

Varieties of Electronic Experience

or,

What Should an Electronic Text Be Like?

Jeffery A. Triggs
Oxford English Dictionary

1994

1 Introduction

The Tenth Waterloo Conference may seem an odd time to ask a naïve question like the one in my title¹, but it is an occasion for some retrospection, and as must be clear to anyone working “in the field”, in spite of these ten years spent considering, writing about, and using it, our working notions of electronic text remain as various and perhaps as confused as ever. The sudden blossoming of interest in the Internet threatens to open up what was once a snug little club of people dealing with electronic text; our numbers are likely to be overwhelmed, and our often unspoken assumptions challenged. If the history of such “mobbings” is any indication, we may fear that the good is not likely triumphantly to drive out the bad. (I would add parenthetically that this affects not only electronic text, but a whole range of issues such as “netiquette”.) Thus it behooves us, while we still can, to review and rearticulate our assumptions, and perhaps give some shape to any future, unsponsored debate.

The term “electronic text” is very broad, and so we come quickly to the task of refining it. We may take electronic text to suggest *any* representation of text

¹Copyright ©1994 by Jeffery Triggs. All rights reserved. This essay was first given as a paper at the 10th Waterloo Conference for the New OED, October 19th, 1994. An abridged version was subsequently published in *Electronic Texts and the Text Encoding Initiative*, special issue of *Text Technology: The Journal of Computer Text Processing*, 5:3, Autumn, 1995, edited by Lou Burnard.

that can be stored on a computer. As a first, broad discrimination I would like to distinguish electronic texts that are *machine displayable* only from those that are *machine readable*, that is, capable of lower *and* higher level processing. The first sort is typified by the bitmap. Several things might fall under the second category, but I will take as my ideal the ascii file with SGML markup. One way of determining whether an electronic text is machine displayable or machine readable is to consider what might be called its “stored” and “used” forms. A file stored as SGML has many possible used forms, depending on its post-processing, all of which are somewhat different from the stored form. There is no difference, on the other hand, between the stored and used forms of a bitmap. (I am intentionally ignoring the form of data compression here, as it represents essentially a coded copy of the original file.) Another test is the degree to which an electronic text in its stored form may be said to aspire to the condition of paper. The bitmap faithfully reproduces the features of a paper text; an SGML text aims to be something at least somewhat different, and not necessarily paper at all. We can describe most kinds of electronic text as gravitating more or less toward one or the other of these archetypes. With the machine displayable group, for instance, we can place any fixed, typeset texts including Postscript files, troff output files, and even WYSIWYG word processor files (as someone once quipped, WYSIWYG is really WYSIAYG — *What You See Is All You Get*). With the machine readable group we place plain ascii files, files with minimal structural markup like COCOA tags, and files with hypertext markup. We will consider some of these in more detail, as well as various bastardized versions.

2 Some Machine Displayable Types

Anyone who has sent or received a fax knows something of the usefulness and limitations of bitmaps. The bitmap is literally a digital image, in zeros and ones, of each pixel of a page. It doesn't matter if the page includes pictures or smudges, or pencil scribbles in the margins: these are all part of the bitmap. Fonts, point size changes, archaic letters or non-Roman alphabets are copied exactly. It's all that easy, certainly the easiest way to convert a paper text to electronic form, but it is surprising how many people still don't realize that fastest way to search a bitmap is with the human eye. A number of library projects, most prominently the American Memory project at the Library of Congress, are creating bitmapped images on a large scale. At least for the present, however, this is in effect little more than a big, clean, pleasant microfilm project. Even assuming that some sort

of massive OCR processing is possible in the near future, it would still be subject to the shortcomings of that technology. Various schemes have been tried to allow something like searching with bitmaps, but these are inevitably rather awkward workarounds: ascii captions, or even, as in the case of the CORE project, parallel ascii texts. Even allowing for data compression in storage, bitmaps are expensive space hogs. Anyone who has ever waited at a printer queue for someone's half-page, bitmapped screen dump to pass through knows this. A small bitmap may be easily over 100,000 bytes. By contrast, the Postscript version of this paper so far is only 21,697 bytes, the ascii file itself a mere 4,407.

Many texts are actually offered on the network in the form of Postscript files. The reason for this is that the Postscript page description language does what it claims to do: it *prints* reliably on many different machines. Thus, it is an efficient way to get hold of and *print off* certain texts. But it is an extremely inefficient way to store electronic text. Some versions, such as psroff, produce a minimally editable output file, but even this cannot be effectively altered: the used form is the same as the stored form. Other versions are unreadable and uneditable, and thus essentially bitmaps. They cannot even be searched in the primitive way the output of psroff can be. Though the Postscript file itself is ascii, it requires so much verbose description that it is inevitably bulky and wasteful compared with a simple ascii text. Raw troff output, not to be confused with high-level troff code in an ascii text, is less bulky than Postscript (this text is at present 9,157 bytes in troff), but is not editable or even readable except as printed.

I know of at least two dictionary projects that have been done completely with Word Perfect. It is perhaps understandable, in that these were complex projects being worked on by people with limited computer experience who thought of their texts' structures primarily in terms of traditional typesetting features. In other words, like so many publishers and unlike, I am proud to say, the new *OED*, they were "print-driven". They thought of their electronic text mainly as a means of producing a paper text for publication. Because of this, two fascinating and very valuable electronic texts are being held hostage by Word Perfect's typesetting codes. To be sure, there is a deceptive feeling of the freedom we associate with electronic texts: they can be looked at, searched, and printed by those with the proper and properly configured release of Word Perfect... They can even be "saved" as ascii and given to the rest of us, but then all the magic is gone. The typesetting features, on which our whole sense of the structure depends, disappear; the special characters disappear. If you look at these texts as they *really* are, you find a horrible mess of print instructions, page breaks, potential hyphenation points, etc., all encoded with non-printing characters. (See Appendix 1)

3 Some Machine Readable Types

It is obvious that to have fully processable, that is, software and device independent electronic text, we need to begin with plain ascii text, either raw or marked up in some way.

The simplest of these texts have an interesting relation to the bitmaps, with which the discussion of machine displayable texts began. These are the raw products of OCR, or Optical Character Reading. OCR, of course, begins with the creation of a bitmap of a page. This bitmap, however, is subjected to a program that tries to assess its fonts and create ascii character equivalents for the letters of the page. It all happens rather like a complicated Xerox process, and thus seems a beautifully simple and direct way of converting whole libraries of paper text to electronic form. There's a rub in all this, however. Even the best OCR software presently available is not really good — it fails miserably on historical texts printed in the lovely old variable typefaces of letterpress days; it is especially bad on texts with complicated typographical arrangements, such as dictionaries; it loses any distinctions of font or point size even in what it does read properly; it introduces a whole new class of typo — things like “slie” for “she”, “hiin” for “him”; it produces a text with no actual structure at all. One might point out, with telling irony, that OCR works best on the simpler laser printed texts. A recent review of OCR techniques concludes that “OCR applications...are cost effective only because of the way that universities allocate resources. It is frequently much more difficult to obtain cash from administrations and granting agencies for keyboarding than it is to be awarded Research Assistant effort and the cost of purchasing and maintaining an OCR device. Unfortunately...the combination of OCR and graduate assistant labor is far less efficient than keyboarding at producing high volumes of very accurate text data.” (Olsen and McLean 127/2) Yet OCR remains a sort of “great white hope”, and we continue to see such texts pouring onto the market. They are indeed somewhat useful just as they are, for they can be searched and converted into used forms very different from the forms in which they are stored. But without structure, and with typographical incidentals, such as “catchwords”, “running titles”, or end-line hyphenation, caught up uninterpreted as it were in the net, they can hardly be said to be in finished form. It is amazing, however, how many of these texts are being dumped “as is” onto CD's and marketed in that form. (See Appendix 2)

It is not uncommon to find ascii texts keyed with word processors or text editors like vi and emacs. These are more difficult to produce than scanned texts, but also allow for the possibility of more subtle markup such as special symbols keyed

to represent accented letters or font changes, or patterns indicating the boundaries of certain structures. Indeed, a text keyed in this way is sometimes the easiest way to begin the transformation to SGML, as what one might call a “creational”, or pre-stored form. The problem is that the “markup” is non-standard and often insufficient. There is usually no formal document type description, and so when dealing with such texts we are often left to our imaginations to sort out what is what — Does this curly brace really mean the beginning of italics? or that Greek text will follow in some scheme of transliteration? Does this “aa;” in the middle of a word mean an acute accent over the letter preceding or following? or perhaps an “aacute”? Does that “@4” begin a small caps font, and “<” a paragraph? Is this octal “\234” a pound sign in the PC Character Set? If so, it simply produces a blank, a sort of electronic black hole, anywhere else. A veritable electronic archeology is necessary to bring some of these texts into the light of modern day, and there is often no Rosetta Stone. There are texts still floating around from the stone age of computers all in caps. We can only thank the often anonymous “stone age” contributors who thought to mark the true caps in some of these with at signs or hash marks. Such texts are still worth converting. (See Appendix 3)

Although we have entertained at least the theoretical notion of hypertext since as long ago as 1965 when Ted Nelson introduced the term, it is “hotter” and “newer” now than it has ever been. This is largely because of two areas of burgeoning growth: CD-ROMs and the World Wide Web, or WWW as it is known, on Internet.

On the surface at least, hypertext sounds as if it should be the archetypical electronic text. As Nelson put it (in what is now the first quote for the term in *OED Additions*, Vol. 2), “Let me introduce the word ‘hypertext’ to mean a body of written or pictorial material interconnected in such a complex way that it could not conveniently be presented or represented on paper.” Certainly, here is text that does not, in its stored or its used forms, aspire to the condition of paper. And yet, in many of its earlier manifestations, hypertext might almost be grouped with the machine displayable texts. If it did not try to look like a certain paper text, it was dedicated to a certain set of device-dependent screens (something virtually as limiting). If it allowed for exciting explorations across the boundaries of texts, these were necessarily chaperoned trips, tourism with options (or should we, remembering Barrymore on footnotes, term them “interruptions”) — a network of paths through a forest but *already laid out*. In other words, its used form, however complex, was identical with its stored form. The prepackaged experiences available on CD-ROMs carry on this tradition of hypertext, offering the user what one might call “virtual freedom”.

The good news is that hypertext has come to the Internet via World Wide Web in a form that is attractively clean and device independent. The Hypertext Markup language, or HTML is a plain ascii descriptive markup that approaches and seeks compatibility with SGML. At the same time, HTML is immediately functional with the Web's powerful clients, the "browsers" such as Mosaic and lynx, on various platforms. In other words, anybody with a text editor and a browser (freely available on the net) can start playing today, even those hitherto limited to non-UNIX operating systems. When applied to texts through a link-creating filter, one might even say that, refreshingly, HTML's beginnings do not know its ends. But as John Price-Wilkin has pointed out, (Price-Wilkin 5-21), there are a number of drawbacks to HTML as a markup language, notably its relatively "impoverished tag set" and its difficulty defining structures and bounded segments. Price-Wilkin has found, however, that such limitations can be largely circumvented if the text is stored as SGML, processed structurally by a search engine like Pat, and only filtered into HTML for processing by the Web's clients. In other words, the HTML is treated essentially as a typesetter, like troff or TeX, for post-processing, one of many possible used forms rather than a stored form.

Even though HTML+ promises a much improved tag-set in the near future, this seems quite a reasonable approach to any large scale use of HTML, and I believe we should keep it paradigmatically in mind to guard against any temptation to treat HTML as a sufficient stored form in itself, a kind of easy substitute for SGML. Considering all the neophytes coming onto the Internet where they are likely now to encounter HTML before SGML, I think we are not unjustified in entertaining as a public worry at least the notion that HTML, because of its relative simplicity, and through the Web, its accessibility, could wind up becoming a sort of de facto standard markup language. But more on this later.

I will not go into a great deal of detail about SGML at this time. With the recent publication of the TEI guidelines, there is no lack of information about SGML specifics available. What I would like to do is emphasize a couple of salient points that will be of use later on in considering various mistakes.

SGML can take many forms and can embrace differing degrees of detail, but its most important function is to mark up the essential structures of a text. SGML may describe meaningful typesetting structures, such as italics for emphasis or to indicate a title, as well as a printer's decorative use of a typeface. It may indicate essentially different letters like the thorn or yogh or Greek letters, as well as historical printing oddities like the long-S. The better electronic text, however, will not allow these accidental structures to obscure the essential structures or to get in the way of machine processing. This is a difficult (perhaps a debatable) line to draw,

but ultimately, in considering what is involved in the structure of an electronic text as such, we need to confront the fact that the use of a thorn makes something a different form of a word, whereas the long-S does not, but is essentially a printer's device.

All this touches on what should be considered an essential tenet of SGML: that all processing should be post-processing. This can take any number of forms. It may be an X-window text display system, such as Lector, a translation into nroff for display on ascii terminals, or translation into troff, or TeX, or Postscript for printing, or a translation into HTML for distribution through the Web. It could also involve processing into word-lists, filtering into a skeletal structure, or specialized tagging for grammar. The important thing is that none of these used forms is identical with the stored form.

4 Varieties of Mistakes

Electronic text has been traditionally a field where in practice at least anything goes. After all, for so long there was so little that to ask rude questions was rather like looking the proverbial gift horse in the mouth. In addition, our notions of what an electronic text should be have been evolving notions, and as we have seen several different notions of electronic text have evolved separately. Furthermore, with a few notable exceptions, the production of electronic text was not driven by the standards of professional publishing, but was often a more or less informal scholarly enterprise.

We need to begin taking a more critical approach to electronic texts as they grow more common and more widespread. In the following sections we will consider a number of systematic mistakes in the production of electronic texts today. Most of these involve in some way the fallacy of devotion to a specimen of printed text.

Before we congratulate ourselves too glibly on the acceptance of SGML and The publication of the TEI guidelines (a long-awaited and important event in the history of electronic text), we need to consider that for a lot of people out there — the folks with PCs and Powerbooks who want some kind of electronic encyclopedia for their kids and maybe to have access to a newswire — electronic text is likely to look a lot more like that produced by Project Gutenberg or the Internet Wiretap. When people new to all this ask me about electronic text or suggest some they've heard about, it is almost always Project Gutenberg and never the TEI. If I may make the same point more empirically, a look at the lists of elec-

tronic books on the net quickly reveals that the vast majority of these are plain (and inert) Gutenberg-style texts. Out of the nearly 800 texts listed on the “Alex” gopher (not to be confused with an http Web server...), only about 85 from the Oxford Text Archive are in TEI SGML. This ratio, of about one in ten actually represents an improvement. Such an imbalance is fine — if we consider these texts to be in a creational as opposed to a stored form, that is, as raw material for future conversion to SGML. But if we consider them finished products — and the fact that the providers are not changing them, but offering them *in place of* and possibly *in preference to* SGML versions available from OTA would suggest this — the prospect is discouraging. No one is in a hurry, it would seem, to superannuate these texts, though a look with a superior browser like Mosaic, where they appear only as large chunks of unformatted typewriter font text, quickly reveals their inferiority. (See Alex, [gopher://rsl.ox.ac.uk:70/11/lib-corn/hunter](http://rsl.ox.ac.uk:70/11/lib-corn/hunter))

So what does a raw Gutenberg-style text look like? And what’s wrong with it? Well, sort of like a printed page — that is, no formal structure and odd bits of print-hangover like end-line hyphenation. The paragraphs are identifiable, if at all, by a indent (often of varying size), or a blank line (occasionally two blank lines). Chapters are marked only by a centered heading, or sometimes simply a centered roman numeral, with perhaps a couple of extra blank lines thrown in. Lists or indented quotes of verse are marked only by indents; if a line of verse was too long for a printed page and was carried on underneath, the electronic line is carried under as well with an extra few spaces of indent. (See Appendix 4) In the case of certain Elizabethan texts, a word too long to fit is often printed, padded by a few spaces, after the line above. This convention is carried over literally in some electronic texts. Often they’ll have done things like convert all non-roman typefaces to capital letters. Other times, and not always described that way, italics will be indicated by underscores. Sometimes both methods are intermixed, the underscores saved in the case of italicized instances of the nominative personal pronoun. Footnotes and other non-textual data are sometimes included without formal distinction at the bottom of a “page”. (On paper, if you were a student assigned to read 400 pages of Philip Sidney, you would be grateful to find modernized spelling and footnotes at the bottom of each page. In electronic text, these same features carried over literally are at best a headache and at worst a nightmare — consider the grungy task of separating out and marking off those 400 plus notes.) In certain poetry texts line numbers taken from a particular edition appear next to the lines — literally next to them after some padded spaces. I can remember one text (in fairness not one from Gutenberg) where the creator announced proudly in a preface (not separated formally from the rest of the text) that he had carefully

removed all the control characters that had generated italics so that users would be able to search more easily with their word processors. This same text included stage directions on the same lines as the text separated by a few padded spaces and carrying down through several lines. Indeed, on a quick glance it had the look if not the feel of print. But one wonders how he imagined a computer was ever going to *read* this text? (See Appendix 5) Mistakes of this sort are clearly the result of naïve thinking about what an electronic text should be. They can seem quite amusing. But this sort of thinking is alive and well and even aggressive. Correspondents have complained to me about not being able to “see the italics” on screen, but only messy tags. (This is a complaint that the more widespread use of Web browsers should be able to silence.) Perhaps more seriously, a recent posting to the net took the Oxford Text Archive to task for not preserving the “lineation” of a particular print edition of an electronic text. Whatever one might say about the lineation of a print edition, it is not an important structural element of electronic text. (See Appendix 6) Indeed, a good many large full-text corpora, including the University of Virginia Electronic Text Center and the *OED*’s historical corpus, do not keep newlines at all in the stored form.

The fallacy of devotion to a specimen of printed text also makes itself felt in the otherwise sensible, embracing environment of SGML. SGML, of course, can have by definition many different flavors, as it needs to do if it is to embrace the wide variety of different possible texts. We should still be able to cite certain efforts as being in better style than others. As I indicated above, we need to distinguish between essential and incidental or accidental structures. SGML can be used to describe both of these, but its efforts are wasted on the latter, and may indeed be considered counterproductive. The “catchwords” and “running titles” used in a particular early edition of a text are not essential structures of an electronic text based on it. Nor are the hyphenations of words on a title page with a very large point size. Some texts even go so far as to use “I” in place of “1” in numerical dates simply because a seventeenth century printer had it that way. What can a computer ever make of that? It certainly won’t treat it as a date. Coding such things, no matter how intricately we do it, can never get us as intimately near the print edition as a simple bitmap. In fact, such markup gets in the way of innately computational processing, as surely as the picturesque stage directions that I cited above. As Robert Amsler once said, an electronic text should be “the mutable document which underlies the static output rendering of it on paper.” Thus the electronic text must bear in its relations to any printed version something of the remoteness of a Platonic abstraction or ideal.

5 Some Conclusions

The electronic sins outlined above might be classified as “sins of ignorance” and “sins of expediency”. The latter, of course, are by far the more dangerous, as they threaten to undermine the whole enterprise of standardizing around SGML. If we can still agree that some form of SGML is the best way to store electronic text, we need now to encourage such projects as make SGML immediately functional and therefore expedient. We should take away, in other words, the expediency of storing texts in one or another of the typeset formats.

The explosion of new CD ROMs with slick interfaces and cheap text is a somewhat disconcerting phenomenon, at least if you’re the type who believes the real future of electronic text lies not in the proliferation of stamped plastic artifacts, but in Infobahning around the network, in having information stored discretely around the globe and globally, fluidly available. I’m afraid that CDs, with their more familiar packaging, make publishers, at least for the present, far more comfortable. But there is also hope that the groundswell of desire for information on the Internet is becoming audible.

The Web and HTML are by far the most exciting things to happen recently in electronic text — the first application in some time that makes you want to fasten your seatbelt. It actually does, and fairly brilliantly at that, a number of the real world things that SGML has only promised it could someday accommodate. And with lots of interest in development, it promises to get better. Even now it offers a logical format for incorporating images and the products of other media into a text — without resorting to the proprietary embrace of a program like Framemaker. Thus, we are beginning to see real electronic publishing in the WWW/HTML format: first the *Encyclopedia Britannica*, and soon, perhaps, other very large scale reference works as well.

This could be the most encouraging news ever for SGML, and perhaps for the future of electronic text, but as I suggested above we need to keep the horse before the cart, that is, to remember that HTML is at best a type or screensetting subset of SGML, and not, in itself, a suitable stored form. HTML no more adequately describes the structure of a text than troff. Bluntly put: we need to control the development of HTML — so that it does not control us. This could be SGML’s “killer app”, or, if we’re not careful, simply its killer. That said, we should resolve to encourage projects like the Web that willingly bring some form of SGML to the people. Nothing but popular pressure will force the already popular word processors to pay more than lip-service to SGML, but this is exactly what needs to be done.

I'm afraid at this point I'm going to have to confess openly that, in spite of having experienced some moving "spots of time" while working on the electronic versions of *Lyrical Ballads* and other texts, when I really want to *read* something I don't yet curl up with a Powerbook. The 20,200 lines, 143,448 words, and 853,318 bytes of my electronic *Far From the Madding Crowd* do not have the same comfortable heft and feel and smell that the 465 printed pages of my Penguin edition do, not to mention the lovely painting by David Farquharson reproduced on the cover. And I find that the "bullets" of the AP newswire do not take the place of the smudgy old *New York Times*, as read over the morning's first cup of coffee. Perhaps it is only a matter of coming up with better screens. Perhaps more than that is needed to make electronic text truly domesticable. A hypertext link to a bitmap might solve the second of my misgivings about reading the Hardy novel on screen, but not the first. The Web may soon change all this dramatically, and indeed even now it may not hold for *all* genres. Reference books, for instance, seem to be considerably further along the way to being replaced by electronic texts. It is an embarrassingly long time since I actually hoisted a paper volume of the *OED*.

Perhaps a different word altogether is needed to describe what humans do with electronic texts: we might not so much "read" them as "browse" or "consult" them. In any event, we need to remember that so far at least the thing that "reads" our electronic texts most happily is a machine, and that our interest in electronic text remains essentially more utilitarian than our interest in books. The electronic text of Hardy, for instance, can tell me in a matter of seconds, that it contains an earlier use of the term "late comer" than anything else in the *OED*'s paper citation files, as well as a very early example of our new sense of "dearly". And while I'm about collecting these quotes, I might enjoy a bit of the exquisite "context". The point is that we still process electronic texts mainly for such information, and not for pleasure, except perhaps accidental pleasure or the different pleasure in the processing itself. Therefore, we must prepare our electronic texts to yield the maximum amount of information. This seems patently obvious, but the real problem with some of the practices outlined above is that by confusing electronic text and print, they prevent the electronic text from offering what we need from it: information, not so much about any *printed* text, as intrinsically about *itself*.

6 Works Cited

Nelson, Ted. *Proc. 20th Nat. Conf. Assoc. Computing Machinery*, (1965), 96.

Olsen, Mark and Alice Music McLean. "Optical Character Scanning: A Discussion of Efficiency and Politics." *Computers and the Humanities*, 27:2, (1993), 121-27.

Price-Wilkin, John. "Using the World-Wide Web to Deliver Complex Electronic Documents: Implications for Libraries." *The Public-Access Computer Systems Review* 5, no. 3 (1994): 5-21.

7 Appendices

7.1 Appendix 1

WYSIWYG As It Really Is

A sample from a file in Word Perfect with non-ascii control codes rendered to represent hex codes. What you see here is what you have: first raw Word Perfect, then the same file in “typesetting” SGML tags, and then rendered further into structural SGML.

7.1.1 RAW WORD PERFECT TEXT

(filtered through a program to print non-ascii as hex):

```
{d4}{1}{c}{0}{0}{6}{89}{0}?{0}{8}{7}{c}{0}{1}{d4}{c3}{c}{c\3}muurkas{c4}{c}{c4}\
/"my:rkas/ {c3}{8}{c3}n.{c4}{8}{c4} Also {c3}{c}{c3}meerkus{c4}{c}{c4}, and
dim. form {c3}{c}{c3}muurkassie{c4}{c}{c4}.
Pl. {d}{d4}{1}{c}{0}{0}{6}{89}{0}?{0}{d0}{7}{c}{0}{1}{d4}{a\
9}{c3}{c}{c3}te{c4}{c}{c4}. [Afk. {c3}{8}{c3}muur{c4}{8}{c4} wall +
{c3}{8}{c3}kas{c4}{8}{c4}
cupboard.] Esp. in traditional Cape{d}Dutch houses: a cupboard built into an
interior wall, usu. for{d}{d4}{1}{c}{0}{0}{6}{89}{0}?{0}'
{c}{0}{1}{d4}holding china and ornaments. See also {c3}{f}{c3}kas{c4}{f}{c4}.
{d4}{1}{c}{0}{0}{6}{89}{0}?{0} (
{c}{0}{1}{d4}{c1}{2}{8}{7}{8}{7}{f}{0}{c1}{c3}{c}{c3}1949{\c4}{c}{c4}
{c3}{f}{c3}L.G. Green{c4}{f}{c4} {c3}{8}{c3}In Land of Afternoon{c4}{8}{c4}
198 The reception room, {d}{d4}{1}{c}{0}{0}{6}{89}{0}?{0}{f0}
{c}{0}{1}{d4}with its stinkwood muurkas and pieces of Delft.
{c3}{c}{c3}1949{c4}{c}{c4} [see
{c3}{f}{c3}kas{c4}{f}{c4}]. {d}{d4}{1}{c}{0}{0}{6}{89}{0}?{\
0}{b8}{b}{c}{0}{1}{d\
4}{c3}{c}{c3}1963{c4}{c}{c4} {c3}{f}{c3}R. Lewcock{c4}{f}{c4}
{c3}{8}{c3}Early 19th C. Architecture{c4}{8}{c4} 344 Where
eighteenth{aa}century furniture was often built{a9}in (the 'muurkassie') or
at{d}least kept small and neat (the Georgian book{a9}case) the
new{d}{d4}{1}{c}{0}{0}{6}{89}{0}?{0}{10}{e}{c}{0}{1}{d4}furniture was heavy
and dominant. {c3}{c}{c3}1964{c4}{c}{c4} {c3}{f}{c3}J. Meintjes{c4}{f}{c4}
{c3}{8}{c3}Manor House{c4}{8}{c4} 66{d}In the long dining
room there were two muurkaste or
built{a9}in{d}{d4}{1}{c}{0}{0}{6}{89}{0}?{0}{a0}{f}{c}{0}{1}{d4}cabinets in
black stinkwood with gabled tops.
{c3}{c}{c3}1965{c4}{c}{c4} {c3}{f}{c3}M.G.
Atmore{c4}{f}{c4}{d}{d4}{1}{c}{0}{0}{6}{89}{0}?{0}h{10}{c}\
0}{1}{d4}{c3}{8}{c3}
```

}Cape Furn.{c4}{8}{c4} 201 Wall Cupboards or Cabinets (muurkaste).
 {c3}{c}{c3}1972{c4}{c}{c4}
 {c3}{8}{c3}S.{d}{d4}{1}{c}{0}{0}{6}{89}{0}?{0}0{11}{c}{0}{\1}{d4}Afr. Garden & Home{c4}{8}{c4} Feb.{a9}Mar. 57 The living{a9}room, for example, has{d}wall cupboards of the Cape Dutch Muurkas type, filled with hand{aa}{d4}{1}{c}{0}{0}{6}{89}{0}?{0}{c0}{12}{c}{0}{1}{d4}\
 }painted crockery. {c3}{c}{c3}1975{c4}{c}{c4} {c3}{8}{c3}Cape Times{c4}{8}{c4} 7 Jan. 8 This year closed with{d}the same auctioneers selling a fine example of an 18th{a9}century{d}Cape stinkwood{a9}framed 'Muurkas', bought for R5{a0}500 for 'Nova{d}{d4}{1}{c}{0}{0}{6}{89}{0}?{0}{18}{15}{c}{0}{1}{d4}\
 }Constantia' a historic house in the Constantia valley.
 {c3}{c}{c3}1985{c4}{c}{c4}
 {c3}{8}{c3}Style{c4}{8}{c4}{d}Dec.{a9}Jan. 193 The Young Cape Wine Farmers are..forever throwing{d}open wildly old houses full of meerkuses and wakuses and wildly{d}{d4}{1}{c}{0}{0}{6}{89}{0}?{0}p{17}{c}{0}{1}{d4}old bottles of wine grown on the estate. {c3}{c}{c3}1987{c4}{c}{c4} {c3}{f}{c3}J. Kench{c4}{f}{c4} {c3}{8}{c3}Cottage{d}{d4}{1}{c}{0}{0}{6}{89}{0}?{0}8{18}{c}{0}{1}{d4}Furn.{c4}\
 {8}{c4} 41 As with other kinds of furniture, yellowwood and{d}stinkwood versions of cupboards were made, including the imposing{d}{d4}{1}{c}{0}{0}{6}{89}{0}?{0}{c8}{19}{c}{0}{1}\
 {d4}Cape 'kaste' and 'muurkaste'. {c3}{c}{c3}1987{c4}{c}{c4} {c3}{f}{c3}G. Viney{c4}{f}{c4} {c3}{8}{c3}Col. Houses{c4}{8}{c4} 40 Behind{d}{d4}{1}{c}{0}{0}{6}{89}{0}?{0}{90}{1a}{c}{0}{1}{d4}4)the {c3}{8}{c3}voorhuys{c4}{8}{c4} was the {c3}{8}{c3}galdery{c4}{8}{c4} with two

7.1.2 INTERMEDIATE STEP - TYPESETTING TAGS:

<e>muurkas /"my:rkas/ <i>n.</i> Also meerkus, and dim. form muurkassie. Pl. &en.te. [Afk. <i>muur</i> wall + <i>kas</i> cupboard.] Esp. in traditional Cape Dutch houses: a cupboard built into an interior wall, usu. for holding china and ornaments. See also <sc>kas</sc>. <qp>1949 <sc>L.G. Green</sc> <i>In Land of Afternoon</i> 198 The reception room, with its stinkwood muurkas and pieces of Delft. 1949 [see <sc>kas</sc>]. 1963 <sc>R. Lewcock</sc> <i>Early 19th C. Architecture</i> 344 Where eighteenth-century furniture was often built&en.in (the &oq.muurkassie&cq.) or at least kept small and neat (the Georgian book&en.case) the new furniture was heavy and dominant. 1964 <sc>J. Meintjes</sc> <i>Manor House</i> 66 In the long dining room there were two muurkaste or built&en.in ,cabinets in black stinkwood with gabled tops. 1965 <sc>M.G. Atmore</sc> <i>Cape Furn.</i> 201 Wall Cupboards or Cabinets (muurkaste). 1972 <i>S. Afr. Garden & amp.

Home

Feb.&en.Mar. 57 The living&en.room, for example, has wall cupboards of the Cape Dutch Muurkas type, filled with hand-painted crockery. 1975 <i>Cape Times</i> 7 Jan. 8 This year closed with the same auctioneers selling a fine example of an 18th&en.century Cape stinkwood&en.framed &sq.Muurkas&cq., bought for R5,500 for &sq.Nova Constantia&cq. a historic house in the Constantia valley. 1985 <i>Style</i> Dec.&en.Jan. 193 The Young Cape Wine Farmers are..forever throwing open wildly old houses full of meerkuses and wakuses and wildly old bottles of wine grown on the estate. 1987 <sc>J. Kench</sc> <i>Cottage Furn.</i> 41 As with other kinds of furniture, yellowwood and stinkwood versions of cupboards were made, including the imposing Cape &sq.kaste&cq. and &sq.muurkaste&cq.. 1987 <sc>G. Viney</sc> <i>Col. Houses</i> 40 Behind the <i>voorhuys</i> was the <i>galdery</i> with two wall cupboards (<i>muurkaste</i>).</qp></e>

7.1.3 STRUCTURAL TAGS:

<e><hg><hw>muurkas</hw> <pr><ph>"my:rkas</ph></pr> <ps>n.</ps> </hg><vf>Also <v>meerkus</v>, and dim. form <v>muurkassie</v></vf>. Pl. &en.te. <etym>Afk. <cf>muur</cf> wall + <cf>kas</cf> cupboard.</etym> <s4>Esp. in traditional Cape Dutch houses: a cupboard built into an interior wall, usu. for holding china and ornaments. See also <x>kas</x>. <qp><q><d>1949</d> <a>L.G. Green <w>In Land of Afternoon</w> 198 <qt>The reception room, with its stinkwood muurkas and pieces of Delft. </qt></q> <q><d>1949</d> &osb.see <x>kas</x>&csb.. </q> <q><d>1963</d> <a>R. Lewcock <w>Early 19th C. Architecture</w> 344 <qt>Where eighteenth-century furniture was often built-in (the &sq.muurkassie&cq.) or at least kept small and neat (the Georgian book-case) the new furniture was heavy and dominant. </qt></q> <q><d>1964</d> <a>J. Meintjes <w>Manor House</w> 66 <qt>In the long dining room there were two muurkaste or built-in, cabinets in black stinkwood with gabled tops. </qt></q> <q><d>1965</d> <a>M.G. Atmore <w>Cape Furn.</w> 201 <qt>Wall Cupboards or Cabinets (muurkaste). </qt></q> <q><d>1972</d> <w>S. Afr. Garden & Home</w> Feb.&en.Mar. 57 <qt>The living-room, for example, has wall cupboards of the Cape Dutch Muurkas type, filled with hand-painted crockery. </qt></q> <q><d>1975</d> <w>Cape Times</w> 7 Jan. 8 <qt>This year closed with the same auctioneers selling a fine example of an 18th-century Cape stinkwood-framed &sq.Muurkas&cq., bought for R5,500 for &sq.Nova Constantia&cq. a historic house in the Constantia valley. </qt></q> <q><d>1985</d> <w>Style</w> Dec.&en.Jan. 193 <qt>The Young Cape Wine Farmers are..forever throwing open wildly old houses full of meerkuses and wakuses and wildly old bottles of wine grown on the estate. </qt></q> <q><d>1987</d> <a>J. Kench <w>Cottage Furn.</w> 41

<qt>As with other kinds of furniture, yellowwood and stinkwood versions of cupboards were made, including the imposing Cape &oq;kaste&cq. and &oq;muurkaste&cq.. </qt></q> <q><d>1987</d> <a>G. Viney <w>Col. Houses</w> 40 <qt>Behind the <i>voorhuys</i> was the <i>galdery</i> with two wall cupboards (<i>muurkaste</i>).</qt></q></qp> </s4></e>

7.2 Appendix 2

OCR Output

This is a sample of OCR output from a relatively clean, modern text, an issue of *The Literary Review*. The program outputs a tilde when it does not know what to make of a character. Note the running title and hyphenation, the loss of distinction of open and close quotes, and the loss of font characteristics.

The Liter~y Review

He smiled, looked at me, and I knew
it was my mind he led me through. (236)

The larger memory of mankind consists in the interlinking memories of the generations, in the connections that we make and maintain with past and future. This theme, so prominent in Berry's writings, is once more poignantly suggested here. And how typical of Berry that it should involve the memory of daily work, the kind of thing some people would consider "drudgery." But it is with such "drudgery" that "we enact and understand our oneness with the Creation" (~e Unsettling ~Ame~a 138). The narrative continues now with a number of episodes like this, as Owen Flood moves Berry

through all the fields of our lives,
preparations, plantings, harvests,
crews joking at the row ends,
the water jug passing like a kiss.

He spoke of our history passing through us,
the way our families' generations
overlap, the great teaching
coming down by deed of companionship . . .

The encounter with and contemplation of death is actually the contempla-
tion of life itself, of which death is an inevitable part. Berry tells us that Owen Flood's "passion" was "to be true / to the condition of the Fall-
/ to live by the sweat of his face, to eat / his bread, assured that the cost was paid" (237). This is Berry's highest compliment.

The fifth section of the poem deals with "the time of [Owen Flood's] pain," when in spite of "the sweet world" about him, "his strength failed / before the light." In one of the most moving passages,

Berry for the first time senses mortal weakness in a man whose strength he had always taken for granted:

Again, in the sun
of his last harvest, I heard him say:
"Do you want to take this row,
and let me get out of your way?"
I saw the world ahead of him then
for the first time, and I saw it
as he already had seen it,
himself gone from it. It was a sight
I could not see and not weep.

This sudden, often untimely fading of strength that lies at the center of our wondering about death is perhaps, to the human imagination, its most tragic quality. Berry the philosopher knows what to say next about it, but Berry, the poet in the grip of his vision, cannot help pausing to

| | Seen | Rejected | Pctg |
|------------|------|----------|-------|
| Characters | 1855 | 5 | 99.73 |

7.3 Appendix 3 Stone Age Text

No reflection on Milton is intended...

<A MILTON><P T. N. CORNS>

<N OF REFORMATION>

<D 1641>

<E YALE1>

SIR,

AMIDST THOSE DEEP AND RETIRED THOUGHTS, WHICH WITH EVERY MAN CHRISTIANLY INSTRUCTED, OUGHT TO BE MOST FREQUENT OF GOD, AND OF HIS MIRACULOUS WAYS, AND WORKS, AMONGST MEN, AND OF OUR RELIGION AND WORSHIP, TO BE PERFORMED TO HIM; AFTER THE STORY OF OUR SAVIOUR CHRIST SUFFERING TO THE LOWEST BENT OF WEAKNESS, IN THE FLESH, AND PRESENTLY TRIUMPHING TO THE HIGHEST PITCH OF GLORY, IN THE SPIRIT WHICH DREW UP HIS BODY ALSO TILL WE IN BOTH BE UNITED TO HIM IN THE REVELATION OF HIS KINGDOM: I DO NOT KNOW OF ANY THING MORE WORTHY TO TAKE UP THE WHOLE PASSION OF PITY, ON THE ONE SIDE, AND JOY ON THE OTHER: THAN TO CONSIDER FIRST, THE FOUL AND SUDDEN CORRUPTION, AND THEN AFTER MANY A TEDIOUS AGE, THE LONG-DEFERRED BUT MUCH MORE WONDERFUL AND HAPPY REFORMATION OF THE CHURCH IN THESE LATTER DAYS. SAD IT IS TO THINK HOW THAT DOCTRINE OF THE GOSPEL, PLANTED BY TEACHERS DIVINELY INSPIRED AND BY THEM WINNOWED AND SIFTED FROM THE CHAFF OF OVERDATED CEREMONIES, AND REFINED TO SUCH A SPIRITUAL HEIGHT, AND TEMPER OF PURITY, AND KNOWLEDGE OF THE CREATOR, THAT THE BODY, WITH ALL THE CIRCUMSTANCES OF TIME AND PLACE WERE PURIFIED BY THE AFFECTIONS OF THE REGENERATE SOUL, AND NOTHING LEFT IMPURE, BUT SIN; FAITH NEEDING NOT THE WEAK AND FALLIBLE OFFICE OF THE SENSES, TO BE EITHER THE USHERS OR INTERPRETERS OF HEAVENLY MYSTERIES SAVE WHERE OUR LORD HIMSELF IN HIS SACRAMENTS ORDAINED; THAT SUCH A DOCTRINE SHOULD THROUGH THE GROSSNESS AND BLINDNESS OF HER PROFESSORS, AND THE FRAUD OF DECEIVABLE TRADITIONS, DRAG SO DOWNWARDS, AS TO BACKSLIDE ONE WAY INTO THE JEWISH BEGGARY OF OLD CAST RUDIMENTS, AND STUMBLE FORWARD ANOTHER WAY INTO THE NEW-VOMITED PAGANISM OF SENSUAL

7.3.1 Trollope example

This has something to grab hold of at least. Note the octal symbol used for the pound sign.

@1CHAPTER I^

@4THE TWO SISTERS@1^

When Egbert Dormer died he left his two daughters utterly penniless upon the world, and it must be said of Egbert Dormer that nothing else could have been expected of him. The two girls were both pretty, but Lucy, who was twenty-one, was supposed to be simple and comparatively unattractive, whereas Ayala was credited---as her somewhat romantic name might show---with poetic charm and a taste for romance. Ayala when her father died was nineteen.

<\We must begin yet a little earlier and say that there had been---and had died many years before the death of Egbert Dormer---a clerk in the Admiralty, by name Reginald Dosett, who, and whose wife, had been conspicuous for personal beauty. Their charms were gone, but the records of them had been left in various grandchildren. There had been a son born to Mr Dosett, who was also a Reginald and a clerk in the Admiralty, and who also, in his turn, had been a handsome man. With him, in his decadence, the reader will become acquainted. There were also two daughters, whose reputation for perfect feminine beauty had never been contested. The elder had married a city man of wealth---of wealth when he married her, but who had become enormously wealthy by the time of our story. He had when he married been simply Mister, but was now Sir Thomas Tringle, Baronet, and was senior partner in the great firm of Travers and Treason. Of Traverses and Treasons there were none left in these days, and Mr Tringle was supposed to manipulate all the millions with which the great firm in Lombard Street was concerned. He had married old Mr Dosett's eldest daughter, Emmeline, who was now Lady Tringle, with a house at the top of Queen's Gate, rented at \$\backslash\$2341,500 a year, with a palatial moor in Scotland, with a seat in Sussex, and as many carriages and horses as would suit an archduchess. Lady Tringle had everything in the world; a son, two daughters, and an open-handed stout husband, who was said to have told her that money was a matter of no consideration.

<\The second Miss Dosett, Adelaide Dosett, who had been considerably younger than her sister, had insisted upon giving herself to Egbert Dormer the artist, whose death we commemorated in our first line. But she had died before her husband. They who remembered the two Miss Dosetts as girls were wont to declare that, though Lady Tringle might, perhaps, have had the advantage in perfection of feature and in unequalled symmetry, Adelaide had been the more attractive from expression and brilliancy. To her Lord Sizes had offered his hand and coronet, promising to abandon for her sake all the haunts of his matured life. To her Mr Tringle had knelt before he had taken the elder sister. For her Mr Program, the popular preacher of the day, for a time so totally lost himself that he was nearly minded to go over to Rome. She was said to have had offers from a widowed Lord Chancellor and from a Russian

prince. Her triumphs would have quite obliterated that of her sister had she not insisted on marrying Egbert Dormer.

7.4 Some Electronic Dictionaries

An example of various typesetting schemes that form part of the output of a popular program at Bellcore.

```

W7
buck 1  n
'b*k
    [italic pl]
bucks 1l
ME, fr. OE [italic bucca] stag, he-goat  akin to OHG
[italic boc]#
he-goat, Mir [italic bocc]
1  [italic pl]
buck 2l
1  n a male animal  [italic esp] : a male deer or
antelope
2 a  n a male human being : [mini MAN]
2 b  n a dashing fellow : [mini DANDY]
3  [italic pl]
buck 2l
3  n [mini ANTELOPE]
4 a  n [mini BUCKSKIN]  [italic also] : an article made
of buckskin
4 b  [italic slang]
4 b  n [mini DOLLAR] 3b
short for [italic sawbuck]
5  n [mini SAWHORSE]
6 a  n a supporting rack or frame
6 b  n a short thick leather-covered block for gymnastic
vaulting
buck 2  vi
1  [italic of a horse or mule]
1  vi to spring with a quick plunging leap
2  vi to charge against something as if butting
3 a  vi to move or react jerkily
3 b  vi [mini BALK]
4  vi to strive for advancement sometimes without
regard to ethical#
behavior
1  vt to throw (as a rider) by bucking
2 a  [italic archaic]
2 a  vt [sup 1][mini BUTT]
2 b  vt [mini OPPOSE], [mini RESIST]
3  vt to charge into (the opponents' line in football)
4  vt to pass esp. from one person to another

```

bucker 4 n
 buck 3 aj
 prob. fr. [^{sup 1}][*italic buck*]
 0 aj of the lowest grade within a military category <~
 private>
 buck 4 n
 short for earlier [*italic buckhorn knife*]
 0 n an object formerly used in poker to mark the next
 player to#
 deal [*italic broadly*] : a token used as a mark or
 reminder
 buck 5 av
 origin unknown
 0 [*italic South & Midland*]
 0 av [mini STARK] <~ naked>

RHD

(buk), v.i., n. Anglo-Indian. [1] bukh.

(buk), Brit. Dial. [1] n. lye used for washing clothes.
 [2] clothes washed
 in lye. [3] v.t. to wash or bleach (clothes) in
 lye.[1350-1400; ME bouken
 (v.); cf. MLG buken, b{be}ken to steep in lye, MHG
 buchen, bruchen]

(buk), n. Slang. [1] a dollar.[1855-60, Amer.; perh.
 {92}{c4}{93}{a2}{81} in
 sense "'buckskin'"; deerskins were used by Indians and
 frontiersmen as a
 unit of exchange in transactions with merchants]

(buk), adv. Informal. [1] completely; stark: buck
 naked.[1925-30, Amer.; of
 obscure orig.]

(buk), n. [1] Poker. any object in the pot that reminds the winner of some
 privilege or obligation when his or her turn to deal next comes. [2] pass
 the buck, to shift responsibility or blame to another person: Never one to
 admit error, he passed the buck to his subordinates. [3] v.t.<gv450,159> to
 pass (something) along to another, esp. as a means of avoiding
 responsibility or blame: He bucked the letter on to the assistant vice
 president to answer.[1860-65; short for buckhorn knife, an object which
 served this function]

(buk), n. [1] the male of the deer, antelope, rabbit, hare, sheep, or goat. [2] the male of certain other animals, as the shad. [3] an impetuous, dashing, or spirited man or youth. [4] Often Disparaging. a male American Indian or black. [5] <gv450,34> buckskin. [6] bucks, casual oxford shoes made of buckskin, often in white or a neutral color. [7] adj. Mil. of the lowest of several ranks involving the same principal designation, hence subject to promotion within the rank: buck private; buck sergeant.[bef. 1000; ME bukke, OE bucca he-goat, bucc male deer; c. D bok, G Bock, ON bukkr; def. 5, 6 by shortening; buck private (from ca. 1870) perh. as extension of general sense "'male,'" i.e., having no status other than being male]

(buk), v.i. [1] (of a saddle or pack animal) to leap with arched back and come down with head low and forelegs stiff, in order to dislodge a rider or pack. [2] Informal. to resist or oppose obstinately; object strongly: The mayor bucked at the school board's suggestion. [3] (of a vehicle, motor, or the like) to operate unevenly; move by jerks and bounces. [4] v.t. to throw or attempt to throw (a rider or pack) by bucking. [5] to force a way through or proceed against (an obstacle): The plane bucked a strong headwind. [6] to strike with the head; butt. [7] to resist or oppose obstinately; object strongly to. [8] Football. (of a ball-carrier) to charge into (the opponent's line). [9] to gamble, play, or take a risk against: He was bucking the odds when he bought that failing business. [10] to press a reinforcing device against (the force of a rivet) in order to absorb vibration and increase expansion. [11] buck for, to strive for a promotion or some other advantage: to buck for a raise. [12] buck up, to make or become more cheerful, vigorous, etc.: She knew that with a change of scene she would soon buck up. [13] n. an act of bucking.[1855-60; verbal use of {92}{c4}{93}{a2}{81}, influenced in some senses by {92}{c4}{93}{a2}{83}]

(buk), n. [1] a sawhorse. [2] Gymnastics. a cylindrical, leather-covered{83} block mounted in a horizontal position on a single vertical post set in a steel frame, for use chiefly in vaulting. [3] any of various heavy frames, racks, or jigs used to support materials or partially assembled items during manufacture, as in airplane assembly plants. [4] Also called door buck. a doorframe of wood or metal set in a partition, esp. one of light masonry, to support door hinges, hardware, finish work, etc. [5] v.t. to split or saw (logs, felled trees, etc.). [6] buck in, Survey., Optical Tooling. to set up an instrument in line with two marks.[1855-60; short for {c2}{91}{c6}{92}{c4}{93}{a2}]

CED

#Hbuck#5%1 (b^k) #6n.

@n#1\$D. a. #5the male of various animals including the goat, hare, kangaroo, rabbit, and reindeer. @n#1b. #5(#6as modifier#5)#6: a buck antelope.

@n#1\$D. #6Informal. #5a robust spirited young man.
 @n#1\$D. #6U.S. slang, derogatory. #5a young male Indian or Negro.
 @n#1\$D. #6S. African. #5an antelope or deer of either sex.
 @n#1\$D. #6Archaic. #5a dandy; fop.
 @n#1\$D. #5the act of bucking.
 @m#1?~#6vb.
 @n#1\$D. #5(#6intr.#5) (of a horse or other animal) to jump vertically, with legs stiff and back arched. @n#1\$D. #5(#6tr.#5) (of a horse, etc.) to throw (its rider) by bucking. @n#1\$D. #5(when #6intr., #5often foll. by #6against#5) #6Informal, chiefly U.S. #5to resist or oppose obstinately: #6to buck against change; to buck change. @n#1\$D. #5(#6tr.; usually passive#5) #6Informal. #5to cheer or encourage: #6I was very bucked at passing the exam.
 @n#1\$D. #6U.S. informal. #5(esp. of a car) to move forward jerkily; jolt. @n#1\$D. #6U.S. #5to charge against (something) with the head down; butt. @m#1~#5See also #1buck up. @n#5[Old English #6bucca #5he-goat; related to Old Norse #6bukkr, #5Old High German #6bock, #5Old Irish #6bocc#5] @m?~#1#!buck#+er@f#6n.
 #Hbuck#5%2 (b^k) #6n. U.S. and Austral. slang. #5a dollar. @m[C19: of obscure origin]
 #Hbuck#5%3 (b^k) #6n. @n#1\$D. #6Gymnastics. #5a type of vaulting horse. @n#1\$D. #5a U.S. word for #1sawhorse. @m#5[C19: short for #7sawbuck#5]
 #Hbuck#5%4 (b^k) #6n. @n#1\$D. #6Poker. #5a marker in the jackpot to remind the winner of some obligation when his turn comes to deal. @n#1\$D. pass the buck. #6Informal. #5to shift blame or responsibility onto another. @m[C19: probably from #6buckhorn knife, #5placed before a player in poker to indicate that he was the next dealer]
 #HBuck #5(b^k) #6n. #1Pearl S#5(#1ydenstricker#5)#1. #51892@=1973, U.S. novelist, noted particularly for her novel of Chinese life #6The Good Earth #5(1931): Nobel prize for literature 1938.
 Macquarie
 \H buck\$s11]
 \P /i[/f/fbxk/i],
 \D1 /in. /1 /nthe male of certain animals, as the deer, antelope, rabbit, or hare.
 \D2 /1 /na young man viewed as a sexual animal; a fop; dandy.

\D3 /1 /iU.S. Colloq. /n(/iderog./n) a male Indian or Negro.

\D4 @/iadj. /1 /n(/iderog./n) male: /ia buck nigger.

\E [ME /ibukke, /ncoalescence of IE /ibucca /nhe-goat and /ibucc
/nmaale deer, c. G /iBock/n]

\H buck\$s12]

\P /i[/f/fbxk/i],

\D1 /iv.i. /1 /n(of a saddle or pack animal) to leap with arched
back and come down with head low and forelegs stiff, in order to
dislodge rider or pack.

\D2 /1 /iColloq. /nto resist obstinately; object strongly: /ito buck
at improvements.

\D3 /1 /iColloq. /nto hurry (fol. by /iup/n)

\D4 /1 /iColloq. /nto become more cheerful, vigorous, etc. (fol. by /iup/n).

\D5 @/iv.t. /1 /nto throw or attempt to throw (a rider) by bucking.

\D6 /1 /iColloq. /nto resist obstinately; object strongly to: /ito
buck the system.

\D7 /1 /iColloq. /nto force or urge (someone) to hurry (fol. by /iup/n).

\H buck\$s12]

\D8 /1 /iColloq. /nto make more cheerful, vigorous, etc. (fol. by /iup/n).

\D9 @/in. /1 /nan act of bucking.

\D10 /1 give it a buck, /iColloq. /nto make an attempt; chance.

\D11 /1 have a buck at, /nto try; make an attempt.

\E [special use of /sbuck\$s11]]

\H buck\$s13]

\P /i[/f/fbxk/i],

\D1 /in. /1 /iPoker. /nany object in the kitty which reminds the winner that he has some privilege or duty when his turn to deal next comes.

\D2 /1 pass the buck, /iColloq. /nto shift the responsibility or blame to another person.

\E [orig. uncert.]

\H buck\$14]

\P /i[/f/fbxk/i],

\D1 /in. Orig. U.S. Colloq. /1 /na dollar.

\D2 /1 a fast buck, /nmoney earned with little effort, often by dishonest means.

\E [shortened form of /sbuckskin, /nan accepted form of exchange in the U.S. frontier.]

7.5 Appendix 4

Gutenberg-style Text

THE LOST WORLD

I have wrought my simple plan
If I give one hour of joy
To the boy who's half a man,
Or the man who's half a boy.

The Lost World

By SIR ARTHUR CONAN DOYLE

COPYRIGHT, 1912

Foreword

Mr. E. D. Malone desires to state that both the injunction for restraint and the libel action have been withdrawn unreservedly by Professor G. E. Challenger, who, being satisfied that no criticism or comment in this book is meant in an offensive spirit, has guaranteed that he will place no impediment to its publication and circulation.

Contents

CHAPTER

- I. "THERE ARE HEROISMS ALL ROUND US"
- II. "TRY YOUR LUCK WITH PROFESSOR CHALLENGER"
- III. "HE IS A PERFECTLY IMPOSSIBLE PERSON"
- IV. "IT'S JUST THE VERY BIGGEST THING IN THE WORLD"
- V. "QUESTION!"
- VI. "I WAS THE FLAIL OF THE LORD"
- VII. "TO-MORROW WE DISAPPEAR INTO THE UNKNOWN"
- VIII. "THE OUTLYING PICKETS OF THE NEW WORLD"

- IX. "WHO COULD HAVE FORESEEN IT?"
- X. "THE MOST WONDERFUL THINGS HAVE HAPPENED"
- XI. "FOR ONCE I WAS THE HERO"
- XII. "IT WAS DREADFUL IN THE FOREST"
- XIII. "A SIGHT I SHALL NEVER FORGET"
- XIV. "THOSE WERE THE REAL CONQUESTS"
- XV. "OUR EYES HAVE SEEN GREAT WONDERS"
- XVI. "A PROCESSION! A PROCESSION!"

THE LOST WORLD

The Lost World

CHAPTER I

"There Are Heroisms All Round Us"

Mr. Hungerton, her father, really was the most tactless person upon earth,--a fluffy, feathery, untidy cockatoo of a man, perfectly good-natured, but absolutely centered upon his own silly self. If anything could have driven me from Gladys, it would have been the thought of such a father-in-law. I am convinced that he really believed in his heart that I came round to the Chestnuts three days a week for the pleasure of his company, and very especially to hear his views upon bimetallism, a subject upon which he was by way of being an authority.

For an hour or more that evening I listened to his monotonous chirrup about bad money driving out good, the token value of silver, the depreciation of the rupee, and the true standards of exchange.

"Suppose," he cried with feeble violence, "that all the debts in the world were called up simultaneously, and immediate payment insisted upon,--what under our present conditions would happen then?"

I gave the self-evident answer that I should be a ruined man, upon which he jumped from his chair, reproved me for my habitual levity, which made it impossible for him to discuss any

reasonable subject in my presence, and bounced off out of the room to dress for a Masonic meeting.

At last I was alone with Gladys, and the moment of Fate had come! All that evening I had felt like the soldier who awaits the signal which will send him on a forlorn hope; hope of victory and fear of repulse alternating in his mind.

7.6 Appendix 5

Some Dramatic Texts

Beggar's Opera. The bits in curly braces are footnotes.

Scene 5, Peachum's Lock.

A Table with Wine, Brandy, Pipes, and Tobacco.

LOCKIT. The Coronation Account{91}, Brother Peachum, is of so intricate a nature, that I believe it will never be settled.

PEACHUM. It consists indeed of a great Variety of Articles.----It was worth to our People, in Fees of different kinds, above ten Instalments.----This is part of the Account, Brother, that lies open before us.

LOCKIT. A Lady's Tail{92} of rich Brocade----that, I see, is dispos'd of.

PEACHUM. To Mrs. Diana Trapes, the Tally-Woman, and she will make a good Hand on't in Shoes and Slippers, to trick out young Ladies, upon their going into Keeping----

LOCKIT. But I don't see any Article of the Jewels.

PEACHUM. Those are so well known that they must be sent abroad----You'll find them enter'd upon the Article of Exportation.----As for the Snuff-Boxes, Watches, Swords, &c.----I thought it best to enter them under their several Heads.

LOCKIT. Seven and twenty Women's Pockets complete; with the several things therein contain'd; all Seal'd, Number'd, and Enter'd.

PEACHUM. But, Brother, it is impossible for us now to enter upon this Affair.--We should have the whole Day before us.----Besides, the Account of the last Half Year's PLate is in a Book by itself, which lies at the other Office.

LOCKIT. Bring us then more Liquor.----To-day shall be for Pleasure----Tomorrow for Business--Ah, Brother, those Daughters of ours are two slippery Hussies----Keep a watchful eye upon Polly, and Macheath in a day or two shall be our own again.

Air XLV.--Down in the North Country, &c.

LOCKIT.

What Gudgeons{93} are we Men!

Ev'ry Woman's easy Prey.

Though we have felt the Hook, agen

We bite and they betray.

The Bird that hath been trapt,

When he hears his calling Mate,

To her he flies, again he's clapt

Within the wiry Grate.

PEACHUM. But what signifies catching the Bird, if your Daughter Lucy will set open the Door of the Cage?

LOCKIT. If Men were answerable for the Follies and Frailties of the Wives and Daughters, no Friends could keep a good Correspondence together for two Days.----This is unkind of you, Brother; for among good Friends, what they say or do goes for nothing.

Enter a Servant.

SERVANT. Sir, here's Mrs. Diana Trapes wants to speak with you.

PEACHUM. Shall we admit her, Brother Lockit?

LOCKIT. By all means,----She's a good Customer, and a fine-spoken Woman---- And a Woman who drinks and talks so freely, will enliven the Conversation.

PEACHUM. Desire her to walk in.

[Exit Servant.]

The Country Wife. This has minimal markup, but many print hangovers. Note the bits after the pipes. Curly braces here indicate italics.

<S Hor.> Doctor, there are Quacks in love, as well as Physick, who get but the fewer and worse Patients, for their boastings a good name is seldom got by giving it ones self, and Women no more than honour are compass'd by bragging : Come, come Doctor, the wisest Lawyer never discovers the merits of his cause till the tryal; the wealthiest Man conceals his riches, and the cunning Gamster his play; Shy Husbands and Keepers like old Rooks are not to be cheated, but by a new unpractis'd trick; false friendship will pass now no more than false dice upon'em, no, not in the City.

{Enter Boy.}

<S Boy.> There are two Ladies and a Gentleman coming up.

<P 3>

<S Hor.> A Pox, some unbelieving Sisters of my former acquaintance, who I am afraid, expect their sense shou'd be satisfy'd of the falsity of the report. | <Z Enter Sir Jasp. Fidget, No--this formal Fool and Women{!} | Lady Fidget, and Mrs.

<S Qu.> His Wife and Sister. | Dainty Fidget.

<S Sr. Jas.> My Coach breaking just now before your door Sir, I look upon as an occasional reprimand to me Sir, for not kissing your hands Sir, since your coming out of {France} Sir; and so my disaster Sir, has been my good fortune Sir; and this is my Wife, and Sister Sir.

<S Hor.> What then, Sir?

<S Sr. Jas.> My Lady, and Sister, Sir.----- Wife, this is Master {Horner.}

<S La. Fid.> Master {Horner}, Husband {!}

<S Sr. Jas.> My Lady, my Lady {Fidget}, Sir.

<S Hor.> So, Sir.

<S Sr. Jas.> Won't you be acquainted with her Sir?
 [So the report is true, I find by his coldness or aversion to the Sex; but I'll play the wag with him.] <D Aside.>
 Pray salute my Wife, mt Lady, Sir.

<S Hor.> I will kiss no Mans Wife, Sir, for him, Sir; I have taken my eternal leave, Sir, of the Sex already, Sir.

<S Sr. Jas.> Hah, hah, hah; I'll plague him yet. <D aside.>
 Not know my Wife, Sir?

<S Hor.> I do know your Wife, Sir, she's a Woman, Sir, and consequently a Monster, Sir, a greater Monster than a Husband, Sir.

<S Sr. Jas.> A Husband; how, Sir?

<S Hor.> So, Sir; but I make no more Cuckholds, Sir. <D makes horns.>

<S Sr. Jas.> Hah, hah, hah, {Mercury, Mercury.}

<S La. Fid.> Pray, Sir {Jaspar}, let us be gone from this rude fellow.

<S Mrs. Daint.> Who, by his breeding, wou'd think, he had ever been in {France?}

<S La. Fid.> Foh, he's but too much a French fellow, such as hate Women of quality and virtue, for their love to their
 <P 4>
 Husbands, Sr. {Jaspar}; a Woman is hated by'em as much for loving her Husband, as for loving their Money: But pray, let's be gone.

7.7 Appendix 6

Lination — To the Human Eye and a Computer...

A sample of text from the First Quarto of Hamlet, as it appears to the human eye and to a machine trying to process it.

```

2 Enter two Centinels. [B1
4-5 1. Stand: who is that? 2
9 2. Tis I. 3
10 1. O you come most carefully vpon your watch,4
16-7 2. And if you meete Marcellus and Horatio,5
17 The partners of my watch, bid them make haste. 6
19 1. I will: See who goes there. 7
18 Enter Horatio and Marcellus. 8
20 Hor. Friends to this ground. 9
21 Mar. And leegemen to the Dane, 10
23 O farewell honest souldier, who hath releued you? 11

```

The text here cleverly displays two sets of line numbers from different editions. The problem with the numbers in this form, however, is that a computer will see this text as follows:

```

2 Enter two Centinels. [B1 4-5 1.
Stand: who is that? 2 9 2. Tis
I. 3 10 1. O you come most carefully vpon your watch,4
16-7 2. And if you meete Marcellus and Horatio,5 17 The partners
of my watch, bid them make haste. 6 19 1. I will: See who goes
there. 7 18 Enter Horatio and Marcellus. 8 20 Hor. Friends
to this ground. 9 21 Mar. And leegemen to the
Dane, 10 23 O farewell honest souldier, who hath releued you? 11

```

as a bit stream with a bunch of digits, tabs, and characters (perhaps including newline characters). Here it is with the tabs made visible:

```

^I2^I^IEnter two Centinels.^I[B1 4-5^I1. Stand: who is that?^I^I2 9^I2. Tis
I.^I^I3 10^I1. O you come most carefully vpon your watch,^I4 16-7^I2. And if
you meete Marcellus and Horatio,^I5 17^IThe partners of my watch, bid them
make haste.^I6 19^I1. I will: See who goes there.^I7 18^I^IEnter Horatio and
Marcellus.^I8 20^IHor. Friends to this ground.^I 9 21^IMar. And leegemen to
the Dane,^I 10 23^IO farewell honest souldier, who hath releued you?^I 11

```

You'll notice that nothing really separates the two numbers (without the perilously ephemeral newlines...). The tabs work to separate number-text-number for the human eye, but they are not identified as distinct structures. In SGML, one might do it something like this with double attributes:

```

<l n=2 id=[B1]><stage type='entrance'>Enter two Centinels.</stage></l>
<l n=4-5 id=2>1. Stand: who is that?</l> <l n=9 id=3>2. Tis I.</l>
<l n=10 id=4>1. O you come most carefully vpon your watch,</l>
<l n=16-7 id=5>2. And if you meete Marcellus and Horatio,</l>
<l n=17 id=6>The partners of my watch, bid them make haste.</l>
<l n=19 id=7>1. I will: See who goes there.</l>
<l n=18 id=8><stage type='entrance'>Enter Horatio and Marcellus.</stage></l>
<l n=20 id= 9>Hor. Friends to this ground.</l>
<l n=21 id= 10>Mar. And leegemen to the Dane,</l>
<l n=23 id= 11>O farewell honest souldier, who hath releued you?</l>

```

so that a machine reading this will have less trouble:

```

<l n=2 id=[B1]><stage type='entrance'>Enter two Centinels.</stage></l>
<l n=4-5 id=2>1. Stand: who is that?</l> <l n=9 id=3>2. Tis I.</l> <l n=10
id=4>1. O you come most carefully vpon your watch,</l> <l n=16-7 id=5>2.
And if you meete Marcellus and Horatio,</l> <l n=17 id=6>The
partners of my watch, bid them make haste.</l> <l n=19 id=7>1. I
will: See who goes there.</l> <l n=18 id=8><stage
type='entrance'>Enter Horatio and Marcellus.</stage></l>
<l n=20 id= 9>Hor. Friends to this ground.</l> <l n=21 id=
10>Mar. And leegemen to the Dane,</l> <l n=23 id= 11>O
farewell honest souldier, who hath releued you?</l>

```

The bits inside tags are quite easy now to isolate and process.