Computer Aided Editing: Arguments Pro and Contra

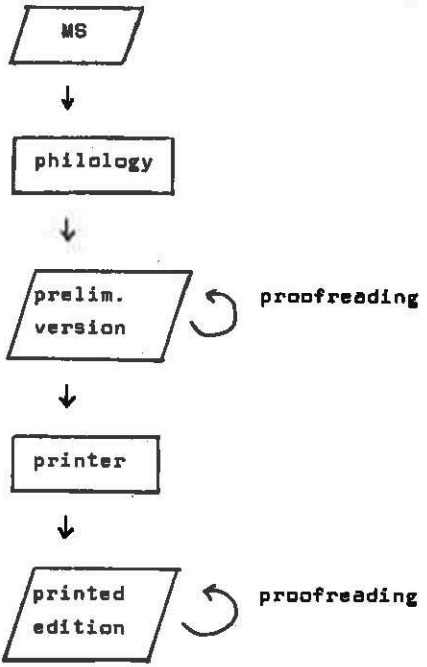
by Andrea van Arkel

Those of us who have in recent years published editions or indeed printed books of any kind, will be aware of the fact that the printing trade has become computerized to a large extent. Just now there seem to be only two possibilities:

1. the typewritten (and subsequently offset or photocopied) book, and
2. the computer typeset book.

I expect, however, that the typewritten book will be replaced within the next few years by what I will call the home-made computer typeset book, where the author types his text into a microcomputer and prints the book in a more or less sophisticated way depending on the quality of the available printer. Possibilities here range from simple matrix printers via daisy wheel printers and golf-ball typewriters to advanced programmable laser printers. As future editions of any type will very likely get into contact with a computer at some stage, with all the problems and advantages involved, it may be worthwhile to consider how to integrate the computer into the editing process.

In order to trace the advantages and disadvantages of using a computer in the editing process let us analyze what happens when one makes an edition in the classical way.



Starting of from the manuscript(s) one writes down the text; according to the type of edition this can be a narrow or broad transcription of the MS, with or without indication of abbreviations and the like, normalized to various degrees, or critical. When after several rounds of corrections and changes a final text is produced, one has usually a typewritten version with a considerable number of handwritten corrections. Retyping to produce a better looking script is usually avoided as much as possible, as this tends to introduce fresh errors in the text. This typewritten version is then sent to the printer. There the text will be retyped into a computer (this apparatus goes under various names and is usually part of a larger set-up, but essentially it is a microcomputer with some special functions). To the text as it stands, markers are added to indicate such things as new paragraphs, change of font, font size, bold face, italics etc. Commands to decide width and length of the printed page and leading are also inserted. This version will be stored on some sort of storage medium like paper tape, magnetic tape, floppy disks or fixed disks. A special

printing program converts this version to instructions for the actual printing machinery. The crucial point is of course the retyping that is needed. After this retyping the text which was carefully checked several times and compared with the MS, corrected and corrected time and again and finally deemed as free of errors as humanly possible, will sport any number of fresh errors. Repeated proofreading will be needed. For a short text this is not catastrophic, but for a longer text it is a major undertaking, requiring a lot of time. Alternatively, a clean nicely formatted typewritten version has to be produced for offsetting or xeroxing; here, again, new errors will creep in. Corrections of those will often involve retyping whole lines, with the risk of still more errors. Thus a "classical" edition comes thus into being in two completely separate steps, one by the philologist, one by the printer (or in the case of a typewritten book the typist). In both stages extensive proofreading is necessary.

Having described in outline how a "classical" edition comes into being (a more detailed description of the philological processes involved will be given below), I will now turn to the "computer aided " edition.

In the first place no such thing as the computer aided edition exists. Any edition which at some stage has used a computer even for very minor aspects can be called computer aided. As nearly every edition nowadays will be computerized at the printing stage, the term used in this wide context means hardly anything. But even dismissing these cases and counting only those editions as computer aided that were already computerized at an earlier stage, there are many different possibilities. I will describe two of them: the "computer as a glorified typewriter" approach and the fully fledged computerized edition. The difference between these two approaches lies in the first,

philological, stage of the edition. Though the schematic representations of both types of computer aided editions look the same as the one given for the "classical" edition, the contents of the box marked printer are different: no retyping has to be done, only replacing codes by other codes and inserting typesetting commands. As replacement will be the same everywhere, a test page with all possible signs in all possible combinations can be drawn up. Once this is correct, the text itself will be fine too. It will not need to be proofread in detail, only to be checked for eventual faulty typesetting commands. As these will result in the sort of printing error that is easy to spot, like whole paragraphs in the wrong font or size, this will not be much work.

A. The computer as typewriter

Although I personally think that this is a waste of the possibilities, I must stress that even this approach has a big advantage over the classical one. The text (transliteration, literal transcription, diplomatic, normalized or critical) is not typewritten (or even handwritten) but typed into a computer (micro, mini or mainframe). A printout is made and checked. The corrections are entered into the computer. Now a fresh printout can be made, and so on. At any stage of the work an updated draft copy is available. When the text is considered correct, a copy on floppy disc, magnetic tape or paper tape can be sent to the printer, who will convert the code to the conventions of his computer and add the typesetting markers.

The advantages are clear: at any stage an updated "clean" text can be produced, and there is hardly any human interference in the transmission to the printed text. That is to say, maybe the conversion will not be perfect, but then it will result in a systematic error, which will be spotted more easily than an accidental error. Moreover, as the text is already computerized, it is easy to draw up lists of all occurring signs

and combinations of signs in order to simplify conversion. This, however, is already starting to explore the possibilities of the computer beyond the mere typewriting and storage device as dealt with in this section.

B. Computerized edition

To appreciate the advantages of the computer more fully, it is necessary to go deeper into the philological process that leads from the manuscript to the intermediary text which will go to the printer. Of course, this process is different for every edition and very different for the various types of edition. As an example, I will try to analyze the process for a diplomatic edition, where abbreviations will be expanded and marked by italics in the printed text. ħ ẽ m̃ ſtoꝛríkr will be printed as hann er maðr storríkr. As a first step the text has to be deciphered: h with cross stroke, blank, e with a tittle over, blank, m with a superscript r, blank, long s,t, o, r rotunda, r, i with accent, k,r. That which is written or typed will probably not be that which one sees and reads, but what one wants to print: hann er maðr storríkr (with the letters to be italicised in print underscored). In fact, many decisions about the edition have to be taken at this transcription stage: Should ħ not be transcribed as hᴀɴ , if hᴀɴ is the prevalent writing in the MS? Should the pair ſ-s be kept apart as such, or should ſ be transcribed as s and s as ss? Should normal r and r rotunda be kept apart? What about the different shapes for y and v? Etc. etc. Depending on the type of edition and the manuscript the problems will vary, but almost certainly one will be confronted with problems around :

1. assigning graphs to graph classes
2. word division. (keep as in the MS or rearrange?)
3. compounds (theoretical difficulties)
4. small capitals (keep or treat as an abbreviation mark)
5. initials

6. layout of the MS page
7. expansion of abbreviations: italicise everything not written explicitly in the MS or everything not written on the base line (f^{a} as f<u>ara</u>, f<u>a</u>ra or f<u>ra</u>), normalize the expanded forms according to the norm in the handbooks or internally, and if so, should one look at the whole MS or rather at the part where the word occurs, at occurrences of this word only or at similar words also?
8. emendations.

This approach almost certainly involves having to go over the text several times. Some things can be decided beforehand, like what to do with small capitals. Other decisions (expansion, emendation) depend very much on practice elsewhere in the MS. Actually, before a transcription in this form can be made, a transcription should be already available! Of course this is a contradiction and so the editor spends much of his time removing inconsistencies. Another problem is that as soon as one starts expanding, some information available in the MS will be lost, i.e. different spellings in the MS will result in the same spelling in the transcript. Where a manuscript consists of different parts with their own history, this may result in an inaccurate estimate of relationships between the various parts, and therefore in less than satisfactory decisions about further expansions.

With a computerized edition one can postpone these sorts of decisions until a later stage and start by making a faithful transcription of the text, i.e. a copy of the MS sign by sign, word by word, line by line and page by page, where all abbreviation marks are kept and all mistakes. I will not recommend a graphetic transcription, unless one intends to study the paleographic aspects to some depth, but when necessary a transcription as outlined here can be updated comparatively easily to a graphetic one. This can be done directly into the

computer or, but this I would recommend only for texts that are very difficult to read, first on paper. When the complete text has been entered, printed out, compared to the MS and corrected, one can start removing the abbreviations systematically, one by one. For example h will probably be a frequent abbreviation. In nearly all cases it will stand for hann. To check this one can either look up all places with a search command in an editor program and check them, or write a small program which lists all relevant places. For a decision which expansion is the right one, it is necessary to look at the context: therefore it will be a good idea to include the contexts in the list. When an edition with expansion according to internal criteria is planned, it will be necessary to look up all other spellings of hann as well. As this has to be done for every abbreviation in the MS, unless the text is very short, it will probably be better not to do this by means of an editor program or special programs for each particular abbreviation, but to make a concordance first. Then one can study the concordance, decide on the expansion and carry it out.

Again, unless the text is a very short one, it will be a good idea to keep copies not only of the original text and the last version, but also of all the commands and operations leading to the last version. This will not only enable you to render an account of the editing methods followed (a thing most desirable but hardly ever done), but also to review earlier decisions as one's understanding of the underlying orthographic system improves, or when one has to take the publishers policy into consideration. In this way, it will be clear, all necessary decisions can be taken after consulting all relevant material, so they can be well founded. They will also be reversible, if one wants to go back on an earlier decision, for example if one wants to change the type of edition. A useful corollary for some texts could be the simultaneous production of two editions, for example a facsimile print (to use the term coined by Hans Fix) and a normalized edition.

The major difference with the approach described under A is that in A the text entered is the text-as-one-wants-it-to-be (an as yet non-existent text), while in B the text is the text-as-it-stands-in-the-MS. Proofreading in A involves therefore not only comparing the entered text with the MS (as the two are intrinsically different, the comparison is not straightforward), but also with the imaginary goaltext (and as this goal exists only in the philologist's mind, this comparison is not easy either). In B the text as entered has to be compared with the MS, but this comparison is comparatively easy as the transcript follows the layout of the MS. When a sophisticated output is produced, for example a plotted version which can reproduce the 2-storey system used in Old Icelandic MSS, missing words or characters will be spotted easily. From this stage on, changes in the text can be monitored automatically. A concordance will almost certainly result as a by-product of this approach. Various lists (alphabetic wordlists, retrograde wordlists, frequency lists) can be made with little cost or trouble. Lemmatizing the word forms will make the text available for more general morphological and syntactic analysis.

A computer aided edition along the lines proposed has structural advantages over editions produced in the classical way and by the computer-as-a-typewriter-method, namely

1. consistency,
2. the second proofreading has become rudimentary (this also applies to A),
3. the first proofreading is much simpler,
4. availability of the complete material during the editing process,
5. future availability on tape at the Computer Tape Bank, and for example as a concordance with frequency lists.

There are, however, some snags. The disadvantages can be summarized under the following headings:

1. risks,
2. programs,
3. input/output.

In the first place there is the inherent risk involved in handling texts by computer and storing them on magnetic media (oldfashioned papertape or punchcards are far less risky):

1. Magnetic disks and tape can develop faulty areas which cannot be read any longer. In this way, part or the whole of a text can be irretrievably lost.
2. One can by mistake and sometimes literally by pushing one button throw away the work of weeks, months or even years.
3. Programs sometimes do not act as intended. If the file resulting from the program overwrites the original file, then repeating the last step with a corrected program will not be possible.
4. Handling text with an editor program may result in changes which are unwanted or not repeatable, especially when the user has little experience.

Therefore one should always take safety precautions :

1. Make backups (copies) of important versions, preferably more than one.
2. Never use programs that update files, or replace an old version by a new one before the new file has been checked carefully.
3. Use a comparison program to compare old and new versions, so that you can check all changes.
4. When using an editor program, store the revised text under a different name.

Further, as already stressed above, it would be wise in order to avoid accidental errors and to keep the process repeatable, to produce the subsequent versions of the text not by making use of an "editor" (or wordprocessor), but by (batch) programs. These programs have to be written and tested, which is rather time consuming. For shorter texts it will probably not be worth the

extra effort. But if, as I hope, future editions will be made in this way, it might be possible to set up some standards, which would make it possible to use existing programs. For longer texts the development of programs will pay off, but even there usage of existing program packets would be an advantage. What I have in mind is a set of flexible programs which can be used for a choice of formats and editing conventions and user-definable replacement operations. This will require however some amount of standardization of editing practice.

These two disadvantages, the risk and the investment (either in time or money) in programs exist for any type of data handled by the computer whether they are library catalogues, recipes for cookies or your income tax data. But handling Old Icelandic texts, especially close transcripts of manuscripts, brings more specific problems as well. Special characters are needed: , þ,ð ,æ ,ø etc., so problems of representation and reproduction arise. In principle most computers support a character set of 256 characters. This is not enough to accommodate all characters encountered in a manuscript (undercast, capitals, small capitals, initials, superscripts, ligatures, abbreviation signs etc), the numbers and various other signs used in an edition (insertion marks, marks for emendation etc.), as a number of characters can not be used as they are reserved for internal use. In most cases only about 90 to 96 characters can be entered directly from the keyboard, while the others can be used only by tricks like entering their hexadecimal code or using combinations of the ESCAPE or CONTROL key with other characters. Of the characters available on the keyboard however, several will not be used like #, $, &, @ . These can be used to accommodate the specific Icelandic characters. To give an example, any MS will use þ and Þ , but w and W will be probably free. Using the w and W keys for þ and Þ is therefore possible. With some careful manoeuvering, most of the frequent

characters will find a place; the more infrequent ones can be represented by a pair of characters. The use of special character sets like initials or superscripts can be marked more clearly by a special initial marker (or superscript marker) with the normal letter than by a whole set of new (and most cases unprintable) characters. In the accompanying Coding Proposal I have drawn up a proposal for the encoding of Old Icelandic manuscripts, which is meant to standarize the texts stored at the Computer Tape Bank. It is NOT meant to be a standard for actual work,i.e where the proposal stores the tittle as %, this does not mean that everyone actually has to use % for tittle when working on his/her own computer. If for some reason or another you find it more convenient to use # or ^, then do so, but do not forget to convert your # or ^ to % before storing your text in Copenhagen. The proposal can, however, be used as a starting point, to which you can make additions or small changes to suit the needs of your special MS and your special computer/screen/keyboard. Though in principle the problem is now solved, in practice it is not. You will have to remember to type w for þ , @ for æ, X for ñ to name but a few. The text on your screen and on your printer will look more like the result of an inebriated work session than Old Icelandic. This can be improved however. Most screens have possibilities to redefine characters, either by program or by replacing an EPROM. Possibilities for dead keys are usually limited however, so you may have e^{7} on the screen, not e with the tittle superimposed. Producing good paper output is more complicated. Even producing printout with accents over the letters instead of next to them may be a major undertaking. Some printing programs while allowing for accented vowels bug up completely when one wants to print an accent over v (not uncommon in Old Icelandic manuscripts). Lineprinters and the like are made for mass production, not for the good quality work required in this field. Laserprinters have the potential, but computing centres

have bought them for their speed. Therefore they will not encourage you to fiddle around with the software, nor will they allow you to bring their printing operations to a near standstill in order to produce a facsimile-like reproduction of a manuscript page. A good matrixprinter might do the job, but with much difficulty. The best idea (still) will be to use a plotter, which is very flexible and can reproduce any shape at all and in all sizes and also combine the pictures in the same way as in the MS so that superscripts, subscripts and ligatures can all be represented. Here I have in mind the intermediary stages, especially the production of a text to be compared with the MS. For the final product again two possibilities are again available: the plotter and computerized phototypesetting. The latter method, especially the one that makes use of digitized characters, gives high quality print and the advantage of being able to define literally any shape wanted.

When you have decided to do a computer edition, much will depend on the equipment you have got and the programming expertise you have or can lay hands on. Those who want to have a go at programming themselves should know that recently a SNOBOL-compiler for IBM-compatibles has become available. But whatever the external circumstances, a workplan for an edition along the lines indicated above will probably look more or less like this:

1. **List of signs.** Make a provisory list of the signs occurring in the MS. It will not be necessary to go over the whole of the MS,a fair sample will do. Anything not yet included will probably be rare.
2. **List of unused keyboard signs.** Make a provisory list of the keys on the keyboard you will not use. Remember that you will need some to indicate paleographic peculiarities or as marker for sets of characters like initials.

3. Coding system. Using the lists above (and eventually the Copenhagen Coding Standard) make a coding system. Try to give frequent signs a single key representation. Remember that MSS usually do not use arabic numerals, so within text parts numerals are free to be used for example for abbreviation marks. Try to make to code easy to memorize. (4 for ꝑ is a good mnemonic).
4. Adapt screen. If this is possible, fine, but remember that unless you want to do nothing other than Old Icelandic with the screen, there should be a possibility for the "normal" character set as well. If adaptation is not possible, bad luck, but when you have typed your code for two or three weeks, you will have no difficulty whatsoever to decipher your screen.
5. Enter text, preferably directly from the photographs of the MS. Intermediate handcopying or typewriting will only enlarge the number of copying errors. Enter the text exactly as in the MS (complete with line changes, scribal errors, abbreviation marks, etc).
6. Plotting program. Write the plotting program (or better: have it written for you, just draw the special signs you need on a large scale). It should restore all the information coded to its proper format (superscripts as superscripts, ligatures as ligatures, etc). If pre-editing markers have already been added, these should be suppressed. Do not make concessions: If your plotted text really looks like the MS, transcription errors like skipped characters, replacement of characters by abbreviation marks and vice versa will jump to the eye. Do not forget that the idea is to skip the proofreading of the printed text, so errors have to be caught now.
7. Plotting. When the plotting program has been extensively tested, the text should be plotted. This can be a time consuming undertaking, but it is unlikely that you can correct the pages much faster than they are plotted.

8. **Compare plotted text with MS.** Probably you will first compare with the photographs, until you are satisfied and then once with the MS.
9. **Correct the text.**

   Steps 7–9 have to be repeated until no more errors are found. When a facsimile print is intended, only footnotes have to be added and the text can be directly printed. Otherwise the following steps have to be taken.
10. **Pre-editing.** Some things like markers to split pseudo-compounds or to join compounds can be added directly at the entering stage or later in the corrected text. From a systematical point of view the latter is better, but from practical experience I feel that early addition of these markers gives more chance to spot inconsistencies, while for comparison with the MS they can be suppressed.
11. **Concordance.** Make a concordance;don't write a program, use existing packages like OCP. Study the concordance and take decisions about abbreviations.
12. **Expansion.** Carry out the decisions taken.
13. **Emendation.** Add emendations and generate footnotes.