

Review: Oxford Electronic Text Library Edition of the Complete Works of Jane Austen

Reviewed Work(s): Oxford Electronic Text Library Edition of the Complete Works of Jane Austen by

Review by: Eric Johnson

Source: *Computers and the Humanities*, Vol. 28, No. 4/5, Humanities Computing in Norway (1994/1995), pp. 317-321

Published by: Springer

Stable URL: <http://www.jstor.org/stable/30204601>

Accessed: 20-02-2018 16:21 UTC

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <http://about.jstor.org/terms>



JSTOR

Springer is collaborating with JSTOR to digitize, preserve and extend access to *Computers and the Humanities*

Price: \$49.95.

Title: Malloy, Judy. *It's Name Was Penelope* (version 1.0). 1990.

Environment: *Storyspace Reader*, 1992.

System Requirements: MacIntosh Plus or higher, with 2M memory and hard disk, or IBM-PC 386/486 compatible, MS Windows 3.1 and hard disk.

Company:

Eastgate Systems, Inc.
134 Main Street
Watertown, MA 02172
(800)562-1638 or (617)924-9044.

Price: \$24.95 for the MacIntosh, \$19.95 for the IBM PC.

Title: Guyer, Carolyn. *Quibbling*. 1993.

Environment: *Storyspace Reader*, 1990.

System Requirements: MacIntosh Plus or higher, with 2M memory and hard disk.

Company:

Eastgate Systems, Inc.
134 Main Street
Watertown, MA 02172
(800)562-1638 or (617)924-9044.

Price: \$19.95.

Oxford Electronic Text Library Edition of the Complete Works of Jane Austen

by Eric Johnson

Scholars in the humanities today are routinely doing textual and linguistic research that a generation

Eric Johnson is professor of English and Dean of the College of Liberal Arts at Dakota State University. He is the Editor of TEXT Technology: The Journal of Computer Text Processing, and he is the author of more than sixty publications about computers, writing, and literary study.

ago would have been impossible or would have required the dedication of a lifetime. Such research is now feasible because humanists use computers and because texts of major writers are available in electronic form. The *Oxford Electronic Text Library Edition of The Complete Works of Jane Austen* (OETL Austen) is exactly the kind of electronic text that modern scholars need. It is an accurate rendering of R. W. Chapman's *Oxford Illustrated Jane Austen*, the standard scholarly edition of Austen, and it contains a wealth of useful information encoded in Standard Generalized Markup Language (SGML). The *OETL Austen* is distributed in both MS-DOS and Macintosh formats, and a site license is available. It will be used in a multitude of ways by students of Austen for years to come.

Figure 1 shows the first 18 lines of text of the *OETL Austen* file of *Pride and Prejudice*. Since this is an SGML-conformant document, it contains tags (enclosed in angle brackets <>) and references (preceded by an ampersand &) that supply various kinds of information about the text. The SGML tags and references can be removed from the text, of course, and necessary substitutions made in order to format the text into a readable form; the result of such formatting is shown in Figure 2.

The formatted version of the text is useful to scholars, since they can load large or small parts of it directly into most word processors. A writer who wants to quote the well-known first two paragraphs of *Pride and Prejudice*, can paste them into a document with a few mouse clicks. The advantage of electronically inserting part of one document into another is not only the saving of time, but the assurance of accuracy; a writer on Austen can confidently assert that all of the included quotations conform to the Chapman edition.

The formatted version of the text can also be searched using a word processor or commonly available software. In this way, a researcher can find that the context of the first occurrence of "invitation" in *Pride and Prejudice* is "This was invitation enough." The problem, however, with searching a formatted text is that it is difficult to identify just where in a text an occurrence is found, and it is nearly impossible to catalog multiple occurrences spread over many pages.

SGML-conformant documents (such as the *OETL Austen*) contain tags that indicate the divisions and parts of documents: volumes, chapters, pages, lines, and so on. As can be seen in Figure 1, each line break in the novel is identified with a tag at the start of the line: the numbers at the end of the tag correspond to the

<lb n=P3.1>It is a truth universally acknowledged, that&H3 a single man
 <lb n=P3.2>in&H4 possession of a good fortune, must be in&H4 want&H0 of a wife.</q></p><p><q
 who=PP0>
 <lb n=P3.3>However little known the feelings or views of such a
 <lb n=P3.4>man may&H1 be on&H4 his first entering a neighbourhood, this
 <lb n=P3.5>truth is so&H51 well&H5 fixed in&H4 the minds of the surrounding
 <lb n=P3.6>families, that&H3 he is considered as the rightful property of
 <lb n=P3.7>some one or other of their daughters.</q></p><p><q who=PPD>
 <lb n=P3.8>&dq;My dear&H21 <name who=PPC>Mr&point;&sp;Bennet</name>,&dq;</q><q who=PP0>
 said his lady to&H4 him one day,</q><q who=PPD>
 <lb n=P3.9>&dq;have you heard that&H3 Netherfield&sp;Park is let at last&H0;?&dq;</q></p><p><q
 who=PP0>
 <lb n=P3.10><name who=PPC>Mr&point;&sp;Bennet</name> replied that&H3 he had not.</q></p><p>
 <q who=PPD>
 <lb n=P3.11>&dq;But it is,&dq;</q><q who=PP0>returned she;</q><q who=PPD>&dq;for&H3 <name
 who=PPV>Mrs&point;&sp;Long</name> has just&H5;
 <lb n=P3.12>been here, and she told me all about&H4 it.&dq;</q></p><p><q who=PP0>
 <lb n=P3.13><name who=PPC>Mr&point;&sp;Bennet</name> made no&H2 answer&H0;.</q></p><p>
 <q who=PPD>
 <lb n=P3.14>&dq;Do not you want&H1 to&H9 know who&H61 has taken it?&dq;</q><q who=PP0>cried
 <lb n=P3.15>his wife impatiently.</q></p><p><q who=PPC>
 <lb n=P3.16>&dq;<hi r=Italic>You</hi> want&H1 to&H9 tell me, and I have no&H2 objection to&H4;
 <lb n=P3.17>hearing it.&dq;</q></p><p><q who=PP0>
 <lb n=P3.18>This was invitation enough.</q></p><p><q who=PPD>

Fig. 1. SGML-conformant content of an *OETL Austen* file.

It is a truth universally acknowledged, that a single man
in possession of a good fortune, must be in want of a wife.

However little known the feelings or views of such a
man may be on his first entering a neighbourhood, this
truth is so well fixed in the minds of the surrounding
families, that he is considered as the rightful property of
some one or other of their daughters.

“My dear Mr. Bennet,” said his lady to him one day,
“have you heard that Netherfield Park is let at last?”

Mr. Bennet replied that he had not.

“But it is,” returned she; “for Mrs. Long has just
been here, and she told me all about it.”

Mr. Bennett made no answer.

“Do not you want to know who has taken it?” cried
his wife impatiently.

“You want to tell me, and I have no objection to
hearing it.”

This was invitation enough.

Fig. 2. Formatted content of an *OETL Austen* file.

Words	References (page number.line number)
amiable	7.14; 17.10; 26.21; 43.4; 54.29; 55.26; 75.36; 92.18; 96.19; 104.16; 111.13; 124.20; 138.10; 141.34; 148.12; 148.15; 149.18; 149.19; 149.22; 160.10; 161.19; 181.8; 191.14; 197.6; 204.16; 243.7; 250.14; 285.1; 328.2; 433.1; 438.21; 450.2; 462.30; 462.38; 474.33
agreeable	11.4; 33.2; 34.35; 42.5; 42.7; 42.17; 47.23; 53.7; 54.28; 75.33; 82.28; 90.7; 111.32; 113.15; 116.6; 139.12; 149.20; 149.33; 150.9; 165.11; 169.30; 171.23; 171.27; 176.36; 191.5; 192.15; 192.37; 194.36; 196.17; 201.2; 202.8; 203.22; 212.26; 221.5; 232.38; 250.17; 281.20; 281.31; 292.33; 303.23; 308.37; 332.2; 367.21; 368.3; 381.25; 381.27; 387.8; 399.31; 444.6; 476.13
delicacy	51.6; 136.34; 149.21; 167.24; 179.31; 199.21; 226.28; 287.9; 348.12; 421.24; 439.21; 442.38; 447.11; 448.16; 463.20; 478.8
manners	6.15; 23.13; 24.1; 33.24; 33.26; 34.12; 34.29; 42.6; 48.33; 54.27; 56.11; 65.25; 92.17; 92.34; 93.9; 100.24; 111.36; 112.16; 118.21; 129.15; 134.20; 134.31; 135.29; 136.29; 149.20; 149.30; 167.9; 169.34; 169.36; 192.15; 194.28; 212.28; 212.29; 262.10; 270.32; 271.2; 272.15; 278.24; 281.33; 284.13; 310.29; 320.4; 321.18; 364.30; 382.2; 386.11; 396.34; 426.28; 426.37; 428.18; 440.25; 445.36; 459.30

Fig. 3. Index for *Emma* giving references for selected words.

page and line number in the printed Chapman edition. Using the full *OETL Austen* texts, a researcher can search for specific words, and when they are located, record the references.

On page 149 of *Emma*, Mr. Knightly makes a distinction important in the novel between being “amiable” and “agreeable.”¹ To be “amiable” is to show a true “delicacy towards the feelings of other people,” but someone can be “agreeable” by merely having “very good manners.” A researcher interested in Knightly’s differentiation might want to have an index of all occurrences in *Emma* of the key words “amiable,” “agreeable,” “delicacy,” and “manners.” Figure 3 gives exactly such an index. It was generated by processing the full *OETL Austen* text of *Emma*, and each time one of the four key words was found, an entry of the corresponding page and line numbers was put into a table. It would be rather straightforward to build a key-word-in-context concordance in a similar manner.

Tags in the *OETL Austen* indicate direct (and indirect) quotation, and the speaker for each. For example, the tag at the end of line seven, <q who=PPD>, indicates that what follows (up to </q> in line nine) is

a quotation, and the code PPD identifies the speaker, Mrs. Bennet (files that accompany the *OETL Austen* texts list the correspondences between codes and full character names). Therefore, a researcher can isolate and analyze the dialogue of each character.² It is interesting to compare how much the heroine of *Emma* and the heroine of *Pride and Prejudice* talk: Emma Woodhouse speaks about 27 percent of the number of words of dialogue and Elizabeth Bennet speaks about 15 percent. Elizabeth speaks only about one-third as much as the narrator in her novel, but Emma talks almost as much as the narrator in hers.

It can be helpful to a scholar to know how many nouns, verbs, adjectives, and so on are in a text, and to observe how they are used. Of course, a form such as “well” commonly functions as a noun (“get water from the well”) and as an adverb (“it was well done”). Therefore, the *OETL Austen* texts contain reference codes to identify the function of homographic forms such as “well.” Line five in Figure 1 contains references (each begins with an ampersand) that identify “so” and “well” as adverbs and “in” as a preposition (files included with the *OETL Austen* list the correspon-

dences between the homograph codes and categories such as noun, verb, and so on). With such references to disambiguate forms, it is possible to make counts. Although the first chapters of all Austen novels are similar, the first chapter of *Sense and Sensibility* has a higher percentage of nouns and a lower percentage of verbs than the first chapter of *Mansfield Park*; they have an identical percentage of prepositions.

Thus, researchers can use the *OETL Austen* to collect the dialogue of each speaker, and they then can analyze occurrences of specific words and phrases as well as the word forms used by each in order to differentiate and characterize a wide range of Austen's speakers. An index of occurrences of words (like that shown in Figure 3) could be restricted to those occurrences in Emma's dialogue – or to those of any specific characters. It might be found that some characters use far more adjectives and adverbs than others. Such analyses can be more sophisticated and far more detailed than would be feasible without the *OETL Austen*.

The electronic edition upon which the *OETL Austen* is based was prepared by John Burrows and Alexis Antonia; conversion of their encoding scheme was done by Lou Burnard. They should be proud of their work. Although, as it is currently shipping, it does not contain the *Minor Works*, the *OETL Austen* appears to be a complete and accurate electronic rendering of the six major novels published in the Chapman *Oxford Illustrated Jane Austen*. Occasionally a space is omitted from the text following a reference code, but that is a slight flaw that affects only the appearance of a formatted text and does not alter the identification of words or codes.

Marketing electronic text files is a new idea, and a book publisher may be a little uncertain about how to do it. The *OETL Austen* contains only files of texts; there is no software included to use with the texts. (The analyses mentioned in this review were produced by programs created by this reviewer.) Oxford University Press would probably find the *OETL Austen* would appeal to a larger market if it included some kind of software: perhaps a search program similar to the one that produced the output shown in Figure 3, or a formatting program like the one that produced the output shown in Figure 2. Of course, since the texts are encoded with SGML, they should be able to be used with software designed to process SGML – such as *Intellitag* (from WordPerfect) or *DynaText* (from Electronic Book Technologies).

The twenty-five pages of printed documentation give a tantalizingly brief account of how the electronic texts were prepared and used by Burrows. However, the texts as Burrows used them were in a format different from the *OETL Austen* SGML format. Since the documentation mostly describes Burrows' format, the documentation contributes almost nothing to an understanding of the complex form of the *OETL Austen* texts. For example, page 13 says, "To assist in the study of sentence-lengths, we used % (as in Mr%) to distinguish abbreviation-points from full-stops." The *OETL Austen* uses *&point* for such a purpose (see line 8 in Figure 1). Sometimes the documentation seems to give details of the conversion to the SGML format, but it is not reliable. As part of the description of marking homographic forms, page 17 says that two numbers, "were separated by a sharp sign" but "this has now been replaced by a dot." There is, in fact, no dot (see line 5 in Figure 1), and, to add to the confusion on this subject, on page 18, a verb code is mistakenly cited for a word that Jane Austen would never have used as a verb. Nevertheless, the documentation is a small part of the package, and perhaps it will be revised along the lines of the documentation for the *OETL Poetical Works of Samuel Taylor Coleridge* which is quite helpful.

Oxford University Press deserves praise and support for publishing the *OETL Austen* – as well as similar editions for other major English writers. The *OETL Austen* is an accurate version of the standard scholarly printed text, and it contains a good deal of useful information precisely encoded in SGML. It is specifically the kind of electronic text that will both enable and stimulate future textual and linguistic research.

Title: *The Oxford Electronic Text Library edition of The Complete Works of Jane Austen.*

Category: SGML-conformant electronic (machine-readable) text.

System Requirements: Currently (without the *Minor Works*) the *OETL Austen* consists of 77 files which require approximately 6.7 MB of disk space; additional space is required to concatenate the files into useful entities. Available in Macintosh disk format or either 5.25-inch or 3.5-inch MS-DOS disk format.

Documentation: Twenty-five page pamphlet.

Company:

Oxford University Press
Electronic Publishing
200 Madison Ave.
New York, NY 10016
(212) 679-7300 ext. 7390.

Price: \$95.00; site license: \$295.00.

Notes

¹ This distinction is discussed in Stuart M. Tave, *Some Words of Jane Austen*, Chicago, 1973, 116-57.

² Exactly such analysis is the content of J.F. Burrows, *Computation into Criticism: A Study of Jane Austen's Novels and an Experiment in Method*, Oxford, 1987. The electronic editions of the Austen novels that were used for his research are the basis of the *OETL Austen*.