



Universal Multiple-Octet Coded Character Set
International Organization for Standardization
Organisation Internationale de Normalisation
Международная организация по стандартизации

Doc Type: Working Group Document
Title: Proposal to add two characters for Middle Scots to the UCS
Source: Michael Everson
Status: Individual Contribution
Date: 2019-04-25

This proposal requests the encoding of two characters used in the orthography of Middle Scots (1450 to 1700). If this proposal is accepted, the following characters will exist:

	A7F0	LATIN CAPITAL LETTER MIDDLE SCOTS S
	A7F1	LATIN SMALL LETTER MIDDLE SCOTS S

• used in Middle Scots for *s*, *ss*, *sis*, etc.

1. MIDDLE SCOTS S. A letter unique to Middle Scots which has been presented in editions of Middle Scots texts is not encoded in the UCS. It derived originally either from a ligature of *s* and long *s* or from a swash final form of long *s*. It is polyvalent; it can be read as a single *s*, as a double *ss*, as a syllable *is* or *sis*, or as a logogram for *schir* or *sir*. Its use did not survive the Middle Scots period, and its status in Middle Scots texts is not the same as that of any ligature of *fs* elsewhere in Britain or in Scotland after 1700. By that time print technology had taken hold, and the ordinary sequence *fs* was used purely mechanically, with *s* being reserved to absolute final position. In the English- and Scots-speaking world, no ligature of *fs* is common during the period where long *s* is current, except occasionally in italic type.

In Scots texts, however, the polyvalence of the character has been respected in many printed editions, though not in those which replaced it editorially with *s*, *ss*, or *sis*. In those which do, there three glyphs have been observed, as shown in the figures below. The most natural of these looks like a long *s* with a sort of extended cedilla or vertical tilde hanging from it. This is a reasonably fair typographic representation of the forms the letter takes in Scots manuscripts, and is the form suggested for the code chart glyphs. In his introduction to Scots handwriting (1973, reprinted with corrections 2009), Grant Simpson describes the MIDDLE SCOTS S:

The letter *s* had two entirely distinct forms, one being more or less the modern form, the other a long vertical stroke with a curved stroke added to the top. This second form of long-*s* looks like an *f* without its cross-stroke. (For both forms use in one word, see **3**, like *2*, *presentes*.) In the later middle-ages and thereafter there was a marked tendency to use long-*s* initially and in the middle of a word and to employ the other form as a final letter only, but this was not an invariable rule. In vernacular texts some words may end in a long-*s* which has a curl attached to it. In form this addition may be either a curl backward and below the line (e.g. **10**, line 5, *Burges*'), or an *s*-shaped curl (e.g. **12**, line

25, *als*)... Occasionally the sense demands that it be taken to mean -s or -is (e.g. 29, line 8, *houssis*), but this is unusual.

The examples Simpson cites are given in Figures 5, 6, and 29.

A similar description is given by W. Mackay Mackenzie (1932, reprinted with corrections 1960) the editor of William Dunbar's poems (Dunbar is the Scottish Chaucer):

The upright *s* with an ornamental curl... is usually printed as *ss*. This seems to be unjustifiable, and it is here represented by a single *s*. If this results in such spellings as "pas," "las," glaidnes," it must be added that it also spares us such as "thuss," "thiss," "wass," while "pas," "sadnes," and "gladnes" do occur in these spellings both in MS. and in later printed texts, e.g. *Philotus* (1603), and we have a similar form in "princes" for "princess," while "wilfulnes" with the final ornamental *s* in one MS. of a poem is in another spelled in the same way with an ordinary *s*. Where this form of the letter occurs initially, it can scarcely stand for *ss* and give "sservis" (5.12). In two or three instances, the ornamental *s* seems to stand for *is*, e.g. 1. 16, 17.

Mackenzie's edition of Dunbar's poetry is for the general reader, and his treatment of the MIDDLE SCOTS *s* is not much different from that of editors (like Mackenzie himself) who substitute *th* for *p* or *y* for *ȝ*. For an accurate representation of Middle Scots orthography, however, the MIDDLE SCOTS *s* is most definitely required. Note that Mackenzie, like other authors, recognizes the polyvalence of the character (standing for *s*, *ss*, *is*).

2. Relation to the LATIN LETTER SHARP S. The Middle Scots *s* is not identical to the German sharp *s*. Mark Jamra 2006 has described the evolution of the German ß:

In the time between AD 750 and 1500, Old High German and subsequently Middle High German had two s-sounds:

- 1) one like the *s* in *Gāst* [*Eng.* guest] and *Maus* [*Eng.* mouse] (long-*ſ* and short-*s* were both in use), and
- 2) a slightly lisped *s* spoken against the teeth and usually spelled with *z* in words like *eẏzen* [*contemp. Ger.* essen; *eat*] and *uẏ* [*aus*; *out*]. At the same time, the letter *z* was also used to denote the "ts" sound, which is its function in German today. Therefore, in an Old High German word like *ſiẏzan* [*sitzen*; *sit*], one couldn't see from reading *z* whether it was pronounced "sis-san" or "sit-san." To remedy this situation, scribes began as early as the 9th century to place an *f* before the *z* to indicate the "ss" pronunciation. For example, *gruẏ* became *gruẏf* [*groß*] and *daẏ* became *daẏf* [*daß*; *that*]. These two letters were eventually combined into a ligature and thus the name "eszett" [Fig. 1]. Interestingly, the lisped *s* of Old- and Middle High German is no longer spoken and so the character ß (*fz*) is actually obsolete.

In terms of the representation of the Middle Scots *s* in the UCS, two things can be observed.

1. The glyph shapes for the Middle Scots character are not suitable for use in German. There are many (indeed *very* many) glyph variants of the German character; four distinct variants of ß are in use in Antiqua fonts:

- fs fß fb/fz B

ß

The use of “alþ” in comparisons in Middle Scots.
THE USE OF “ALÐ” IN COMPARISONS IN MIDDLE SCOTS.
THE USE OF “ALÐ” IN COMPARISONS IN MIDDLE SCOTS.

... s < s << S < s << S < **ß** << **B** < f < f < f << Σ ...

Page 3

- Simpson, Grant S. 2009. *Scottish handwriting 1150-1650: An introduction to the reading of documents*. Edinburgh: John Donald. ISBN 978-1-906566-11-1.
- Skeat, W. W., ed. 1870. *The Bruce; or, The Book of the most excellent and noble prince, Robert de Broyss, King of Scots, compiled by Master John Barbour, Archdeacon of Aberdeen, A.D. 1375*. Part I. (Early English Text Society; Extra Series: 11). London: Kegan Paul.
- Skeat, W. W., ed. 1894. *The Bruce; or, The Book of the most excellent and noble prince, Robert de Broyss, King of Scots, compiled by Master John Barbour, Archdeacon of Aberdeen, A.D. 1375*. Volume I. (Scottish Text Society). Edinburgh and London: William Blackwood and Sons.
- Smith, G, Gregory. 1902. *Specimens of Middle Scots*. Edinburgh and London: William Blackwood and Sons.

8. Figures.

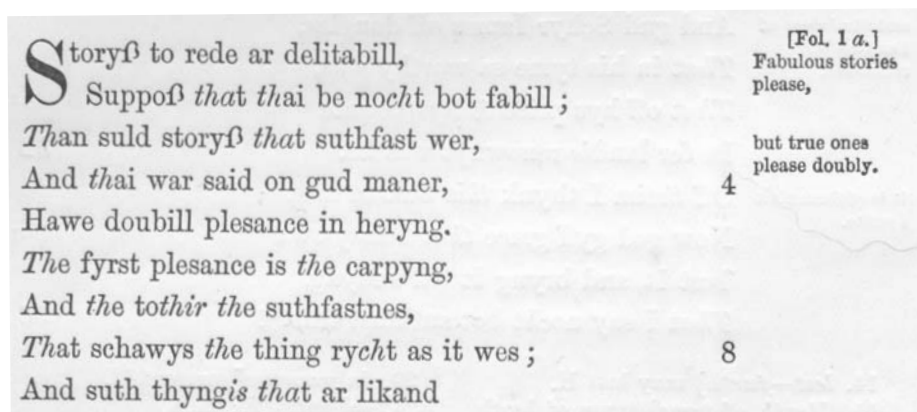


Figure 1. Example from Skeat 1870 showing LATIN SMALL LETTER MIDDLE SCOTS S in the first three lines. This is the most appropriate typographic form to use for the character, as it exemplifies best what is actually in the manuscripts. The type for this was cut in London; I have not seen an earlier example of it.

The MSS. are carefully followed, except that I have, of course, supplied the punctuation. The symbol “f” is employed to represent a very similar symbol in the MS. Its proper signification is “ss”; but it is frequently used in place of final s, as in ll. 1, 2, and 3. In a few cases, it has to be taken to represent a final “sis,” as in the case of “houf”=“housis,” vii. 163, viii. 514. When this symbol stands *alone*, it represents “schir” or “sir”; and I have so expanded it. The symbol *lk* is used for *kk*

Figure 2. Example from Skeat 1894:lxviii showing LATIN SMALL LETTER MIDDLE SCOTS S in the first three lines. As in Figure 1, the font here has a long-s shape with a sort of wiggle dangling from it. This glyph is good; it also cannot be confused with the German ß. This type may have been cut in Edinburgh but matches the glyph of the 1870 edition.

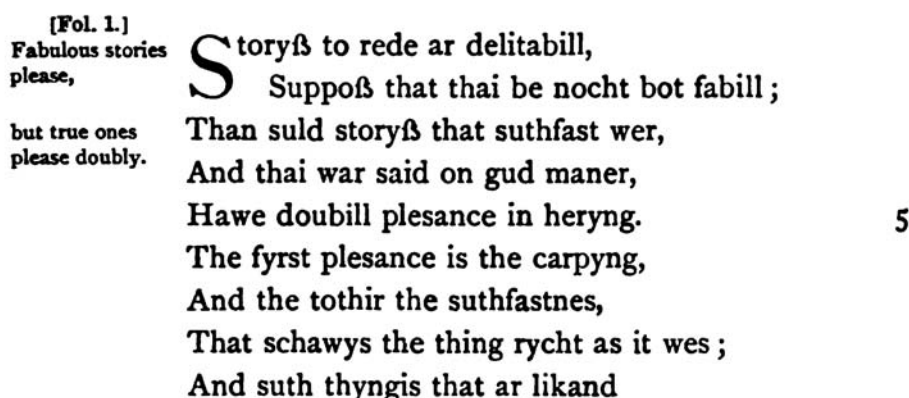


Figure 3. Example from Skeat 1894:119 showing LATIN SMALL LETTER MIDDLE SCOTS S in the first three lines. Here the glyph is very much different from what is in the manuscripts; this image is not very clear but in Figure 4 below a clearer version can be seen. Here the typesetter tried for a compromise ligature of long-s and s fs; this isn’t so satisfactory as it differs from the manuscripts and looks a bit too much like a German ß—though this glyph would hardly be acceptable for German.

27. 'ß' is an orthographical device without any phonological value. As a sign it is nothing but an elaborately formed *s*, for which letter it generally stands. Occasionally it must be interpreted as *s's*, but the symbol does not mean that. Cf. *chesß* = *chess's*, 244. 15. In the Adv. Lib. MS. of Bellenden's *Livy* the scribe writes *housß*; in the Boyndlie MS. it is *housis*. Sometimes it is equivalent to *se*: cf. Knox (190. 1, 11, &c.), where *horsemen* and *horßmen* are used indiscriminately. When there is a double *s* the 'peculiar' is confined to the second place. It is desirable to retain it in reprints of old texts for philological purposes, in order to avoid dispute as to its value in special cases. In this volume the double form is printed 'sß.' The usual custom (in the E. E. T. S. and S. T. S.) is to print 'fß,' but this appears inconsistent when the obsolete long *s* is modernised in all other positions. The 'ß' is retained to show that it is a separate form.¹

Figure 4. Example from Smith 1902:xxx showing LETTER SMALL LETTER MIDDLE SCOTS S. Here the example from Skeat 1894 (Figure 3) can be seen more clearly. Note in particular the lack of harmonization with the serifs of lowercase *s* in the same font:

s ß

The typescutters could certainly have used a German glyph if they thought that this was suitable; they did not, and the glyph here is not one of the glyphs that could be used for German without attracting attention and criticism. The MIDDLE SCOTS S and the SHARP S should not be unified. Note too that the typescutters did not cut italic type for the glyph. A good recommended glyph for the two is this:

chesß chesß horßmen horßmen
CHESß CHESß HORßMEN HORßMEN

Compare this with the ordinary standard German ß:

groß groß Heßmann Heßmann
GROß GROß HEßMANN HEßMANN

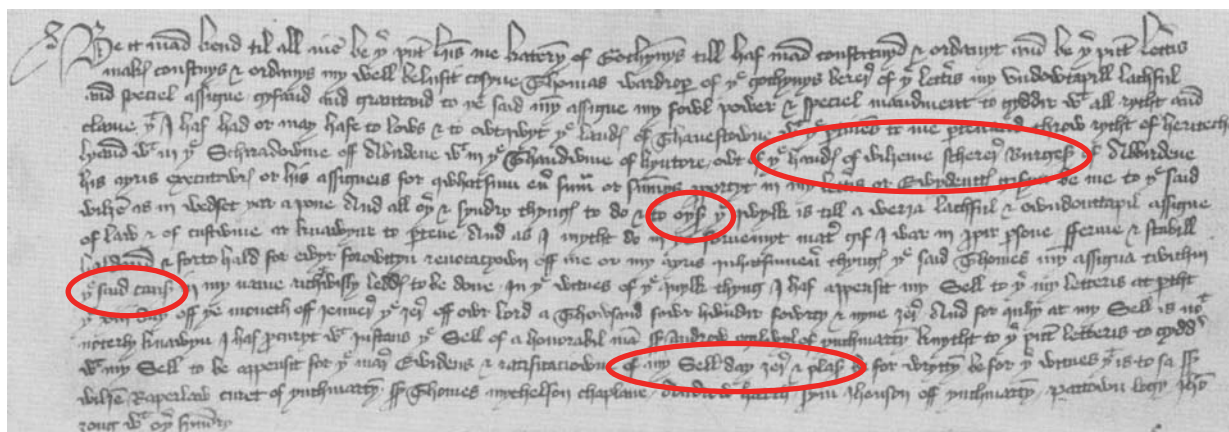
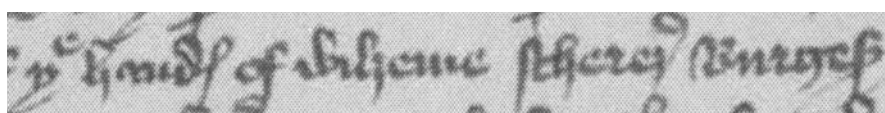
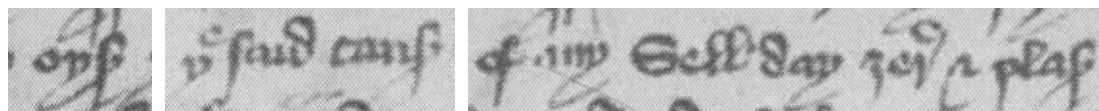


Figure 5. Example from Simpson 1998: Plate 10 (a letter, dated 1449), showing MIDDLE SCOTS S alongside U+A76D LATIN SMALL LETTER IS used for -is and alongside ordinary -is. Date 1449/50. Here are some close-ups:



Line 5: p^e handf of wilzeme fcherer Burges. ‘the hands of William Scherer Burges’. The glyph here is interesting; it is quite similar to some glyphs used for U+1E9C LATIN SMALL LETTER LONG S WITH DIAGONAL STROKE used for *spir/sper* in Latin, though it would not make sense to give a palaeographic reading with that character. Clearly this is a differently modified long s in the same tradition with the other Middle Scots manuscripts. There is also an example of the COMBINING OVERCURL here.



Line 7: oys ‘use’. Line 10: p^e said caus ‘the said cause’. Line 13: of my Sell day 3er 7 plaf ‘of my selling day year and place’. There is also an example of the COMBINING OVERCURL here.

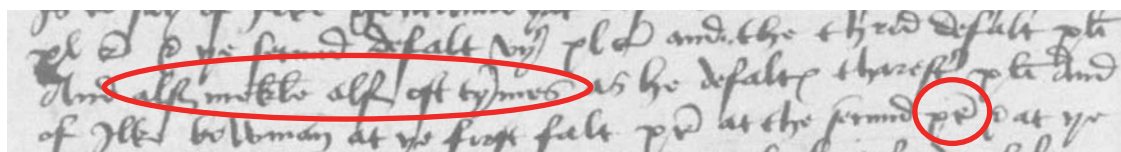
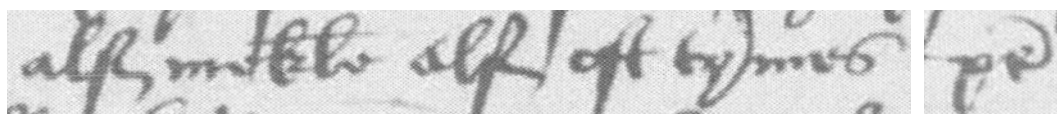


Figure 6. Example from Simpson 1998: Plate 12 (an act of parliament, 1491), showing MIDDLE SCOTS S alongside U+A76D LATIN SMALL LETTER IS used for -is and alongside ordinary -is. There are also examples of the COMBINING OVERCURL used with a solitary s to indicate *shilling(s)* and with y to indicate *ym*. Here are some close-ups:



Line 2: als mekle als oft tymes ‘as much as oft times’; line 3: x s ‘10 shillings’.

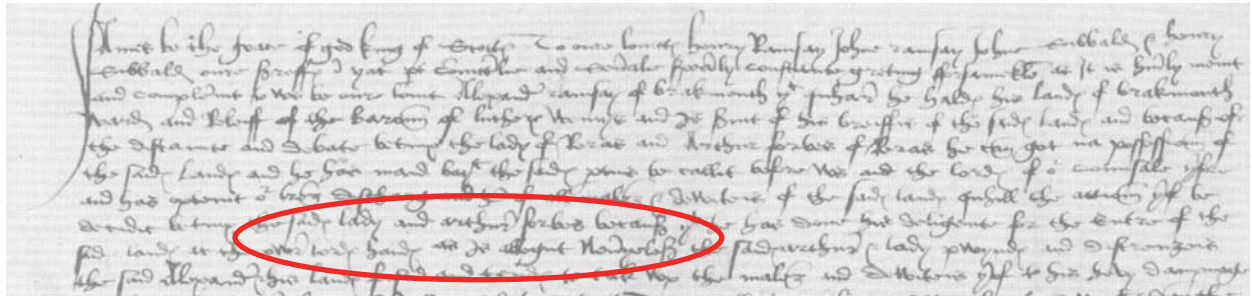
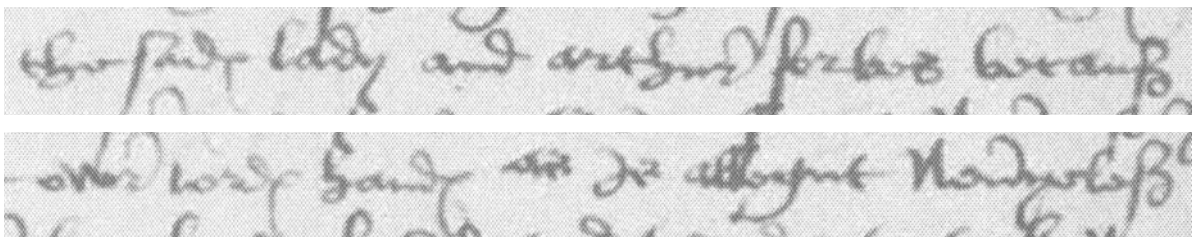


Figure 7. Example from Simpson 1998: Plate 13 (a letter by James IV, 1494), showing MIDDLE SCOTS s alongside U+A76D LATIN SMALL LETTER IS used for *-is* and alongside ordinary *-is*. Here are some close-ups:



Line 5: the said^l lady and arthur^r forbes becau^ß. ‘the said lady and Arthur Forbes because’. Line 6: ow^r lord^l hand^l as Is allegiit Neüpeleß ‘our lord’s hands and is alleged nevertheless’. The final round flourish on the glyph here is interesting; this text was written in 1494, and the shape of the MIDDLE SCOTS s is *accidentally* similar to the Sulzbacher sharp s (ß) which was devised only in 1879.

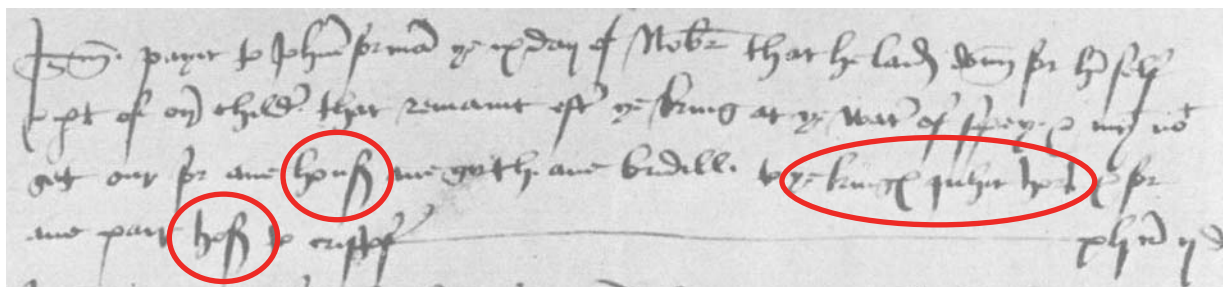
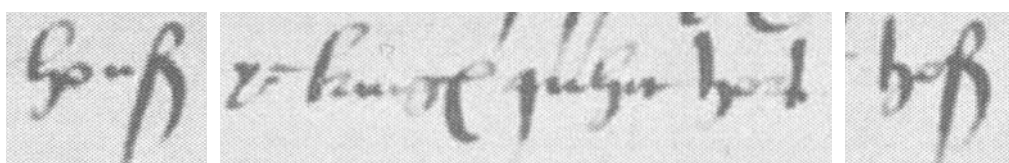


Figure 8. Example from Simpson 1998: Plate 14 (treasurer’s account, 1505). The text here is not easy to read, but says:

ITm̄. payit to John^e formā^e þe ix day of No^bz that he laid^e down^e for hīself | ȳ pt of oþ^r child^e that zemanit eff^e þe king at þe water of Spey. ȳ my^t not^e | get our for ane hous^e ane girth. ane b2idill. to þe king^e qhuit horf ȳ for | ane pair hos^e to cristof^r

“Item: paid to John Forman the 9th day of November that he laid down for himself and part of other children that remained after the king at the water of Spey, and might not get over for one house one garden, one bridle, to the king’s white horse and for one pair horse to Christopher”

Here are some close-ups:



Line 3: hous^e ‘house’; þe king^e qhuit horf ‘the king’s white horse’; line 4: hos^e ‘horse’.

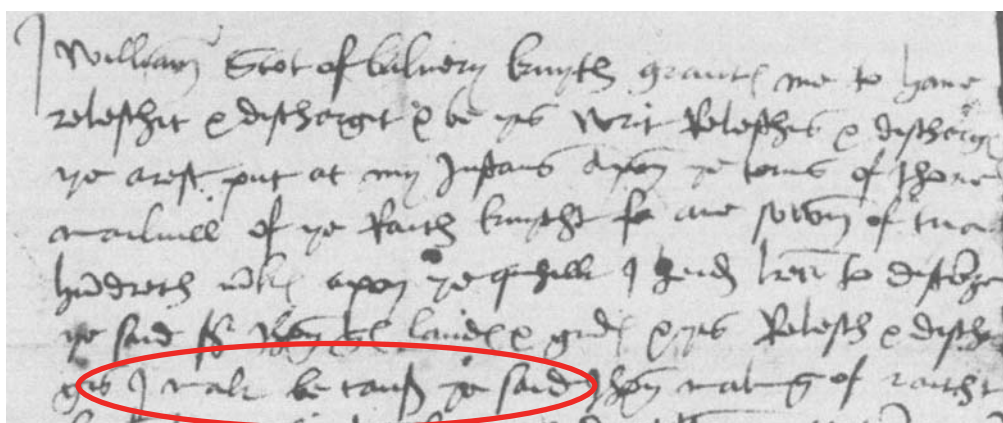
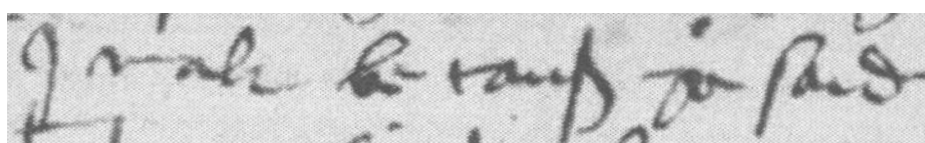


Figure 9. Example from Simpson 1998: Plate 15 (a prisoner discharge, 1516), showing MIDDLE SCOTS s. Here is a close-up:



Line 7: I mak be causß þe said. ‘I make because the said’. Here the descending second half of the glyph is quite long indeed.

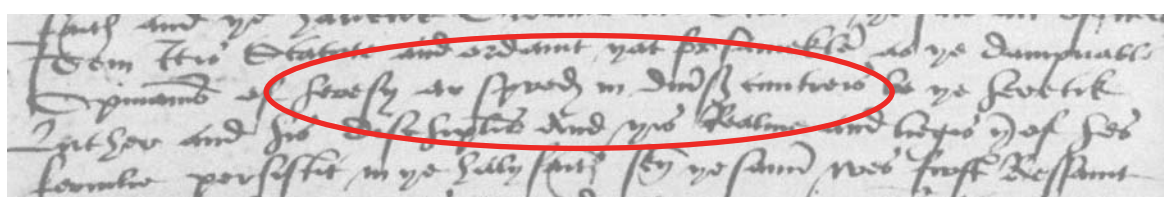
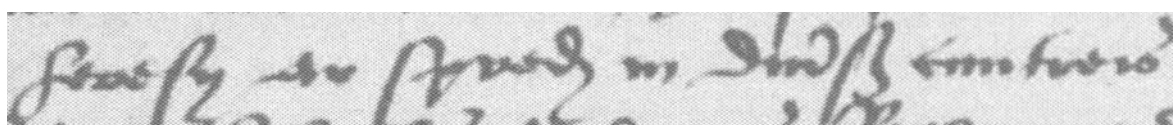


Figure 10 Example from Simpson 1998: Plate 16 (an act of parliament against Luther's heresy, 1525), showing MIDDLE SCOTS s. Here is a close-up:



Line 2: [Opinionis of] heresy ar spred in diuß cuntreis. ‘[opinions of] heresy are spread in diverse countries’.

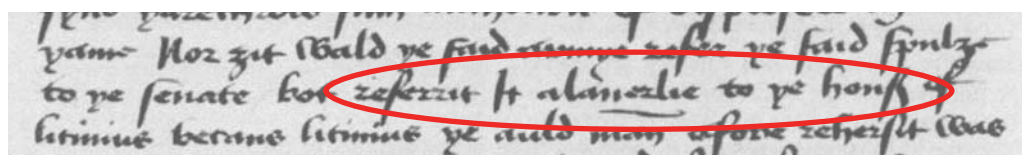
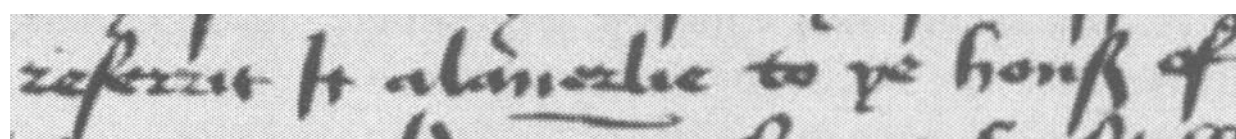


Figure 11. Example from Simpson 1998: Plate 17 (a translation of Livy's *History*, 1540), showing MIDDLE SCOTS s. Here is a close-up:



Line 5: referrit It alanerlie to þe hous of [licinius]. ‘referred it solely to the house of [Licinius]’.

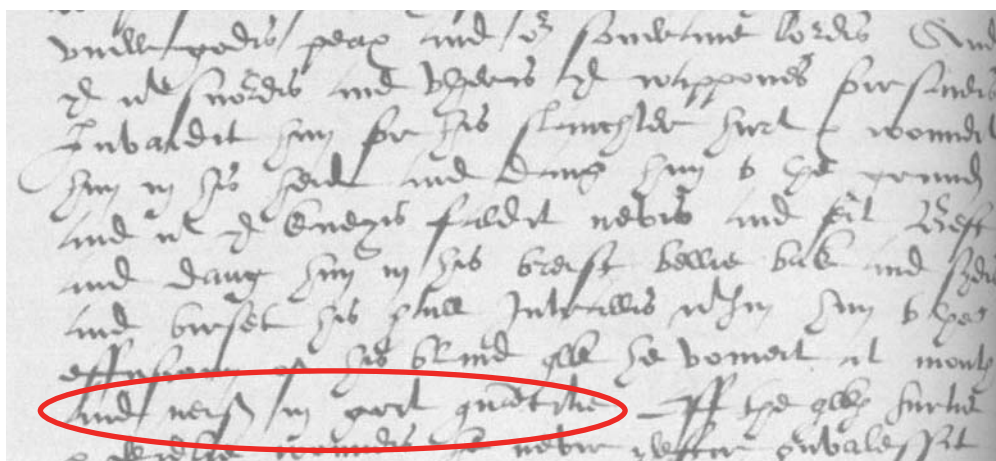
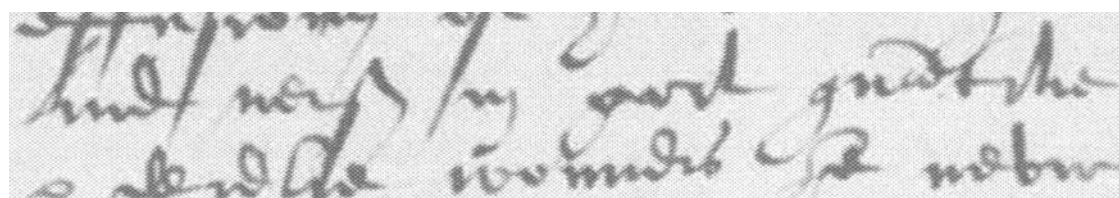


Figure 12. Example from Simpson 1998: Plate 25 (legal text describing a murder, 1609), showing MIDDLE SCOTS S. The text is rather gruesome: And I þ w^t fuordis and vtheris þ wappones foirfaidis I Invaidit him for his flauchter huzt ⁊ woundit him in his heid and dang him to the grund ⁊ and w^t þ kneysis faldit nevis and feit Beft I and dang him in his bæist bellie bak and sydis I and burfet his haill Intrallis wthn him to the I effufioun of his bluid qlk he vomeit at mouth I and neiß in grit quātitie Off the qlk^j hurtis I ⁊ deidlie woundis he nevir þeſter conualeſſit.

“And there with swords and others their aforesaid weapons invaded him for his slaughter hurt and wounded him in his head and beat him to the ground and with their knees clenched fists and feet struck and beat him in his breast, belly, back, and sides and burst his whole entrails within him to the effusion of his blood which he vomited at the mouth and nose in great quantity. Of the which hurts and deadly wounds he never thereafter convalesced.”

Here is a close-up:



Line 5: and neiß in grit quātitie. ‘and next in great quantity’.

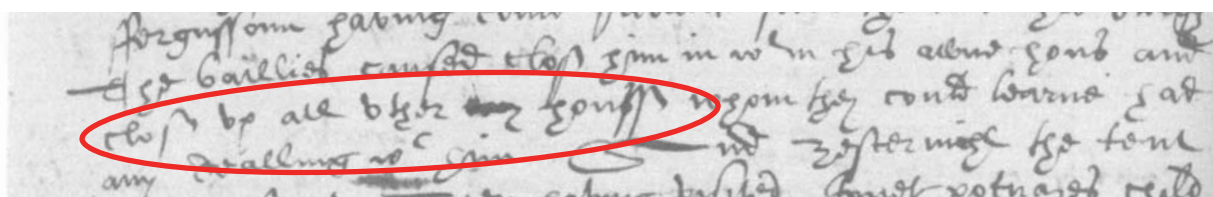
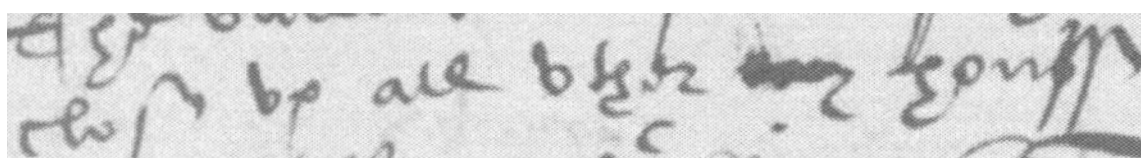


Figure 13. Example from Simpson 1998: Plate 29 (town council minutes, 1647), showing MIDDLE SCOTS S. Here is a close-up:



Line 5: cloß vp all vther houß ‘close up all other houses’. The word after “vther” is a deleted error.

A. Administrative

1. Title

Proposal to add two characters for Middle Scots to the UCS

2. Requester's name

Michael Everson

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

Individual contribution.

4. Submission date

2019-05-25

5. Requester's reference (if applicable)

6. Choose one of the following:

6a. This is a complete proposal

Yes.

6b. More information will be provided later

No.

B. Technical – General

1. Choose one of the following:

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

No.

1b. Proposed name of script

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

Yes

1d. Name of the existing block

Latin Extended-D

2. Number of characters in proposal

2.

3. Proposed category (A-Contemporary; B.1-Specialized (small collection); B.2-Specialized (large collection); C-Major extinct; D-Attested extinct; E-Minor extinct; F-Archaic Hieroglyphic or Ideographic; G-Obscure or questionable usage symbols)

Category A.

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?

Michael Everson.

5b. If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools used:

Michael Everson, Fontographer.

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. 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.

8. Submitters are invited to provide any additional information about Properties of the proposed Character(s) or Script that will assist in correct understanding of and correct linguistic processing of the proposed character(s) or script. Examples of such properties are: Casing information, Numeric information, Currency information, Display behaviour information such as line breaks, widths etc., Combining behaviour, Spacing behaviour, Directional behaviour, Default Collation behaviour, relevance in Mark Up contexts, Compatibility equivalence and other Unicode normalization related information. See the Unicode standard at <http://www.unicode.org> for such information on other scripts. Also see Unicode Character Database <http://www.unicode.org/Public/UNIDATA/UnicodeCharacterDatabase.html> and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.

See above.

C. Technical – Justification

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

No.

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.)?

No.

2b. If YES, with whom?

2c. If YES, available relevant documents

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

Germanicists, Anglicists, dialectologists, lexicographers, and Scots.

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

Common in Middle Scots.

4b. Reference

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

Yes.

5b. If YES, where?

Various publications.

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

Yes.

6b. If YES, is a rationale provided?

Yes.

6c. If YES, reference

Accordance with the Roadmap. Keep with other Latin characters.

7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?

No.

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

No.

8b. If YES, is a rationale for its inclusion provided?

8c. If YES, reference

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

No.

9b. If YES, is a rationale for its inclusion provided?

No.

9c. If YES, reference

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

Yes.

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

Yes.

10c. If YES, reference

Discussuon of the similar but unrelated German sharp s is given above.

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)?

Yes.

11b. If YES, is a rationale for such use provided?

No.

11c. If YES, reference

11d. Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?

No.

11e. If YES, reference

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

No.

12b. If YES, describe in detail (include attachment if necessary)

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

No.

13b. If YES, is the equivalent corresponding unified ideographic character(s) identified?