

(R)Unicode: Encoding and Sustainability Issues in Runology

Elisabeth Maria Magin¹, Marcus Smith²

¹Museum of Cultural History, University of Oslo

²Swedish National Heritage Board

Abstract

In this article, the basic premises of reading a runic inscription and transferring the data into digital formats are discussed. Particular attention will be on the Runic block in the Unicode standard for digital character encoding, which currently does not suit the needs of the academic runological community. The proposed solution builds upon the existing standard without adding unnecessary additional characters, but with the ability to encode form-variants on top of the base character while retaining backwards compatibility, an approach from which other archaic scripts like Cuneiform or Khitan that suffer of similar encoding issues, could also benefit.

Keywords

Runes, Unicode Runic, Runology, Ancient scripts, Digitisation

1. Introduction

While most commonly associated with the Vikings in the popular consciousness, the term “runes” is used for at least four closely related variations of the same alphabetic writing system used across a geographic area from Ukraine to Greenland, from the 3rd to 19th centuries CE. The lion’s share of surviving runic inscriptions are found in Sweden, most often carved into stones, and Norway, where more than half of the extant runic corpus was carved into small everyday objects, often wood or bone. During the approximately 1700 years of active use, the original 24-character runic row referred to as the Older Fuþark (alternatively futhark) underwent several major changes. In the British Isles, the character inventory was expanded (Anglo-Saxon Fuþorc), while in Scandinavia, only 16 runes of the original 24 remained in use during the Viking Age (Younger/Viking-Age Fuþark). During the High Middle Ages, these 16 runes were modified and the inventory expanded to properly represent the phonemes in later Old West Norse (Younger medieval Fuþork). There is also a post-Reformation runic tradition, which is, however, not the topic of this article [introductions in 1, 2, 3].

Studying the main variations of runic writing and their development and derivatives remains however challenging even with digital methods to support research, a problem also frequently


DHNB2023 | Sustainability: Environment - Community - Data. The 7th Digital Humanities in the Nordic and Baltic Countries Conference. Oslo – Stavanger – Bergen, Norway. March 8–10, 2023.

✉ e.m.magin@khm.uio.no (E. M. Magin); marcus.smith@raa.se (M. Smith)

🌐 <https://www.khm.uio.no/english/about/organisation/archaeology-department/staff/elismari/index.html>
(E. M. Magin)

🆔 0000-0002-7612-1234 (E. M. Magin); 0000-0002-7438-5498 (M. Smith)

© 2023 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 DHNB Publications, DHNB2023 Conference Proceedings, <https://journals.uio.no/dhnbpub/issue/view/dhnb2023>

encountered by scholars working with other archaic scripts like Cuneiform or Khitan. In recent years, this has become a serious drawback in runology, as digital searches would for example enable runic scholars to conduct macro- instead of the currently most common microstudies of runic inscriptions. It is currently also impossible to search runic inscriptions *as texts* without relying on the current workaround, transliterations into Roman letters.

Getting the runes into the computer has, however, never been a straightforward endeavour. Several of the difficulties encountered can be attributed to encoding runic characters using the Runic block, added to the Unicode standard in 1999 to allow digital representation of runes. The initiative was driven by several established and senior runologists, for example Helmer Gustavsson (Sweden), James Knirk (Norway), Klaus Düwel (Germany), Ray Page (Great Britain) and Marie Stoklund (Denmark) [4]. It was only accepted as a revised proposal, which led to a code block Runic being created, then featuring 81 runes from the Older Germanic, the Anglo-Frisian, the Viking Age and the medieval Futhorks. It also includes characters of what are generally considered derivatives of runic writing, referred to as “rune rows”, in this case the long-branch and short-twig variations. These appear as allographs during the Viking Age, although some of them are later distinguished and become separate characters in the medieval Futhork [for example 5, 6].

As this short introduction exemplifies, the development and reciprocal influence of different runic scripts is by no means clean-cut and unambiguous. Within the field, there is general agreement that the standardised Futhorks presented in handbooks and introductions to runes are simplifications of a far more complex situation [7]. They certainly serve a purpose [7, 6-8] and one may argue that character encoding and Unicode is an area where simplification and standardisation are precisely what is required. The original code block Runic (not the later additions from the Franks Casket and Tolkien runes, which have caused much consternation to those working with runes professionally – in the first case, the runes are allographs of already encoded characters, in the second, they are part of a fantasy script, not actually part of any historical runic script) to a certain extent follows these principles. The code block, due to inconsistent encoding, still leaves much to be desired for the runologist, however, especially when the aim is to conduct research into the genesis and use of the different runic rows; as such its use is mostly confined to non-runologists. It is the authors’ contention that the Unicode Runic block does not serve the needs of runic scholars, following no clear principles in which characters are encoded and which are considered as allographs and therefore not encoded.

Following a short introduction into the differences between the four main runic rows on a graphemic and linguistic level, this article proceeds to the different stages in the process of deciphering a runic inscription. It will explain how runologists currently remedy the issues with Unicode Runic when encoding runic inscriptions and why these are unsustainable, before examining a potential solution making use of Unicode Form Variation Selector and Stylistic Sets in OTF fonts in an attempt at a more sustainable approach to runic encoding better aligned with the needs of runic scholarship in the twenty-first century. Since most runologists work with inscriptions belonging to one of these four Futhorks, focus here rests on the encoding of these, although runes were actively used in some areas until the 18th century.

2. Futhorks and rune rows: runic writing systems

The first potential, albeit debated, runic inscription is a fibula from Meldorf, Germany, dated to the first half of the 1st century AD. Equally debated other early inscriptions originate from Vimose, Denmark, and date to the 2nd and 3rd centuries AD. The currently most widely accepted explanation is that the Older Futhork was invented somewhere in the regions of Northern Germany or Denmark at some point during the first two centuries AD, inspired by the Roman or other Italic alphabets, and then used in areas settled by Germanic tribes. This oldest version consists of 24 signs divided into three ættir with eight runes each: $\mathfrak{F}\mathfrak{R}\mathfrak{U}\mathfrak{N}\mathfrak{I}\mathfrak{S}\mathfrak{T}\mathfrak{A}\mathfrak{R}\mathfrak{J}\mathfrak{K}\mathfrak{L}\mathfrak{M}\mathfrak{N}\mathfrak{O}\mathfrak{P}\mathfrak{Q}\mathfrak{R}$

These stayed in use until around 800 AD, when the character inventory was (perhaps gradually) reduced to just 16 runes around beginning of the Viking Age on the continent [e.g. 8]. In Britain, to where the Angles, Saxons and Jutes had exported the Older Futhork, the opposite development took place: the inventory was gradually expanded to 32 runes owing to changes in the spoken language, which gained more distinct speech sounds. No scholar has been able to find an explanation for these contradictory developments, since the language changes on the continent also led to an increase in distinct speech sounds [e.g. 3]. As runes are a phonemic script, meaning each character (grapheme) is tied to a distinct speech sound or sounds in spoken language (phoneme), the reduction of the rune inventory on the continent during the Viking Age continues to be a topic of scholarly debate. This situation is compounded by the development of no less than three derivatives of the Younger Futhork, referred to as long-branch, short-twig and staveless runes. None of these are considered to be separate Futhorks, instead the runes are considered to be allographs in derivative systems, the reasons for which become obvious when comparing them:

Long-branch: $\mathfrak{F}\mathfrak{R}\mathfrak{U}\mathfrak{N}\mathfrak{I}\mathfrak{S}\mathfrak{T}\mathfrak{A}\mathfrak{R}\mathfrak{J}\mathfrak{K}\mathfrak{L}\mathfrak{M}\mathfrak{N}\mathfrak{O}\mathfrak{P}\mathfrak{Q}\mathfrak{R}$: $\mathfrak{F}\mathfrak{R}\mathfrak{U}\mathfrak{N}\mathfrak{I}\mathfrak{S}\mathfrak{T}\mathfrak{A}\mathfrak{R}\mathfrak{J}\mathfrak{K}\mathfrak{L}\mathfrak{M}\mathfrak{N}\mathfrak{O}\mathfrak{P}\mathfrak{Q}\mathfrak{R}$: $\mathfrak{F}\mathfrak{R}\mathfrak{U}\mathfrak{N}\mathfrak{I}\mathfrak{S}\mathfrak{T}\mathfrak{A}\mathfrak{R}\mathfrak{J}\mathfrak{K}\mathfrak{L}\mathfrak{M}\mathfrak{N}\mathfrak{O}\mathfrak{P}\mathfrak{Q}\mathfrak{R}$ Short-twig: $\mathfrak{F}\mathfrak{R}\mathfrak{U}\mathfrak{N}\mathfrak{I}\mathfrak{S}\mathfrak{T}\mathfrak{A}\mathfrak{R}\mathfrak{J}\mathfrak{K}\mathfrak{L}\mathfrak{M}\mathfrak{N}\mathfrak{O}\mathfrak{P}\mathfrak{Q}\mathfrak{R}$: $\mathfrak{F}\mathfrak{R}\mathfrak{U}\mathfrak{N}\mathfrak{I}\mathfrak{S}\mathfrak{T}\mathfrak{A}\mathfrak{R}\mathfrak{J}\mathfrak{K}\mathfrak{L}\mathfrak{M}\mathfrak{N}\mathfrak{O}\mathfrak{P}\mathfrak{Q}\mathfrak{R}$: $\mathfrak{F}\mathfrak{R}\mathfrak{U}\mathfrak{N}\mathfrak{I}\mathfrak{S}\mathfrak{T}\mathfrak{A}\mathfrak{R}\mathfrak{J}\mathfrak{K}\mathfrak{L}\mathfrak{M}\mathfrak{N}\mathfrak{O}\mathfrak{P}\mathfrak{Q}\mathfrak{R}$

This approach is also owed to the fact that runecarvers could freely choose which variation they were using; inscriptions frequently mix and match the systems, excepting staveless runes [9, 10]. Why runecarvers would make use of characters from either system or whether they even considered them to be distinct systems is not fully understood.

Disregarding developments of runic script post-15th century, one more change in the runic writing system takes place around c. 1100 with the end of the Viking Age. Potentially owing to the problem of having to use the same rune to express several different speech sounds and the resulting confusion, the Medieval Futhork starts to de-unify runes and adopts, for example, long-branch \mathfrak{f} and short-twig \mathfrak{f} as separate characters. New runes are also added to the repertoire [for example 11]. Use of these extra characters is, however, not mandatory; the characters in an inscription archaeologically dated to the middle of the 13th century AD can therefore look precisely the same as those in an inscription dating to the 9th century AD.

3. The three (or more) stages of reading a runic inscription

As described in a variety of (runic) handbooks [12, 1, 2, 3], the reading or interpretation of a runic inscription generally proceeds in three steps following an initial examination. The first of these – unless drawings and images are used instead – is most often referred to as “transcription” or “transrunication” and describes the process of standardising the shape of

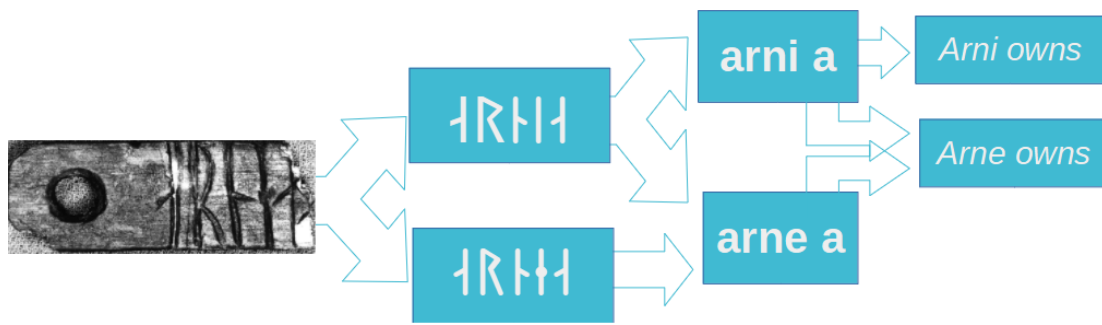


Figure 1: The three stages of interpreting or reading a runic inscription

the runes on an object (more rarely in a manuscript) into what various authors have referred to as “idealruner” [9] or “print runes” as illustrated by the second “column” in 1. At its most basic, this is the same process as typing up handwritten notes on a keyboard, just with runes instead of Roman letters. However, other than the well-standardised Roman alphabet and despite the fact that standardised Fuparks exist, transcribing runes is somewhat more complicated since runic writing exhibits a variety of peculiarities that frequently render the process less than straightforward. Firstly, while the direction of writing is generally left-to-right, inscriptions in the Older Fupark in particular can be written left-to-right or right-to-left, or sometimes both in alternating lines, a phenomenon called “boustrophedon” (as the ox plows).

To complicate matters further, even single runes within one inscription can be carved against the prevalent writing direction, becoming “Wenderunen” (turned/flipped/reversed/inverted runes). This happens as late as the Middle Ages, where a prominent example is N 737 [13], reading 1ᚠᚱᚢᚦ, interpreted either as the name Ljótr or Þjóðarr. If the latter is correct, the use of 1 instead of Þ is a little surprising, but not impossible. However, 1, the t-rune, can also be interpreted as ᚠ, the l-rune, written against the otherwise prevailing left-to-right writing direction in the inscription. The same phenomenon can be observed for ᚢ (short-twig n) and ᚠ (short-twig a), ᚦ (long-branch n) and ᚠ (long-branch a), ᚢ (o) and ᚠ (o) and others, for example ᚠ (k), which can be flipped upside-down: k. The impact of this phenomenon on reading and interpretation varies. As N 737 shows, a Wenderune can lead to a completely different reading of a character sequence, whether that be a name or another type of word. Other Wenderunen can appear both as Wenderunen and as characters in their own right: ᚢ (o), when turned around, is also used to spell the letter “b” in Viking Age and medieval inscriptions. At this point, runologists are already faced with a difficult decision: in print, do they represent the runes as they see them or as they interpret them?

Seim [1] is not the only runologist arguing for a strict division between describing the purely visual and the interpretation of what the scholar sees. A variety of publications, for example Waldispühl [14], Palumbo [6], Nowak [15], Spurkland [16] analyse the relationship between graphemes, graphtypes, graphtype variants and glyphs. The studies, for the most part, define “grapheme” as the abstract character a given sign is supposed to represent, which is in turn connected to a specific sound value in phonemic scripts like runes. Graphtypes represent

the idealised visual representation of that abstract character, glyphs are the actual characters on objects or in manuscripts, basically the “handwritten” forms that are subject to a range of influences. Graphtype variants group the “handwritten” runes together based on visual similarities, roughly equivalent to allographs; they represent a level between the single actual written runes and the idealised graphtypes. For example, during the Viking Age, ᚠ and ᚡ could be considered allographs or graphtype variants of a graphtype ᚠ, with ᚡ and ᚢ being further Wenderunen allographs of this graphtype, while in themselves also representing a separate graphtype ᚢ with the short-twig allograph ᚣ. This shall be returned to in 4.

The second step is usually called “transliteration” and describes the process of equating runes with Roman letters. Again, practices can differ. Some runologists equate rune=Roman letter by way of the phoneme or sound value of the respective rune/letter, whereas others use Roman letters as a stand-in for the graphical form of the rune which may also signify a letter-sound relationship, but does not have to [17, 307]. It is illustrated in the second column in 1. The main problem at this stage arises from the fact that the speech sounds Roman letters are ascribed and the reconstructed speech sounds for runes do not always align – and that in specific Fuparks, one and the same rune can be used to express several different speech sounds. One example of this is illustrated in 1, where the first sequence of runes can be transliterated as both **arnia** and **arnea**, since ᚠ is a multi-valued rune used primarily to express the speech sound /i/, but with a secondary value /e/. However, around c. 1100, several multi-valued runes acquire a diacritic marker, a dot, to differentiate between the primary and the secondary sound value. This is illustrated in the first column, where the first transcription shows ᚠ while the second uses ᚡ (since the dot is usually created by simply sticking the point of whatever sharp object – mostly a knife – is used to carve the inscription, it can be fairly hard to decide whether a rune is dotted or not). This rune is unambiguous inasmuch as it is never used for /i/, so when it appears in an inscription, interpretation is fairly clear. Technically, this would solve some problems for modern rune-readers and runologists *if* the system of using dots as diacritic markers to distinguish between primary and secondary sound value was used consistently – which it was not. Instead, using the system was up to the single runecarver’s preferences.

Another complicating factor arises in the fact that the short-twig **h**-rune looks exactly alike. So interpretation is clear only when ᚡ appears in a context where interpretation as short-twig ᚠ would make no sense, but reading it as /e/ does.

As mentioned above, the phenomenon of multi-valued runes is particularly observed in the 16-character Viking Age Fupark, although since use of the system is optional, the problem continues in medieval inscriptions. The diacritic dot also fails to solve the problem when a rune has more than a primary and secondary sound value, like ᚠ. In this case, the dot generally only serves to exclude the primary sound value from consideration, in this case /u/. ᚠ can still signify /o/, /y/, /v/, /w/ or /f/.

Different runologist communities follow different traditions regarding transliterations to handle the problem. Today, many runologists follow the custom of transliterating runes with the same Roman letter regardless of supposed sound value in any given instance [for example 18, 105]. This, however, was not the case for earlier publications or, for example, many German runologists, whose transliterations frequently aim to represent the suspected pronunciation of any given inscription rather than following a consistent transliteration system. This leads to the interesting phenomenon that a transliteration of ᚠᚠᚠ may look like **iuar** when transliterated

by a runologist following the first approach, but **ioar** when transliterated by one following the latter. Furthermore, the sequence may then be normalised by the first as *Ívarr*, while the second may instead choose to normalise as *Jóarr* – which are two different names with different etymology and meaning (although they can also be variants of the same name) [19, on different transliteration conventions see].

This last step of the process, normalisation, mainly aims to render the text of the runic inscription according to the modern-day standards of spelling the different languages appearing in runic inscriptions, for example proto-Norse, Old East Norse or Old West Norse. The main issue here arises once more precisely from the fact that multi-valued runes can be interpreted as a variety of speech sounds. This in turn can lead to differences in final conclusion like *Ívarr*/*Jóarr* or, as in 1, a reading of the name as Arni versus Arne. While this may in the first instance seem to be a minor issue, for Arni and Arne are, after all, just two spellings of the same name, this uncertainty can become a much greater problem when, for example, linguists attempt to use runic inscriptions to trace and date sound changes in a given language, as is the case for the evolution of Old West Norse into Middle Norwegian between 1350 and 1550. Although scholars mainly use manuscripts for the purpose, indications for the change also appear in runic inscriptions, specifically the inscriptions found in Bryggen, Bergen, Norway [13]. Encoding runic inscriptions in such a way that linguists can use the resulting data as well, while retaining the information necessary for them to decide whether a particular inscription is relevant to their study, is therefore not just a desideratum, but a requirement.

4. Current encoding solutions

This, however, is where the problems begin, since runologists are more often than not interested not only in runes as strings of characters, but in what the runes in any given inscription look like. It should be clear from 2 that the specific form of a rune carries importance, whether in terms of dating, the potential origin of the runecarver or simply just to decide which particular speech sound the rune represents. Runologists are therefore not only in need of *one* transcription; in theory (although this is currently not established as a custom), they are in need of transcriptions at *two* levels, one simply representing the purely visual, the second representing the interpretation of the purely visual. The code block Runic in the Unicode character encoding, despite being developed by runologists, is unfortunately inconsistent in how and which runes are encoded, more often than not mixing the different levels, which has led to the code block not being used at all by the academic runologist community.

Instead, they currently fall back on one of two solutions to remedy the issues with Unicode Runic. The first is to ignore the runes as characters completely, instead presenting only a transliteration into Roman characters, supplemented with photographs or drawings of the inscription. This approach is complicated not only by the different approaches taken to transliterations within the academic runologist community, but also the fact that this means the runes are often not represented as runes *at all* in the given publication. For someone specifically aiming to examine the use of different graphtypes and graphtype variants, this approach is therefore less than useful.

Alternatively, runologists rely on bespoke typefaces allowing them to represent a broader

range of the visual forms of runes more accurately. This does allow for representation of both graphtypes and graphtype variants, but creates another problem, namely at which level of accuracy the runes are represented (3). Runic scholars rarely specify whether they are using a purely visual approach and represent the runes in print exactly as they see them on the object, Wenderunen and graphtype variants included, or whether they are choosing to print their interpretation of what it is they see, using graphtypes instead of graphtype variants, or even a mix of both approaches. It is generally through the comparison with the images provided that one is able to determine which kind of transcription one is dealing with.

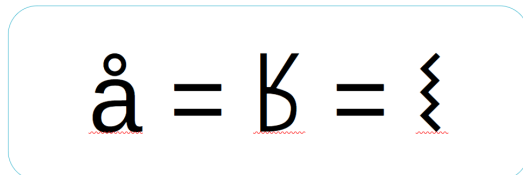


Figure 2: U+00E5 printed using a Unicode font and the non-Unicode runic fonts Gullskoen and Gullhornet

Neither method is particularly sustainable in the long run. Images for many runic inscriptions are often lacking completely or of poor quality and hindered by copyright issues, and transliteration customs vary from country to country and by runic writing system. Custom runic typefaces rely on everyone making use of the data having access to the same typeface, as the underlying characters remain encoded as plain ASCII, unreadable without the font. This is illustrated in 2, where the

image of the character å, which should be encoded by the codepoint 00E5, has simply been replaced by, in the Gullskoen font, a flipped version of the r-rune, whereas the Gullhornet font uses the same codepoint for a variation of the Older Fupark s-rune.

While acceptable for print purposes, this solution is ill-suited for long-term storage purposes, the documents thus created are virtually unsearchable since they encode runic characters as Roman letters. Even worse, if more than one font was used for document and the fonts are inconsistent in their use of codepoints, like Gullskoen and Gullhornet, anyone trying to conduct a document search has to basically know which fonts were used for document creation and which codepoints they respectively use for different runic characters.

But even if one were to accept the use of already existing, bespoke codepoints for characters from a completely different script, the use of fonts and graphtype variants creates other issues for the kind of analyses runologists or linguists may want to conduct. While the visual form of a rune is highly relevant to especially graphemic analyses of runic corpora, for analyses of combinations of different runes or potential changes in the spoken language, the precise visual form is irrelevant to begin with. However, if fonts are used at the graphtype variant level, this means that every graphtype variant is also encoded using a separate codepoint. In turn, this means that to find and compare inscriptions including the same word or name, all different possible combinations of the graphtype variants need to be searched for, a task that quickly becomes unmanageable when single graphtypes can span 10 or more graphtype variants. Perhaps this would not be as much of a problem as it is if the different runic scholarship traditions did not follow different customs where transliteration and normalisation are concerned. The situation is what it is, however, and from a purely methodological perspective, it is also inadvisable to conduct searches on what are higher-level interpretations of a piece of writing, which transliterations and normalisations are.

The question may be asked why runologists do not make use of Unicode Runic instead of

fonts then. The simple answer to that is that Unicode Runic currently suffers of precisely the same problem as fonts in that respect that it is inconsistent in how graphtypes and graphtype variants are encoded. For example, while there is an argument to be made that † and ‡ do represent different graphemes starting around c. 1100 AD, when † starts to be used to represent /æ/ and ‡ continues to represent /a/, long-branch † and short-twig † are also each given their own codepoint, completely disregarding the fact that either rune always and consistently (unless it is a Wenderune) represents /n/.

Other issues arise from the fact that some of the characters come with an inherent dating and geographical placement. Older Futhork ᚠ and Anglo-Frisian ᚠ look exactly alike, the only argument for encoding them as separate codepoints being that the latter also comes to represent /æ/ from a certain point in time. When precisely is still a topic of discussion amongst runologists; but even if that were not the case, objects and people were mobile. Using the Anglo-Frisian ᚠ to encode an inscription that could also have been carved by someone using the Older Futhork – far from an unlikely scenario, the two Futhorks being in parallel use and the geographic areas being in contact – automatically hardcodes an interpretation in what *should* be a neutral description of a visual sign. Encoding at the graphtype level by using Unicode Runic is therefore problematic from a methodological point of view; for encoding at the graphtype variant level, however, Unicode Runic simply lacks too many runes, and this approach additionally results in the same problem as using fonts does by giving each graphtype variant its own codepoint.

To the best knowledge of the authors of this article, the academic runologist community have therefore avoided actively using Unicode Runic whenever possible. There are either too many or too few encoded characters to be of practical use for the kinds of analyses runologists and scholars from adjacent fields like linguistics or archaeology would like to run.

It is instructive here to compare the digital encoding of runes with similar ancient and epigraphic scripts. Perhaps the closest ancient script is the insular Celtic Ogham, in use from around the 4th century CE, and continuing in scholastic use into the 9th century. Like runes, Ogham forms were optimised for carving on wood, metal, stone, or bone, with orthogonal and diagonal tally-like forms. The simple and uniform nature of Ogham does not lend itself to orthographic variation, and allographic deviations are rare. The Ogham Range was added to Unicode in 1999, in the same revision as Runic, and provides adequate coverage for the digital encoding of Ogham texts. Examples of current digital Ogham corpora successfully employing Unicode include *OG(H)AM* from the University of Glasgow, and the Research Squirrel Engineers' linked open data research hub [20]. In other cases, however, the present state of digital script encoding is in a similarly unsatisfactory state to that of runes, and potential solutions may be applicable to both. This applies to Cuneiform, which (much like runes) is an umbrella term for several variants of the same writing system used to write different languages and showing temporal and geographical variations, and Khitan [21, providing an overview of the encoding process].

4.1. Problems with Unicode Runic

In conclusion of the above, Unicode Runic range and its character coverage are – on their own, and in their current form – not fit for the purpose of transcribing runic inscriptions. In addition to inconsistencies making the correct method of accurate and consistent text encoding uncertain

or in some cases impossible, there are also a number of aspects of the standard that are to be found wanting. Firstly, there are a number of runes that are not covered by the standard, which it is thus not possible to reliably encode at the moment. Examples include the runes of the post-reformation Dalecarlian tradition (*Dalrunorna*) used in inscriptions in central Sweden (and, in one notable case, in North America) up until the late 19th century, as well as important variants from the Viking-Age and medieval rune-rows such as the so-called “Greenlandic”-**r** (which in fact occurs just as frequently in non-Greenlandic inscriptions from the period). Also awkward to encode are the staveless runes, a form of the Younger Futhork frequently used as a form of runic short-hand on ephemeral inscriptions on wood and bone, as well as some monumental inscriptions, especially in the Hälsingland province of Sweden. In the case of the staveless runes, they have in fact been deliberately omitted from the standard, and are considered instead – correctly, in the authors’ opinion – to be allographs of the runes of the Younger Futhork, and should thus be encoded using the short-twig characters of that rune-row, and displayed using an appropriate font [22, ch. 13 Archaic Scripts, specifically 13.3 Runic, 342]. However, this intention is not clear from the code block itself nor widely known among runic scholars: at least one proposal to encode the staveless runes separately has been submitted to the Unicode Consortium apparently in ignorance of this, and rejected [23]. A related problem is that it is not always clear to working runologists not intimately acquainted with Unicode which character should be used in the first place even for characters that do exist. The glyphs provided on the code charts are representative examples only, but for a runologist wishing to encode a short-twig **o** from an inscription in the Man-Jæren group, it would be easy to mistakenly encode a **b** instead, since the forms of those runes are swapped from their usual values in many of those inscriptions.

In addition to runic characters absent from the standard – thus unencodable – which characters *are* encoded, and how, is frequently inconsistent. The long-branch and short-twig allographs of the Younger Futhork are encoded as separate characters, but as noted above, the staveless allographs are not. It would have been preferable to have treated all such allographs in the same way, either encoding them all or, ideally, encoding only a single character for each and treating the allographs as variant forms. In some cases, separate encodings do make sense, such as the above-mentioned long-branch **†** and short-twig **†** (4); but in most cases there is no such justification for allographs being allocated distinct codepoints, such as single- and double-barred varieties of Older Futhork **h** (𐐃𐐃). For other runes, significant allographs are not given their own code points: the closed “maskros” form of Younger Futhork **ʝ** is absent, as is the aforementioned “Greenlandic”-**r**. Once again, some sort of consistent approach would have been preferable here: either encode such variations for all runes, or for none. Perhaps most egregiously, some variant forms occurring in a single inscription have been assigned their own codepoints. This is the case for the five variants of the Anglo-Saxon **o**, **i**, **e**, **a**, and **æ** runes which appear on the Franks Casket [24]. Other such idiomatic glyph variants are not afforded the same luxury; thus, for example, it is not possible to encode the elaborately embellished forms which occur on IMM MM111 Andreas V and GR 1 Kingittorsuaq. Other aspects of encoding the runic script not directly related to character coverage also remain problematic and are largely unaddressed by the Unicode standard; examples include bind-runes (runic ligatures, which might be better modelled a Zero-width Joiners character sequences), and mirrored, flipped, or inverted runes.

4.2. Transcription, glyphs, graphemes (and phonemes?)

As explained above, the grapheme is almost always important, and indeed it is some approximation of the grapheme, guided by a supposed sound value, that is recorded in runic transliteration. Furthermore, often – but not always – the specific glyph that has been used, the form variant, is also important to record. However, as we have seen, in Unicode Runic this distinction between grapheme and glyph inconsistently implemented, making accurate and consistent transcription difficult, sometimes impossible. The rune encoding problem is not really a new one, but rather derives, and is inseparable from, the general problem of transcription and transliteration. On some level, an agreed-upon set of standardised forms (glyphs and glyph-variants) is required, and ideally a set of characters (for our purposes, more or less synonymous with graphemes) that those forms represent. This is complicated by the fact that the correlation between glyphs and graphemes is often not straightforward. Multiple rune forms may denote the same grapheme within a given graphemological system (these are allographs). Conversely, multiple graphemes may be denoted by the same rune forms across graphemological systems (these are homographs). Such distinctions are frequently elided in Unicode Runic. The most accurate representation of a runic inscription – and particularly the forms of the runes themselves – is often a photograph or a drawing, but these are of limited value when trying to encode the texts *as text*. After a graphemic representation, a text transcription is the next level of useful recording with minimal interpretation, and it is this level of recording that Unicode Runic simultaneously seems to seek to enable and fails, disappointingly, to achieve.

So, there is a need in runic transcription to encode not only the grapheme, but also the glyph or form variant. Furthermore, when choosing which runes and rune-forms to encode, there is a choice between taking a maximalist or minimalist approach. With a maximalist approach, one would attempt to assign each identified form variant its own character. With a minimalist approach, one would assign characters only to distinct graphemes, viewing form variants as allographs with distinctions in form to be handled only at the font rendering level. The Unicode Runic range tries to be maximalist in some cases, but minimalist in others, apparently unsystematically, with unsatisfactory results. There are, in any case, issues with both the maximalist and minimalist approaches. While a maximalist character set would allow all identified form variants to be unambiguously encoded at the character level, it would make such encoding, and the use of such an encoded text, tedious indeed. Consider a hypothetical runic database or corpus encoded with Unicode Runic; how would one search for a character sequence in Younger Futhork inscriptions including the rune **b**? In this case, we're not interested in the form variant of the rune, just the occurrence of the grapheme itself. As things stand at the time of writing, one would have to provide at least two alternatives in the search parameters, to cover cases of either long-branch and short-twig **b** occurring, since these have separate codepoints. But that's just for one grapheme: if we wanted to search for longer strings, the number of necessary alternatives would quickly become unmanageable, as the number of character variations for each rune multiply. In a hypothetical truly maximalist approach, we might have to deal with ten possible alternative forms for **b** alone! Conversely, for the encoder, they would always have to explicitly encode not just e.g. a **b**-rune, but which specific form that **b** took. Being able to encode characters alone, and not the specific graphical representation of each one, is something we take for granted in many other scripts. But there are problems,

too, with a minimalist character set, chief among them that it becomes impossible to include information about specific form variants in plain text where that *is* of interest – additional levels of encoding, such as markup or rich text formatting are required on top of the text. So the problem remains: how can we digitally encode both graphemes and form variants for runes and runic texts, without creating separate characters for everything?

5. What can be done?

Having established that Unicode Runic does not live up to the needs of the runologist user community, and given the limitations of the existing Runic range, what might be done to address these shortcomings? There are a few options open to runologists feeling frustrated with this unsatisfactory state of affairs and considering effecting positive change:

5.1. Option 1: Do nothing.

- Accept that Unicode Runic is not suitable for encoding old inscriptions, and should only be used to formulate new texts.
- Rely solely on photos, drawings, transliteration, and some idiosyncratic rune fonts.

This option is not a particularly appealing one, but it is nonetheless a strong possibility, given the limited action runologists as an academic community have taken regarding the standardisation of the script to which they devote their study since the Runic block was first published in 1999. The position that Runic should only be used to formulate new texts is also, sadly, in line with Unicode’s original stated remit, explicitly prioritising characters for modern and future use over “preserving past antiquities”: “Beyond... modern-use characters, all others may be defined to be obsolete or rare; these are better candidates for private-use registration than for congesting the public list of generally-useful Unicodes” [25, 5]. (One might however speculate that this attitude has softened over the years, given that the standard now includes numerous characters – including, regrettably, several in the Runic range! – that occur only in single historical texts or inscriptions.)

5.2. Option 2: Expand the Runic block to cover all rune-forms

- Add new explicit codepoints for Dalecarlian runes, staveless runes, and all the other observed form variants of existing characters.

This maximalist approach would involve radically expanding the Runic block well beyond its current bounds with dozens of new characters, aiming for much broader coverage of variations in observed runic forms, and encoding them as distinct characters. It goes against Unicode’s principle of encoding characters, not glyphs, but in a sense that ship has already sailed (see 4.1. It would also be consistent with the current state of the block, encoding each standardised form separately and explicitly. However, as noted above, this would make text encoding unmanageable for any but the shortest of texts, and would render searching and indexing on the character/grapheme level much more difficult, requiring all form variants to be covered for

grapheme-level searches, and complex rules for character-folding and text normalisation. It would, if anything, complicate the kinds of corpus, correspondence and statistical text analyses that scholars may wish to carry out, rather than facilitating them. The authors do not believe that this option would be either desirable or sustainable – it would make an already bad situation worse. A set of standardised forms *is* necessary. But Unicode is explicitly not for encoding glyphs as characters; something we do not question for other scripts.

5.3. Option 3: A minimal character set + markup

- Add missing characters
- But otherwise reduce to a minimal set of distinct characters. (Albeit nonetheless with potentially multiple glyphs at the font level!)

This is the minimalist approach. Removing characters from the existing standard is not possible, and would be highly undesirable even if it were, since there already exist texts encoded using those characters. It may, however, be possible to change character semantics of some characters to declare existing variants deprecated (unlikely) or canonically equivalent for normalisation (more likely, but admittedly not by much). Here we would propose to declare a number of characters representing duplicates and variants of a single grapheme to be canonically equivalent, and one of the pair to be deprecated for use in new texts. Thus through Unicode normalisation rules, the deprecated variants might be normalised to their canonical counterparts. Examples of such pairs include, but are not limited to, the Franks Casket forms (U+16F4–U+16F8), Older Fuþark ƒ (U+16A8) and Fuþorc ƒ (U+16AB), and long-branch/short-twig pairs where the forms never represent distinct graphemes, favouring the existing long-branch character as the canonical one. This approach would require glyph variations (allographs) to be encoded at the markup level, or as rich text, not as separate characters. This has the benefit of preserving reserves graphemic integrity, while allowing for glyph-level encoding, in much the same way as for any other script. However, texts would no longer be able to be accurately encoded as plain-text only. This approach would require runological consensus on a minimal set of graphemes, as opposed to forms, and also a parallel standardisation effort to ensure comprehensive coverage of alternative glyph variants in font support.

Such a minimal character set based upon a subset of the Unicode Runic block has in fact been implemented as a proof of concept: Elisabeth Magin’s database of the medieval inscriptions from Bryggen, Bergen [26] has already applied this approach with attribute-based (non-XML) markup, suffixing each runic character with a code specifying the observed form-variant. It’s possible to search the database by grapheme, but also by specific form-variants.

5.4. Variation selectors

None of these options are completely satisfactory, with both the maximalist and minimalist approaches forcing compromises in one way or another. However, there may be another way to enable encoding of both the character and the form variant as plain text while keeping the two elements distinct. The Unicode standard itself offers a potential solution, in the form of *variation selectors*.

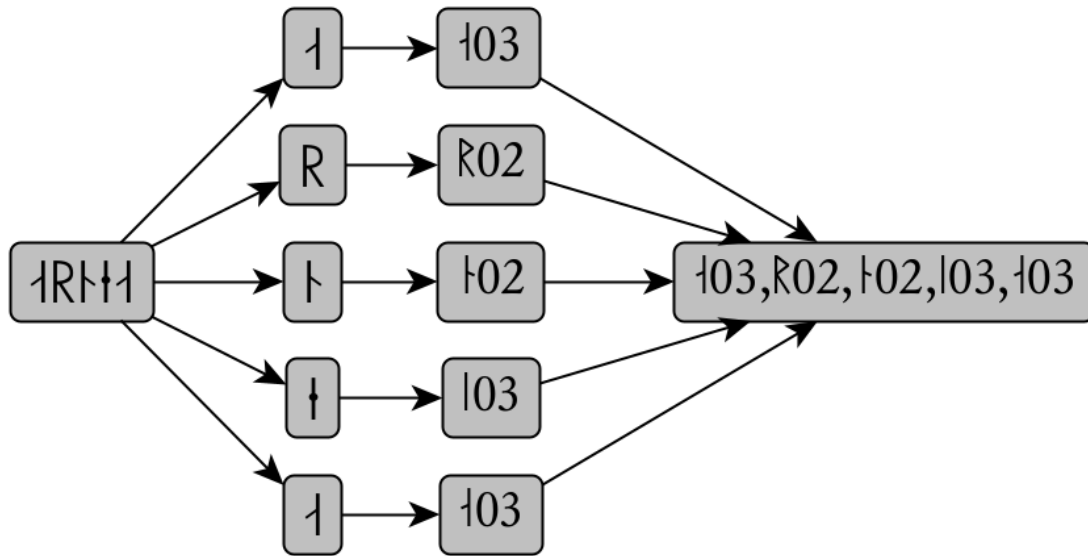


Figure 3: Possible encoding of the inscription from 1 using a minimal character set and attribute-based non-XML markup for graphemic variations of the runes.

Variation selectors are, “non-spacing combining marks. They have no graphic shape of their own; instead they function to pick out a particular, defined subset of potential graphic presentations for the base character to which they are applied” [27]. They provide a way to indicate what form-variant a glyph should take for display, encoded as part of the text. Variation selectors are (relatively) new in Unicode, but fairly widely deployed. They have been used in the Emoji space to distinguish between text and emoji forms of certain characters, for example, as well as to describe historic, obsolete, and other variant forms for Chinese characters. Other applications include specifying a barred variant form for the digit zero. Variation selectors are *modifiers* to the base character. As such, they provide a hint to the rendering engine, that this character should be displayed using a particular form variant, if possible (font support permitting) while not changing the underlying character. Variation selectors degrade gracefully; that is, if font or software support is lacking for a particular form-variant, the correct character will still be displayed, albeit with its “base” or default glyph. To use a hypothetical runic example, you might not be shown a “Greenlandic”-*r* if your font doesn’t support it, but you’ll still see an *r*. Because the underlying character doesn’t change, compliant applications can ignore variation selectors for the purposes of search and indexing; but because the variation selector is still present at the text level, they can also be searched for explicitly as part of a string if desired. This addresses the limitations that would otherwise be concomitant with a minimal character set, and allows for a fourth possible way forward.

5.4.1. Option 4: A minimal character set + variation selectors

- A reduced character set, as Option 3, but instead of allographs at the markup level, define a set of Variation Selectors for each character as part of the Unicode Standard

This would require the same consensus on a minimal set of graphemes, and the same additional standards work to define a set of standard forms as in Option 3 above, but now divided between both the Unicode Consortium for the character/glyph set, and efforts from other bodies for font support for contextual alternate glyphs and stylistic sets for character strings using the variation selectors. Crucially however, the glyph variants would be part of the standard, and no special markup would be required to use and display them – just plain text. Both text encoding and processing, e.g. search, would have the flexibility to work both at the generic character (grapheme) level, *and* at the form variant level, as required. Instead of markup, form-variants could be specified as modifiers to the base character, in plain text, using the appropriate variation selector. Fonts supporting that form would display the appropriate glyph, recognising the string of base-character + variation selector, but fonts without such support would ignore the variation selector and fall back to a generic form for the base character. Similarly, other applications could choose to respect or ignore the variation selector depending on context and the users’ intentions.

This would be the authors’ preferred solution, and is proposed here as a potentially fruitful avenue for future work. However, it should be noted that despite the existence of Variation Selectors as entities explicitly intended to address the kinds of text encoding needs present in digital runology, the Unicode Consortium have in recent years not accepted proposals defining new variation sequences (e.g. <http://www.unicode.org/cgi-bin/GetMatchingDocs.pl?L2/11-059>); whether such an approach might have realistic prospects for approval by the Consortium remains to be seen.¹ This does not mean, however, that variation sequences for runes could not be defined for runological use and data interchange within the domain, outside of but in complement to the Unicode standard.

6. Future work

Runologists, as a user-community, frequently lack familiarity with technology and new digital approaches, and have been more or less inactive with regards to Unicode since the Runic block was published. The fact that very few runologists actively use the Unicode Runic block in their work, almost 25 years after it was created, suggests that it does not meet their needs, and this paper has attempted to explain why, and to propose ways to address those shortcomings. Further inaction and failure to engage with the digital standardisation of the script from working runologists will inevitably lead to one of two outcomes. Either no one else will do anything either, the problem will remain unaddressed, and we *still* won’t be able to satisfactorily encode runes; or someone else from outside of runic research will do something and the fate of the script will be out of our hands, guided by the whims of individuals’ particular interests – or worse.

With this in mind, the authors propose that runologists, as a user community, discuss the alternative options presented in this paper, and take action. We favour Option 4: a reduced, minimal character set, declaring current duplicates deprecated or canonically equivalent, adding characters that are genuinely not represented, and specifying a set of glyph variants for allo-graphs to be defined as variation sequences. This would require the field to unite in standards

¹Sentence adjusted and quote added 24.11.2023.

work, whether it be in collaboration with or outside Unicode. It would require work within the field to pin-down and reach consensus on a minimal comprehensive set of graphemes, and of allographs. It would then require font support for those glyph variants. Nonetheless, the authors are of the opinion that such work, while potentially challenging and onerous, would greatly benefit the field of runology and allow greater integration with other fields of digital scholarship – including digital scholarship on ancient scripts in general.

References

- [1] K. F. Seim, Runologi, in: O. E. Haugen (Ed.), *Handbok i norrøn filologi*, 2 ed., Fagbokforlaget, Bergen, 2013, pp. 128–193.
- [2] M. P. Barnes, *Runes: a handbook*, Boydell Press, Woodbridge, 2012.
- [3] T. Spurkland, *Norwegian runes and runic inscriptions*, Boydell & Brewer, 2005.
- [4] W. Lundström, Proposal to ISO/IEC JTC1/SC2/WG2 Concerning Inclusion into ISO/IEC 10646 of the Repertoire of Runic Characters, Technical Report, Central Board of National Antiquities Sweden, 1995.
- [5] T. Spurkland, Kriteriene for datering av norske runesteiner fra vikingtid og tidlig middelalder, *Maal og Minne* (1995) 1–14.
- [6] A. Palumbo, *Skriftsystem i förändring: En grafematisk studie av de svenska medeltida runinskrifterna*, Uppsala: Institutionen för nordiska språk vid Uppsala universitet, 2020. URL: <http://www.diva-portal.org/smash/record.jsf?pid=diva2:1509013&dswid=-3719>.
- [7] M. P. Barnes, Standardised fuþarks: A useful tool or a delusion?, *Arkiv för nordisk filologi* (2006) 5–22. URL: <https://journals.lub.lu.se/anf/article/view/11743/10422>.
- [8] M. Schulte, The rise of the younger fuþark, *NOWELE* 60-61 (2011) 45–68. doi:10.1075/nowele.60-61.03sch.
- [9] K. F. Seim, *De vestnordiske futhark-innskriftene fra vikingtid og middelalder – form og funksjon*, Ph.D. thesis, NTNU Trondheim. Det historisk-filosofiske fakultet. Institutt for nordistikk og litteraturvitenskap, 1998.
- [10] S. Fridell, Graphic variation and change in the younger fuþark, *NOWELE* 60-61 (2011) 69–88. doi:<https://doi.org/10.1075/nowele.60-61.04fri>.
- [11] M. P. Barnes, On the status and transliteration of the additional characters of medieval scandinavian runic writing, *Amsterdamer Beiträge zur älteren Germanistik* 67 (2011) 9–21. doi:10.1163/9789401200783_003.
- [12] K. Düwel, *Runenkunde*, 4 ed., Metzler, Stuttgart, 2008.
- [13] A. Liestøl, I. S. Johnsen, J. E. Knirk, *Norges innskrifter med de yngre runer: Bryggen i Bergen*, number VI in *Norges innskrifter med de yngre runer*, Kjeldeskriftfondet, Oslo, 1980-1990.
- [14] M. Waldspühl, *Schreibpraktiken und Schriftwissen in südgermanischen Runeninschriften*, number 26 in *Medienwandel - Medienwechsel - Medienwissen*, Chronos, 2013.
- [15] S. Nowak, *Schrift auf den Goldbrakteaten der Völkerwanderungszeit: Untersuchungen zu den Formen der Schriftzeichen und zu formalen und inhaltlichen Aspekten der Inschriften = Writing on the Migration-Period Gold Bracteates*, Ph.D. thesis, Georg-August-Universität

- zu Göttingen, Philosophische Fakultät, 2003. URL: <http://resolver.sub.uni-goettingen.de/purl/?webdoc-512>.
- [16] T. Spurkland, En fonografematisk analyse av runematerialet fra Bryggen i Bergen, Ph.D. thesis, Institutt for nordistikk og litteraturvitenskap, Institutt for arkeologi, kunsthistorie og numismatikk, Universitetet i Oslo, 1991.
- [17] L. Peterson, Scandinavian runic-text data base: a presentation, in: B. Ambrosiani, H. Clarke (Eds.), *Developments around the Baltic and the North Sea in the Viking Age*, volume 3 of *Birka studies*, Viking Congress, Birka project for Riksantikvarieämbetet and Statens Historiska Museer, Stockholm, 1994, pp. 305–309.
- [18] H. Gustavson, S. Jörsäter, Runes and the computer, in: C. W. Thompson (Ed.), *Proceedings of the First International Symposium on Runes and Runic Inscriptions*, number VII in *Michigan Germanic Studies*, Department of Germanic Languages and Literatures, The University of Michigan, Ann Arbor, Michigan, 1981, pp. 98–106.
- [19] C. W. Thompson, On transcribing runic inscriptions, in: C. W. Thompson (Ed.), *Proceedings of the First International Symposium on Runes and Runic Inscriptions*, number VII in *Michigan Germanic Studies*, Department of Germanic Languages and Literatures, The University of Michigan, Ann Arbor, Michigan, 1981, pp. 89–97.
- [20] A. Doyle, Digital ogam: Implementation and implications, 2022. URL: <https://ogham.glasgow.ac.uk/index.php/2022/05/23/digital-ogam-implementation-and-implications-guest-blog-by-adrian-doyle/>.
- [21] A. West, Babelstone: Khitan scripts, ??? URL: <https://babelstone.co.uk/Khitan/index.html>.
- [22] The Unicode Consortium, J. Becker, M. Davis, M. Everson, A. Freytag, J. Jenkins, E. Muller, L. Moore, M. Suignard, K. Whistler, *The Unicode Standard, Version 4.0*, Addison-Wesley Professional, 2003. URL: <https://www.unicode.org/versions/Unicode4.0.0/ch13.pdf>.
- [23] M. Deroñ, Proposed additions to the runic range, l2/09-312, 2009. URL: <http://unicode.org/L2/L2009/09312r-runic-additions.pdf>.
- [24] Unicode, *The unicode standard, version 7.0, runic range: 16a0–16ff*, 2014.
- [25] J. D. Becker, *Unicode 88 (standard)*, 1988. URL: <https://www.unicode.org/history/unicode88.pdf>.
- [26] E. M. Magin, *Runes, Runic Writing and Runic Inscriptions as Primary Sources for Town Development in Medieval Bergen, Norway*, Ph.D. thesis, University of Nottingham, School of English, Nottingham, 2021.
- [27] Unicode, *Glossary of unicode terms*, 2023. URL: <https://www.unicode.org/glossary/>.