	remarks, clipping source
		squiggleVertical		braceCurlyClose	
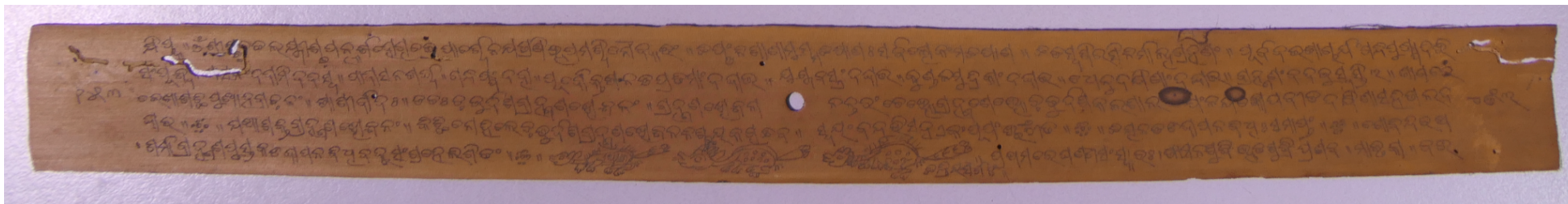
`<lb n="153v5"/>`śarmāvrāhmaṇapustaka Udyāpanavidhi va`<unclear>`hu`</unclear>`saṁgrahe likhitaṁ|| `<g type="floret"/>` `<g type="beast"/>` `<g type="beast"/>` `<g type="beast"/>`|| `<add place="below">`Indriyasparśaḥ||`</add>` `<foreign xml:lang="ori-Latn">`prathamare ghaṇaṭī saṁskāraḥ||`</foreign>` Āsanaśuddhi bhūtaśuddhi praṇava|| mātr̥kā|| vaR̥

Unclear and unknown symbols

– when a symbol is definitely present, but it is so damaged (or unclear in your surrogate for a reason other than damage) that it cannot be assigned to any of the primary categories, use "indistinct" as a genus token, e.g. `<g type="indistinct"/>`
– when a symbol can be assigned to one of the primary categories but due to damage (or any kind of unclarity in your surrogate) it cannot be identified more precisely, do either of the following:
  – simply classify it using the applicable genus token on its own, e.g. `<g type="spiral"/>`
  – add "indistinct" as a species to the genus token, e.g. `<g type="circleIndistinct"/>`
    – this latter method is recommended if your impression is that the symbol is not a typical representative of the genus
– when you are working from a previous edition without any visual aid, and the previous editor(s) only report the presence of a symbol at a certain point but do not describe it (or describe it insufficiently), then, analogously to the above case, use whichever of the following options best suits your case:
  – use `<g type="unknown"/>` to encode a symbol that you cannot assign to a genus
  – use a genus without a species (e.g. `<g type="spiral"/>`) if you can assign a symbol to a genus but have no further information
  – use a genus with the species "unknown" (e.g. `<g type="circleUnknown"/>`) if your impression is that the symbol is not a typical representative of the genus

Multiple symbols

– when several symbols (of a different kind, the same kind, or any combination) appear together, create a separate `<g/>` element for each symbol
  – the only exception to this rule of thumb is the double *daṇḍa* which, as covered under 2 above, is often a special sign (i.e. different

from two separate *daṇḍa*s) and is therefore normally encoded as a single `<g/>` element

- when multiple signs are used together, there is usually no need to add a . within any of the `<g/>` elements involved in the encoding, since such combinations are not normally used for punctuation in the narrow sense (as defined in EGD §4.2.4)
  - but if it is your judgement that such a combination serves for punctuation, then feel free to include a . character within any or all of the `<g/>` elements involved
- for example, the symbols in the image on the right may be encoded as follows:


- ```
  <g          type="dandaSerif"/><g          type="dandaSerif"/><g
  type="dandaSerif"/><g                       type="circleLarge"/><g
  type="dandaSerif"/><g          type="dandaSerif"/><g          type="dandaSerif"/><g          type="circleSmall"/><g
  type="gomutraFinal"/>
  ```

Discussion: Spirals

Arlo: this is a discussion section. I don't mean it to be part of the final Supplement, so no need to correct my typos in it :) just let's make a decision.

I've drawn specimens of what I see as the basic varieties of spirals, see the image.

A, "spiral," a normal spiral (with more or fewer complete turns)

B, "curl," a spiral with a tail (with more or fewer complete turns in the spiral part, and the tail may be longer or wavy or may curl at the end)

C, "hook," a spiral reduced to a curl (which may have a tail bending as in B)

Each of the above may be right-turning or left-turning, and each may face in any direction; I've drawn up four orientations, with the outer end of the line facing the bottom, left, top or right.

Thus, 24 basic varieties of spirals.

I'm sure it would be overkill to create separate tokens for all of these, and to make encoders pore over a table finding the right token.

So how to reduce this?
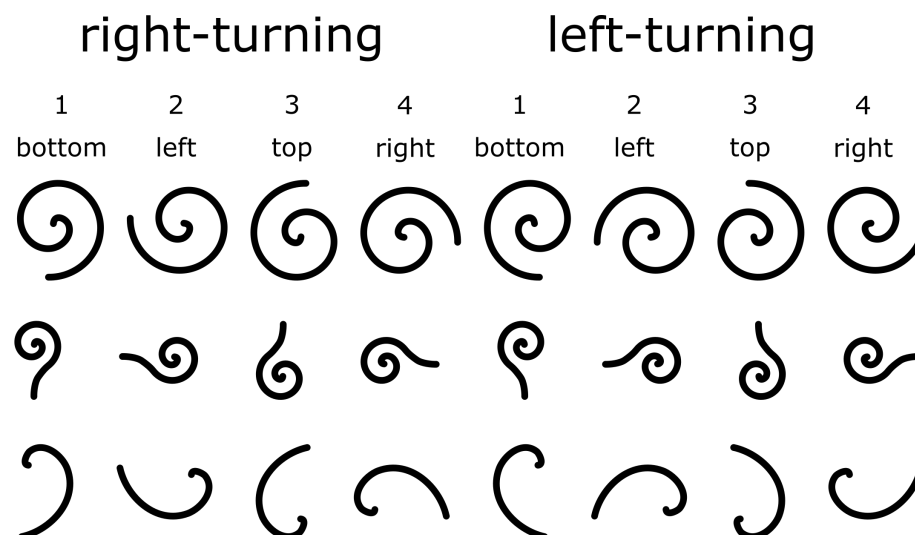
I'm also sure it would be good to have a finer distinction than just spiralL and spiralR.

The distinction between type A and types B/C is in my opinion important.

Is it important to distinguish B from C or shall we merge those into a single category?

Both Sander and Roth show very few L-spirals in their tables: as best I can see, CL1 and CL3 are the only forms attested there (possibly BL1 and BL3 too, depending on how sharp a line we want to draw between B and C). But I'm sure I've seen L-spirals of the A type in my experience.

I don't think the orientation could possibly matter much in the A type, but who knows.

Given the above considerations, perhaps a typology along these lines may work:

- spiralR for all of class AR
- spiralL for all of class AL
- curlRV for the vertically-oriented ones in classes BR and CR collectively (BR1, BR3, CR1, CR3)
- curlLV for the vertically-oriented ones in classes BL and CL collectively (BL1, BL3, CL1, CL3)
- curlRH for the horizontally-oriented ones in classes BR and CR collectively (BR2, BR4, CR2, CR4)
- curlLH for the horizontally-oriented ones in classes BL and CL collectively (BL2, BL4, CL2, CL4)

Even this may be too fine a distinction, but on the other hand, it may be interesting to distinguish 4 orientations in curls, and to distinguish curls from tailed spirals.

There is also the slight problem that some of the horizontally-oriented curls actually match some examples of gomūtras provided by Chloé.

Finally: are there any further spiral forms that you need to distinguish in texts I'm not familiar with? I've just learned from Google that there is such a thing as a "Cham punctuation spiral", which would by the above be encoded as <g type="spiralR">.</g> (assuming it is right-turning) and distinguished from spirals as maṅgala symbols by the presence of . in that tag.

Continuing: spirals and siddham signs

Premises. I accept your opinion that gomūtras should not be merged into "miscellaneous" but kept as a fully fledged separate genus. I maintain my own opinion that siddham signs should not be merged into gomūtras. I'm still not sure how best to proceed, but since we agree that our terminology should aim to describe shape (and not function or native name), I think that 1) "gomutra" is an OK token for describing that particular kind of squiggly line with diminishing oscillations; 2) spiral-based signs that we tend to call siddham should be described using some sort of spiral terminology. Do we agree so far? If yes, then I think the following problems remain: A) sorting out the exact details, including partial overlaps such as gomutras involving a spiral component versus spirals with long tails; B) dealing with signs that we might call siddham, but which are not included in my spiral permutations above, such as the one in your Phuoc Thien stele. To my mind, B) is not really an issue: there is no need to put it in the same genus as spiral-shaped "siddham" symbols, so if it, or something like it, needs to be encoded, we can simply give it a unique token, anything at all, and let it go at that.

As for A), the messy details, here are some thoughts.

If form is to be given priority over supposed function or "glyphological" identification, then some of the gomūtras from Chloé's original table may need to be reclassified, though I have now rephrased the definition to include stylised renderings of the presumed basic form.

It appears from Chloé's illustrations that these signs come in quite a large variety, and we probably don't want to encode them in more detail than "initial" and "final" species, as originally proposed by her. But I've proposed an alternative classification, with some subspecies, under 10 above.

I still don't know what to do about the variety of spirals. I now think the terms "spiral", "conch" and "curl" could apply very nicely to my rows A, B and C (respectively) of the spiral table above, so we could use these three genus names for them (and not just "spiral" for A and "curl" for B and C together, as I proposed above). But distinguishing 8 species in each of those three genera may be too much of complexity - while not distinguishing those species may be too much simplification. In your comment above you say you like the division into 24 species, so if you are sure about that, I'm happy to accept your opinion, with the twist that in my opinion those 24 species should be in 3 genera, not in one. I also agree that adding subspecies may be a good idea (though it will complicate matters even further), so long as we keep gomūtras separate from spirals.

Discussion: Dots and Circles

See the image on the right (unless it moves).

My problem is that the symbols in row A (dots) are identical, or at least very similar, in their variation and function to those in row B (small circles), and are in fact interchangeable at least in some cases. They may also be impossible to tell apart in some cases, unless you have a very good facsimile of a pristine inscription. Those in row C (big circles) are, however, different in function, and have a very different assortment of variants.

Therefore I think it would be a very bad idea to call those in A "dot" and all of those in B and C "circle".

I'm not sure, however, which is better: use three separate genus names (A, "dot"; B, "ring"; C, "circle") for these, or make A and B a single genus. You mention above that you are in favor of allowing distinction between dot and circle. But making A and B into a separate genus would imply that, e.g. the first item in A, which we might then call "dotHigh", is a particular species of dot, essentially different from the first item in B ("ringHigh"), which is a particular species of ring. Merging A and B into one genus (to be called e.g. "bullet") would get rid of this implication. We could still distinguish specimens of A from specimens of B when this is desired and clear in our surrogate, by
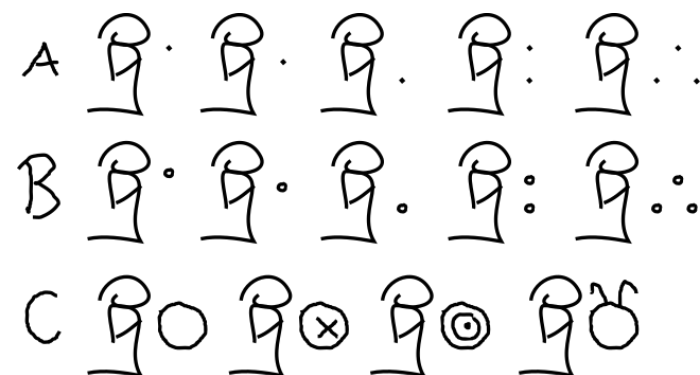
introducing a subspecies. Thus, if A and B together were to become the genus "bullet", which could have species such as bulletHigh, bulletMid, bulletLow, bulletDouble, bulletTriangle (in this order in the image), then these species could all have subspecies, e.g. dotHighPoint and dotHighRing to distinguish A1 from B1.

However, on the other side, now that I'm proposing "comma" for what we used to call half daṇḍas, we come to the point that some of those commas are also going to overlap with, and be interchangeable with, dots as in row A here. And even dashes may come into the picture, since dots could manifest as points, little rings, little horizontal strokes, or little vertical strokes. So, all in all, perhaps keep "dot" and "ring", as well as "dash" and "comma" separate, and forget about my worry that some of these can be the same character manifest as slightly different glyphs. I definitely do not, however, want to put rings (B) in the same genus as circles (C).

And an hour or two later (see how I vacillate?): now that the term "bullet" has occurred to me, I'm thinking it may be a good idea after all to use "bullet" for A and B together (e.g. bulletHigh, bulletMid, bulletDouble, etc., adding "ring" or "dot" as a subspecies if someone really wants to distinguish one from the other), and keeping comma and dash completely separate from these.

Thoughts?

Discussion: space filler, margin equaliser, word break sign

−Ryosuke in email (20200908):
  −What is indicated by the sign [¦] is 'a mark consisting of two short vertical lines one above the other, indicating a word divided between lines' (Salomon 1998: 67) as I cited in my article on the Rajibpur plates (p.40). I herewith give you one example, the end of line 7, where vigrahapala(8)deva is bridged by this sign.
  −[DB note: see Ryosuke's correction below; the clip actually continues vigrahapāla(8)s, so the sign is inside a non-compound word]

−Arlo in email (20200908)
  −The way Salomon defines its function ('indicating a word divided between lines') is a bit different from ours, but the sign in question is the Bengal manifestation of what we call the 'space filler' sign for which we propose the transliteration symbol § (TG 4.2.2).

- You can see two typical examples of the Javanese shape of this sign at the ends of lines 2 and 4 in the attached photo — the first of which **does not fall inside a word**, as least not in my understanding of word-boundaries in Old Javanese (where the matter may be a bit less evident than in Sanskrit). In my view, the Bengal and the Java shapes are clearly related palaeographically, and so are the functions of these signs — what we know of historical relations between Bengal and Java may imply that the sign's very common use in Java is due to influence from Bengal, even though the local scripts belong to different branches of Brahmi.
- Salomon's definition is too narrow even for India. From an 11th-c. Bengal manuscript, I can quote the following line
- <lb n="21r4">tena sahaikībhūyaḥ sarvatathāgatādhiṣṭhānena§ ○ mahān sarvākāśasamavasaraṇapramāṇapañcasūcikavajravigraho bhūtvā | punaḥ pūrvvahṛdgatavajrapramāṇam e
- which you see in the other attached photo. The sign stands not at line end but before the string-hole and after a full, decline, word.
- Dan: I suggest we rewrite the TG to include the idea of word division as often (though not always) relevant, and add the Salomon quotation that Ryosuke gives. I also suggest we might (in due course) want to reduce the insistence on variety of shape, because I do not expect there to be a great variety and I don't think it will lend itself to interesting study. We can also include Ryosuke's image (that I will forward to you) by way of illustration; if so, I can also try to furnish a nicer Javanese example.
- Dan in email (20200908)
- I suggest that in the line defining "space filler" in the TG and EG (" symbols whose function is clearly and unambiguously to fill up space in a line to the binding-hole or margin "), we add something like this: "including, but not limited to, symbols used before line ends or binding holes to indicate a word divided by the break" and leave it at that. I see no need for the Salomon citation.[1] We do already have images in the EG, so there is no need for you to find a nicer Javanese one, but I could add Ryosuke's vigrahapāla as yet another illustration. As for encoding the variety of shapes, don't forget Manu's example of gomutra symbols used as space fillers. We should not be hasty in reducing the variety of encoding here, especially since our TG already allows you to use a § without a <g> tag as shorthand for a filler of the Javanese form. I also have some possible space fillers in my Cālukya texts, though I'm not sure about any of them. Several were read as superfluous visargas by previous editors and may be the same sign as that in the Bengal inscriptions; one is a small circle at baseline level.
- That said, I'm not sure our current recommendations are the best there can be, and I'm open to improving them. I have no knowledge of the Javanese and Bengali texts and can't tell whether these symbols are ever iterated, or are always used singly. My impression is that in fact we have two different functions here. One is that of actual space fillers, which can be a variety of shapes and are often iterated to fill up a line that ends well before the margin as a sort of paragraph marker. That is to say, the penman started a new line for a new semantic unit, but for aesthetic reasons filled up the incomplete line with these symbols. And the other is that of the Javanese and Bengali sign, which I believe is never or at least rarely iterated, and its function seems rather to be the justification of the margin in situations that do not involve the end of a semantic paragraph, i.e. where the break is within a semantic unit. So the idea here would be that a line ends shortly before a margin (or hole) because the next character is too large to fit in the

---

[1] To this, Arlo in email (20200911) said, "I would really like to have the Salomon quote included, because it is one of the very few mentions of this sign in the sec. lit. that I'm aware of." - so let's put it in when we finalise the text.

space left there, and the penman wants to tell the reader that although the line is not filled to the end, we do _not_ have a semantic break here. I can well imagine that these signs were originally used only in split words, and then generalised to split sentences / semantic units. It is of course entirely possible that it was then generalised further to become iterated space fillers of the first kind, but as I said, I don't know about that.

- Given all of that, how could we improve our transliteration/encoding strategy so that we avoid unnecessary complication but hopefully do something useful for future research? It may be best to forget about fillers as a distinct category and just call all of them symbols. We could then say § is the recommended shorthand for <g type="squiggleVertical"/> (and not <g type="squiggleVertical">§</g> as currently conceived), and if Ryosuke needs it, we could add ¦ as recommended shorthand for <g type="commaDouble"/> (or whatever token we decide on). Manu's iterated gomutras would then have to be encoded as a series of <g type="gomutraFinal"/> elements rather than <g type="gomutraFinal">§§§§§</g>. Thus, our encoding would contain no explicit indication that the editor considers something to be a "space filler". However, research questions on fillers could still be asked by combing our XML files for any <g/> elements that occur immediately before an <lb/> or a <space type="binding-hole"/>. Some of those would be punctuation marks that by random chance occur at line ends, but those could be excluded on the basis that they are not empty <g/>-s but contain a . character.
- If we don't want to eliminate the category of space fillers, then we may need to make a distinction between the two types about which I mused above, perhaps choosing a name better than "filler" for the § and ¦ type (e.g. call them margin equalisers as proposed in the related discussion on the Indic texts markup list?) and merging actual fillers of the iteratable type (such as Manu's gomutras) back into generic symbols.
- I'm not sure what would be best, but the more I think about it, the more I feel that we should try to minimise both the complexity of the encoding and the level of editorial interpretation. For every distinction we add to the guidelines, e.g. that between fillers and everything else, or between fillers and margin equalisers, etc., new cases are going to pop up sooner or later that don't quite fit. I'd better end on that note.
- Ryosuke in email (20200910)
- You are correct in differentiating the sign encountered in Bengali (and Javanese) inscriptions from those space filler signs: the sign in Bengal inscriptions 'mostly' functions to indicate continuity of the words, between the elements of compound or not. Here I cite some cases from the same inscription:
  - satya¦(5)-tapasas (connecting elements of a compound)
  - citrāṅśukā¦(14)yāḥ|| (connecting a word. Actually the same as the case of vigrahapala, which I mistakenly put as vigrahapala-deva. Actual case is śrīmad-vigrahapāla¦(8)s).
- On the other hand, there is a case in which it is put at the proper end of a word:
  - acchrattanyāḥ ¦
  - (23) prāleyādreḥ
- In any case, it seems not to be a space filler, so we may differentiate it from the latter after gathering more cases to confirm its

function and frequency. I provisionally keep it in my encoding, to be replace later if necessary.

Discussion of *tha* as symbol

Daniel writes: "I've re-checked what Salomon says about this and he explicitly mentions tha. You'll have to make a somewhat subjective editorial decision here."

Wurudu Kidul (844 Śaka) has a closing symbol classified below as "tennisball":
it is perhaps but not evidently intended to have the same shape as the *th* in

2 x tithi
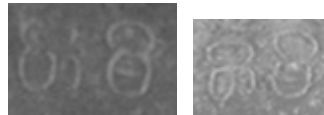
2 x yathānya
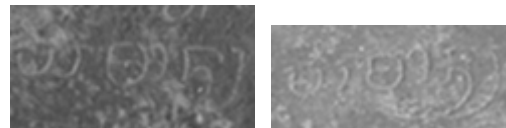
Arthahetoḥ
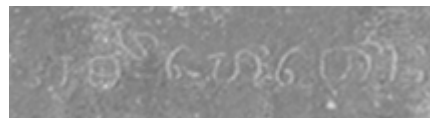
In Mantyasih I (829 Śaka, the same closing symbol has this shape:
and here it does seem intended to have the same shape as the *th* in

tithi

yathānyan·

yathā

A search in my digital version of H.B. Sarkar, Corpus of the Inscriptions of Java, which covers inscription dating until 928 CE discovered until the mid-20th century, did not yield any other cases of a free standing "tha" interpreted/interpretable as closing symbol. And I don't recall having encountered any others. So the phenomenon seems to be very rare on Java, and the above data not entirely conclusive, but tending to support the assumption that akṣara *tha* could serve as closing symbol.