Proposal to encode the Grantha script in Unicode

Shriramana Sharma (jamadagni-at-gmail-dot-com), India
2009-Oct-24

Author's Note

This proposal started out as a joint effort of Elmar Kniprath of Germany and myself. However, Mr Kniprath has decided for personal reasons to retire shortly before its submission. This proposal as it stands contains a lot of material contributed by him, but I assume full responsibility for any errors or omissions. I however retain the words "we", "us" etc in their generic sense.

On the need for a separate proposal

We are aware that a proposal for encoding Grantha (L2/09-141) has been submitted by Naga Ganesan in April 2009. We have thoroughly studied it and compared it with a parallel proposal on which we was working since summer 2008. We decided to continue our work and submit an independent proposal for the following reasons:

1. Mr Ganesan's proposal contains several false assertions. We have submitted our objections as L2/09-316 "Comments on Mr Ganesan's Grantha Proposal".

2. Mr Ganesan's proposal does not consider the use of Grantha for Vedic Sanskrit. Since a large part of the current user community of Grantha is the Vedic scholars of Tamil Nadu, it is very important that Vedic Sanskrit be provided for in encoding.

3. Efforts to work together with Mr Ganesan to submit a unified proposal failed.

§1. Introduction

The Grantha script is an Indic script descended from Brahmi, still being used in its modern form in parts of South India, especially Tamil Nadu and to a lesser extent in Sri Lanka and other places.

Grantha is used to write the Sanskrit language, which includes Vedic Sanskrit. It is also used to write the Sanskrit words of Tamil Manipravalam, which is a mixed Sanskrit-Tamil language that has mostly been used to write works which are Tamil-based commentaries of earlier Sanskrit texts but also some independent ones. Though Manipravalam has become rare, the usage of Grantha for Sanskrit proper continues to this day.

There are no clear statistics on the userbase of Grantha. Our inquiries with publishers of Grantha books suggest that about 10,000 Vedic scholars and students in Tamil Nadu and about 15,000 Hindu temple priests in Tamil Nadu, Sri Lanka and elsewhere do use the script daily. A much larger number of other people interested in Sanskrit and Vedic have learnt the script to varying extents and use it occasionally. The total size of the userbase is however certainly less than 50,000. [Whether to classify this script as "Category A, Contemporary" due to its continuing daily use, or to classify it as "Category B1, Specialized Small" due to its limited user community is a decision we will leave to the UTC. At any rate, this script is certainly not an extinct script (category C) and should not be portrayed as such. Any description of this script in the Unicode Standard should clearly note its continuing daily use as a live script albeit in limited contexts.]

Apart from this "native user community", others have good reason to be exposed to Grantha. Due to a reawakening of interest in ancient Indian heritage and culture, a lot of ancient Indian texts are being unearthed. Many such works from South India are in the Grantha script. Very many manuscripts written in Grantha are housed in indology-related libraries all over the world. At least hundreds of such works have been printed in Grantha in the past century. Many archaeological inscriptions are in Grantha. Thus enabling the standardized usage of Grantha in computers would be of great use to those interested in these works for reading/transcribing them.

Of more importance to the common people, "adding Grantha to computers" will simplify the task of publishing in the script thereby benefiting many thousands of ordinary people in Tamil Nadu desirous of learning Sanskrit or reading religious texts composed in that language.

The Tamil script does not easily lend itself to writing Sanskrit, due to its insufficient repertoire of characters. The usage of diacritic marks such as superscripted numbers has not been a grand success, since readers (especially those whose formal knowledge of Sanskrit is insufficient) often disregard the diacritics and read the script as if it denoted Tamil words, which leads to improper pronunciation. This is undesirable, especially in religious mantra-s which should be pronounced correctly and according to tradition.

Grantha, on the other hand, with its distinct characters for all the sounds of Sanskrit, enables proper representation of that language. It is also easier than Devanagari for Tamilians to learn, due to the high glyphic similarity and even unity of many characters between Tamil and Grantha. A Tamilian concerned with proper pronunciation of Sanskrit words will, while learning to read and use Grantha, stand better chance of being exposed to the distinctness of Sanskrit sounds.

Thus the Grantha script still has potential for widespread use for Sanskrit in Tamil Nadu. A number of people of Tamil Nadu interested and involved in Sanskrit (and not just Vedic Sanskrit), when informed of our ongoing effort to "add Grantha to computers", were very enthusiastic about it especially because of it being still useful for Sanskrit in Tamil Nadu.

Due to this need for Grantha, some interested people in Tamil Nadu have created non-Unicode fonts and software for Grantha, with the obvious disadvantages. Thus the inclusion of Grantha into Unicode is necessary, for which this proposal in intended.

Though there have been earlier versions of the script along its history, as we discuss presently, in this proposal we consider only the shapes and behaviour of the modern form of Grantha. Whether the archaic versions of the script may be considered distinct scripts in their own right to merit separate encoding or whether they should be unified with the proposed encoding of (modern) Grantha is not discussed here.

§2. History

The Grantha script is historically traced as ancient, transitional, medieval and modern Grantha. Ancient Grantha or Pallava Grantha is exemplified by the inscriptions of the early Pallava kings who ruled over parts of what is currently northern Tamil Nadu and southern Andhra Pradesh. It is of the time period 300 CE ~ 650 CE. Transitional Grantha was used by the later Pallava kings between about 650 CE and 950 CE. Medieval Grantha is from about 950 CE to 1250 CE and was used by the Chola kings of central Tamil Nadu. Modern Grantha belongs to the period after this.

After printing presses came into vogue, modern Grantha was further refined and de facto standardised by the publications of books which had hitherto been contained only in ancient media like palm leaves. However, after India's attaining independence in 1947, Hindi was promoted as the national language and hence the Devanagari script started receiving more attention even in South India. More and more Sanskrit books were printed in Devanagari by publishers desiring to cater to a nation-wide audience, and thus the usage of Grantha for the purpose decreased, even in South India.

Recently, Grantha has witnessed a reawakening of interest and a number of manuals for learning the Grantha script have been published. As mentioned above, some interested people have created fonts and composition/rendering programs for Grantha. Based on this, work has begun on digitizing works which were published in the Grantha script in the pre-computer era.

§3. The character repertoire

3.1. Grantha repertoire for Sanskrit

(As to "Extended Grantha" see §8.)

The only language Grantha has been attested to represent is Sanskrit. Therefore the basic character repertoire for Grantha can easily be listed as comprising 67 characters:

1.	independent vowel A – 1
2.	independent vowels and vowel signs for AA, I, II, U, UU, Vocalic R, Vocalic RR, Vocalic L, Vocalic LL, (long) EE, AI, (long) OO, AU – 13×2 = 26
4.	class consonants KA-NGA, CA-NYA, TTA-NNA, TA-NA, PA-MA – 25
5.	"semi-vowel" consonants YA, RA, LA, LLA, VA – 5
6.	fricative consonants SHA, SSA, SA, HA – 4
7.	anunasika sign, anusvara, visarga, ardhavisarga, virama, avagraha – 6

Apart from this, in Grantha there is an archaic sign for pluti or extra-normal length of a vowel. This must be encoded separately.

Further, there are two forms each of the vowel signs for Vocalic L, Vocalic LL and AU. The two forms need to be mutually disunified as we shall presently show in §3.5.

Therefore, 4 more characters are needed for Sanskrit in general making a total of 71.

The following 21 characters seen in Grantha texts are mostly limited to Vedic Sanskrit:

8.	Vedic anusvara-s – 2
9.	Rig/Yajur Vedic anudatta, svarita and dirgha svarita svara markers – 3
10.	Sama Vedic svara markers – all superscript combining:
	a.	Digits 0-6 – 7
	b.	A, KA, NA, VI, Tamil PA – 5
	c.	Asterisk (*), Dash (¯), Caret (^) – 3
11.	Sama Vedic special punctuation superscript double bar (‖) – 1

As to punctuation, dandas are used as in Devanagari. Modern printings also use Latin punctuation. Digits, numbers and fractions as the same as in Tamil.

A table showing all the above characters is shown on the next page.

Vowels			Consonants				Various		Numerals	
a	அ	-	ka	க	da	உ	anunasika sign	ఀ	0	௦
ā	ஆ	ா	kha	வ	dha	ய	anusvara	் ௦	1	௧
i	இ	ி	ga	ம	na	ந	visarga	் 8	2	௨
ī	ஈ	ீ	gha	வ	pa	ெ	ardhavisarga	் ჯ / ் ჯ	3	௩
u	உ	ு	ṅa	ஐ	pha	ெ	avagraha	ऽ	4	௪
ū	ஊ	ூ	ca	ச	ba	ப	danda-s	। ॥	5	௫
ṛ	ஃ	்	cha	ௐ	bha	ஐ	pluti (archaic)	३	6	௬
ṝ	௸	்	ja	ஜ	ma	உ	vedic anusvara	�~	7	௭
ḷ	ள	்ள / ்	jha	௹	ya	ய	vedic double anusvara	ౠ	8	௮
ḹ	௺	்ி / ்	ña	ஞ	ra	ர	Yajur Vedic svarita	్	9	௯
ē	எ	ெ	ṭa	ட	la	ல	Rig Vedic svarita	౹	10	௰
ai	ஐ	ை	ṭha	௦	ḷa	ழ	anudatta	္	100	௱
ō	ஒ	ொ	ḍa	உ	va	வ	Sama Vedic		1000	௲
au	ஔ	்ள / ெள	ḍha	உ	śa	ஶ	superscript combining digits	௦கஉஙசருசூ	Fractions	
virama	-	்	ṇa	ண	ṣa	ஷ			வ, உ, ஜ etc	
			ta	த	sa	ஸ	other svara markers and punctuation	சுகூநவிபட— ∧ ∗ ॥	Modern punctuation	
			tha	ய	ha	ஹ			. , ? ! ' " etc	

3.2. Attestation of the repertoire

The general repertoire (common to Vedic and non-Vedic) is variously attested by refs 2, 3 and 4, which are all of the nature of primers of Grantha or reference works.

The anunasika sign and the Vedic characters are variously attested directly from printed Vedic source texts such as refs 5, 6, 7 and 9. As we are not aware of any formal documentation describing the Vedic characters in the context of the Grantha script, we provide detailed descriptions of the Vedic characters and attestation samples for them from those source texts in §4 of this proposal. We have verified the details with Vedic scholars of Tamil Nadu who are both native users of Grantha and also well-versed in their respective subjects.

As the occurrence of characters in printed texts is an important source for establishing a character inventory, these details are to be considered sufficient attestation for characters for encoding in Unicode despite the absence of any other scholarly publications regarding the same.

3.3. Characters which will be reused and not disunified

The digits 0-9 and numbers 10, 100 and 1000 are the same as in Tamil, and already encoded at 0BE6–0BF2. Also, since the danda-s used for punctuation in Grantha are in no way consistently distinct from the danda-s used in Devanagari or other Indic scripts, in accordance with section F.7.1 of the P&P document, the danda-s encoded at 0964 and 0965 will be used for Grantha too. The following samples prove that the shapes of the numerals in Grantha are the same as in Tamil and those of the danda-s are the same as in other Indic scripts:
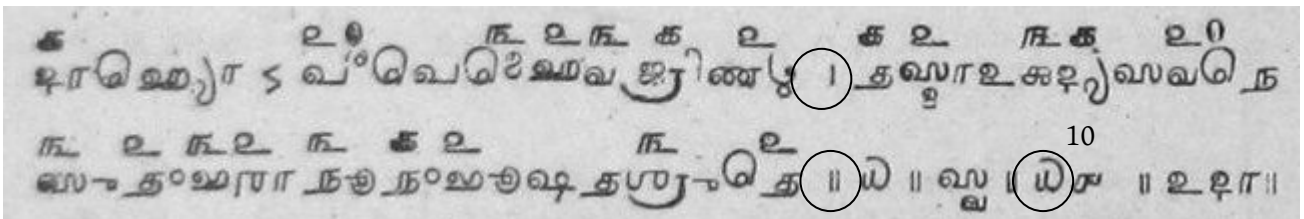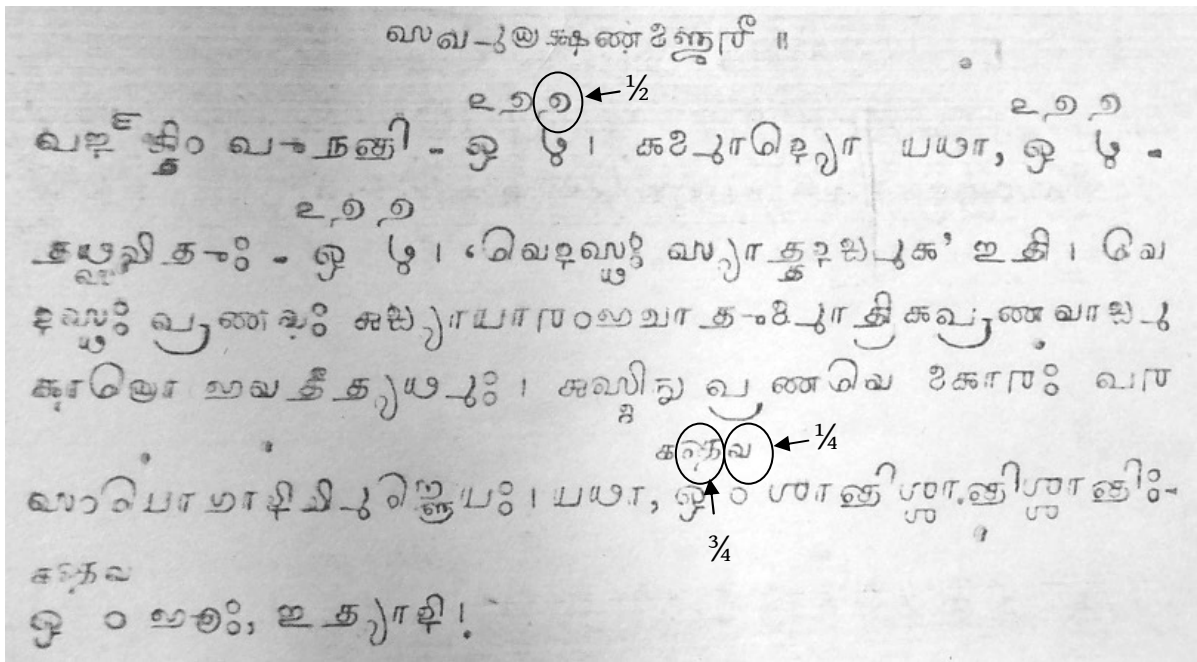

ABOVE:
From ref 6 Ārcika p 15. Shows usage of single and double danda-s, and usage of number 10 (among others).

LEFT:
From ref 10, illustrating all the digits 0-9 being used in the place value system to denote page and line numbers.

One also comes across characters for the major fractions – quarter, half and three-quarters in printed Grantha texts. These are used in texts related to phonetics to indicate the length in mora-s of the sound represented by the character above which they are placed. Such a character

placement is to be handled by ruby notation and not by encoding separate above-combining forms of these characters, since these characters retain their regular meanings when so placed above the main line of text. This is quite unlike the case of the above-combining digits used as Sama Vedic svara markers which do not retain their numeric values then. Therefore it is quite appropriate to use ruby hence making it sufficient to encode normal (non-combining) fractions.

Since these fractions belong to the set of numerals common to both Grantha and Tamil, it is appropriate to encode them along with the other unified Grantha-Tamil numerals in the Tamil block. We have separately submitted a proposal (L2/09-___) for encoding these characters in the Tamil block at the codepoints 0BFB, 0BFC and 0BFD.

The svara markers used in Grantha texts for anudatta, and Rig Vedic svarita and Yajur Vedic dirgha svarita are already encoded at 0952 and 0951 respectively. These characters are generic Vedic accents used for Vedic texts in all Indic scripts and hence it would be appropriate to give them normative aliases indicating this. We have separately submitted a document (L2/09-339) asking for this and also aliases for the danda-s at 0964 and 0965 indicating their generic nature.

The svara marker for Yajur Vedic svarita is sometimes seen in Devanagari texts printed in Tamil Nadu also and therefore a proposal (L2/09-344) has been separately submitted by us to encode it in the generic Vedic Extensions block as 1CF4 Vedic Tone Candra Above.

The Sama Vedic svara signs superscript dash and caret are already present as 1CD0 Vedic Tone Karshana and and 1CD2 Vedic Tone Prenkha. The Sama Vedic special punctuation double danda is a non-oblique glyphic variant of 1CD3 Vedic Sign Nihshvasa. The superscript asterisk is already present as 20F0 Combining Asterisk Above.

The ardhavisarga is the same as 1CF2 Vedic Sign Ardhavisarga, though its usage is not limited to Vedic (either in Grantha or elsewhere). There is a rotated form of ardhavisarga sometimes used in clear semantic contrast with 1CF2, and we have separately submitted a proposed (L2/09-343) to encode it, preferably at 1CF3.

All these characters will be reused and not disunified for Grantha.

It is worthy to note here that some have suggested that even the dependent vowel signs for Vocalic L/LL in Grantha not be disunified from the independent vowels, since they are glyphically identical and both are ordinary spacing characters. (While in the preceding chart we have shown distinct glyphs for the independent and dependent forms of Vocalic LL, either form is valid for either requirement, as will be shown below.) However, the dependent vowel signs of Vocalic L/LL are attested to ligate with their base consonant, as will be shown in §5.2. The independent vowels will certainly not ligate this way. Therefore this disunification is justified.

3.4. Glyphic variants which will not be disunified

As in all scripts, there are many glyphic variants for the characters of Grantha, especially when one considers manuscripts. It is however out of the scope of this document to consider all such archaic variants, for which one should resort to sources like ref 4. We therefore note here only a few variants that are seen even in printed matter:

3.4.1. Vocalic LL

The independent and dependent vowel forms of Vocalic LL have both the same two glyphic variants, as is seen in the following attestation from refs 4, 15 and 16 (from left to right):
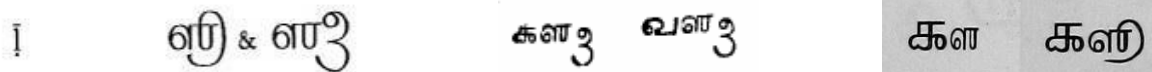
To distinguish these two for the code chart, we have chosen for the independent vowel the form similar to the independent vowel Vocalic RR in having the 3-like "lengthening" mark and for the dependent vowel the other form that has the extra curve.

### 3.4.2. Independent vowels O and AU

The glyph representing independent vowel OO and the left-side glyph in the independent vowel AU are normally the same. There are two "modern" variants of this glyph as seen in ref 2 and ref 6:

The first pair of glyphs from ref 2 are preferred as representative of these characters, due to wider usage and distinctness from the Tamil equivalents.

3.4.3. Avagraha

Grantha has two attested forms of the Avagraha. One looks like an angular S and another like the small letter z. Here are samples from ref 6 Gana p 3 and ref 7 p 576:
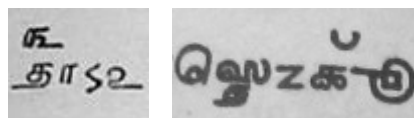
We prefer the angular S for the representative glyph for its distinctness from sans-serif Latin z.

3.5. Disunification of alternate forms of some vowel signs

In Grantha, the vowel signs for Vocalic L, Vocalic LL and AU have alternative forms. However, neither can these alternative forms be considered glyphic variants, nor is it technically viable to unify them and suggest that the alternative forms be handled at the font level.

3.5.1. Dependent vowel signs Vocalic L and Vocalic LL

The dependent vowel signs for Vocalic L and LL are normally written in the same glyph form as the independent vowel placed to the right of the base consonant. However, even in printed books we occasionally come across a variant form where the glyph is instead placed below the base consonant. This variation is seen in these samples from page 4 of ref 15 and page 4 of ref 17:

Dr Ehlers of the Orientabteilung, Staatsbibliothek zu Berlin, who is an expert on manuscripts, confirmed to me in his email dated 2009-Sep-28 15:16 +0530 that "Vocalic L in Grantha is sometimes placed below, sometimes to the right" and he later confirmed in his mail dated 2009-Oct-19 16:17 +0530 that this is true for Vocalic LL as well. Therefore these sub-base forms such as seen in ref 17 are not to be considered a typographic aberration.

In fact, a glance at *all* the other Indic scripts shows that the same glyphs as used for the independent vowels L and LL denote the dependent vowels when placed below a consonant. So one is tempted to say that sub-base forms are "the original form" of the dependent vowels in Grantha.

However, the fact remains that contemporary printing practice consistently places the glyphs *after* the base consonant, and hence that should be retained as the "standard" form of the dependent vowels. This post-base form is also the one conducive to archaic vowel sign ligatures, which we shall discuss in §5.2. Hence this is the "standard" form in Grantha.

Now since the same glyphs are used in both forms, these are not "glyphic variants" but rather "positional variants", since the difference is in the position. Therefore it is inappropriate to suggest that this difference should be handled at the font level, especially since they would differ in their GC (Mc against Mn). If it were not for the requirement for Indic vowel signs to have CCC=0, they would differ in that also. Thus due to an difference in Unicode character properties, these alternative forms of the vowel signs Vocalic L and LL should be disunified.

However, native users of Grantha regard these as merely different forms of the same vowel sign and therefore they expect that searching for one should produce the other also. This is because the only language Grantha is used for is Sanskrit and Sanskrit is both a script-agnostic and multi-script language. That is, Sanskrit words produce their meaning based solely on how they are pronounced and not how they are written (or whether they are written at all), and Sanskrit is written in very many Indic scripts, not to say of transliterations like ISO, IAST etc. This results in users searching for words in Sanskrit text by phonological content and not by orthography.

The reasons for disunification are technical in nature, and not semantic. Ordinary users (of Grantha) will certainly be entirely unaware of these technical reasons or the resulting distinct encoding nor should they need to be aware. Therefore they would certainly expect words like kḷpta presented in either way to turn up in searches.

The only solution is to encode the two forms separately but provide a canonical decomposition from the old-style to the new-style form which would then be the "default" (in the sense of being isomorphically placed with the corresponding characters of the other Indic scripts). Since normalization is done before search operations, the decomposition would achieve the desired result of searching for either alternative returning the other as well.

Thus in order to preserve the language-script relationship, we request that disunification be done but with a decomposition. Disunification should not be done without a decomposition.

In modern printings the dependent vowel AU is mostly written as a single sign placed after the consonant. However in old-style orthography it is sometimes written as a two-part vowel sign as in current Tamil orthography. as seen in pages 5 of both refs 2 and 15:


Unification of the two forms is not possible here since the old-style form must get a decomposition, and the new-style form is a part of the decomposed sequence. Thus these two forms of the vowel sign cannot be canonically equivalent. However, collation should treat these as equivalent. As for searching, if there is a mechanism by which searching for either of the weakly equivalent pairs ä/ae etc in German text turns up the other, a similar mechanism can be used.

3.6. The pluti mark

We said in §3.1 that Grantha texts use a unique sign for pluti and that it needs separate encoding. Pluti means extra-normal length for a vowel, which usually is 3 or 4 moras. Ref 13 (a compilation of manuscripts) shows a distinct sign for denoting this, as seen in the following sample:

whereas in most contemporary printings the actual number of moras is indicated in Grantha-Tamil digits within parantheses, as seen in the following sample from ref 5 p 162:
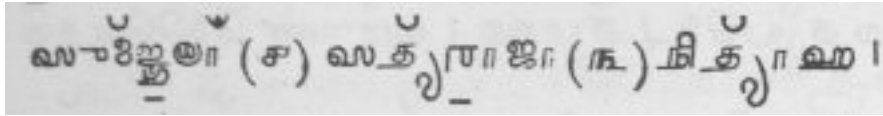
The archaic sign must be encoded as a separate character since it has been attested and cannot be considered a glyphic variant of any other character. It is also not to be considered a purely Vedic character despite the preceding samples being provided from Vedic texts, since non-Vedic Sanskrit also occasionally indicates pluti (in modern orthography by the digits three and four), as seen in the following sample from page 365 of ref 10. If a user so wishes, s/he may easily use this character to denote the pluti in such a non-Vedic Sanskrit text as well:
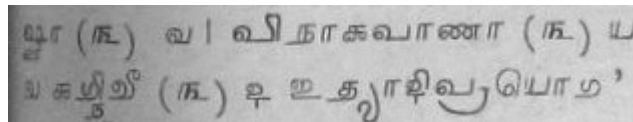
For the two reasons that this is not a Vedic-specific character and that it has not been attested in any other script, it should be encoded in the Grantha block and not the generic Vedic Extensions block. We point out that ref 4 which describes the character forms and signs used in manuscripts and prints of a number of South Indian scripts has attested this character specifically for Grantha and not for any other script. Therefore, this character should not be over-genericized to pan-Indic status like the Kannada Signs Jihvamuliya and Upadhmaniya were (requesting the de-genericization of which we have separately submitted L2/09-342).

3.7. Characters within Grantha which are mutually glyphically similar

In Grantha there is a peculiar characteristic (some would say weakness) by which a large number of characters within the script itself are mutually glyphically similar and differ only by a few small features. Here we list some (not all) such similarities we have observed, and the distinguishing factors between such similar characters. Those reading, writing or making fonts for Grantha need to be aware of these similarities and differences and diligent in preserving them in the glyphs.

1. Independent Vowel A ଅ, Consonant KA க, Digit One க

The vowel A has a line jutting up the top-right quadrant and a rectangle in the bottom-right. KA has no jutting line and has a circle for the rectangle. It also has a small "pinch" in the bottom left. The digit one is identical to Tamil KA and does not have the pinch or the circle in the bottom right.

2. Independent Vowel U உ, Digit Two உ

The difference is visible in the top-left quadrant in the above shapes. Sometimes this distinction is not observed in handwriting and disambiguation is to be done by context.

3. Consonant LLA ளூ, Independent vowel I இ, Consonant NGA ஞ

In LLA there is a horizontal line in the bottom-right, which lengthens and moves to the bottom of the glyph in the vowel I. To this, a small vertical line in the bottom is added for NGA.

4. KHA லூ, PA லூ, VA ல, CA ஸூ

"nose"   "nose"

In KHA, the "head" on the top-left bulges, and creates a "nose". This bulge and nose are missing in PA which is about the same width as KHA. VA is a horizontally contracted form of PA. In CA, the circle on the left moves above (compared to VA) and a pointed "nose" is added.

5. PHA லூ, CHA உஇ

The difference between these two is the same as that of VA and CA above, except that PHA is of the width of PA and not that of VA.
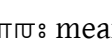
6. Independent vowel OO ஒ, Consonant DA உ

In vowel OO, the head on the top bulges, and the circle at the left side of this head lies upon on the baseline. DA has a smaller head, and the entire head including the circle floats above the nose.

7. THA யூ, DHA யூ, YA யூ

THA has a circle on the top right which is missing in DHA. THA and DHA have a "belly" which is missing in the "trim" (and mostly rectangular) YA.

8. Consonant TTHA ௦, Anusvara ௦, Digit Zero ௦

TTHA is of the same height as the other consonants whereas the anusvara is clearly smaller. For example see the word ṭhaṃkāraḥ ௦௦கார: meaning "a sound like ṭham" (as of something metallic falling). The digit zero is not consistently glyphically different from TTHA (though this font shows it somewhat oblate) and hence disambiguation should be done by context.

The difference is visible in the bottom left quadrant. NYA has a curve towards the circle in the top-left quadrant whereas N·TA does not have a curve (but a straight line) and the circle from the top-left instead somewhat descends towards the line.

3.8. Relationship to Tamil and the case for Tamil-Grantha disunification

From the preceding sections, it is evident that the script historically, culturally, and geographically closest to Grantha is Tamil. Due to this Tamil is also the script glyphically closest to Grantha.

Many characters are glyphically identical in Grantha and Tamil. These are the independent vowels U, UU, the consonants NNA, TA, NA, YA and VA, the vowel signs AA, I and II and all numerals (including the fractions mentioned in §3.3). The Tamil vowel sign AU is the same as the old-style vowel sign AU in Grantha. The Grantha vowel signs for long EE and OO are identical to those of short E and O in Tamil.

The consonants JA, SSA, SHA and HA are also identical between the scripts, though Tamilians explicitly label these as "Grantha letters", excluding them from being native parts of the Tamil script. The "Grantha letter" SA in Tamil is the same as an archaic variant of SA in Grantha.

Some Grantha characters are very close to their Tamil counterparts – the independent vowels A, AA, II, OO and AU, the consonants KA, TTA, RA and SA. The ligature K·SSA is also very similar to that used in Tamil.

Despite these glyphic uniformities, Grantha and Tamil deserve to be considered as distinct scripts. For one, native users perceive them as so. For another, most people who can read Tamil cannot read Grantha and some people who can read Grantha cannot read Tamil properly. Tamil is a phonemic script where many letters denote phonemes with more than one allophone. Grantha is a phonetic script where most letters denote individual phones directly.

It may be said, however, that these factors need not affect encoding in Unicode which mostly considers only the visual aspect of a script. Then let us consider the visual aspect alone. Only 17 characters out of 64 characters comprising the vowels and consonants of modern Grantha are also identically seen in modern Tamil. And as for the 9 characters that are similar, it is unacceptable to present the glyphic distinction at the font level since in Grantha these characters consistently have their particular representative glyphs distinct from the Tamil equivalents.

Further, there are many behavioural differences between Grantha and Tamil consonants. Tamil consonants exhibits ligatures with the vowel signs U/UU whereas modern Grantha does not. Tamil does not stack consonants, does not have a "repha", "ra-vattu" or "ya-phalaa" and uses only the single ligature K·SSA. Grantha, however, regularly stacks consonants, uses the "repha", "ra-vattu" and "ya-phalaa" consistently and has very many ligatures apart from K·SSA.

It has been suggested that these behavioural differences be resolved by using ZWJ and disunifying the Grantha virama which results in all these distinct behaviours. However, the fact remains that a large number (73%) of Grantha vowels and consonants are distinct from Tamil.

If it is suggested to distinctly encode these extra vowels and consonants alone and not disunify the 17 characters that are identical, then the reply is that this does not fit within the pan-

Indic Unicode model. The case of Kaithi is a fitting illustration of how a script, despite sharing most vowel signs and some consonants with an existing script (Gujarati) is disunified in Unicode.

Thus it is entirely justified to disunify Tamil and Grantha. (We have also submitted L2/09-324 in reply to L2/09-277 which asks for Tamil-Grantha unification.)

§4. Vedic characters

4.1. Vedic Anusvara

Ref 8 details the four different types of anusvara-s occurring in the Krishna Yajur Veda. Of these the Agama Anusvara and Lupta Agama Anusvara are denoted by this symbol:
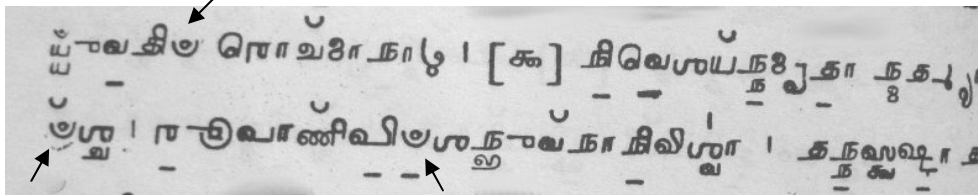
This sample from p 382 of ref 5 shows this letter with svara marks (see the arrows). It is possible to distinguish from the context which of the two anusvara-s is meant by the symbol. When a consonant cluster follows (as in line 2 of the photo) it is the Lupta Agama Anusvara. Otherwise it is just the Agama Anusvara.

In the Gana part of the Sama Veda (described in §4.4 below), this Vedic anusvara sign is used to denote the normal anusvara, while the normal anusvara sign is only used as a 'shortcut' for the homorganic nasal of the following consonant. A sample from ref 6 Gana p 2 is given above.
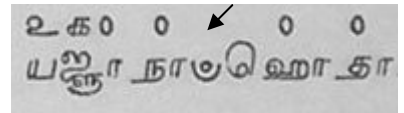
4.2. Vedic Double Anusvara
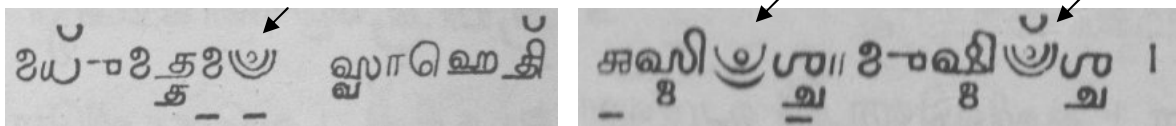
The Dvir Bhuta ("doubled") Lupta Agama Anusvara of the Krishna Yajur Veda mentioned in ref 8 is denoted by this symbol. Samples from pp 75 and 94 of ref 5 are given above.

4.3. Rig/Yajur Vedic Svara Markers
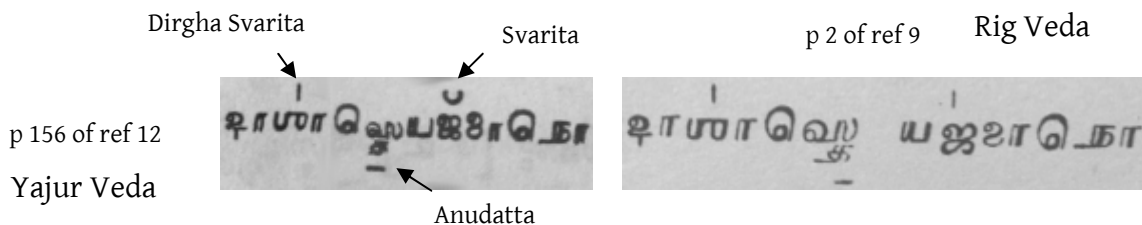
The Krishna Yajur Vedic svara-s anudatta, svarita and dirgha svarita are marked as shown on the sample above left. In the Rig Veda, the same symbol is used for both the svarita and dirgha svarita. Our enquiries revealed that there exist no Shukla Yajur Vedic or Atharva Vedic texts written in Grantha with svaras. Therefore no samples from them are provided here.

4.4. Sama Vedic Svara Markers

As this is a quite complicated matter, we accord it special treatment here. The Sama Veda has two chief parts – the Arcika or poetry part and the Gana or singing part.

The mantra-s which are in the form of poetry are included in the Arcika part with svara-s appropriate for recitation. These svara-s are four – udatta, svarita, anudatta and pracaya.

The Gana part contains these same mantra-s in the form in which they must be sung, and here seven svara-s (comprising a full octave) named krushta, prathama, dvitiya, tritiya, chaturtha, mandra and atisvarya are present.

There are somewhat complicated rules on how the svara-s should be marked. We enumerate them below. All markers are placed on top of the syllables they are applied to ('superscript') unless specified otherwise. To 'add' a marker to a syllable means to place it after an existing marker, or to place it newly on the syllable if no marker previously exists.

We will give rules for the markup of the Arcika part and Gana part separately.

4.4.1. Markup of the Arcika part Of the Sama Veda

UDATTA:

1. The udatta is marked by the number 1.
2. It is marked by the number 2 when it is: a. followed by an anudatta, or b. occurs at the end of a sentence.
3. If there are many sequential udatta-s only the first is marked by rules 1 or 2.
4. If there are many udatta-s followed by an anudatta, a Tamil-like PA is added to the first (the only marked one). [Such an udatta sequence is called shivaka.]

SVARITA:

5. The svarita is marked by the number 2.
6. If it follows a sequence of more than one udatta, a 0 is added to it.
7. Independent svarita-s (a special kind of svarita) are marked by KA+2+0 when they do not follow an udatta, but are: a. followed by an anudatta, b. followed by a pracaya or c. occur at the end of a sentence. (The condition of not following an udatta may be satisfied by their occurring at the beginning of a sentence or after an anudatta.)
8. Independent svarita-s followed by an udatta are called kampa-s. They are marked by 2+KA (without 0) on top and an avagraha and digit 3 on the mainline.
9. If such kampa-s are also preceded by an udatta, the 2 (only) is removed.

ANUDATTA:

10. The anudatta is marked by the number 3.
11. If there are many sequential anudaatta-s, only the first is marked.

PRACAYA:

12. The pracaya is unmarked.

Attestations for the above rules from the Arcika part of ref 6 are given on the top of the next page. The page numbers are given with the samples. One example each of every case in the rules is marked by the rule number.
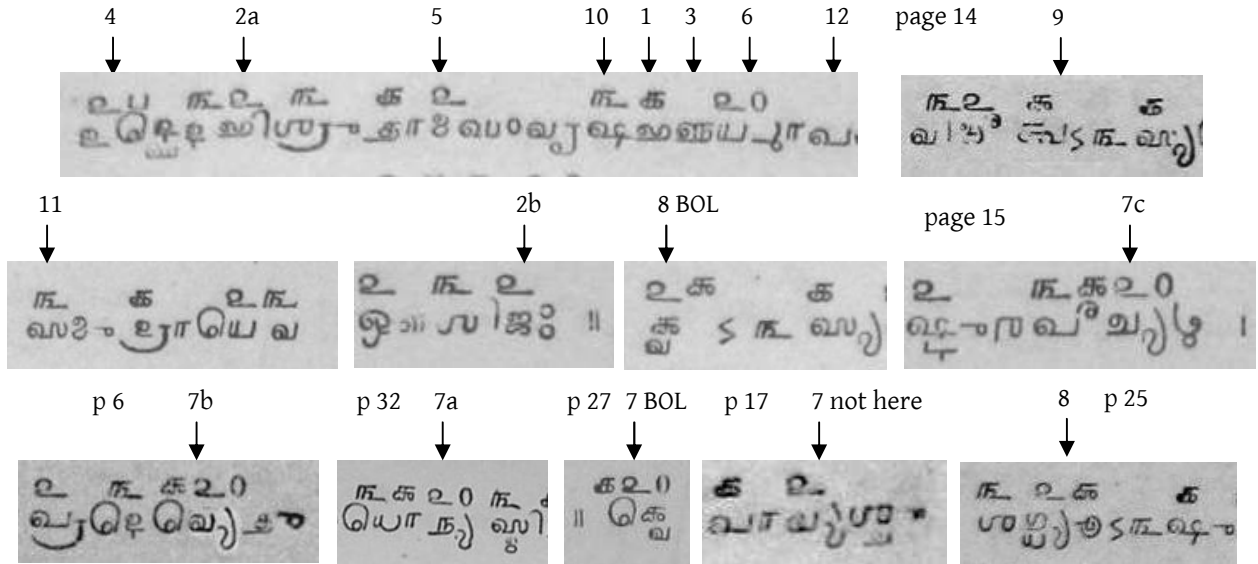
It is worthy to note here that many scholars, including the authors of N3366 which proposed Sama Vedic svara markers for Devanagari, have analysed the superscript KA in Devanagari as an addition to 3 marking an anudatta when such anudatta is followed by an independent svarita or kampa. From the samples we have given here, it is obvious that at least in Grantha the KA does not belong to such anudatta, because it even occurs in the case of independent svarita-s or kampa-s which occur at the beginning of a line (see samples marked BOL) where there is no previous anudatta.

Further, modern printings use an asterisk (*) instead of KA in rule 7. Thus the asterisk indicates independent svarita-s which are not kampa-s.

In summary, the svara symbols used in the Arcika part of the Sama Veda are: the number 1 for the udatta, 2 for the udatta and svarita, 3 for anudatta, 0 as additional sign for some svarita-s, KA as additional sign or sometimes alone for some svarita-s, Tamil PA for shivaka udatta-s, and asterisk (*) for some svarita-s.

4.4.2. Markup of the Gana part of the Sama Veda

1. The normal svara-s krushta to atisvarya are marked by 11, 1, 2, 3, 4, 5 and 6 respectively. Sometimes the 11 is preceded by an asterisk (*).

2. If sequential syllables carry the same svara, only the first syllable is marked.

3. The special svara-s abhigita, namana, vinata, dipta, prenkhana and ayata are different manners of pronouncing the normal svara-s. They are indicated respectively by adding A, NA, VI, Tamil PA, dash (ˉ) and caret (^).

4. Long vowels are by default prolated in singing. Each syllable with a long vowel that should not be prolated gets the number 0 added it.

5. If the vowel of a syllable is continued in other svara-s than the one it first carries (primary svara), the syllable is followed by an avagraha on the main line followed by the numbers marking such secondary svara-s also placed on the main line.

6. If the last secondary svara of a syllable is the primary svara of the next syllable, then the next syllable will NOT be marked as per 1.

7. Vowels are normally pronounced in secondary svara-s for a period of 3 mora-s (traditionally called 'matra'). For each secondary svara that persists only for 1 mora a superscript 1 is placed on top of the number indicating the svara.

Here are attestations from the Gana part of ref 6 marked with the rule numbers and for rules 1, 3 and 5 followed by the svara marker that was placed by the rule:
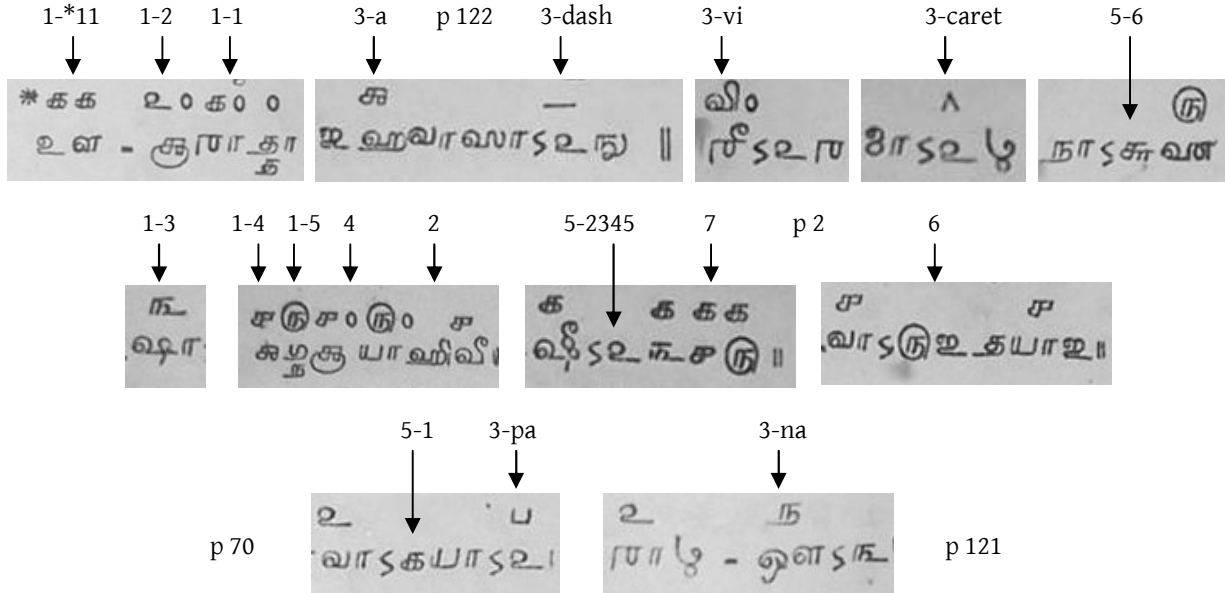

Note that we were not able to get attestation for superscript 6 or mainline 11 in ref 6. However, in N3366 attestation is provided for Devanagari superscript 6 and Grantha superscript 6 will certainly be required if that source is to be written in Grantha. In any case, the requirement for superscript 6 for the atisvarya svara is very real.

In the case of mainline 11, attestation is not a must for the present proposal since mainline svara markers are going to be composed as the regular digits anyway.

Further we note that we do not currently propose the encoding of superscript 7, 8 and 9 as they have no current use. We however reserve space for them.

4.5. Sama Vedic Punctuation

In the Rig Veda, which consists wholly of verses, half-verses and full-verses are ended by a danda and double danda respectively, as is normal for all Sanskrit poetry. In the Krishna Yajur Veda, which comprises mixed verses and prose, the danda is often used throughout without special regard to verses or prose. Sometimes the double danda may be used for full-verses. In the Arcika part of the Sama Veda, their usage is as in the Rig Veda, as it also wholly comprises verses.

page 32 of Ārcika part of ref 6
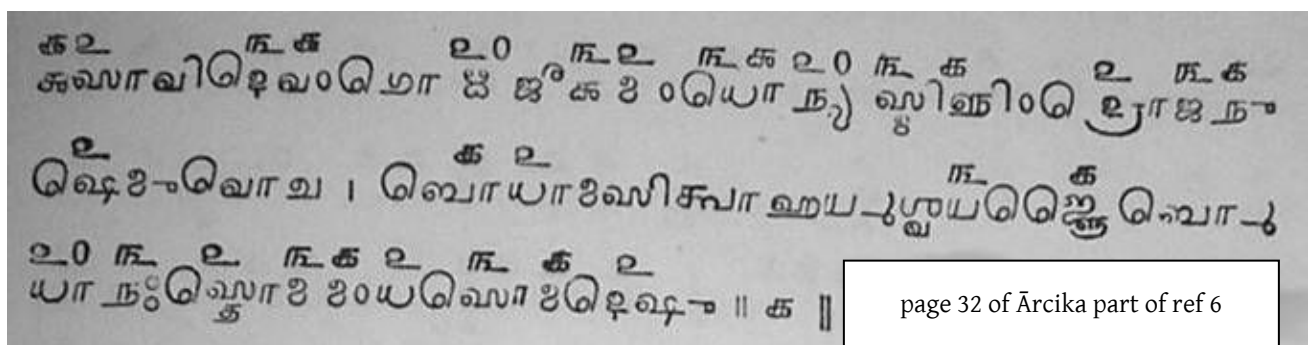
In the Gana part of the Sama Veda, there are special conventions. This part comprises of mantra-s from the Arcika part sung to different melodies called Saman-s. While thus being sung the distinction and identity of the words itself is often lost, what to say of half-verses and full-verses. Thus the subdivision and punctuation are done differently as in this sample from p 3 of ref 6:
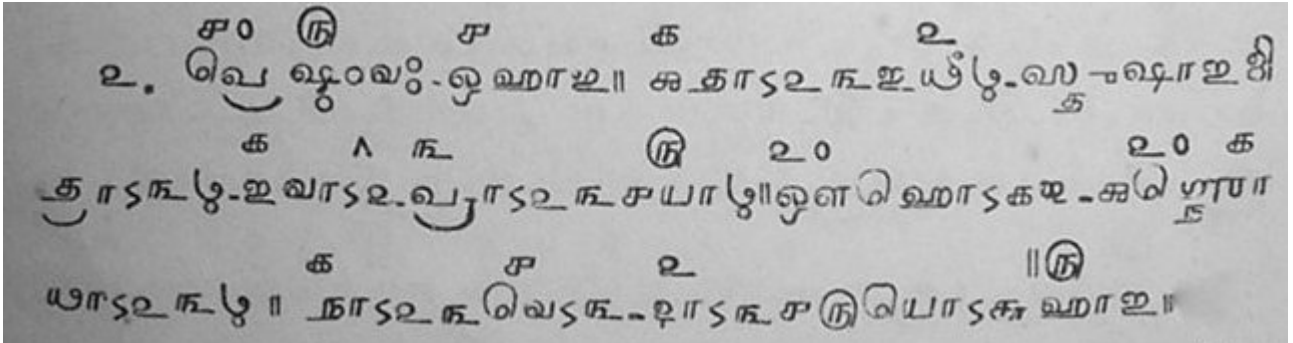
In rituals, each mantra is sung in either five or seven sections called bhakti-s. There are ritualistic specifications as to which of many persons participating in the rituals should sing which bhakti. To denote the end of a bhakti, the double danda is used.

Within each bhakti, there are groups of syllables (let us call them 'phrases') which are to be sung without pausing in between even for taking in breath. The hyphen marks the end of each such phrase where a pause is made and one may take breath. Where the end of a phrase coincides with the end of a bhakti, the pause is indicated by the double danda terminating the bhakti, and so no hyphen is added for that purpose.

4.5.1. The special punctuation superscript double danda

Normally phrases are totally contained within bhakti-s. Sometimes, however, phrases will cross bhakti boundaries. One syllable of a phrase will be in one bhakti and the next in the next. In such cases, it is necessary for the person(s) who are responsible for singing the next bhakti to start precisely at the end of the first syllable pronounced by the person(s) singing the previous bhakti.

At this point, if a normal double danda is placed to mark the end of the first bhakti in the midst of a phrase, it causes the erroneous impression that a pause can be made at that point, since that is how it is used everywhere else. To avoid this error, a double danda is inserted in superscript between the two syllables of the phrase at the precise border of the bhakti. This indicates that the bhakti changes but that no pause is allowed, as the phrase continues beyond this point.

Here we should mention that though until now we have been using the term superscript for non-spacing markers placed above the base character, this 'superscript' character is seen to be spacing, though its advance width is very small (and should be so in fonts designed to support Sama Veda markup). It should also be placed at the precise border of the two bhakti-s.

4.5.2. The existing character for this special punctuation

This superscript double danda is represented by 1CD3 VEDIC SIGN NIHSHVASA, though the representative glyph in the Vedic Extensions block code chart is somewhat oblique rather than upright as in Grantha texts. Whether upright or oblique, the purpose of the character is clear and is one and the same. In Devanagari texts we come across both the upright and oblique forms. To

cater to the expectation of Sama Vedic scholars using Grantha, a Grantha font can show an upright glyph. Therefore these are mere stylistic variations and hence not worthy of disunification.

Here it should be noted that the name of the character 'nihshvasa' (meaning 'taking breath' in Sanskrit) and the corresponding explanation in N3366 (which proposed 1CD3 for encoding) "used to indicate to the performer where a breath can be conveniently taken" are quite misleading. If a person singing the previous bhakti has to also participate in singing the next (which can occur as per the ritualistic rules) then he is prohibited from taking a breath here as that would cause a pause in the middle of a phrase.

The informative alias "vaidika saamagaana yogakaala" merely means in Sanskrit "the timepoint of joining in (i.e. where those who should sing the next bhakti join in) in Sama Vedic singing", which is no doubt quite appropriate. However, this does not clearly indicate that taking a breath here is prohibited, which is indeed the very point of using a superscript double danda rather than a regular double danda. To clarify this, we suggest adding the following annotation: "Separates sections of Sama Vedic singing between which a pause is disallowed".

§5. Orthography

Grantha, being a Brahmic abugida written left to right, exhibits many qualities in common with the other Indic scripts. Consonant symbols have the inherent vowel /a/, which is killed by the addition of a sign called (in Unicode parlance) virama, and replaced by other vowels by the addition of combining vowel signs called matra-s. Consonant clusters are represented by ligatures or conjoining forms. These and other features of Grantha orthography are now discussed.

There is no official prescriptive standard for Grantha orthography. Works like ref 4 merely *describe* the orthography of Grantha as seen in various sources including manuscripts. Therefore, for the purpose of this proposal, the Unicode rendering chapter and resulting implementations, the orthography followed by modern Grantha printing is recommended to be followed since it is "expected" and followed by current users of Grantha, especially the Vedic scholars of Tamil Nadu. This "preferred" orthography is described now, as are some of the most important and common variants, both archaic and contemporary. A rendering model based on this preferred orthography will be described in technical detail in §6. It also provides for the common variants described here.

We also note that in this section and in §6 we distinguish the usage of the terms "cursively connect" and "ligate" as in chapter 16.2 of TUS (5.0 as of this writing). While it may not be fully possible to give objective, satisfactory, unambiguous and mutually distinguishing definitions of these two terms (as even TUS 5.0 does not give such definitions), this distinction is important here and in §6. By "cursively connect" we generally mean that way in which two glyphs physically join, if necessary by means of a single connecting line, so as to retain their individuality in shape and physical region (for the most part). By "ligate" we generally mean a closer form of binding in which the components may change their shapes and overlap each others' regions but still be discernible, or change and merge so much that they may not be discernible at all.

Now one important feature of Grantha is that even contemporary printings of Grantha often show scriptio continua, as seen in the major part of this page from ref 7:

However, it should be noted that this is limited to omitting of whitespace, and punctuation marks are diligently placed, as can be seen in the various other samples throughout this document. In fact, an important reason for the previous page showing scripto continua is that the underlying Vedic passages are in fact "lingua continua", with no pauses occurring throughout! This is quite unlike Latin scriptio continua where pauses are present even if not orthographically indicated. An interesting tidbit here is that one of the tests of mettle of a Vedic scholar is how long a passage he is able to pronounce without pausing to take a breath! The section markers with square brackets do not denote pauses but the end of 50 underlying Sanskrit words.

5.1. Vowel signs

1. Only the vowel signs I and II ever touch their base. They are cursively connected with their base. All other vowel signs do not touch their base.
2. The vowel sign for II and the old-style vowel signs for Vocalic L and LL are non-spacing. All other vowel signs are spacing.
3. The vowel signs for I and II are placed above the base consonant. The old-style vowel signs for Vocalic L and LL are placed below the base consonant.
4. Of the spacing non-touching vowel signs, the vowel signs for EE and AI are reordrant and placed to the left of their base. The vowel sign for OO and the old-style vowel sign for AU are two-part with one part placed to the left of the base and one part to the right. All other (spacing non-touching) vowel signs are "simple" and merely placed to the right of the base.

The images below show the spacing nature and placement of the vowel signs and also the cursively connected forms of the vowel signs I and II with all consonants.

Application of vowel signs to consonants, adapted from ref 15:

க கா கி கீ கு கூ க௃ க௄

கெ கே கை கொ கோ கௌ

கெ கே கா கே கோ

Cursively connected forms of consonants with vowel signs I and II, adapted from ref 15:
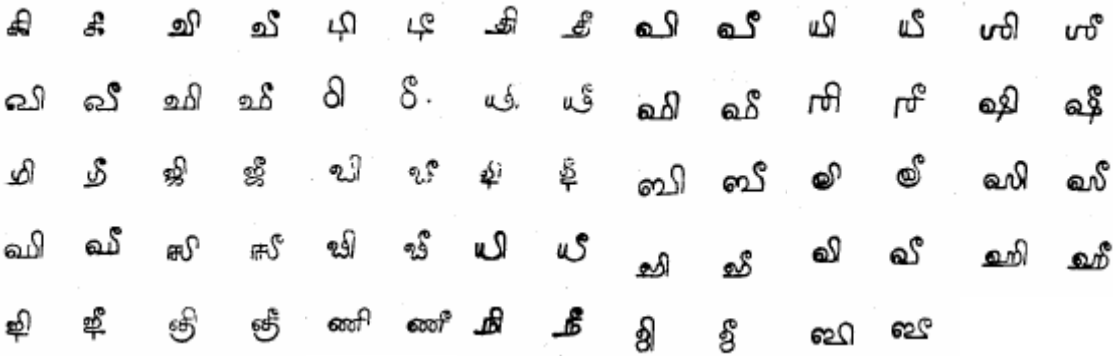
Known aberration: In some printings, the vowel signs for I and II as detached from the base consonant. (This is seen in the first sample of §4.5.) This is considered to be an aberration and possibly results from the printer not having the required connected glyphs in the typeface.

5.2. Archaic vowel sign ligatures

In handwritten manuscripts, one comes across cases where the writer has joined the vowel signs U, UU and Vocalic R and RR (and very rarely, Vocalic L and LL as well) with the base, where such joining is not seen in contemporary printing. These may be variously considered as ligated or cursively joining forms. Some samples from ref 12 are shown below:

Here we remind our readers that had the vowel signs for Vocalic L and LL be solely written below the base consonant (as in other Indic scripts), the ligatures shown above involving them would likely not have happened (at least in the particular way shown above). This should be taken as further evidence that both styles of writing the vowel signs for Vocalic L and LL existed.

5.3. Virama forms

5.3.1. Three different forms of virama

The vowel-killer of abugida scripts is termed 'virama' in Unicode parlance. In contemporary printings of Grantha, the virama sign is most often seen as a spacing character treated in the same way as the other "simple" vowel signs such as that for AA, U, UU etc and placed after the base consonant, to the top-right of the base consonant to be precise. (See the next page for samples attestating this simple spacing form of the virama and the other forms described below.)

Sometimes, the virama cursively joins with ("touches") the base consonant and becomes non-spacing. In yet other cases, the virama ligates with the base consonant, and also effectively becomes non-spacing. In this, the Grantha virama is special among all Indic virama-s in that it produces written forms in three different ways!

Spacing virama forms from ref 2 pp 5-7

"Touching" virama forms in free variation with spacing forms from (same) ref 2 pp 13, 27
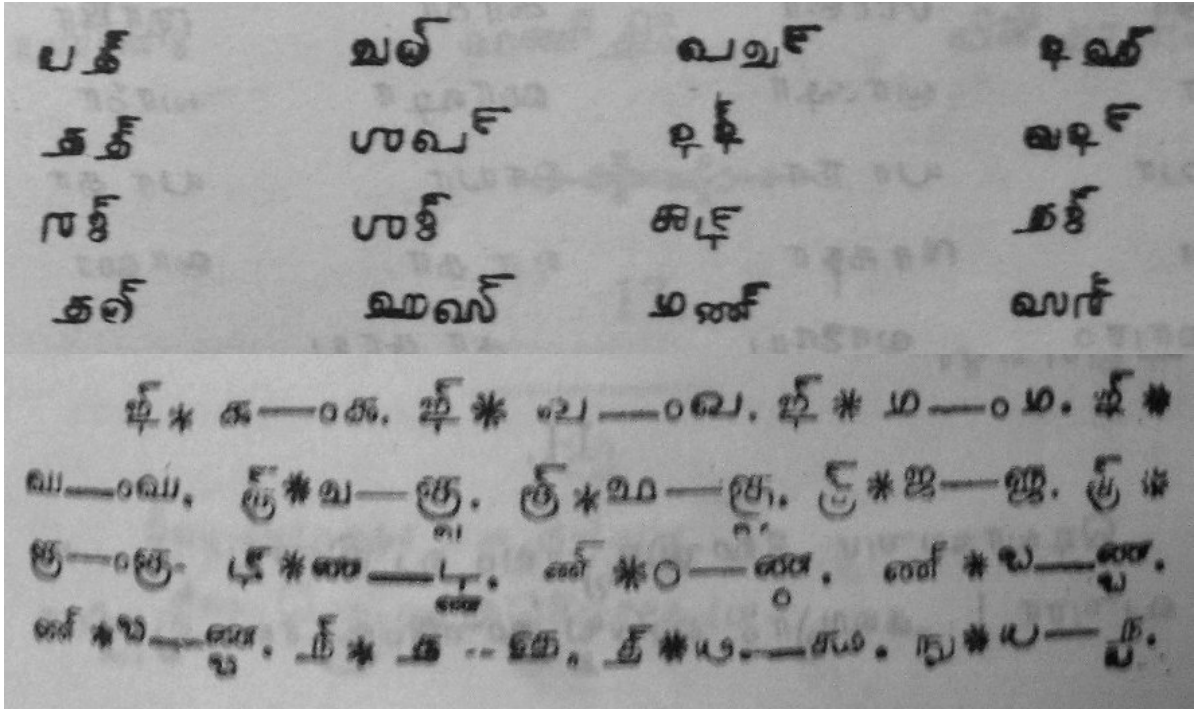
It should however be noted that while all consonants can take spacing or "touching" virama forms, ligated virama forms have not (yet) been attested for all consonants. By consulting manuscripts, ref 4 has attested such ligated forms for only 22 (out of 34) consonants. It is however to be conceded that it is possible that still more ligatures may be present in manuscripts hitherto unexamined by ref 4 or others.

It should also be noted that only 3 of these 22 virama-ligatures – to wit, those of TA, NA and MA – are seen in most recent printings. Those of NGA and TTA are only seen in somewhat earlier printings, and that of KA in still earlier printings. This decrease over time is possibly due to loss of required glyphs in typesetters' lithographic repertoire due to neglect (itself possibly due to decrease in demand).
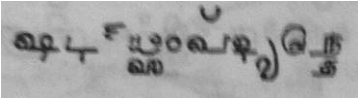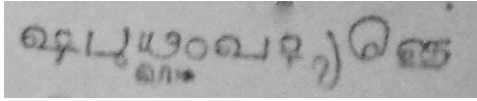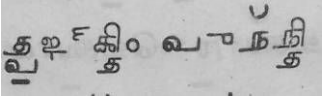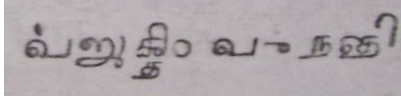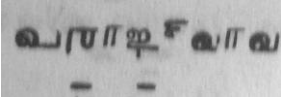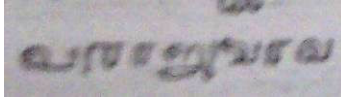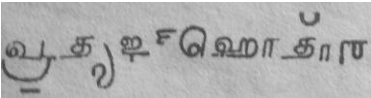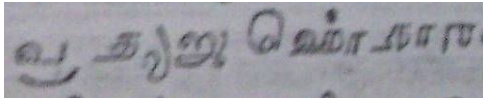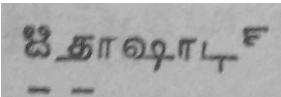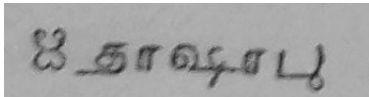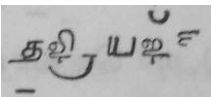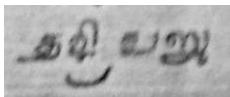
Virama ligatures from ref 4

21

It should be noted that in Grantha, which has Sanskrit (and only Sanskrit) for its native language, all three of these forms of virama are mutually semantically equivalent. This is because, as said before in §3.5.1, Sanskrit is a script-agnostic language and does not distinguish meanings by differences in orthography. The samples shown above for the "touching" virama forms show them in free variation with the spacing forms. The following samples show that the ligated forms also occur in free variation with the spacing forms (thereby establishing three-way equivalence).

For each sample is provided a reference in the format of the reference ID (index to reference list in §14) followed by the page number of the page that shows the sample.

|		
5/808:		18/198:
5/901:		18/276:
7/105:		18/202:
7/673:		18/202:
18-II/078:		18/259:
18-III/116:		18/164:

The respective meanings of the above pairs are: "they become six in number", "they cleanse their line", "moving away indeed", "behind the priest Hotṛ", "he who defeats falsehoods" and "in that direction". (These may be verified from any Vedic scholar who has studied commentaries or translations of the Veda.) In fact, the same words are written in two different ways in the above pairs simply because the virama-ligatures for NGA and TTA fell out of use. The book showing the ligatures was printed in 1908 and the ones without the ligatures were printed in the '80s and '90s.

Further, these virama ligatures have no role in determining word boundaries, despite ref 4 labeling them as "prepausal consonants" due to their being *many times* seen in word-final position. For one, the second example shown above has the single word paṅktim ("line", acc. case) written with the word-medial ṅ alternatively in a ligated and an unligated virama form.

As another example, the following pair shows the 'n' at the end of the word paśumān ("one wealthy in cattle") alternatively represented by a virama-ligature and in combination with the initial consonants of the following word as a consonant stack (described later in §5.4.2):

5/405:		7/698:

The following sample shows a three-fold variation between a virama ligature form in the middle of the word omanvatī ("the protecting ones") as against a consonant ligature (described later in §5.4.1) and consonant stacks.

5/477:   18/259:
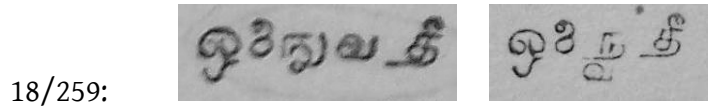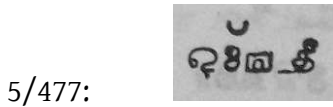
These free variations of virama ligatures with other forms, and the fact that such ligatures can occur both at the middle of and the end of words (or not occur at all) clearly show that the orthography of Grantha does not treat these virama ligatures as "special" in any way.

In Sanskrit, the determination of word boundaries can often only be done by means of knowledge of the language's grammar etc, and not by any such orthographic devices as ligatures, as all Sanskritists know. An important reason for this is that Sanskrit is even now often written in scriptio continua (as shown before) with words overlapping each others' orthographic boundaries.
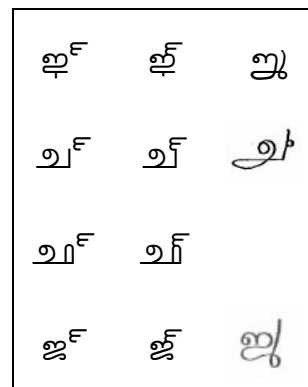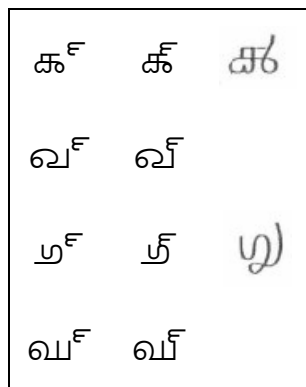
In Devanagari and most other Indic scripts, there are no virama ligatures. When those scripts are fully capable of representing Sanskrit (including Vedic) without contrasting ligated/unligated forms, it is not possible that this change of writing style in Grantha (which is basically what the choice between a ligature and unligated form is) will cause a change in meaning.

Thus we conclude that despite Grantha having these three distinct forms of presenting the virama, there is no semantic difference between the forms.

We also draw the UTC's attention to the unambiguous statement, appended to this document in §17, of Dr R Krishnamurti Shastri, well known Sanskrit scholar and chief trustee of the major Grantha publisher of Tamil Nadu, Heritage India Educational Trust, that using these "special vowelless forms" as he terms them are in no way semantically distinct from adding the "vowel absence marker" that is the virama. Hence there should be no doubt about this matter.

5.3.3. Summary of this section

A chart summarizing the various virama forms of Grantha is shown below. Most pairs of "simple" virama forms and cursively connecting virama forms are not all that mutually different glyphically, except for the former being spacing and latter being non-spacing. However, one must notice that TTA and LLA both develop a curve to join them to the virama, since they do not occupy the top-right quadrant of their bounding box (in font parlance) which is where the virama stands.

꣤ꯆ	ꯆ			ꬺ	ꬺ	
ꯁ	ꯁ	ꭇ		ꩢ	ꩢ	ꭟ
ꬺ	ꬺ	ꭞ		ꫂ	ꫂ	
ꦺ	ꦺ			ꯀ	ꯀ	ꯗ
ꬽ	ꬽ	ꭢ		ꯌ	ꯌ	ꭥ
ꬽ	ꬽ			ꬺ	ꬺ	ꭗ
ꬿ	ꬿ	ꭓ		꧁	꧁	
꫄	꫄	ꭘ		ꬺ	ꬺ	ꭟ
ꭎ	ꭎ			ꭎ	ꭎ	ꭟ
ꫂ	ꫂ	ꭚ		ꭞ	ꭞ	ꭟ
ꦺ	ꦺ			ꭝ	ꭝ	ꭟ
ꭏ	ꭏ	ꭠ		ꭜ	ꭜ	ꭟ
ꭐ	ꭐ	ꭡ		꭛	꭛	

Now it may be suggested: There is no real reason to distinguish the spacing and touching virama forms, since the latter are merely a result of cursive writing and hence tantamount to glyphic variants. Just as in the Latin script, there exist "printed" and cursive forms of all letters, these forms exist in Grantha. Therefore there is no need to specially consider and treat them.

However, this is not so. The distinction between the spacing and touching forms is not merely glyphic, but also behavioural, as we shall show in §5.4.6. Just as spacing and non-spacing forms of the vowel signs for Vocalic L and LL were separately considered in §3.5.1 since rendering engines would have to treat them differently, these also separately considered as they also need to be treated differently by rendering engines to support the orthography indicated in §5.4.6.

5.4. *Consonant clusters*

A phonological consonant cluster is a group of sequential consonants without any intervening vowel. The rules of Sanskrit grammar, to our knowledge, allow for upto as many as *seven* consonants in a cluster (rṭtstry being one such super-size cluster). While such clusters of seven or six consonants are quite rare (while existing nevertheless), five is a more often seen maximum.

Many clusters are (usually) preferably written as ligatures, which then have their own inherent vowel as well, and behave like ordinary consonants for the purpose of application of vowel signs, including the cursively joining vowel signs for I and II.

Some ligatures we have observed in printed matter are shown below. All these represent two-consonant clusters. It is possible that more ligatures of two consonants and even ligatures of three consonants are seen in archaic manuscripts or native users' handwriting.

kta	கௌ	ñja	ஞ	nta	ஞ	nva	நு
kṣa	கௌ	ttha	சு	ntha	நு	mma	ஞ
jja	ஜ	tra	து	nda	நு	śca	பகு
jña	ஞ	tva	சு	ndha	நு	śra	ரு
ñca	ஞு	ddha	து	nna	ஞ	ṣṭa	ஞ
ñcha	ஞு	dva	து	nma	நு	hma	வு

While all these ligatures can easily take spacing or touching (cursively joining) virama forms, only one ligature is known to further ligate with a virama. This is nna, and its virama-ligature is:

எநு

Just like the single-consonant virama-ligatures, the usage of this two-consonant virama-ligature causes no difference in meaning, as seen in the following samples:

	ref 5 p 285
	ref 9 p 36
	ref 7 p 449
	ref 9 p 72
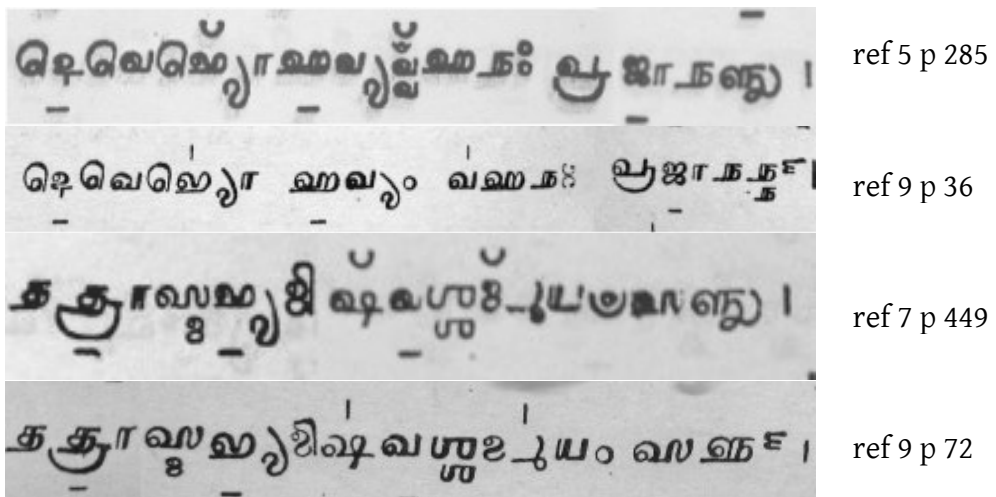
The respective meanings of the two pairs are: "O thou knowledgeable One, take our oblations unto the gods!" and "There may the arrows shelter and protect us". Thus there is no semantic difference. The glyphic difference is merely because of a difference in the printer's choice.

5.4.2. Stacks

If a particular consonant cluster has no ligated form, or if the writer does not wish to form a ligature, then consonant clusters form stacks. These are described now.

The basic principle of stacking is that successive consonants in a consonant cluster are written in their normal form one below the other. ("Normal form" means the form that is normally used to represent the "full" form of the consonant with the inherent vowel.)

In simple **two-consonant clusters**, the first consonant is written in its normal form on the baseline, and the second consonant is written in its normal form below the first consonant, sometimes in a slightly smaller size than usual. Ref 15 shows examples:
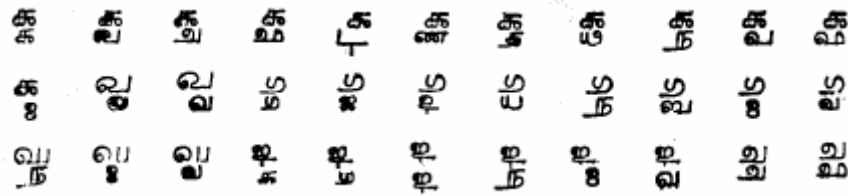

Sometimes, users may choose to form stacks at all, and instead simply write the virama form (whether ligated or not) of the first consonant followed by the second consonant:

7/105:

18/202:
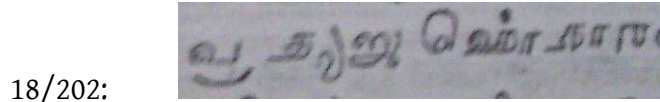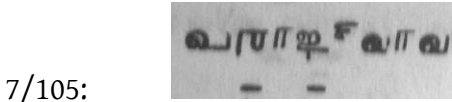
In **three-consonant clusters**, there are many options, two using three-level stacks.

The first option is: The first consonant is written on the baseline, the second below the first (possibly in small size) and the third below the second (again possibly in small size). The second option is: The second consonant is placed on the baseline, the first above the second and the third below the second. In this case all consonants take a small size. This latter option is probably resorted to because it conserves space.

The Yajur Vedic passage shown on page 18 shows both options being used. The third and sixth lines below the header both show both forms of stacking. Here is another passage from ref 7 p 807 replete with three-level stacks (to be precise, many instances of the same three-level stack):
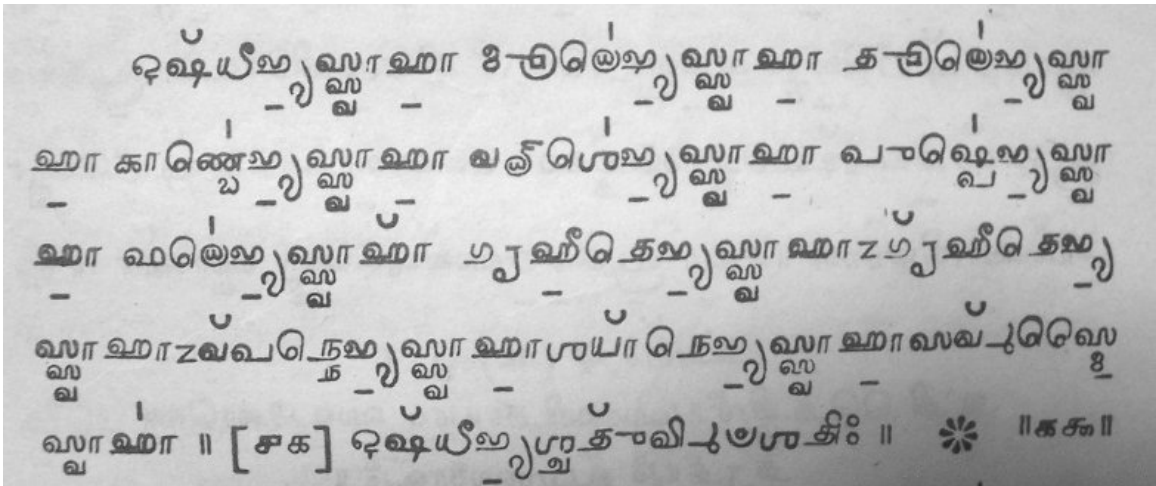

Two more samples from ref 18 p 182 shows many different three-level stacks:


26
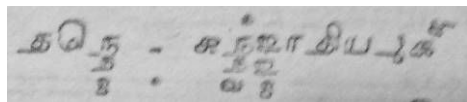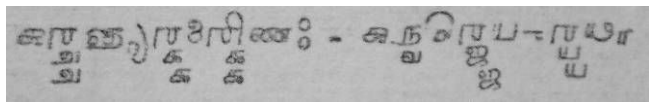
Some of the stacks in these two samples progress slightly to the right and some slightly to the left! These are merely arbitrary variations and not in any way "standard" unlike in Kannada and Telugu where successively stacked consonants are mostly progressively placed to the below and right.

The other option is to not make a three-level stack. In this case, the *first* consonant is pushed outside the stack, as seen in this sample from ref 18 p 197:
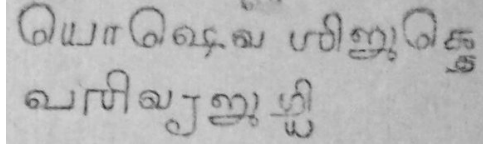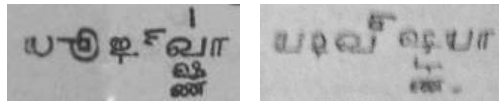

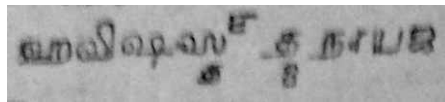Never is seen the case where the first two consonants get stacked and the last is pushed out!

In **four-consonant clusters**, there are very rare instances of full four-level stacking. Here are samples from ref 18, pp 182, 200 and 216 (left to right):
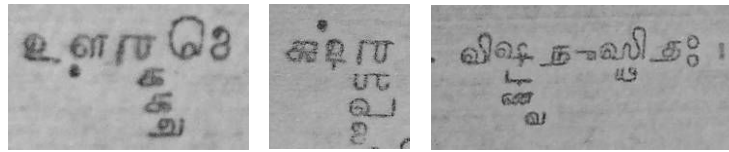
However, most often four-level stacks are not seen, and as before, the first consonant is pushed outside the stack, as in these samples from ref 7 p 577 and ref 18 p 190:

Sometimes, the first two consonants may be pushed out, as in this sample from ref 18 p 190:

Never is seen the case where the last consonant is pushed out!

In **five-consonant clusters**, we have not (yet) seen five-level stacks, or four-level stacks with the cluster-initial consonant pushed out. We have however seen the cluster split up into one two-level and then a three-level stack as in this sample from ref 18 p 202:
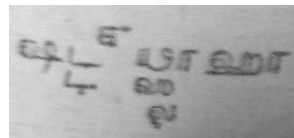
While **six-** and **seven-consonant clusters** (or perhaps even more) are quite possible in Sanskrit, they are to a large extent theoretical and based on phonetic sandhi rules and unlikely (though not impossible) to be attested in printings. Many of them involve special conjoining forms like "repha" and "ya-phalaa" which do not participate in stacking, and which we discuss later. Even if there is a need to stack an arbitrary number of consonants, it can easily be done by the general rule seen above: initial consonants are pushed out and form separate stacks, but subject to the rule that preceding stacks are never of greater size than following stacks.

For all practical purposes, three consonants may be taken to be the maximum stack height in Grantha. Printings usually do not show four-level stacks for the obvious reason that they are quite uneconomical in terms of causing an excessive advance height per line.

In fact, most texts of non-Vedic Sanskrit try to avoid even three-level stacks for this same reason. They are able to be contented with two-level stacks for most cases, since the pronunciation rules are more relaxed than for Vedic Sanskrit resulting in short-cuts being available (such as the one commonly seen in other Indic scripts of writing the anusvara for the homorganic nasal of the following consonant and which is forbidden in writing Vedic Sanskrit because there are strict rules in Vedic on where the anusvara is to be pronounced and where homorganic nasals should occur).

Thus stacking, while theoretically infinitely extendable, is mostly limited to three levels.

### 5.4.3. Stacks with vowel signs and virama

Stacks are equivalent to single consonants for the purpose of application of vowel signs. Reordrant vowel signs reorder around the entire stack, and two-part vowel signs split-position themselves across the entire stack. The cursively connecting vowel signs for I and II connect to the topmost member of the stack. All these are seen in the following sample from ref 7 p 623:
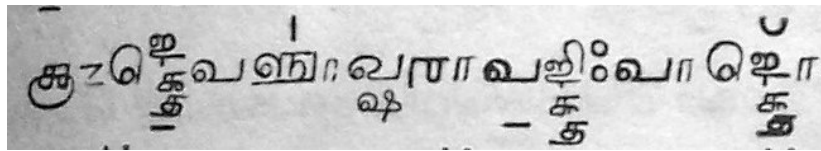
The spacing virama is placed as usual after a stack (as seen in samples in the previous page) and if a virama were to cursively connect or ligate, it would logically do so with the topmost member of the stack (while we we not able to find attestations due to stacks taking virama-s being very rare).

Similarly, if any archaic vowel sign ligatures were to occur, they would also apply to the topmost member of the stack just like the cursively connecting vowel signs I and II. If at all the sub-base old-style vowel signs for Vocalic L and LL are to be applied to stacks, they are to be placed below the lowest member of the stack and above the level of any anudatta accent placed below.

Another important fact is that reordering or split-positioning of vowel signs is always with respect to only the final stack which immediately precedes the vowel sign (in logical order). Any initial consonants pushed out of the stack always take a virama (in any of three forms) and are not affected by vowel signs following the cluster, as seen in ref 18 p 197:
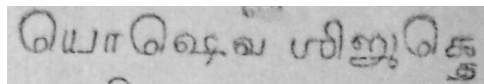
5.4.4. Ligatures in stacks

Hitherto we have shown only simple stacks of consonants. As said before, ligatures are usually preferred to stacks. When a phonological consonant cluster cannot be represented by a ligature alone, it may be written using a combination of ligatures and stacking, as seen in ref 7 p 821:
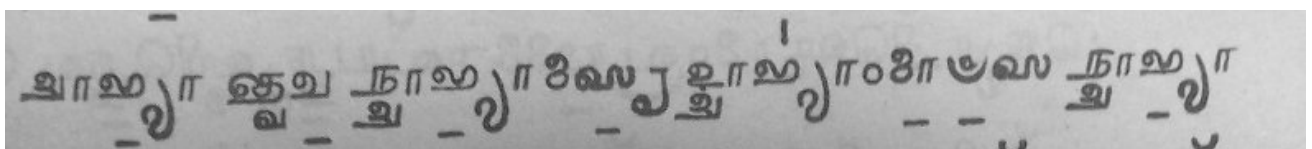
As seen in the given sample, ligatures may take any position in a stack. Further, the choice of which two consonants to represent by a ligature is entirely the writer's. The clusters 'ntv' and 'ndv' seen in this sample can be represented either as NA + sub-base T·VA/D·VA or N·TA/N·DA + sub-base VA. The choice exhibited by the text composer above is quite arbitrary.

When ligatures are used as the topmost member of a stack, they will cursively connect with the vowel signs I and II and possibly also cursively connect or ligate with a virama.

Some ligatures have descenders, and these may affect the number of consonants that can be stacked. For example, in the sample from ref 18 p 196 shown below, the SS·TTA ligature descends below the baseline and hence physically occupies one additional stacking space. Thus, the cluster ṣṭṇ has itself occupied three stacking spaces, leading the initial kh to be pushed out.
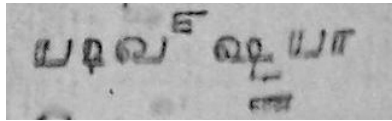
5.4.5. "RA-vattu"

When RA occurs as part of a stack but is not the topmost member, it is replaced by a sign that, in some Indic script terminology, is called "RA-vattu", as shown below from ref 2:
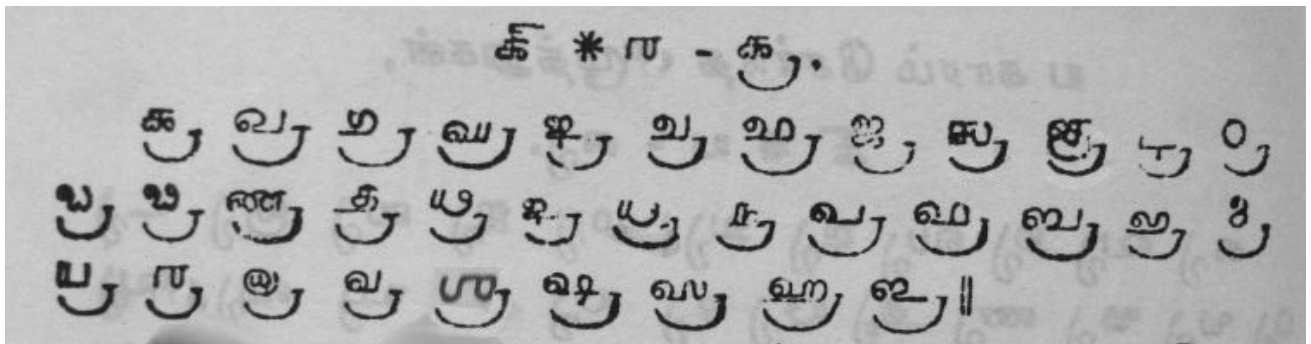
However, occasionally one comes across cases where the normal form of RA is seen in stacks, such as in the following samples from ref 7, pp 56 and 375:
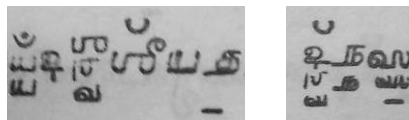
Upon inquiry, Vedic scholars using Grantha inform us that while it is most convenient to write the RA-vattu, the normal form of RA may be used in free variation, and thus, RA-vattu is optional in Grantha unlike in other scripts. We are also told that there is no rule that RA-vattu can only occur at the bottom of the stack, which is why we used the words "not the topmost member" above.

5.4.6. "Repha"

While the word "repha" means in Sanskrit merely the consonant sound /r/, in Unicode parlance it denotes a special conjoining form of cluster-initial RA. (The word "repha" is possibly also used elsewhere in contrast with "rakāra", while these terms are synonymous in Sanskrit.)

In Grantha, the repha takes the form of a spacing sign placed at the end of the consonant cluster. This may mean placement after a single consonant, ligature, or stack, as seen in these samples from refs 15 p 8, ref 5 pp 436 and 477 (from left to right):
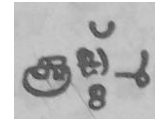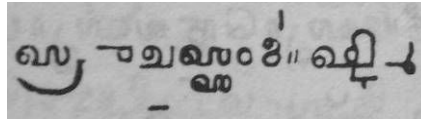
Obviously, using a repha decreases the potential stack height. Thus it is preferentially used when possible, though variations placing the RA at the top of the stack do exist (as seen in various samples in the preceding section on stacking §5.4.2).

When a repha is used, vowel signs for I and II are to be applied to the cluster preceding it, as seen in the above sample from ref 5 p 436. The archaic sub-base vowel signs for Vocalic L and LL and any cursive connection or ligation with virama would also be applied to that cluster.

Further, any spacing vowel signs would come after the repha, as seen in ref 7 p 568:
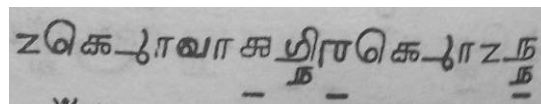
Similarly, spacing virama-s also come after the repha as in the following samples from ref 5 p 678, ref 19 pp 108 and 260 (from left to right):
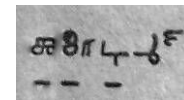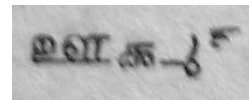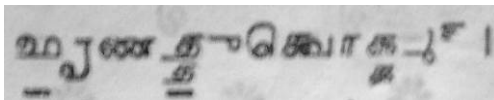
This is only fitting, seeing as the virama is quite often analysed as a vowel sign, to wit, a sign of vowel absence, and since the repha is placed before all the other spacing vowel signs, it is only to be expected that it is placed before the spacing virama as well. This is precisely the reason why we have shown the spacing virama forms as distinct from the touching virama forms in §5.3. The touching (or ligating) virama forms would obviously not allow the repha to intervene.

Further, if a consonant cluster can take a repha and it is split up to decrease stack size, then the repha would gravitate to the initial consonants. For example:

Known aberration: Sometimes the repha is seen after vowel signs, as in ref 6 Arcika p 32:
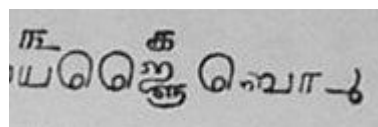
This may be considered an aberration since the repha is *mostly* seen before spacing vowel signs. The same applies to repha with the spacing virama as well.

5.4.7. "YA-phalaa"

The Bengali-derived word "YA-phalaa" is used to denote a post-base form of a cluster-final YA. In Grantha, this takes a form similar to that seen in Kannada and Telugu, as seen in ref 2 p 23:
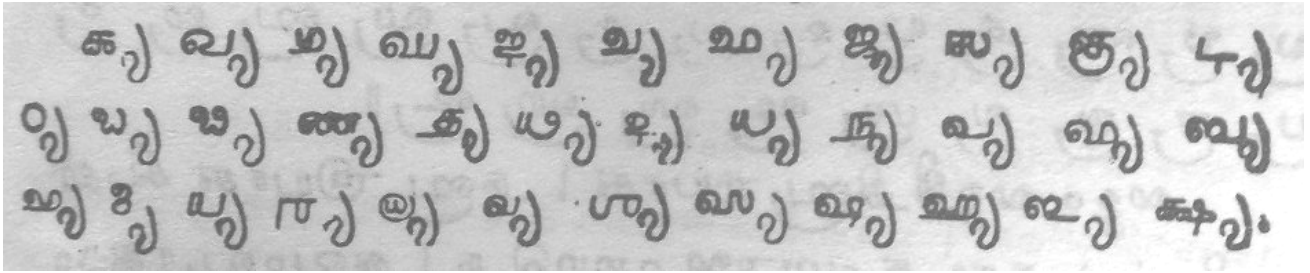
When YA is not cluster-final, YA-phalaa is never seen, as in ref 5 p 105:
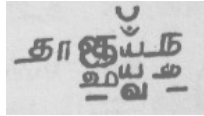
However, even when it is cluster final, especially when after YA, it is optional, as seen by the same Vedic text being written in both ways in ref 7 p 177 and ref 19 p 250 (left and right):
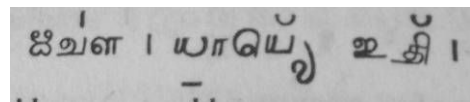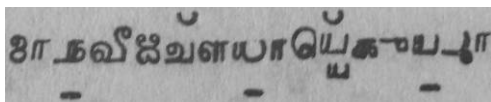
The YA-phalaa is placed before spacing vowel signs (or parts thereof), as seen in ref 7 p 763:
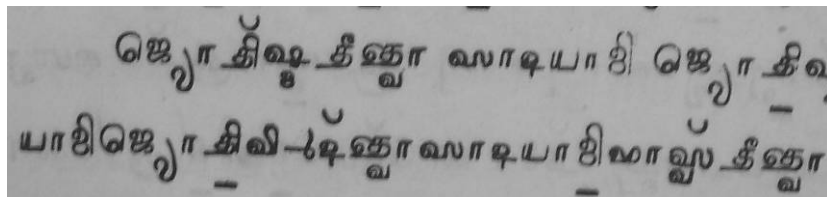
Just as for the repha, cursively joining vowel signs for I, II etc all apply to the main cluster only.

Known aberration: The YA-phalaa may also be sometimes seen to the right of vowel signs but this is to be considered an aberration since as shown above it *mostly* stands before spacing vowel signs.

5.4.8. Repha vs YA-phalaa

There are two possible ways of writing R·YA, one with the initial RA taking the repha form and placed after the normal form of YA, and the other with the final YA taking the YA-phalaa form and placed after the normal form of RA. The form with repha is preferred while the latter is also seen.

In other clusters where both repha and YA-phalaa occur, the repha is placed before the YA-phalaa (which is then placed before spacing vowel signs), as in ref 5 p 679 and ref 7 p 763:
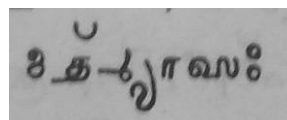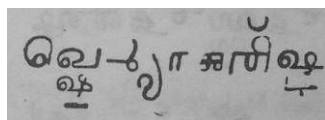
Known aberration: Sometimes the YA-phalaa is seen before the repha, as in ref 6 Arcika p 27 (shown on the right). This is to be considered an aberration since the repha *mostly* stands before the YA-phalaa.
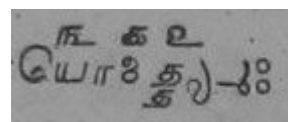
We have remarked that in "standard" orthography, repha, YA-phalaa and spacing vowel signs stand in that order. A further justification for this is that it makes some sense for displayed forms to not excessively deviate from their underlying characters in logical (phonetic) order. Cluster-initial RA and cluster-final YA both occur before vowel sign characters in logical order, and obviously cluster-initial RA occurs before cluster-final YA. It would make sense to choose that visual form which most closely follows this logical order as the "standard" form and consider other forms aberrations. As most contemporary printings conform to this, we call it "standard".

5.5. Typefaces

We here include a few notes on known Grantha typefaces and fonts.

The oldest Grantha book this author remembers personally having seen is a copy of Smṛti Muktāphala (ref 10) printed in some year in the latter part of the 19th century. The same typeface as seen in that book is seen in most of the samples provided throughout this proposal, despite these samples originating from various publishers. In response to our inquiries, the proprietor of one of the few publishing houses providing for Grantha told us that this same (lithographic) typeface has been used by them since 1903 till date for the few Grantha printings they get commissions for (usually a single annual page for religious almanacs).

This typeface is as shown in this sample from ref 10 p 12:
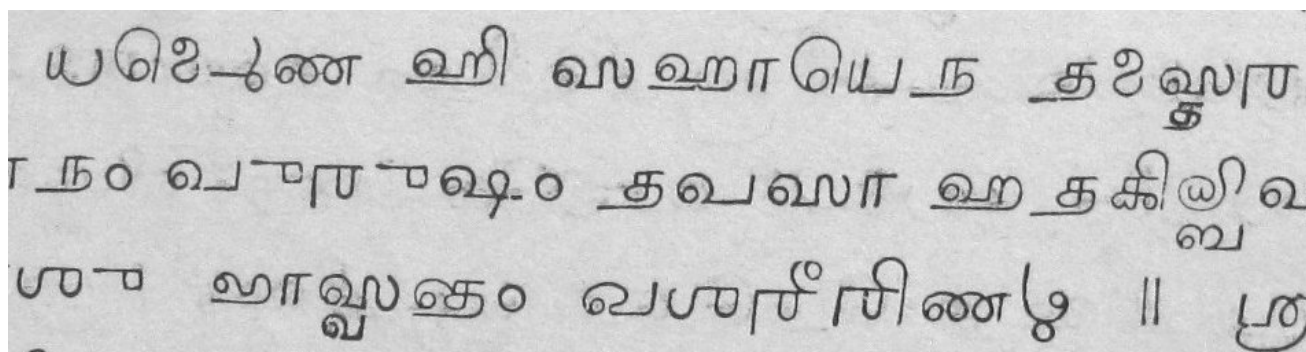
The Vedic scholars of Tamil Nadu perceive this typeface as clear, legible and very aesthetic, with the least ambiguity between the various ambiguous characters of Grantha (see §3.7). It is hence perhaps no wonder, especially considering the limited user base, that no other (lithographic) typefaces were churned out of type foundries (or at least reached widespread use).

Ref 9 is one of the few traditional books using a (slightly) different typeface:
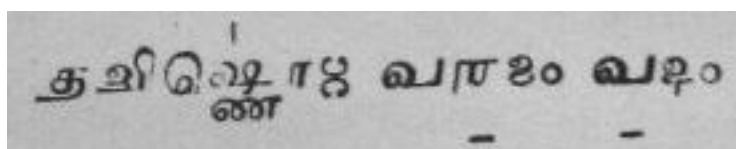
Unlike the consistent thin pen of the previous sample, alternate thick and thin strokes are seen.

IIT Madras has created a non-Unicode font for Grantha. This font iitGranth is available from http://acharya.iitm.ac.in/mirrors/vv/vvcrdt/software.html. It is not as aesthetic (in our opinion) as the traditional typeface, but it is nevertheless legible. A sample is on the top of the next page:
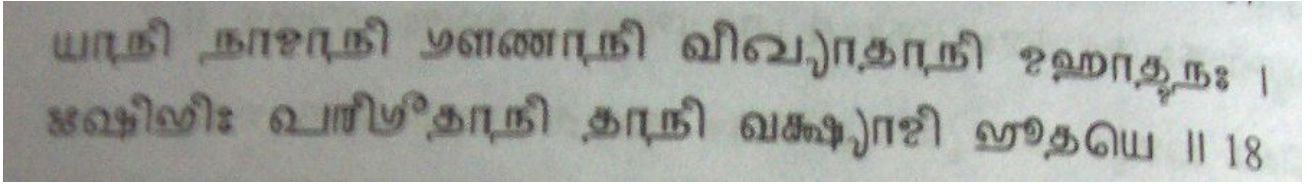
This sample is taken from the Grantha primer book Kranta Lipi Ōr Aṟimukam, published 2004 by one R Sridharan, Chennai. It uses the IIT's font throughout.

James Kass' font Code2000 has (apparently recently added) glyphs for Grantha:

கஒவழஒழஈ

As has been noted elsewhere (http://unifont.org/fontguide/, retrieved 2009-Oct-26), Code2000 is more to be commended for its coderange coverage than its glyphic beauty. However it is to be noted that this font shows all attested virama ligatures and some consonant ligatures also.

Finally, but most important of all, the only "naturally usable" Grantha fonts are e-Grantha OT and e-Grantamil by Elmar Kniprath, publicly available at no charge for non-commercial use from http://www.uni-hamburg.de/Wiss/FB/10/IndienS/Kniprath/INDOLIPI/Indolipi.htm. These are "naturally usable" in that they employ OpenType tables to render Grantha characters when input a Bengali encoding of regular Indic vowels, consonants etc. As such, they help users to immediately start producing Sanskrit text displayed as Grantha, though with a Bengali encoded form. Since the Sanskrit repertoire is common to most Indic scripts, the Bengali encoding is sufficient to (temporarily) store Sanskrit text displayed in Grantha. When Grantha is finally encoded in Unicode, these texts may easily be converted to proper Grantha encoding by virtue of the Grantha characters being placed isomorphically to Bengali (and other Indic blocks).

These fonts are able to properly display most consonant clusters and support ligatures, three-level stacks and special conjoining forms like RA-vattu, repha and YA-phalaa. Further, while e-Grantha OT supports only Grantha, e-Grantamil supports both Grantha and Tamil and hence enables the writing of Sanskrit-only, Tamil-only and also Tamil Manipravalam text. As we describe in §7, Tamil Manipravalam employs mixed Grantha and Tamil characters and hence users of Tamil Manipravalam would certainly benefit from the same font supporting both scripts. (There is even a blog http://zraddhaa.blogspot.com/ which depends on this font for displaying Manipravalam.)

To conclude, while it is obvious that Grantha has not seen very many fontfaces, many efforts are ongoing to create a fully usable Grantha font for modern (computer-based) typography.

§6. Rendering

The orthography of Grantha has been described in much detail in the previous section. We have also considered whether, when there is more than one representation of the same content, the variations are all acceptable, or one is representation is "standard" whereas others are aberrations.

To achieve the standard representation, much is expected of fonts and rendering engines, especially regarding the complications in consonant cluster rendering and with the Vedic svara markers. We wish to emphasize that **it should be possible for the user to see his "default expected behaviour" without needing to input invisible characters like ZWJ/ZWNJ and by merely inputting the "plain" sequence of consonants, vowels etc**, since this is possible for

Sanskrit written in other Indic scripts. However, the most common variations should be provided for by mechanisms using invisible characters if necessary, and we here discuss such mechanisms.

We also discuss some important user expectations regarding combining marks, especially Vedic svara markers, most of which are probably true on a pan-Indic scale, since Vedic texts are written and printed in very many Indic scripts and most renderings have poor support for Vedic even in Devanagari. We however restrict ourselves here to Grantha.

6.1. Archaic vowel sign ligatures

The archaic vowel sign ligatures of Grantha described in §5.2 may be supported in the same way as Bengali vowel sign ligatures.

C + VOWEL SIGN	→	user-preferred vowel sign form for C
C + ZWJ + VOWEL SIGN	→	ligated vowel sign form for C
C + ZWNJ + VOWEL SIGN	→	unligated vowel sign form for C

6.2. Virama forms

6.2.1. Requirement of a ligating virama

We have shown that there is a three-way glyphic and behavioural contrast in Grantha between spacing, touching and ligating virama forms, while there is no semantic contrast between these at all. In particular, the spacing and touching virama forms, while very close glyphically, need to be distinguished because they are behaviourally different in that the spacing form of virama causes the repha to be placed between itself and its base consonant whereas touching (or ligating) virama forms would obviously not allow this.

If there is a need in Devanagari plaintext to distinguish between three different forms of displaying K·SSA – ligated, conjoining and with overt virama – it can equally be argued that in Grantha it would be appropriate to provide a mechanism to unambiguously distinguish between these forms in plaintext encoded representation.

On the other hand, these forms are semantically equal, and Grantha users use any of these three virama forms in free variation for the same language content. Therefore any such mechanism should preserve semantic equality. That is, a text search for a word or phrase containing any of the three virama forms in Grantha should equally return that same word or phrase showing a different virama form (just as in Devanagari all three forms of presenting K·SSA are returned). This is because, as we have repeatedly said, Sanskrit users search for content by phonological content and not by orthography.

In Bengali (and Grantha, as said above) old-style vowel sign ligatures are supported by providing two sequences using ZWJ/ZWNJ to unambiguously denote ligated and unligated forms. The plain sequence CONSONANT + VOWEL SIGN is however left free for the font maker to render as he may require in order to satisfy his target audience. Anyone with the appropriate font will never have to use invisible characters like ZWJ or ZWNJ to achieve his default desired appearance.

Similarly, in the case of the virama forms of Grantha, the plain sequence CONSONANT + VIRAMA should be left free to be rendered as the font maker or user pleases to achieve their desired

"default expected behaviour", and other sequences defined to unambiguously request all three virama forms. These sequences should nevertheless be mutually semantically equivalent.

For this, we suggest the encoding of at least one other virama, which should be non-spacing (as against the spacing nature of the default virama) and *most importantly be canonically decomposable to the default virama in order to maintain semantic equivalence.* We suggest that this virama be called "ligating virama". Using this additional virama one can define:

C + VIRAMA	→	user-preferred virama form for C
C + ZWNJ + VIRAMA	→	spacing virama form for C
C + LIGATING VIRAMA	→	touching virama form for C
C + ZWJ + LIGATING VIRAMA	→	ligated virama form for C

Here in C + ZWNJ + VIRAMA the ZWNJ is used as for Bengali vowel signs, to prevent any combination (cursive connection or ligation) between the consonant and the virama. The term "ligating virama" was chosen because in general usage (i.e. outside this document and outside TUS ch 16.2) the verb "ligate" is often used for both cursive connection and ligation as distinctly defined here. (For example, TUS ch 9.6 Tamil Ligatures uses the word "ligature" for cursively connecting vowel signs I and II in Tamil.) C + LIGATING VIRAMA is defined to cause simple or first-level "ligation" (i.e. cursive connection) and in C + ZWJ + LIGATING VIRAMA the ZWJ promotes this to full ligation.

This four-way mechanism is desired because, apart from the three unambiguous sequences for the three virama forms, the "default" (i.e. user-preferred) virama form for each consonant differs. For example, in the typographic style adopted by the major Grantha publisher, Heritage India Educational Trust, Chennai for their Yajur Vedic books, the consonants TA, NA and MA and the ligature N·NA consistently take a ligated virama, the consonants RA and LA take a touching virama and all other cases take a spacing virama, as seen in these samples from refs 7 and 5:
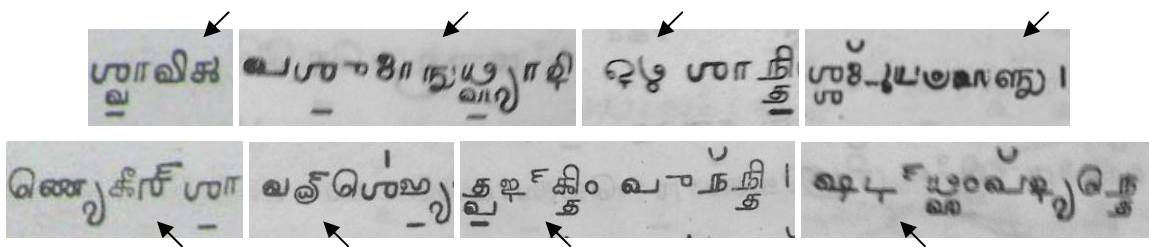
A font catering to this typographic style would automatically provide the appropriate virama form when input the sequence C + VIRAMA and not require the user to input invisible characters.

We note that while using ZWJ/ZWNJ after the Bengali model for vowel signs would be sufficient in Grantha if there had been only two different virama forms, and hence the encoding of a separate ligating virama would be unnecessary, the three-fold contrast between the three virama forms in Grantha cannot be achieved by meaningful and unambiguous sequences using ZWJ/ZWNJ alone. (Please see L2/09-___ "Further clarifications on Grantha virama ligatures" for more details.)

6.2.2. Rendering the sequence C + VIRAMA + ZWNJ

It is important to note that even in rendering systems which do not immediately output a visible virama when input the sequence C + VIRAMA and hence one further needs to input a ZWNJ to force

display of the vowelless form of the consonant, the sequence C + Virama + ZWNJ should still output the same user-preferred virama form that would be output by C + Virama in other systems. This user-preferred form may even be a virama-ligature for a particular consonant C and there is no requirement that it must be an overt (= spacing) virama form.

While TUS 5.0 chapter 9.1 p 301 says that this sequence is used to display an "explicit virama", it is our understanding that here the term "explicit virama" is used only to distinguish forms that can independently represent vowelless consonants from conjoining forms and not distinguish to between various forms of virama which may equally independently represent vowelless consonants. (This is all the more true since most Indian scripts do not have more than one more form of virama and hence chapter 9 does not consider differences in virama forms at all.)

In Grantha, as we have shown before, the relation between ligated and unligated virama forms is exactly that between the ligated and unligated vowel sign forms in Bengali and both forms (along with touching forms) are used in mutual free variation. Thus all three virama forms in Grantha are equally capable of independently representing vowelless consonants and it is permissible for all three forms to be displayed for C + Virama + ZWNJ.

A further justification for this is the following text in the same TUS section quoted above: "... the Unicode Standard adopts the convention of placing the character U+200C Zero Width Non-Joiner immediately after the encoded dead consonant ... In this case, *the virama sign is always depicted as appropriate for the consonant to which it is attached.*" In Grantha, the *appropriate* method of depicting the virama is a matter of typographic style and may be any of the three forms. (This is again similar to the case of the Bengali vowel signs where the appropriate method of depicting vowel signs depends on whether the orthography is old-style or new-style.)

While it is possible that the above wording in TUS was not crafted with Grantha in mind, our understanding is that it is more than coincidental, and this sequence is intended to output the "default" independent vowelless form of a consonant across all Indic scripts. As is being repeatedly said, what the "default" form is (in Grantha) is a matter of choice in typographic style and the ZWNJ after the virama has nothing to do with selecting which default form should be used.

While in other scripts (Malayalam) virama-ligatures have been analysed as conjoining forms and hence C + Virama + ZWNJ must not produce those forms, in Grantha virama-ligatures are *not* conjoining forms. Thus C + Virama + ZWNJ must render exactly in the same way as C + Virama.

6.3. Consonant Clusters

6.3.1. General rules

The orthography of consonant clusters has been described in detail in §5.4. As far as "default" expected behaviour is concerned, based on the patterns observed in the majority of printings, the following rules are provided in order of preference and may be taken to be normative:

1. Ligatures are formed.
2. Cluster-initial RA is shown as a repha.
3. Cluster-final YA is shown as a YA-phalaa.
4. Stacking occurs.
6. Stacks are split up to avoid over-stacking.

5.   RA in stacks when not the topmost member is shown as a RA-vattu.

6.   Individual consonants are shown in their preferred vowelless form.

Here we note that we place ligatures before repha formation because there exist some archaic ligatures with cluster-initial RA. If the repha formation pass is executed before ligature formation, such ligatures would be overlooked. For modern orthography, a font would merely not make the ligatures available and "fallback" to using a repha would happen. Similarly, there are some archaic ligatures with cluster-final YA which would prevent YA-phalaa formation.

Since repha and YA-phalaa have no role in stacking, after repha and YA-phalaa are formed, stacking should occur for the remaining consonants. The relative positioning of repha and YA-phalaa with stacks has been described in §5.4.6-8.

6.3.2. Stack overflow

We have clearly mentioned in §5.4.2 that stacking preferentially occurs in back to front order. As the stack limit for Vedic orthography is normally chosen to be three levels, and that for non-Vedic Sanskrit is two levels, when the number of elements (simple consonants or ligatures) needing stacking overflows this limit, the initial elements are pushed out of the (main) stack and may form their own stacks. We illustrate this in the case of Vedic orthography:

two elements	→	two-level stack
three elements	→	three-level stack
four elements	→	vowelless element + three-level stack
five elements	→	vowelless two-level stack + three-level stack
six elements	→	vowelless three-level stack + three-level stack

An important implication of this pushing out of elements is that it necessarily causes an orthographic cluster break. Thus the elements pushed out of the main stack will form a separate orthographic cluster (or theoretically even more than one), and due to the absence of any vowel they would take their default vowelless representation which may be any of three virama forms.

Thus a single phonological consonant cluster (followed by a vowel) may be represented by more than one orthographic cluster. Any applicable repha would then operate upon the first such orthographic cluster, and any YA-phalaa, vowel signs (and subsequent non-spacing marks like the anunasika sign) would operate upon the last, as in (the largely theoretical example) ūrkkstryeva:

<div align="center">ஊௗ௯௨ர்ᵉ௦௫ு௵வ</div>

Here we note that it is possible that most rendering systems parse consonant clusters for conjoining form formation in logical order. Such systems might require the user to input ZWNJ at the desired place to achieve proper stacking distribution. However, since Grantha orthography unambiguously uses back to front order for preferential stacking, an ideal rendering engine for Grantha should cater to this without needing the use of ZWNJ.

It should also be remembered here that the stack limit may be exceeded leading to stack overflow not only because of the number of stacking elements but also because some ligatures involved in stacks may occupy two stacking positions, as shown in §5.4.4.

6.3.3. Manual override of automatic rendering for consonant clusters

The two joiners ZWJ and ZWNJ can be used as in all Indic scripts to override automatic rendering for consonant clusters. We here discuss their application to Grantha, based on TUS ch 9 and PRI 37.

ZWNJ has one and only one role. It "separates text elements", i.e. splits up the preceding and following sequences and obstructs any normal connecting (cursively joining or ligating) behaviour between them. In effect, it causes either to be rendered as if the other did not exist.

Apart from the use of ZWNJ to prevent archaic vowel sign ligatures in Grantha if a font/rendering engine produced them by default, it may be used to obstruct normal rendering of consonant clusters. When placed after an encoded vowelless consonant i.e. C + Virama, ZWNJ "excludes it from conjunct formation" i.e. removes it from the orthographic cluster. It should be remembered that the virama is then depicted as appropriate for the consonant, and in Grantha, this may be in any of three forms. (We repeat this intentionally, since it needs to be made clear that the "explicit virama" section of TUS ch 9 equally applies to all three virama forms in Grantha.)

In effect, ZWNJ has the effect of splitting up orthographic clusters, especially at points where they would not be split up by default. This means that a sequence C + Virama + ZWNJ has an obligatory following cluster break.

The **other three sequences which request specific virama forms** as defined in §6.2.1, namely C + ZWNJ + Virama, C + Ligating Virama, C + ZWJ + Ligating Virama, all have an obligatory following cluster break as well since they would cause their specific requested virama form to be displayed, thereby precluding ligatures and stacks.

It should be noted that it cannot be predicted from the "plain" sequence C + Virama whether a cluster break would follow or not without knowledge of font internals. This is not specific to Grantha, but is true even for Malayalam which, albeit to a smaller scale than Grantha, also uses ligatures, stacks and stack overflow for consonant clusters.

The application of **ZWJ** to Grantha is the same as prescribed for all Indic scripts. Cluster-initial RA is the only Grantha consonant with a C1-conjoining form. RA in other positions and all other consonants have only C2-conjoining forms. Thus the sequence RA + Virama + ZWJ explicitly requests a repha (which may be needed to prevent archaic ligatures involving cluster-initial RA) and ZWJ + Virama + C explicitly requests C2-conjoining forms of consonants.

ZWJ can also be used in Grantha to effectively choose which of two possible ligatures is to be formed. In a three-consonant cluster, when a three-consonant ligature and two two-consonant ligatures are possible, as in 'ntv', 'nt' and 'nv', ZWJ can be used as follows:

NA + Virama + TA + Virama + VA	→	ஞ்வ
NA + ZWJ + Virama + TA + Virama + VA	→	ந்க்வ
NA + Virama + TA + ZWJ + Virama + VA	→	ஞ்வ

In a case where there is no three-consonant ligature but two two-consonant ligatures are possible, as in 'ndv', a sequence without ZWJ-s will be ambiguous as to which of the two-consonant ligatures is to be chosen. There, the usage of the ZWJ will disambiguate:

NA + ZWJ + VIRAMA + DA + VIRAMA + VA   →   நௗ

NA + VIRAMA + DA + ZWJ + VIRAMA + VA   →   நௗ

NA + VIRAMA + DA + VIRAMA + VA   →   either of the above

ZWJ can further be used here *twice* to cause three-level stacks:

NA + ZWJ + VIRAMA + TA + ZWJ + VIRAMA + VA   →   நௗ

NA + ZWJ + VIRAMA + DA + ZWJ + VIRAMA + VA   →   நௗ

To display C1- and C2-conjoining forms in isolation, the sequences RA + VIRAMA + ZWJ (since the only C1-conjoining consonant in Grantha is RA) and NBSP + ZWNJ + VIRAMA + C may be used.

 It should be noted that the rendering of the sequence C + VIRAMA + ZWJ where C != RA is undefined, since no consonants other than RA have C1-conjoining forms in Grantha. It is suggested that implementations render this as equivalent to C + VIRAMA + ZWNJ i.e. as the default vowelless form of C, to provide *some* meaningful rendering.

6.3.4. Requirement for two variation selector sequences

In §5.4.5 we have noted that RA-vattu occurs in free variation with the nominal form of RA placed in sub-base position. As both are C2-conjoining forms of RA, it is not possible to distinguish these using joiners. These may also not be considered glyphic variations but behavioural variations since RA-vattu is a behavioural pattern of C2-conjoining RA seen in almost all Indic scripts. Usage of the regular RA in sub-base form is also attested in Grantha and hence must be provided for. It cannot be considered an aberrant rendering since it conforms to the general logic of writing the nominal form of consonants in sub-base position to form stacks, in contrast with other aberrations in Grantha not conforming to any logic but in fact going against logic, as shown in §5.4.9.

 While RA-vattu has been taken to be the default rendering, a variation selector sequence is requested for displaying the nominal form of RA in sub-base position.

 Similarly, in §5.4.7, we have noted that a behavioural variation exists between post-base YA-phalaa formation and the nominal form of YA being used in sub-base position. While YA-phalaa has been taken to be the default rendering, a variation selector sequence is requested for displaying the nominal form of YA in sub-base position.

6.3.5. Summary of consonant cluster rendering

It is evident that the task of rendering Grantha consonant clusters is quite complex. It would be a challenge for any rendering engine to properly render them in the "default expected" form without requiring the user to input invisible characters. Apparently some rendering engines do not even support more than a small number of consonants in clusters, even for Devanagari. This would seriously hinder support for Sanskrit, especially Vedic, where a number of unexpected sequences can and will occur. Considering the fact that Grantha is solely devoted to Sanskrit, and its importance for Vedic in particular, a good rendering engine for Grantha should be designed keeping in mind all the nuances (and idiosyncrasies) of cluster orthography we have described.

6.4. Combining marks

6.4.1. Anusvara, visarga and ardhavisarga

The anusvara and visarga are distinct sounds in Sanskrit. The anusvara is defined to be a "pure" nasal sound with no oral articulation. Well-trained Vedic scholars do realize this sound correctly in pronunciation of Sanskrit while various aberrant pronunciations may exist. The visarga is a voiceless glottal fricative (whose voiced equivalent is the consonant HA). Both these sounds are represented by distinct written characters of the same name in most Indic scripts.

The ardhavisarga is a written character representing the two sounds jihvamuliya and upadhmaniya (the voiceless velar and bilabial fricatives respectively). The written ardhavisarga is pronounced unambiguously as either the jihvamuliya or upadhmaniya as per Sanskrit grammar. Sometimes 'ardhavisarga' may be used as a generic name for the two sounds themselves.

Despite these characters thus denoting distinct sounds, these characters are commonly considered combining marks, perhaps after the native tradition of appending aṃ and aḥ to the list of vowels. However, there are many situations in Sanskrit where these characters are written separate from any preceding syllables. Here are samples from ref 6, pp 6 and 3, ref 5 p 484:
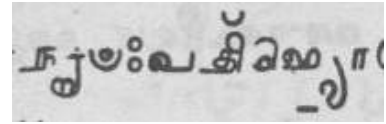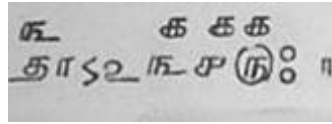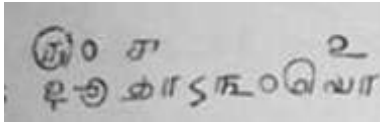
These are only some examples of situations in Sanskrit (and not only Vedic) where these characters need to be used independently in totally unexpected contexts. Therefore, it seems to us that it would be useful to encode these characters as independent letters (GC=Lo) and not as combining marks (GC=Mc) especially since many rendering engines have the annoying habit of churning out dotted circles with Indic combining marks in an unexpected (!= incorrect) sequence.

While we have encoded the Grantha anusvara and visarga with GC=Mc in keeping with the pattern of existing Indic scripts, we request the UTC to consider whether GC=Lo may not be advisable. While it has been considered "destabilizing" to change the GC of existing Indic characters, perhaps these yet-to-be-encoded characters may be safely allotted GC=Lo. We request the UTC to do this *if and only if* it is judged that it would not create problems in round-trip conversion of Sanskrit texts to and from Grantha.

We have included this discussion in this section on rendering since if it is judged to keep GC=Mc for the Grantha anusvara and visarga, makers of rendering engines need to be aware of the need to support the use of these characters in such expected sequences in Vedic Sanskrit.

6.4.2. Non-spacing marks

The non-spacing marks needed for Grantha orthography are the anusvara and the Vedic svara markers. These are normally applied to independent vowels, consonants (with the inherent vowel) and consonants with vowel signs. Sometimes they are also applied to consonants with a virama, as we shall show presently. They are all preferably placed horizontally centered relative to the syllable, but this may not be possible in some cases, such as when there are more than one mark, or the ascender of vowel signs hinders such placement. In any case, they should not go beyond the

horizontal span of their base so as to be clearly and unambiguously associated with that base. Vertically, they are ideally placed below the level of Vedic svara markers since this is how this is seen in Grantha texts, as in this sample:
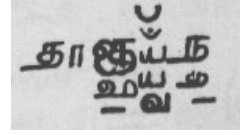
We will discuss the anunasika sign and the various Vedic svara markers separately.

6.4.3. The anunasika sign

In Sanskrit, the three consonants Y, L and V have two 'forms' each, nasal and non-nasal. (We do not discuss here whether these are allophonic variations or not.) In writing, the nasal sounds are represented by the symbols for the non-nasal sounds with an anunasika sign on top, as follows:

<p style="text-align:center">யீ ய்ீ ஞ் ஞ்ீ வ்ீ வ்ீ</p>

These nasal vowelless forms would be encoded as C + VIRAMA + ANUNASIKA. In some esoteric texts on Vedic phonetics other (non-nasal) consonants also take an anunasika sign to denote features like nasal release. Hence any sequence C + VIRAMA + ANUNASIKA should be rendered properly as shown. The other sequences used to request specific virama forms must also be supported this way.

It is to be noted however that sequences such as C + ANUNASIKA or C + VOWEL SIGN + ANUNASIKA are ambiguous as to whether the nasality belongs to the vowel or to the consonant. This ambiguity is inherent in the abugida nature of the Indic scripts and cannot be rectified.

<p style="text-align:center">ய்ீ = யீ + அ்ீ / ய்ீ + அு, ய்ா = யீ + ஆ்ீ / ய்ீ + ஆு</p>

However, such consonants are always vowelless in Sanskrit and hence ambiguity is not seen in real world usage. Even then, when the nasal consonants Y/L/V get stacked, the anunasika sign is placed at the top of the stack in "standard" orthography, and the ambiguity surfaces once again:

<p style="text-align:center">ததுஜெய்ீ்ொ: ததுய்ீஜெொ:</p>

<p style="text-align:center">ஈஉாஜெ்ொகாநு ஈஉாய்ீஜெொகாநு</p>

<p style="text-align:center">ஈஉாவ்ீ்ாஅழ ஈஉாவ்ீவாஅழ</p>

The forms on the right are the unambiguous representation of the phonetic content, but are not commonly seen in printings, since it is assumed that whoever is reading the text is sufficiently knowledgeable in Sanskrit so as to properly identify the anunasika sign as applying to the cluster-initial consonant. While people ignorant of Sanskrit phonetics would certainly misconstrue the anunasika sign in these sequences as applying to the vowel based on its usage elsewhere, those in the know will certainly encode these clusters as C + VIRAMA + ANUNASIKA + C + VOWEL SIGN. Therefore rendering engines must ensure that these *correct* sequences result in the forms shown on the left. Anyone desirous of the unambiguous representation shown on the right can use a ZWNJ to split up the orthographic cluster. Consonants other than Y/L/V may also be treated likewise.

Here are some more samples from ref 5 showing the desired rendering of some more consonant clusters beginning with nasal Y/L/V:
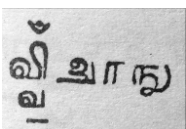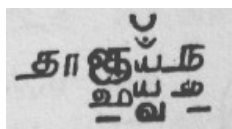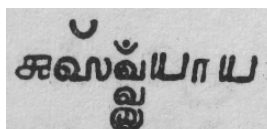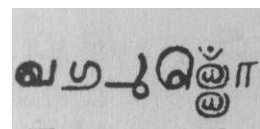
 p 239   p 105   p 98   p 205

<u>6.4.4. The Rig/Yajur Vedic svara markers</u>

The Rig/Yajur Vedic svara markers can generally be applied to all syllables. The anudatta is displayed below its base. The other two are displayed above, and above any anunasika sign (or ascenders of vowel signs).

The anudatta and svarita marks can also be applied to the two Vedic anusvara-s. For samples see §4.1. They can also be applied to vowelless consonants rendered in any virama form. Some samples: (from ref 5 pp 126, 120 and 114)

Therefore the sequence VOWELLESS CONSONANT + SVARA MARK should be rendered properly. Here VOWELLESS CONSONANT is any one of the four sequences defined in §6.2.1 for virama forms.

Further, in the Rig Veda, the sequences TAMIL DIGIT ONE/THREE + DIRGHA SVARITA + ANUDATTA are used to indicate kampa-s (a kind of svarita). The Yajur Veda also uses TAMIL DIGIT ONE/TWO + ANUDATTA for some kampa-s. Samples are shown below:
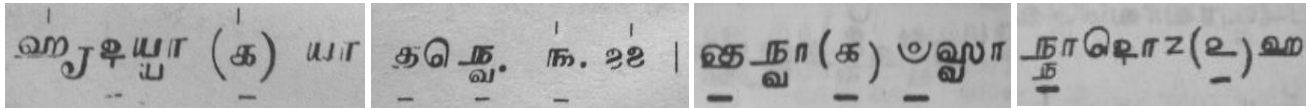
ref 9 p 3   ref 9 p 6   ref 5 p 733   ref 5 p 877

So generic support for svara marks applied to the unified Tamil-Grantha digits should be provided.

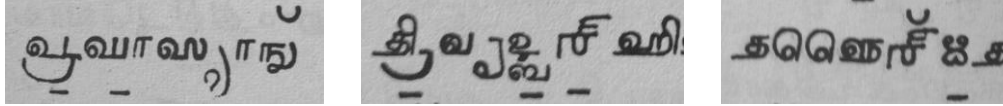<u>6.4.5. The Sama Vedic svara markers</u>

The Sama Vedic svara markers are all placed above their base and above the level of any anunasika sign (or ascenders of vowel signs). They can be applied to syllables and digits. For samples see §4.4. As the rules of Sama Veda svara markup are already complicated, it is best to generically provide support for rendering Sama Vedic svara markers as applied to syllables and digits without going into details like which markers are valid for syllables, which for digits etc. It should also be remembered that not all Sama Vedic svara markers are encoded in the Grantha block.

Another important thing about Sama Vedic svara markers is that they are used in various combinations with each other. In these combinations, they should be placed side-by-side and not stacked. The following list is a guide to the various possible combinations:

In the Arcika part of the Sama Veda:

1. 0 and 2 and KA are used in the combinations 2, 0, 2+0, ka, ka+2+0 and 2+ka.
2. Modern printings have *+2+0 and ka+2 instead of the last two.
3. The Tamil letter PA is used either alone or after 2 (as 2+PA).

In the Gana part of the Sama Veda:

1. All markers except the asterisk can occur alone.
2. The digit 0 can occur after all other markers.
3. One and only one marker from point 10a of the repertoire other than 0 can be followed by one and only one marker from point 10b/c. (Repertoire = §3.1.)
4. The digit 0 can occur after such a combination.

5.   The sequences 1-1, *-1-1, 1-1-PA and *-1-1-PA can occur.

Support for these combinations must be provided in the font and rendering engine. Generic support for placing other as-of-now unforeseen combinations of such marks may also be provided.

§7. Manipravalam

As said in the introduction, Grantha has solely been used for the purpose of writing Sanskrit. However, there exists the mixed language called Tamil Manipravalam in which Tamil and Sanskrit words are used mixedly. In it, Tamil words are written in the Tamil script and Sanskrit words in Grantha. Sometimes words are formed by Sanskrit roots taking Tamil suffixes. Even in such cases, the appropriate script is used for the appropriate part of the word.
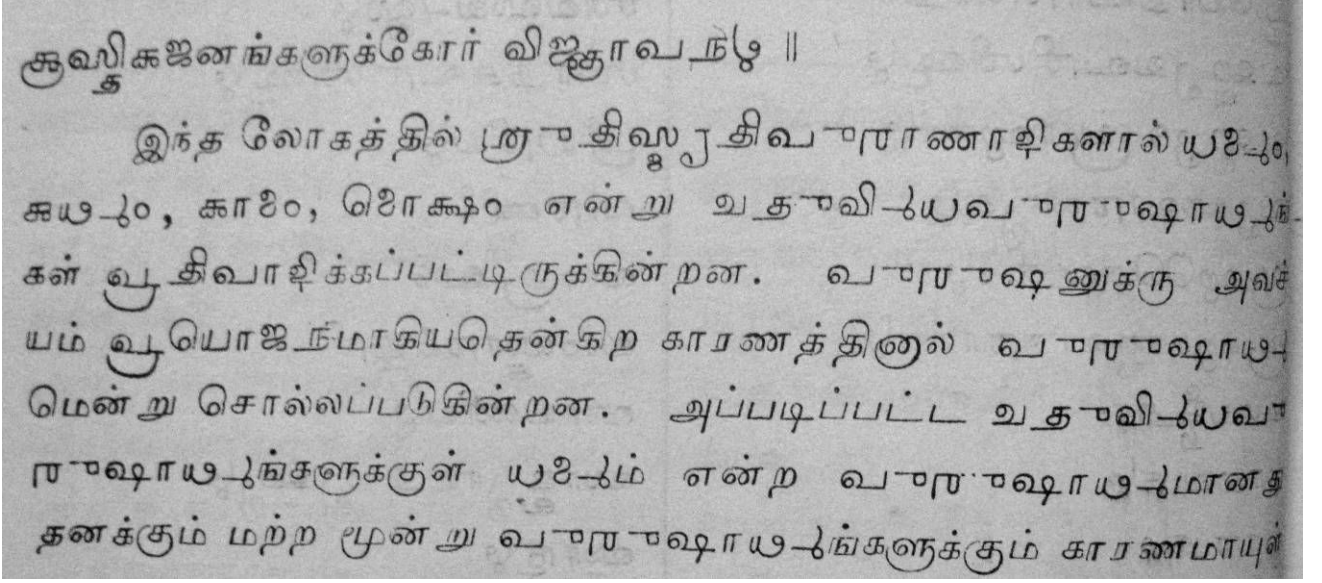
sample from p 24 of the introduction of ref 10 showing usage of Grantha and Tamil scripts for Manipravalam

Since Tamil does not use ligatures or conjoining forms for consonant clusters, even if there occurs a consonant cluster at such a Sanskrit-Tamil border within a word, the Sanskrit part is written as if the Tamil part did not exist and vice versa, meaning that the consonants take their respective final and initial forms. This does not cause any new issues in rendering since the Tamil-Grantha boundaries are clear and no Grantha characters take up Tamil behaviour or vice versa.

There is also a lesser-known case where the Tamil and Sanskrit scripts are used mixedly, where Grantha characters do take up some Tamil behaviour (but the converse is not true). This is shown in the sample on the top of the next page.

Here the language is totally Sanskrit and not Manipravalam. Grantha characters have been imported into the Tamil script to achieve phonetic accuracy in reading Sanskrit while keeping it easy for those who know Tamil but not Grantha.

Grantha characters are used only for the following: 1) independent vowels and vowel signs for vocalic R etc, 2) class consonants KA CA TTA TA and PA when they take vowel signs vocalic R etc, 3) all other class consonants. The rest is all as in Tamil.

Grantha consonants take a Grantha virama and Tamil consonants take a Tamil virama. Consonant clusters are as in Tamil with virama applied to all the vowelless consonants. For the

43
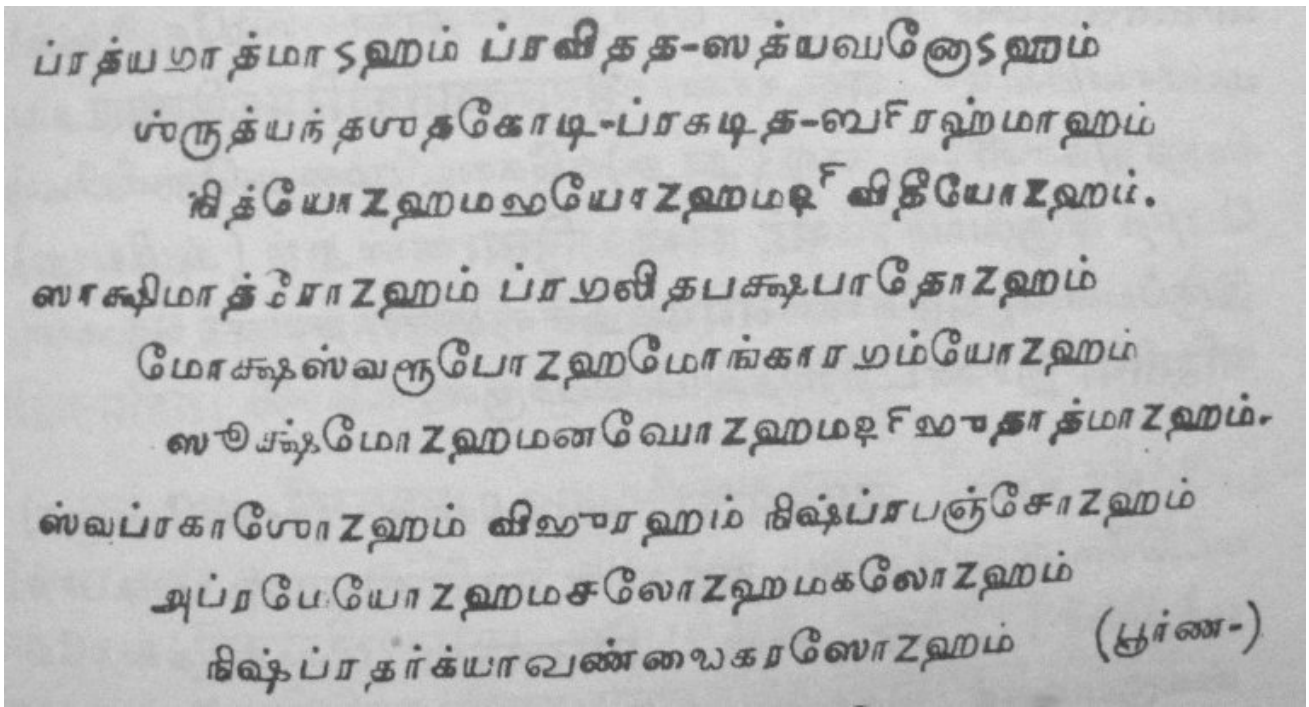
p 65, Śiva Mānasa Pūjā, Kīrtana-s and Ātma Vidyā Vilāsa of Śrī Sadāśiva Brahmendra, 1951, Kamakoti Koshasthanam, Chennai

most part, only spacing virama-s are used. The only ligature is K·SSA which is used in its Grantha variant though the (only slightly different) Tamil variant could also be used.

This particular sample falls short of phonetic accuracy since it uses the Tamil character NNNA to instead of NA in many places, after the nature of the Tamil language in which the NA does not occur as an unclustered consonant except in word-initial position. It also does not use the Grantha anusvara where one is needed. However this sample is indicative of the general principle of using Grantha characters to extend the Tamil script to denote Sanskrit and of how the Grantha and Tamil scripts have been freely mixed over the years.

There have also been printings where the Tamil virama (pulli) has been used for Grantha consonants as well, as in the following sample from Bhoja Charitram by T S Narayana Sastri, 1916, (page 26 of PDF, http://www.archive.org/stream/bhojacharitrama00sastgoog):

It is obvious that this 'pseudo-Manipravalam' writing will be useful as a bridge to true Grantha and therefore it also has scope for real-world use. The only new rendering support that is required is to apply the Tamil vowel signs EE, AI and OO to Grantha consonants. For the other Tamil-style vowel signs, the Grantha equivalents can be used with Grantha consonants since they appear the same. For Tamil-style vowel sign AU, the Grantha old-style vowel sign AU can be used.

It should however be considered whether the Grantha characters that are visually identical to Tamil characters should be provided canonical decompositions to them for security and identity in text search, since in Manipravalam these may be quite validly mixed up as seen above.

§8. 'Extended Grantha'

Though Grantha is traditionally used only for Sanskrit, if the use of Grantha to write Sanskrit is to be perpetuated, and newer works in Sanskrit (such as textbooks for secular subjects) are to be created, the script should be extended to support transliteration of some sounds not existing in Sanskrit but occurring in other-language names of people and places which may practically need to be mentioned in Sanskrit texts. Since in Tamil Nadu (where Grantha is used the most) words from (Indian) English and from Indian languages are the ones which practically need to be mentioned, characters must be added to Grantha to enable it to represent the sounds of these languages. It is however to be held in mind that such extending must be done in a way that will not affect the way Grantha represents Sanskrit, since Sanskrit is the native language for Grantha.

8.1. Nukta

First, we first propose the usage of the nukta just as is used in other Indic scripts for the purpose of extending the alphabet to denote non-native sounds. The suggested representative glyph is two dots placed below the consonant but above the level of the anudatta sign. Some usage examples are shown on the right.

ஜ஼ �'t'஼ வ஼

za fa wa

If a consonant with a nukta occurs in a consonant cluster, it is preferred that it should not participate in stacking in order to highlight its foreign nature (and since it would be a further complication for rendering engines). The explicit vowelless forms are to be used if needed.

It may be suggested that 0CBC KANNADA SIGN NUKTA be used for this purpose, but we request the encoding of a separate nukta for Grantha because the two-dots-below shape we have suggested is purely arbitrary. It is possible that the Grantha community would like to use a different shape for the nukta in the future and in which case a request for change of representative glyph may be submitted. Grantha users would not like to be "locked in" to the Kannada form of the nukta which is decided on by a different community. If they are forced to use the Kannada character, it would unnecessarily necessitate the conversion of data to the new separate Grantha nukta. Therefore we feel it is advisable to provide a separate nukta for Grantha right now at the point of encoding.

Encoding a separate nukta is also called for to to complete the set of consonant-modifier signs, just as the anunasika sign, Tamil vowel signs AA I II etc are separately encoded for Grantha. This will also ensure that implementers will not overlook having to provide support for the nukta.

8.2. Chandra

The chandra sign used for svarita can be used as a modifier on Grantha vowels EE and AA to denote the English-style long vowels [æ:] and [ɔ:].

ஹ்ˮ கூˮ பெ்ˮட கூாˮட

8.3. Short Vowels E and O

In common with the previous proposal for encoding Grantha submitted by Naga Ganesan (L2/09-141), we support the encoding of four new characters, two independent vowels and two vowel signs for the short vowels E and O common in South Indian languages and also in some North Indian languages such as Bhojpuri. The short vowel E is also common in Indian English.

Here, we give brief remarks about what these vowels are intended for and also what they are *not* intended for.

8.3.1. Transliteration of Tamil, not a valid case

The transliteration of Tamil is not a valid cause to encode short E and O in Grantha, since the Tamil and Grantha scripts are so similar and the mixed usage of those scripts has been attested previously in the case of Manipravalam. Anybody who can read Grantha can reasonably be expected to read Tamil (though the converse is not true). Therefore Tamil words which need to be inserted in Sanskrit text written in Grantha can be written in their own native script.

Further, Grantha is a phonetic script whereas Tamil is a phonemic script. Tamil class consonant letters denote both voiced and voiceless stops whereas the Grantha ones denote only one sound each. Tamil vowel sign U denotes both normal [u] and 'shortened U' whereas Grantha denotes only normal [u]. Thus Grantha is a bad representative of Tamil phonemes and so Tamil written in Grantha would be highly awkward to read. There can be no valid case for writing Tamil in Grantha, especially seeing as most, if not all, Grantha users can read Tamil also.

8.3.2. Transliteration of Malayalam, a partially valid case

Malayalam is a mixed phonetic-phonemic script, because the Malayalam language has two kinds of words – Sanskrit-based and Tamil-based. Thus going by the previous arguments, and since Malayalam is not as similar to Grantha as Tamil, Malayalam should properly be transliterated using a mixture of Tamil and Grantha, much like in Manipravalam. We give an example:

ശ്രീരാമചരിതത്തെ   അടിസ്ഥാനമാക്കി   ആത്മതത്ത്വത്തെക്കുറിച്ചു   വിവരിക്കുന്നതിനാൽ

അദ്ധ്യാത്മരാമായണം എന്നു പറയുന്നു.

ஸ்ரீராஜசரிதத்தெ   அடிஷ்டாநமாக்கி   ஆதததூத்தெக்குறிச்சு   விவரிக்குந்நதிலுல்

ஸுஜுாதூராஉயணழ எந்நு பறயுந்நு.

Note that this still does not require the encoding of any new characters in Grantha.

8.3.3. Transliteration of Kannada and Telugu, a fully valid case

Kannada and Telugu, while totally phonetic scripts, do have the short vowels E and O, and to transliterate these, one requires Grantha characters for those vowels, both independent and dependent. The shapes proposed for these characters are the same as those of the long vowels with a "pulli" added on top. The Tamil script formerly marked its short vowels thus (as shown below from ref 12 pp 1 and 2) and the same would be appropriate for Grantha today.

Tamizh	அ	ஆ	இ	௱ா	உ	௨ா					எ	எ	ஜ	௧	௭	௧ள
Tamizh	க	கா	கி	கீ	கு	கூ					கே	கெ	கை	கோ	கொ	கௌ

The "pulli" can be in the form of a dot or a ring. While the ring form is shown above, the dot form is also seen as in the following passage from page 52 of ref 13:

எனவே ஏகார ஓகாரங்கட்குப் புள்ளி
எ-ஒ-என வரும்.
இஃது உயிர்மெய்க்கும் ஒக்கும்.

The choice between the dot and ring forms is a matter of personal taste. We prefer the ring form.

Examples of transliterating Kannada and Telugu names of towns and villages using these new characters are given below:

ಎಡ್ಪಾಡು   ഞ॒ಬ್ಬೆಾಟಾ   ಒಂಗೋಲು   ഀಂ ಒஈ஌

ಆಗುಂಬೆ   ஆ஗ுஈஂ஖ெ   ಹೊಸನಗರ   ஌஑ாஸஶஹஶ

Of course, words from Indian English with short E would also be transcribed in this way.

8.4. Dravidian Consonants LLLA and RRA

As clarified above, among Dravidian languages, Grantha will only be used for transliteration of Kannada and Telugu. Both these languages have between them the Dravidian consonants LLLA and RRA, though their usage is rare. If at all it is required to denote them, the corresponding Tamil characters with Tamil behaviour should be used as is from the Tamil block. As shown above in §7, Tamil and Grantha characters are known to be used mixedly, so this is merely another such usage. Using the nukta is also an option.

If such consonants occur as part of clusters, some users prefer not to see Grantha-style stacking to highlight the foreign nature of these characters. ಹಟಗನ್ನಡ   ஔஶஉஶஎ஖ಲ

Others prefer stacking to maintain visual uniformity. ಗುಟ್ಟಮು   ஈாஂ஋஍ஶ஌/ஈாீ஍஌ To satisfy both usage styles, the Grantha virama may be used if stacking is desired and the Tamil virama used otherwise. Rendering engines should then provide appropriate support for the usage of Grantha virama with Tamil consonants.

By using the existing Tamil characters, no characters are disunified.

§9. Encoding in Unicode

9.1. Name of the Block

As we mentioned at the outset, in this proposal we have concerned ourselves only with the repertoire and orthography of the "modern" version of the Grantha script. Whether historic versions of the script may be disunified from the modern version or not is not discussed in this proposal. However, even if earlier versions are desired to be disunified, this block may still be named merely "Grantha" without any adjectives as this is the name by which both native users and international scholars (like Dr Gerhard Ehlers of Germany whom we have repeatedly consulted for this proposal) refer to the contemporary form of the script. Historic versions may be qualified by adjectives as in "Pallava Grantha" etc.

The name 'Tamil Grantha' has been suggested, but the adjective 'Tamil' is misleading as it is not clear what 'Tamil' is supposed to mean. It certainly cannot refer to the Tamil language. If it is suggested that it refers to the current form of the script which is similar to Tamil in many ways, we point out that ancient versions of the Tamil script may also be considered similar to ancient versions of Grantha, so the adjective does not really qualify the noun. Native users of Grantha also express distaste at labeling the script 'Tamil Grantha' and feel that it would amount to ignoring the individuality of the script.

Therefore we request for the block to be named merely 'Grantha'.

9.2. Placement in the SMP

Grantha has been allocated the block 11300-1137F in the roadmap of the SMP. While the Unicode Standard 5.0 ch 2.8 says that the SMP is used for 'historic' or 'invented' scripts as against 'modern' scripts, Grantha is by no means a 'historic script' in the sense that it is not 'not used today' and it is certainly not 'invented'. However, it is recognized that the presumed distinction between the BMP and SMP in being used for 'modern' versus 'ancient or invented' scripts no longer exists and that the descriptions of the BMP and SMP in the Unicode Standard are due for an update.

Therefore we do not object to encoding the Grantha script in the SMP, especially seeing that there is no contiguous space for comfortably encoding the entire script in the BMP. We will therefore follow the roadmap.

All the same, we wish it to be verbally recorded in the description of the Grantha script in any future version of the Unicode Standard that "the Grantha script is still in contemporary use in parts of South Asia, especially Tamil Nadu, for the writing of Sanskrit" or equivalent text which will convey clearly and prominently that the script still lives and is not extinct.

9.3. The assignment of code-points

The Grantha script will as far as possible be encoded isomorphically with the other Indic scripts encoded in the BMP. Thus the independent vowels, consonants, dependent vowel signs (including virama), the anunasika sign, anusvara, visarga, nukta and avagraha are all placed isomorphically to Devanagari (and hence other scripts as far as applicable). The old- and new-style dependent vowel signs for AU are placed at 1134C and 11357 parallel to Malayalam 0D4C and 0D57.

The codepoints for the danda-s are reserved with a recommendation to use the generic Indic danda-s at 0964 and 0965 to maintain uniformity with all Indic scripts. The codepoint corresponding to OM is reserved for future encoding if any unique Grantha OM is found attested.

The superscript digits 0 to 6 used as Sama Vedic svara markers are also placed at 11366 to 1136C parallel to the normal digits from the Tamil block 0BE6 to 0BEC, though this is not vital.

The characters which are separately encoded for technical reasons are the ligating virama and the old-style non-spacing vowel signs for Vocalic L and LL. The ligating virama is placed next to the default virama for visibility. The vowel signs are arbitrarily placed at the end of the block.

The remaining characters have no equivalents in the other Indic blocks and are hence laid out as per convenience. Some unused codepoints with equivalents in other blocks are reserved.

§10. Collation

As with all Indic scripts:

<p align="center">INDEPENDENT VOWELS < CONSONANTS < DEPENDENT VOWEL SIGNS</p>

The order of independent vowels (as in all other Indic scripts) is:

<p align="center">A < AA < I < II < U < UU < VOC R < VOC RR < VOC L < VOC LL < E < EE < AI < O < OO < AU</p>

That of the dependent vowel signs AA etc is the same with VIRAMA being last. This means:

<p align="center">KA + KA < KA + AA-SIGN + KA < KA + I-SIGN + KA < ... < KA + VIRAMA + KA</p>

Important: Note that the old and new-style vowel signs for AU should be treated equivalently for the purposes of collation. It should also be remembered that the old-style vowel sign AU is

decomposable, and the new-style vowel sign AU forms part of the decomposition. Further, since the ligating virama decomposes to the default virama and since the old-style vowel signs for Vocalic L and Vocalic LL decompose to the new-style ones, these will automatically be treated equal for collation.

The consonants are ordered as in the code-chart with the only exception being LLA:

$$\text{KA} < \text{KHA} < ... < \text{LA} < \text{VA} < ... < \text{HA} < \text{LLA}$$

No exception is made for K·SSA and J·NYA. That is, they are not considered consonants in their own right, and are hence sorted like every other consonant cluster.

Consonants with nukta should be treated as a single unit and come after 'pure' consonants. Among consonants with nukta, the base consonant should determine sorting. Thus:

$$\text{KA} < ... < \text{LLA} < \text{KA} + \text{Nukta} < \text{KHA} + \text{Nukta} < ... < \text{LLA} + \text{Nukta}$$

Further:

$$\text{Anusvara} < \text{Vedic Anusvara} < \text{Vedic Double Anusvara} < \text{Visarga} < \text{Ardhavisarga} < \text{Consonants}$$

Which may be reduced to:

$$\text{KA} < \text{KA} + \text{Anusvara-Family} < \text{KA} + \text{Visarga-Family} < \text{KA} + \text{KA}$$

The avagraha also sorts before the consonants but after anusvara etc, so:

$$\text{KA} + \text{Avagraha} + \text{KA} < \text{KA} + \text{KA}$$

The pluta sign should be sorted after the consonants, so:

$$\text{KA} + \text{AA-Sign} < \text{KA} + \text{AA-Sign} + \text{KA} < \text{KA} + \text{AA-Sign} + \text{Pluta} < \text{KA} + \text{AA-Sign} + \text{Pluta} + \text{KA}$$

This may probably be summarized as:

$$\text{Independent Vowels} < \text{Anusvara Etc} < \text{Avagraha} < \text{Consonants} < \text{Pluta Sign} / \text{Vowel Signs}$$

Now we turn to the nonspacing combining marks. First, with the anunasika sign:

$$\text{Base} < \text{Base} + \text{Anunasika}$$

Non-nasal syllables (or consonants Y L and V) should sort before the nasal counterparts. Thus:

$$\text{KA} + \text{KA} < \text{KA} + \text{Anunasika} + \text{KA}$$

$$\text{YA} + \text{Virama} + \text{YA} < \text{YA} + \text{Virama} + \text{Anunasika} + \text{YA}$$

In the Rig Veda, Yajur Veda and the Arcika part of the Sama Veda, the various svara markers should be reduced to their underlying svara and then sorting must be done in the order:

$$\text{Udatta} < \text{Svarita} < \text{Anudatta} < \text{Pracaya}$$

In Rig and Yajur Veda texts in Grantha, the udatta and pracaya are not distinguished, but these are self-sorting due to the grammatical rules behind their usage. In the Sama Veda Arcika, distinction exists in markup so identification is not a problem. Further, between the two Yajur Vedic svarita-s:

$$\text{Svarita} < \text{Dirgha Svarita}$$

Among the Sama Vedic svarita-s:

$$\text{Ordinary Svarita} < \text{Independent Svarita} < \text{Kampa Svarita}$$

As regards the Gana part of the Sama Veda, the svara marker sequences *-1-1 and 1-1 both form a single collation-unit and have equal weight. (Effectively, the * should be disregarded.) Apart from this, the general rule is that each additional combining mark increases sorting weight. The markers of the seven svara-s sort in descending order of the tone denoted:

$$11 < 1 < 2 < 3 < 4 < 5 < 6$$

Special care should be taken when handling Sama Vedic svara markers since the rules governing them are quite complicated, as described in §4.4.1 and §4.4.2. Svara markers are often omitted and must be inferred from preceding syllables. The rules however provide for unambiguously determining which syllable carries which svara when they are understood and applied properly.

Further, for the special Sama Vedic svara markers A, VI, NA, Tamil PA, ^ and ⁻, no specific order is defined. From what we were able to learn from Sama Vedic scholars, it is highly unlikely that these markers determine order. If at all required, sorting may be done by codepoint!

Collation of the unified Grantha-Tamil numerals should be done by the numeric value. Old-style numerals are written using the separate characters for 10, 100 and 1000 whereas new-style numerals use the place value system. This should be taken into account.

NOTE: Many of these rules are based on Sanskrit collation tradition, Sanskrit grammar and the general expectations of Sanskrit/Vedic scholars using Grantha about how sorting should be done. We have tried to provide for all theoretical possibilities, though probably not all will be seen. These rules are not documented anywhere else but compiled by us in consultation with scholars.

§11. Linebreaking

Linebreak is not allowed before the pluta sign, though it is given the category Lo.

Printed Vedic texts show linebreaks before the Vedic anusvara-s of Grantha, so it is not required to prevent linebreak before them (unlike in N3383 §3 where it is required to prevent linebreak before the glyphically different Devanagari anusvara-s proposed there).

Linebreak is allowed before or after an avagraha but not between two avagraha-s.

The question of linebreak before combining marks does not arise.

The remaining – independent vowels and consonants – may be treated as in Devanagari.

§12. Unicode Character Properties

The properties are given here on parallel with those of characters from other Indic blocks. The Sama Vedic svara markers are not given digit or numeric values since they are not intended as numbers (and will never be used as such) but only as svara markers.

```
11301;GRANTHA SIGN ANUNASIKA;Mn;0;NSM;;;;;N;;;;;
11302;GRANTHA LETTER ANUSVARA;Mc;0;L;;;;;N;;;;;
11303;GRANTHA LETTER VISARGA;Mc;0;L;;;;;N;;;;;
11305;GRANTHA LETTER A;Lo;0;L;;;;;N;;;;;
11306;GRANTHA LETTER AA;Lo;0;L;;;;;N;;;;;
11307;GRANTHA LETTER I;Lo;0;L;;;;;N;;;;;
11308;GRANTHA LETTER II;Lo;0;L;;;;;N;;;;;
11309;GRANTHA LETTER U;Lo;0;L;;;;;N;;;;;
1130A;GRANTHA LETTER UU;Lo;0;L;;;;;N;;;;;
1130B;GRANTHA LETTER VOCALIC R;Lo;0;L;;;;;N;;;;;
1130C;GRANTHA LETTER VOCALIC L;Lo;0;L;;;;;N;;;;;
1130E;GRANTHA LETTER E;Lo;0;L;;;;;N;;;;;
1130F;GRANTHA LETTER EE;Lo;0;L;;;;;N;;;;;
11310;GRANTHA LETTER AI;Lo;0;L;;;;;N;;;;;
11312;GRANTHA LETTER O;Lo;0;L;;;;;N;;;;;
11313;GRANTHA LETTER OO;Lo;0;L;;;;;N;;;;;
11314;GRANTHA LETTER AU;Lo;0;L;;;;;N;;;;;
11315;GRANTHA LETTER KA;Lo;0;L;;;;;N;;;;;
11316;GRANTHA LETTER KHA;Lo;0;L;;;;;N;;;;;
```

```
11317;GRANTHA LETTER GA;Lo;0;L;;;;;N;;;;;
11318;GRANTHA LETTER GHA;Lo;0;L;;;;;N;;;;;
11319;GRANTHA LETTER NGA;Lo;0;L;;;;;N;;;;;
1131A;GRANTHA LETTER CA;Lo;0;L;;;;;N;;;;;
1131B;GRANTHA LETTER CHA;Lo;0;L;;;;;N;;;;;
1131C;GRANTHA LETTER JA;Lo;0;L;;;;;N;;;;;
1131D;GRANTHA LETTER JHA;Lo;0;L;;;;;N;;;;;
1131E;GRANTHA LETTER NYA;Lo;0;L;;;;;N;;;;;
1131F;GRANTHA LETTER TTA;Lo;0;L;;;;;N;;;;;
11320;GRANTHA LETTER TTHA;Lo;0;L;;;;;N;;;;;
11321;GRANTHA LETTER DDA;Lo;0;L;;;;;N;;;;;
11322;GRANTHA LETTER DDHA;Lo;0;L;;;;;N;;;;;
11323;GRANTHA LETTER NNA;Lo;0;L;;;;;N;;;;;
11324;GRANTHA LETTER TA;Lo;0;L;;;;;N;;;;;
11325;GRANTHA LETTER THA;Lo;0;L;;;;;N;;;;;
11326;GRANTHA LETTER DA;Lo;0;L;;;;;N;;;;;
11327;GRANTHA LETTER DHA;Lo;0;L;;;;;N;;;;;
11328;GRANTHA LETTER NA;Lo;0;L;;;;;N;;;;;
1132A;GRANTHA LETTER PA;Lo;0;L;;;;;N;;;;;
1132B;GRANTHA LETTER PHA;Lo;0;L;;;;;N;;;;;
1132C;GRANTHA LETTER BA;Lo;0;L;;;;;N;;;;;
1132D;GRANTHA LETTER BHA;Lo;0;L;;;;;N;;;;;
1132E;GRANTHA LETTER MA;Lo;0;L;;;;;N;;;;;
1132F;GRANTHA LETTER YA;Lo;0;L;;;;;N;;;;;
11330;GRANTHA LETTER RA;Lo;0;L;;;;;N;;;;;
11332;GRANTHA LETTER LA;Lo;0;L;;;;;N;;;;;
11333;GRANTHA LETTER LLA;Lo;0;L;;;;;N;;;;;
11335;GRANTHA LETTER VA;Lo;0;L;;;;;N;;;;;
11336;GRANTHA LETTER SHA;Lo;0;L;;;;;N;;;;;
11337;GRANTHA LETTER SSA;Lo;0;L;;;;;N;;;;;
11338;GRANTHA LETTER SA;Lo;0;L;;;;;N;;;;;
11339;GRANTHA LETTER HA;Lo;0;L;;;;;N;;;;;
1133C;GRANTHA SIGN NUKTA;Mn;7;NSM;;;;;N;;;;;
1133D;GRANTHA SIGN AVAGRAHA;Lo;0;L;;;;;N;;;;;
1133E;GRANTHA VOWEL SIGN AA;Mc;0;L;;;;;N;;;;;
1133F;GRANTHA VOWEL SIGN I;Mc;0;L;;;;;N;;;;;
11340;GRANTHA VOWEL SIGN II;Mn;0;NSM;;;;;N;;;;;
11341;GRANTHA VOWEL SIGN U;Mc;0;L;;;;;N;;;;;
11342;GRANTHA VOWEL SIGN UU;Mc;0;L;;;;;N;;;;;
11343;GRANTHA VOWEL SIGN VOCALIC R;Mc;0;L;;;;;N;;;;;
11344;GRANTHA VOWEL SIGN VOCALIC RR;Mc;0;L;;;;;N;;;;;
11346;GRANTHA VOWEL SIGN E;Mc;0;L;;;;;N;;;;;
11347;GRANTHA VOWEL SIGN EE;Mc;0;L;;;;;N;;;;;
11348;GRANTHA VOWEL SIGN AI;Mc;0;L;;;;;N;;;;;
1134A;GRANTHA VOWEL SIGN O;Mc;0;L;11346 1133E;;;;N;;;;;
1134B;GRANTHA VOWEL SIGN OO;Mc;0;L;11347 1133E;;;;N;;;;;
1134C;GRANTHA VOWEL SIGN AU OLD;Mc;0;L;11347 11357;;;;N;;;;;
1134D;GRANTHA SIGN VIRAMA;Mc;9;L;;;;;N;;;;;
1134E;GRANTHA SIGN LIGATING VIRAMA;Mn;9;NSM;1134D;;;;N;;;;;
11357;GRANTHA VOWEL SIGN AU NEW;Mc;0;L;;;;;N;;;;;
1135D;GRANTHA SIGN PLUTA;Lo;0;L;;;;;N;;;;;
1135E;GRANTHA LETTER VEDIC ANUSVARA;Lo;0;L;;;;;N;;;;;
1135F;GRANTHA LETTER VEDIC DOUBLE ANUSVARA;Lo;0;L;;;;;N;;;;;
11360;GRANTHA LETTER VOCALIC RR;Lo;0;L;;;;;N;;;;;
11361;GRANTHA LETTER VOCALIC LL;Lo;0;L;;;;;N;;;;;
11362;GRANTHA VOWEL SIGN VOCALIC L;Mc;0;L;;;;;N;;;;;
11363;GRANTHA VOWEL SIGN VOCALIC LL;Mc;0;L;;;;;N;;;;;
11366;COMBINING GRANTHA DIGIT ZERO;Mn;230;NSM;;;;;N;;;;;
11367;COMBINING GRANTHA DIGIT ONE;Mn;230;NSM;;;;;N;;;;;
11368;COMBINING GRANTHA DIGIT TWO;Mn;230;NSM;;;;;N;;;;;
11369;COMBINING GRANTHA DIGIT THREE;Mn;230;NSM;;;;;N;;;;;
1136A;COMBINING GRANTHA DIGIT FOUR;Mn;230;NSM;;;;;N;;;;;
1136B;COMBINING GRANTHA DIGIT FIVE;Mn;230;NSM;;;;;N;;;;;
1136C;COMBINING GRANTHA DIGIT SIX;Mn;230;NSM;;;;;N;;;;;
```

```
11370;COMBINING GRANTHA LETTER A;Mn;230;NSM;;;;;N;;;;;
11371;COMBINING GRANTHA LETTER KA;Mn;230;NSM;;;;;N;;;;;
11372;COMBINING GRANTHA LETTER NA;Mn;230;NSM;;;;;N;;;;;
11373;COMBINING GRANTHA SYLLABLE VI;Mn;230;NSM;;;;;N;;;;;
11374;COMBINING TAMIL LETTER PA;Mn;230;NSM;;;;;N;;;;;
1137E;GRANTHA VOWEL SIGN VOCALIC L OLD;Mn;0;NSM;11362;;;;N;;;;;
1137F;GRANTHA VOWEL SIGN VOCALIC LL OLD;Mn;0;NSM;11363;;;;N;;;;;
```

§13. Acknowledgments

And finally, but most important of all, a very special thank you to Rick McGowan for all the guidance, help and patience shown by him throughout our drafting and submission process.

To anyone whom I have omitted to acknowledge above, my apologies and since thanks.

§14. References

1. Epigraphy – Inscriptions in Grantha Script, website of the Dept of Archaeology, Tamil Nadu Govt, http://www.tnarch.gov.in/epi/ins3.htm, retrieved 2009-Sep-09.

2. Grantha Bāla Pāṭhamālā, Heritage India Education Trust, Chennai, India, 2003.

3. A primer in Grantha characters, Venugopalan, K, St. Peter, Minn: James H. Nye, 1983. http://dsal.uchicago.edu/digbooks/dig_toc.html?BOOKID=PK419.V468_1983, retrieved 2009-Sep-09.

4. South Indian Scripts in Sanskrit Manuscripts and Prints, Reinhold Grünendahl, Wiesbaden, Germany, 2001. ISBN 3-447-04504-3.

5. Taittirīya Brāhmaṇam, Heritage India Education Trust, Chennai, India, 1980.

6. Sāma Veda Saṃhitā, AVMSSGDPS Samithi, Kumbakonam, Tamil Nadu, India, 1985.

7. Taittirīya Saṃhitā, Heritage India Education Trust, Chennai, India, 1980.

8. A Monograph on the Anusvara of the Taittiriya Krishna Yajur Veda, Shriramana Sharma, http://sanskritweb.net/sansdocs/anusvara.pdf, retrieved 2009-Sep-09.

9. Āśvalāyana Śrāddha Prayoga, AVMSSGDPS Samithi, Kumbakonam, Tamil Nadu, 1985.

10. Smṛti Muktāphala – Varṇāśrama Dharma Kāṇḍa, Sharada Vilasa Press, Kumbakonam, Tamil Nadu, 1939.

11. Clarification and Consolidation of the Function of Zero Width Joiner in Indic Scripts, Peter Constable, http://unicode.org/review/pr-37.pdf, retrieved 2009-Sep-09.

12. Ancient and Modern Alphabets of the Popular Hindu Alphabets of the Southern Peninsula of India, Capt Henry Harkness, 1837, Royal Asiatic Society, http://www.archive.org/details/ancientmodernalp00harkrich, retrieved 2009-Sep-09.

13. Tolkāppiyam Eḻuttatikāram, Tirumakaḷ Aḻuttakam, Cuṉṉāgam, Śrī Laṅkā, 1937, http://noolaham.org/wiki/index.php/தொல்காப்பியம்_எழுத்ததிகாரம், retrieved 2009-Sep-09.

14. A Descriptive Catalogue of Manuscripts in the Central Library Baroda. Vol. I Vedic. Compiled by Gajanan K. Shrigondekar and K.S. Ramaswami Shastri Siromani. Nos. 9851 a-f. Baroda 1925.

15. Samskiruta Mutaṟ Puttakam, Civappirakāca Paṇṭitar, Cōtiṭappirakāca Yantira Cālai, Kokku, Śrī Laṅkā, 1991.

16. Sadyojata Samskrita Shiksha - Infant Reader, Sadyojata Samskrita Prakashana, Cuddalore, Tamil Nadu, 2001.

17. Śrī Hālāsya Māhātmyam, http://ambaa.org/pdf/halasya_mahatmyam.pdf, (inquiries as to further publication details of this out-of-copyright book not fruitful)

18. Vyāsa Śikṣā with Veda Taijasa commentary, Brahmananda Press, Tiruvadi, 1908.

19. Taittirīya Saṃhitā Pada Pāṭha (in three parts), Mylapore Veda Adhyayana Sabha, Chennai, 1996.

Dependency chart

This proposal document also relates and refers to other documents that have been submitted to the UTC. It depends for its fulfillment in the full proper representation of *Grantha* on some other proposals for non-Grantha-specific characters. These documents are listed below after the style of a computer software package dependency chart. They are all authored by the present author except when otherwise mentioned.

Depends:

L2/09-344: Request for encoding 1CF4 VEDIC TONE CANDRA ABOVE

L2/09-___: Request for encoding Tamil Major Fractions

Recommends:

L2/09-343: Request for encoding 1CF3 ROTATED ARDHAVISARGA

L2/09-339: Request for normative aliases for some generic Indic characters

L2/09-___: Request for encoding 11355 GRANTHA LENGTH MARK

Conflicts:

L2/09-141: Proposal to encode the Grantha Script, Naga Ganesan

Other references:

L2/09-298: Why the word "nihshvasa" in 1CD3 VEDIC SIGN NIHSHVASA is inappropriate

L2/09-316: Comments on Mr Ganesan's Grantha Proposal

L2/09-324: Comments on Kent Karlsson's document L2/09-277

L2/09-342: Misrepresentation in Unicode of characters related to the Sanskrit sounds jihvamuliya and upadhmaniya

L2/09-___: Why jihvamuliya and upadhmaniya characters are *not* proposed for Grantha

L2/09-___: Further clarifications on Grantha virama ligatures

* * *

§15. Official proposal summary form

A. Administrative

1. Title

   Proposal to encode the Grantha script in Unicode

2. Requester's name

   Shriramana Sharma (jamadagni-at-gmail-dot-com), India

3. Requester type (Member body/Liaison/Individual contribution)

   Individual Contribution

4. Submission date

   2009-Oct-24

5. Requester's reference (if applicable)

6. Choose one of the following:

6a. This is a complete proposal

   Yes

B. Technical – General

1. Choose one of the following:

1a1. This proposal is for a new script (set of characters)

   Yes

1a2. Proposed name of script

   Grantha

1b1. The proposal is for addition of character(s) to an existing block

1b2. Name of the existing block

2. Number of characters in proposal:

   90

3. Proposed category:

   Category A – Contemporary, or Category B1 – Specialized Small Collections

4a. Is a repertoire including character names provided?

   Yes

4b. If YES, are the names in accordance with the "character naming guidelines" in Annex L of P&P document?

   Yes

4c. Are the character shapes attached in a legible form suitable for review?

   Yes

5a. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for publishing the standard?

   Elmar Kniprath (kniprath-at-online-dot-de), Germany, TrueType

5b. If available now, identify source(s) for the font and indicate the tools used:

6. References:

6a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?

Yes

6b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?

> *Yes*

7. Special Encoding Issues: Does the proposal address other aspects of character data processing (if applicable) such as input, presentation, sorting, searching, indexing, transliteration etc. (if yes please enclose information)?

> *Yes, see detailed proposal.*

8. Additional Information:

> *Additional information about properties of the proposed characters and script that will assist in their correct understanding of and correct linguistic processing are provided in the detailed proposal.*

C. Technical – Justification

1. Has this proposal for addition of character(s) been submitted before? If YES, explain.

> *Yes. A proposal for Grantha was submitted by Naga Ganesan (naa -dot ganesan -at gmail -dot com) as L2/09-141, but our proposal was independently developed and differs from that proposal in some important ways. We have submitted to the UTC and also publicly posted to the indic@unicode.org list a document outlining the problems with Naga Ganesan's proposal, L2/09-316.*

2a. Has contact been made to members of the user community (for example: National Body, user groups of the script or characters, other experts, etc.)?

> *Yes*

2b. If YES, with whom?

> *Shriramana Sharma is part of the user community and is in daily contact with people, especially the teachers and students of traditional Vedic schools and other Vedic/Sanskrit scholars who use the script for teaching and studying Sanskrit (classical and Vedic) and for religious purposes. The following Grantha users who are all accomplished Vedic scholars were consulted:*
>
> 1. *Shri Svami Atma Bodha Tirtha, who is a Sama Veda scholar and before taking to holy orders was the principal of the Raja Veda Pathashala, Kumbakonam, Tamil Nadu, one of the oldest thriving pathashala-s of India established in the 15th century CE.*
>
> 2. *Mahamahopadhyaya Dr R Krishnamurti Shastri, who is an expert in the Yajur Veda, Vedanta and traditional commentaries on the Veda-s; retired principal of Madras Sanskrit College, Chennai; well known Vedic/Shastric/Sanskrit scholar in Tamil Nadu and at the Indian national level; chief trustee of Heritage India Education Trust which has been publishing works in Grantha for over twenty years. (See also: an Indian newspaper article on Dr Shastri: http://www.thehindu.com/fr/2009/02/06/stories/2009020651000300.htm.)*
>
> 3. *Dr R Mani Dravid, an expert in Vedanta and Yajur Veda scholar, lecturer of Mimamsa at Madras Sanskrit College, Chennai, and President's Award (Badarayana Puraskar) awardee.*
>
> 4. *Shri Jambunatha Ghanapathi, expert in Yajur Veda and Vedanga-s, principal and managing trustee, Shri CCVV Trust Pathashala, Coimbatore, Tamil Nadu.*
>
> 5. *Shri Shrikrishna Shrauti, expert in Sama Veda and traditional ritualistic procedures, Department of Sama Veda Bhashyam, Shri Venkateshvara Vaidika University, Tirupati, Andhra Pradesh.*

6. *Shri Venugopala Ghanapathi, Rig Veda teacher at Kanchi Shankara Matham, Kumbakonam.*

7. *Shri Parameshvara Dikshitar, scholar of Yajur Veda and Atharva Veda, Chidambaram, Tamil Nadu*

8. *Dr P Sambandham Gurukkal, Shaiva Agama Research Assistant, French Institute for Oriental Research, Pondicherry and founder of Sadyojata Samskrita Prakashana, Cuddalore, Tamil Nadu, which has published a Grantha infant reader and also yearly religious almanacs.*

9. *Dr K Ramasubramanian, Assistant Professor, Department of Humanities and Social Sciences, IIT Mumbai, scholar of the Yajur Veda and various Shastra-s*

10. *many peers of Shriramana Sharma, all part of the Tamil Nadu Vedic scholar community.*

The following Grantha users who hold engineering degrees were also consulted regarding many aspects, especially technical: Vinodh Rajan, Sri Ramadoss M, both of Chennai.

Dr Gerhard Ehlers of the Orientabteilung, Staatsbibliothek zu Berlin, expert in manuscript cataloguing and author of many works on manuscripts also was consulted. (His online page at: http://www.geschkult.fu-berlin.de/e/indologie/mitarbeiter/lehrbeauftragte/ehlers/index.html.)

2c. If YES, available relevant documents:

Appended endorsement of the proposal by Dr R Krishnamurti Shastri as well as words of approval from Dr Gerhard Ehlers.

3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included?

Yes, see detailed proposal

4a. The context of use for the proposed characters (type of use; common or rare)

Used for writing Sanskrit and Sanskrit words of Tamil Manipravalam.

4b. Reference:

5a. Are the proposed characters in current use by the user community?

Yes

5b. If YES, where?

Chiefly in Tamil Nadu but also in other parts of India, Sri Lanka and elsewhere.

6a. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP?

Yes, but as there is insufficient space in the BMP, we agree to encode in the SMP.

7. Should the proposed characters be kept together in a contiguous range?

Yes

8a. Can any of the proposed characters be considered a presentation form of an existing character or character sequence?

No

9a. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters?

No.

10a. Can any of the proposed character(s) be considered to be similar (in appearance or function) to an existing character?

Some letters and vowel signs are similar to those in the Tamil block. The nukta is similar to that in Kannada.

10b. If YES, is a rationale for its inclusion provided?

Removal of those characters would leave subsets of characters in the Grantha block incomplete. Precedent from other Indic and European blocks allows for the separate encoding of characters which are needed to complete subsets like consonants, combining marks etc.

11a. Does the proposal include use of combining characters and/or use of composite sequences (see clauses 4.12 and 4.14 in ISO/IEC 10646-1: 2000)?

No

12a. Does the proposal contain characters with any special properties such as control function or similar semantics?

No

13a. Does the proposal contain any Ideographic compatibility character(s)?

No

	1130	1131	1132	1133	1134	1135	1136	1137
0		ெந	௦	௱	◌ௗ		ஜ௩	◌ (சு)
1	◌		உ		◌ௗ		௴ர	◌ (கூ)
2	◌ం	ஷ்	ஊ	ஐ	◌இ		ெ௴	◌ (ந)
3	◌ః	ஒ	ண	௨	◌ஜ		ெ௵	◌ (வி)
4		ஔ	த		◌ஜ			◌ (ப)
5	சு	கூ	௳	வ				
6	ஆ	வ	உ	��	◌ం		◌ (ஓ)	
7	ஐ	ழ	ட	ஷ	ெ◌	◌ள	◌ (க)	
8	ஈ	வ	ந	ஸ	ெெ◌		◌ (உ)	
9	உ	ஐ		ஹ			◌ (ங)	
A	ஊ	அ	ப		◌ெ�ா		◌ (சு)	
B	ஃ	உ	ஃ		ெ◌ா		◌ (ரு)	
C	ஏ	ஜ	ப	◌	ெ◌ள		◌ (சூ)	
D		௶	ௐ	◌	◌ௗ	�</		
E	ண்	௭	உ	◌ா	◌	అ		◌ (ஏ)
F	ண	ட	ய	◌		ఉ		◌ (ஏ)

Since Grantha reuses many characters from other blocks, the Grantha code chart must contain sufficient annotations to help users to locate their desired characters. Further, in view of the importance of Grantha for Vedic, copious annotations are provided for the Vedic characters.

Various Characters:

11301	$\dot{\circ}$	GRANTHA SIGN ANUNASIKA
11302	$\circ$o	GRANTHA LETTER ANUSVARA
11303	$\circ$ᴐ	GRANTHA LETTER VISARGA

For ardhavisarga, use 1CF2 ○x VEDIC SIGN ARDHAVISARGA or 1CF3 ○x ROTATED ARDHAVISARGA.

Independent Vowels:

11305	அ	GRANTHA LETTER A
11306	ஆ	GRANTHA LETTER AA
11307	இ	GRANTHA LETTER I
11308	ஈ	GRANTHA LETTER II
11309	உ	GRANTHA LETTER U
1130A	ஊ	GRANTHA LETTER UU
1130B	௠	GRANTHA LETTER VOCALIC R
1130C	எ	GRANTHA LETTER VOCALIC L
1130D		<reserved>
1130E	எ்	GRANTHA LETTER E
		* for transcribing short E
1130F	ஏ	GRANTHA LETTER EE
11310	ஐ	GRANTHA LETTER AI
11311		<reserved>
11312	ஒ்	GRANTHA LETTER O
		* for transcribing short O
11313	ஓ	GRANTHA LETTER OO
11314	ஔ	GRANTHA LETTER AU

Consonants:

11315	க	GRANTHA LETTER KA
11316	ఖ	GRANTHA LETTER KHA
11317	౧	GRANTHA LETTER GA

11318	ஙப	GRANTHA LETTER GHA
11319	ஜ	GRANTHA LETTER NGA
1131A	உ	GRANTHA LETTER CA
1131B	உற	GRANTHA LETTER CHA
1131C	ஐ	GRANTHA LETTER JA
1131D	ஈப	GRANTHA LETTER JHA
1131E	ஞ	GRANTHA LETTER NYA
1131F	டஂ	GRANTHA LETTER TTA
11320	௦	GRANTHA LETTER TTHA
11321	உ	GRANTHA LETTER DDA
11322	ஜ	GRANTHA LETTER DDHA
11323	ண	GRANTHA LETTER NNA
11324	த	GRANTHA LETTER TA
11325	ஊ	GRANTHA LETTER THA
11326	உ	GRANTHA LETTER DA
11327	ௐ	GRANTHA LETTER DHA
11328	ந	GRANTHA LETTER NA
11329		<reserved>
1132A	பட	GRANTHA LETTER PA
1132B	படஂ	GRANTHA LETTER PHA
1132C	பப	GRANTHA LETTER BA
1132D	ஔ	GRANTHA LETTER BHA
1132E	உ	GRANTHA LETTER MA
1132F	ய	GRANTHA LETTER YA
11330	௱	GRANTHA LETTER RA
11331		<reserved>
11332	௨	GRANTHA LETTER LA
11333	எஂ	GRANTHA LETTER LLA
11334		<reserved>
11335	வ	GRANTHA LETTER VA

11336	௴	Grantha Letter SHA
11337	ஷ	Grantha Letter SSA
11338	வ	Grantha Letter SA
11339	ஹ	Grantha Letter HA

Various Signs:

1133C	ఁ	Grantha Sign Nukta

    *   for extending the script to new letters

1133D	⸗	Grantha Sign Avagraha

Dependent Vowel Signs:

1133E	௑ா	Grantha Vowel Sign AA
1133F	௑ி	Grantha Vowel Sign I
11340	௑ீ	Grantha Vowel Sign II
11341	௑ு	Grantha Vowel Sign U
11342	௑ூ	Grantha Vowel Sign UU
11343	௑ృ	Grantha Vowel Sign Vocalic R
11344	௑ౄ	Grantha Vowel Sign Vocalic RR
11345		\<reserved\>

Reordrant Dependent Vowel Signs: (stand to the left of the consonant)

11346	௦ఁ	Grantha Vowel Sign E

    *   for transcribing short E

11347	௦ఁ	Grantha Vowel Sign EE
11348	௦௦ఁ	Grantha Vowel Sign AI
11349		\<reserved\>

Split Dependent Vowel Signs: (with two parts standing on either side of the consonant)

1134A	௦ఁா	Grantha Vowel Sign O

    *   for transcribing short O

    ≡ 11346 ௦ఁ 1133E ௑ா

1134B	௦ఁா	Grantha Vowel Sign OO

    ≡ 11347 ௦ఁ 1133E ௑ா

1134C	௦ఁ௎	Grantha Vowel Sign AU Old

* old-style form of the dependent vowel sign AU

→ 11357 ◌ௗ GRANTHA VOWEL SIGN AU

≡ 11347 ெ◌ 11357 ◌ௗ

Various Signs:

| 1134D | ◌ௗ | GRANTHA SIGN VIRAMA |

* removes the inherent vowel A of consonants

| 1134E | ◌ | GRANTHA SIGN LIGATING VIRAMA |

≡ 1134D ◌ௗ

* specifically requests virama-ligated forms of consonants

| 11350 | | <reserved for OM> |

Rig/Yajur Vedic Svara Markers (reserved):

For Rig/Yajur Vedic svara markup, use the generic Vedic characters 0951, 0952 and 1CF4.

| 11351 | | <reserved> |

→ 1CF4 ◌ VEDIC TONE CANDRA ABOVE

→ 0951 ◌ VEDIC TONE SVARITA

| 11351 | | <reserved> |

→ 0952 ◌ VEDIC TONE ANUDATTA

Dependent Vowel Sign:

| 11357 | ◌ௗ | GRANTHA VOWEL SIGN AU |

* modern form of the dependent vowel sign AU

→ 1134C ெ◌ௗ GRANTHA VOWEL SIGN AU OLD

Various Characters:

1135D	ௐ	GRANTHA SIGN PLUTA
1135E	உ	GRANTHA LETTER VEDIC ANUSVARA
1135F	உ	GRANTHA LETTER VEDIC DOUBLE ANUSVARA

Additional Vowels:

11360	௹	GRANTHA LETTER VOCALIC RR
11361	��}ௗ	GRANTHA LETTER VOCALIC LL
11362	◌ௗ	GRANTHA VOWEL SIGN VOCALIC L
11363	◌ௗ	GRANTHA VOWEL SIGN VOCALIC LL

63

Punctuation (Reserved):

For danda punctuation, use the generic Indic characters 0964 and 0965.

11364   <reserved>

              → 0964   ।   INDIC DANDA

11365   <reserved>

              → 0965   ॥   INDIC DOUBLE DANDA

Numerals:

For numerals, use the unified Grantha-Tamil numerals from the range 0BE6 to 0BF2 and the major fractions from 0BFB to 0BFD.

Sama Vedic Svara Markers:

For Sama Vedic svara markers not encoded here, use 1CD0 ◌ VEDIC TONE KARSHANA, 1CD2 ◌ VEDIC TONE PRENKHA and 20F0 ◌ COMBINING ASTERISK ABOVE.

11366   ◌   COMBINING GRANTHA DIGIT ZERO

          *   marks some special svarita-s in recitation

          *   marks vowels that are not prolated in singing

11367   ◌   COMBINING GRANTHA DIGIT ONE

          *   marks udatta in recitation, prathama svara in singing

          *   sometimes indicates vowel length

11368   ◌   COMBINING GRANTHA DIGIT TWO

          *   marks udatta and svarita in recitation, dvitiya svara in singing

11369   ◌   COMBINING GRANTHA DIGIT THREE

          *   marks anudatta in recitation, tritiya svara in singing

1136A   ◌   COMBINING GRANTHA DIGIT FOUR

          *   marks caturtha svara in singing

1136B   ◌   COMBINING GRANTHA DIGIT FIVE

          *   marks mandra svara in singing

1136C   ◌   COMBINING GRANTHA DIGIT SIX

          *   marks atisvarya svara in singing

1136D   <reserved>

1136E   <reserved>

1136F		<reserved>
11370	◌ �…	COMBINING GRANTHA LETTER A

* marks abhigita svara in singing

11371	◌	COMBINING GRANTHA LETTER KA

* marks kampa and other special svarita-s in recitation

11372	◌	COMBINING GRANTHA LETTER NA

* marks namana svara in singing

11373	◌	COMBINING GRANTHA SYLLABLE VI

* marks vinata svara in singing

11374	◌	COMBINING TAMIL LETTER PA

* marks shivaka udatta in recitation, dipta svara in singing

Sama Vedic Special Punctuation:

For separating sections of singing where a pause is disallowed, use 1CD3 ‖ VEDIC SIGN NIHSHVASA.

Additional Vowels:

11360	◌	GRANTHA VOWEL SIGN VOCALIC L OLD

≡ 11362 ◌ GRANTHA VOWEL SIGN VOCALIC L

11361	◌	GRANTHA VOWEL SIGN VOCALIC LL OLD

≡ 11363 ◌ GRANTHA VOWEL SIGN VOCALIC LL

–o–

§17. Feedback from scholars

Mahamahopadhyaya, Veda Bhashya Ratnam
Mullaivasal DR R KRISHNAMURTI SHASTRI
M.A., Ph.D.

No 2/1, Sanskrit College Street,
Mylapore, Chennai – 600 004.
Ph: 2498 4698, 4207 4402.

Long life and prosperity be unto Shriramana Sharma.

In the Grantha script, there are two ways of writing consonant sounds that are not followed by a vowel. Adding the vowel-absence-marker ௭ is one which is well known for all consonants. The other is the usage of a special vowelless form, and is known only for some of the consonants. Whether the vowel-absence-marker is added, or the special vowelless form is used, either way no difference is created in the meaning of words:

Further, when a consonant that has such a special vowelless written form is followed by another consonant to form a group of consonants, there are many possibilities in writing. The first consonant may be written with a vowel-absence-marker or in the special form, with the second in the full form. The consonants may also be written in an vertical manner or form a joint character which you call a ligature. Here are the examples you requested:

Though some of these ways of writing are not common, no other fault can be found with them as readers will have no problem in understanding the proper meaning even from them.

Further, even where such consonant cluster occurs at the border of two words, as in ṣaṭ-saptatiḥ, san-tamasam, san-mārgaḥ, bhagavat-kāryam etc, there is no rule or norm that a specific alternative should be used in a specific context. For

1/2

66

Mahamahopadhyaya, Veda Bhashya Ratnam   No 2/1, Sanskrit College Street,
Mullaivasal DR R KRISHNAMURTI SHASTRI   Mylapore, Chennai – 600 004.
                M.A., Ph.D.   Ph: 2498 4698, 4207 4402.

---

identification of word boundaries in Sanskrit one can only rely on knowledge of the language, and not some factor in orthography, since it is often not possible to indicate word boundaries using orthography.

For about the past twenty years, the Heritage India Education Trust, Chennai has published under my guidance many books related to the Veda-s, Śāstra-s etc in the Grantha script. The book Grantha Bāla Pāṭha used for learning the script has especially been very useful to the public. In all these books we have, as per convenience, interchangeably used any and all of the above alternatives for indicating vowelless consonants. This is true throughout the history of writing and printing Grantha.

So have no doubt that these 'special' vowelless forms of consonants are not different letters, but only alternative ways of writing the same letters when there is no vowel.

Regarding your inquiry about the numerals, I can also say without doubt that all the numerals are currently written and used in Grantha in the same way as in the Tamil script. There is no change in either the shape or the way of using them to denote numbers. There may have been earlier versions of these numerals used in inscriptions etc, but these certainly do not belong to modern Grantha and so are not in current use today.

I have also seen and approve of the other aspects of your document on the Grantha script. Your detailed presentation of Vedic characters, in particular the complicated svara markers of Sama Veda, is to be commended. I am pleased that you have appropriately depicted the the Grantha script and are doing the needful to standardize it in computers.

Wishing you all goodness and success in your efforts,

Professor R Krishnamurti Shastri,
Retired Principal, Sanskrit College, Mylapore, Chennai.
Dated – 2009-Aug-29.

Dr Gerhard Ehlers of the Orientabteilung, Staatsbibliothek zu Berlin, was so kind to give us the following words of encouragement in his e-mail dated 2009-Sep-22 19:40 +0530: (I have included some of the email headers on purpose.)

```
…
Received: by 10.239.149.132 with SMTP id j4cs31686hbb;
        Tue, 22 Sep 2009 07:14:32 -0700 (PDT)
Received: by 10.223.54.15 with SMTP id o15mr281712fag.96.1253628869927;
        Tue, 22 Sep 2009 07:14:29 -0700 (PDT)
…
MIME-Version: 1.0
Content-Type: text/plain;
      charset="iso-8859-1"
Content-Transfer-Encoding: quoted-printable
Subject: AW: AW: Transcription of the pluti mark
Date: Tue, 22 Sep 2009 16:10:36 +0200
Message-ID: <FBC64F6568978B4CB698DCA8BFAD8150ADB7BC@exchangepk4>
In-reply-to: <4AB76CE2.8080307@gmail.com>
X-MS-Has-Attach:
X-MS-TNEF-Correlator:
Thread-Topic: AW: Transcription of the pluti mark
Thread-Index: Aco6tDocVmvPpyt+QgiU5xebmjKTPQA0JabQ
From: "Ehlers, Gerhard" <Gerhard.Ehlers@*****.de>
To: "Shriramana Sharma" <*****@gmail.com>

Dear Mr Sharma,

Thank you very much for providing me with your Grantha proposal for Unicode. It
is really impressive especially the part concerning Vedic characters!

…

Your choice for the middle pair of o and au (p. 7) is felicitous and supported by
the testimony of the manuscripts.

…

I do hope that the 'revival of Grantha' attracts more people to read South Indian
manuscripts.

All the best and good luck,
Gerhard Ehlers
```
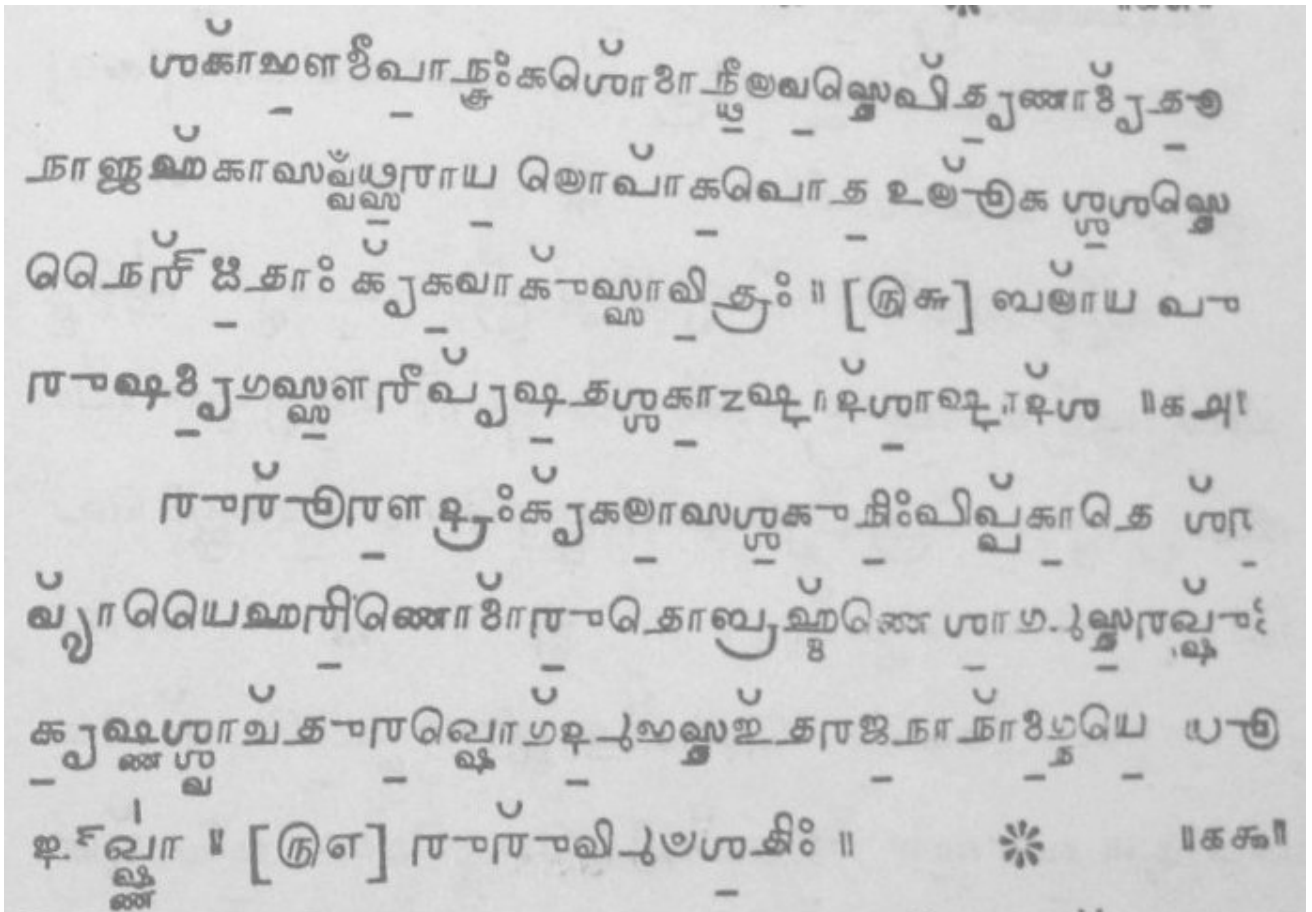
§18. Samples from printed books

Samples for the mixed usage of Grantha and Tamil scripts for Tamil Manipravalam have been given in §7. We have also given many samples while illustrating the various Vedic characters. Here we provide (some more) samples of the usage of Grantha for pure Sanskrit:

'Classical' Sanskrit (ref 10, page 12)

Yajur Veda (ref 7, page 577)