

## WG2: Transcription of Early Letter Forms

### Brian Hillyard

*{This is the first of two or possibly three position papers for this working group—DJL}*

I should explain that quite deliberately I have not gone back through the archives on this since at this stage I wanted to avoid getting bogged down in arguments. What I've tried to do is to assemble as much relevant data as I could, and then at the very end I've indicated where I think all this data leads us.

Transcription as used in implementing DCRB requires instructions covering

- which words are transcribed (i.e. permitted omissions)
- the order in which words are transcribed (i.e. transposition issues)
- what punctuation is transcribed
- how characters themselves are transcribed

This paper tackles the fourth group of instructions, i.e. how on a character-by-character level text<sup>1</sup> from areas 2, 4, 6<sup>2</sup> is transcribed in a catalogue record, but it doesn't tackle this comprehensively – I've concentrated on the I/J and U/V issue.

#### RELEVANT DATA

(1) In the context of shared MARC cataloguing it is not possible (or at least practicable) for characters to be faithfully transcribed in all cases. Examples of what cannot be faithfully transcribed are:<sup>3</sup>

- the contractions and ligatures of 16th- or 17th-century Greek fonts (even if a modern Greek character set is being used)
- upsidedown (“turned”) letters
- “v”s filed down and used side by side to represent “w” (i.e. two pieces of type for one letter)
- more widely a whole range of extremely common ligatures (e.g. “ct”), though there are some (e.g. “æ”) that can be transcribed

---

<sup>1</sup> We cannot say “letterpress” since the instructions cover engraved title pages, and “words” is insufficiently broad. I'm not sure that “data” or “information” excludes (e.g.) printer's rules.

<sup>2</sup> It is very interesting that for the Physical Description area 5B1 instructs “Give arabic and roman numerals as they appear in the publication. Give roman numerals uppercase or lowercase as they appear” (examples given include “XII, 120 leaves” and “Ixiij, [1] p.”) even though the instruction at the beginning of 5B1 is “Give the complete number of leaves ...” and later explains “Giving the complete number as stated above means giving the number on the last numbered page ...” and therefore suggests that the figure given is no more than a method of giving a total. Given the normal DCRB attitude towards capitalisation, it seems a bit odd that it should be thought important to preserve uppercase roman numerals. But I see that this is in AACR 2.17 (though not the “Ixiij” example). I suppose it allows for a book extent “xii, XII, 120 leaves”, though this must be one in a million. Does the serious DCRB approach to this mean that if the majority of pages are numbered in uppercase but some are numbered in lowercase, this should be noted?

<sup>3</sup> There is the additional special case of punctuation (“...” and “[ ]”) that cannot be transcribed as it stands because in catalogue records governed by DCRB rules those signs have special meanings: see OF (final paragraph), instructing that normal transcription is replaced by “—“ and “( )”. This is most unsatisfactory since the replacements might themselves be found on title pages. Clearly there has to be a note, and if there is a note anyway, what is to be gained by transcribing them in this different way?

(2) In addition, while there is no technical problem with transcribing uppercase letters, it is a well-established convention across the whole of cataloguing that when transcribing for a catalogue record the use of uppercase or lowercase is determined by the current conventions of the language concerned and not by the use of uppercase or lowercase as found in the book. It is very important to note this: it is nothing to do with the difficulty of accurately transcribing in lowercase or uppercase as required, but rather is done for readability, esthetics, etc. Similarly in 4D2 the normal approach to roman numerals is “change them to arabic numerals unless they are erroneous or misprinted”.

(3) It follows from (1) and (2) that sometimes parts of the transcription will not be transcription at all as they will include

- many lowercase letters used to transcribe their uppercase equivalents and, though less often, uppercase letters used to transcribe their lowercase equivalents (though on this see ISBD(A) as mentioned in para (9) below). The only reason for doing this is to make it more “comfortable” for us to read. This procedure is carried out completely silently.
- modern forms of letters in place of earlier forms: e.g. modern “s” for the earlier character known as a long “s”, and all ligatures that are not invariable in modern usage.<sup>4</sup> Square brackets are not used, and it seems to be normal not to make notes about this. This practice is simply adopting a pragmatic stance (this seems to me to be slightly different from reflecting the intention of the printer since he was not using the one character intending it to stand for the other).
- “w” for “vv” when two “v”s have been filed down to represent “w” or “hand” for “hand” with the “n” printed upside down. In these cases the transcription will reflect the *intention* of the printer. Square brackets are not used, and it seems to be normal not to make notes about this.
- initial letters of words where the printer has left a blank space for an illuminator. 0G instructs – rather oddly<sup>5</sup> – that in these cases there should be notes but no square brackets.
- characters within square brackets for a variety of purposes from expanding contractions (e.g. “amico[rum]”) to giving a whole word in place of an abbreviation (e.g. “[et]” for an abbreviation signifying that)

In some of these cases notes could be made, but I think that to insist on notes in all the above cases would be found unacceptable

(4) It is a fact that in some printing locations and in all or part of some earlier centuries (e.g. the 16<sup>th</sup> century) the letters “u” and “v” were regarded as two ways of printing the same letter, and that the practice was to use V anywhere in a word when uppercase was being used but “v” at the beginning of a word and “u” medially when lowercase was being used. It follows from the convention described in (2) above that a cataloguer will sometimes convert uppercase V to a lowercase equivalent. If a

---

<sup>4</sup> According to rule 0H the exceptions to the instruction to transcribe a ligature by giving its component letters separately are those cases where modern usage preserves the ligature. The final sentence there, “When there is any doubt as to the correct conversion of elements to modern form, transcribe them from the source as exactly as possible.”, implies that the main objective is to modernise; being faithful to the original comes second.

<sup>5</sup> In the case where no letter is printed at all, not to use brackets seems an extraordinary decision (the case where there is a guide letter is different).

cataloguer followed what “the man on the street” would regard as normal conversion procedure (i.e. “a” for “A”, “b” for “B”, and so on), when he came to transcribe “V” as “v”, he would not always end up with what the printer would himself have printed in lowercase (sometimes “u”, sometimes “v”).

(5) Within the restrictions imposed by the character set (these same restrictions govern the title/word searches with which users interrogate the catalogue), MARC cataloguing provides the facilities for providing added entry access points that free the transcription from the need to take access into account: i.e transcription rules need not be influenced by access needs. (After writing this I wondered about spelling – and keywording searching – in edition statements or imprints.)

(6) One category of catalogue user – the bibliographer – will be aware from such standard books as Bowers, *Principles of Bibliographical Description*, that what is called “simplified” (as opposed to “quasi-facsimile”) transcription has always converted uppercase to lowercase, to accord with modern usage, and followed the principle of using the lowercase letters that the printer would have used. Users in this category *may* presume that this is the convention followed in rare book cataloguing.

(7) There is no evidence known to me for the presumptions of any other category of catalogue user about conventions followed in the converting of uppercase to lowercase, but it would be a reasonable guess that the majority of them will not know the background to this and will assume that in catalogue records we read letters as used by the printer. Appendix A, 0H, says “Make an added entry for the modern orthography for a title proper in which I/J and U/V have been transcribed according to pre-modern conventions, when the modern version would differ from the title as transcribed. If necessary [i.e. if the added entry would turn out different?], also make an added entry for a form of the title in which all letters are transcribed as they appear in the source, but giving only initial letters in uppercase [e.g. IVNIVS would be transcribed “Junius” in the title area but “Ivnivs” in an added entry].” This instruction recognises the needs of uninformed users.

(8) No catalogue user has any way of knowing about any lowercase letter in a transcription that it has been converted from uppercase (see (10.4) below). If (exceptionally) a user becomes aware that letters have been converted from uppercase to lowercase or vice versa, those who have no knowledge of bibliographical description will most of them assume that the current exact equivalent has been used (i.e. “v” represents “V” and “u” represents “U”).

(9) BDRB was published in 1981 and so for over 20 years now in English-speaking countries catalogue records have been created using this simplified transcription. There are in particular a large number of ESTC records that have adopted the current I/J and U/V conventions, though note that ESTC transcribes æ/Æ and œ/Œ. I have not yet discovered clear evidence of practice followed in continental Europe, but it is likely to be similar given that DCRB follows ISBD(A) (2<sup>nd</sup> ed., 1991) 0.8 “In converting capitals to lowercase, the usage (including that of diacritics) in the publication being described should be followed. The following usage is recommended for converting I, J, U, V and VV where practice is not consistent: ... Black letter capitals in the form J or U are transcribed as I or V.” DCRB is in

generally in harmony with ISBD(A) with one exception: ISBD(A) prescribes **“lower case letters are never transcribed into capitals”**.

(10) What are the uses of catalogue records involving I/J and U/V transcription?

(10.1) Those studying the printing history of 15th-17th-century books – i.e. “bibliographers” – search library catalogues to find different editions and/or different issues/states. Such people are quite likely to be acquainted with “simplified” transcription and will presume use of such a system.

(10.2) Other students, who are using 15th-17th-century books as primary sources of information and do not require a “bibliographical” approach, are likely to search for books as they find them described in footnotes or a list of sources consulted or similar, and in searching OPACs they will probably use modernised spelling, though occasionally they may use “simplified” transcription if that is what is in front of them. Of course, this will only matter if they search by title or keyword in title; if they are using controlled vocabulary (names, subjects, etc.), the details of the transcription of the title will make no difference. And if they are searching by title, additional access points (Appendix A, 0H) should ensure success: the details of transcription will be relatively unimportant.

(10.3) Cataloguers, working book-in-hand, may (a) search for records to download/copy for their own catalogue or (b) want to add holdings to bibliographic records in union catalogues (e.g. ESTC). In the case of (a) the details of transcription will be unimportant since the records will be checked against the book and made to conform to the standards the cataloguer is implementing. In the case of (b) the cataloguer is likely to have the advantage of being able to find out the standards followed by the union catalogue, and, in addition, as the years pass, the likelihood of having access to a digitised image of the title page will increase (e.g. a cataloguer would now compare the title page of a pre-1701 ESTC book with the image available through EEBO, though except for cases when the ESTC record is based on the same copy as that filmed for UMI, there is no guarantee that the transcription in an ESTC record is of a title page identical to that in EEBO).

(10.4) It could be argued that for all users for whom the details of the title as found on the title page is important, it is not so much what the standard is that matters as *knowing* what it is. In the course of writing these notes, the experience of examining records for 16<sup>th</sup>-century books to discover the standard used has convinced me that it’s quite impossible to find out in this way because it is impossible to *know* when conversion from uppercase to lowercase has taken place. Among non-librarians, the “bibliographer” is the only user who will want to know what is on the title page, and he will be expecting the system used to be “simplified” transcription.

(10.5) We should also examine this from the point of view of the cataloguer who is required to create catalogue records for books with these transcription problems. Are the instructions for I/J and U/V transcription too difficult to follow? I think they are as clear as they could be. Both ISBD(A) and DCRB emphasise that the “table” is only for use when practice, usage, etc. is not decisive. There is a minor difference in that ISBD(A) refers to “the usage ... in the publication being described” and DCRB to “the pattern ... employed by the particular printer”. When you compare these, it looks

as though DCRB is allowing for other publications to be taken into account when the publication in hand does not allow a practice to be determined: I think this is unhelpful and prefer the ISBD(A) wording.

## CONCLUSION

The key points that emerge are

- there is a very well and long established system (followed by ISBD(A)) of “simplified transcription” that is known to those users to whom this exact transcription of the title pages of 15<sup>th</sup>-17<sup>th</sup> century books is most important
- a large number of MARC records have already been created using this system
- there is no clear view that this system is impossibly difficult to apply
- the present general attitude towards transcription in rare book catalogue records is not sufficiently rigorous so as to apply pressure to moving to a system of transcription that accepts conversion to lowercase but is otherwise exact, e.g. transcribing V as v and U as u regardless of other considerations.<sup>6</sup>

There is nothing here to persuade us to change DCRB’s present approach to I/J and U/V transcription.

6 February 2003

---

<sup>6</sup> However, it is strongly against a change in the practice of transcribing lowercase letters as they stand, e.g. u as u and v as v (i.e. LCRI on u/v should not be applied to pre-1801 publications).