Hypertext and the

Oxford English Dictionary

Darrell R. Raymond and Frank Wm. Tompa

Department of Computer Science
University of Waterloo
Waterloo, Ontario, Canada
N2L 3G1

ABSTRACT

Hypertext databases can be produced by converting existing text documents to electronic form. The basic task in conversion is fragmentation, which is the identification of independent and discrete nodes and links in continuous text. We illustrate that fragmentation is not always a straightforward process with an analysis of the Oxford English Dictionary. The results of the analysis suggest that dynamic fragmentation may be more suitable for some texts than static fragmentation.

INTRODUCTION

Low-cost high-capacity storage devices are making widespread computerized access to existing reference texts a reality. Documents currently available in electronic form include Bowker's *Books In Print*, Microsoft's *Bookshelf* (a compendium of ten works including *Roget's Thesaurus*, *Bartlett's Familiar Quotations*, the *World Almanac*, and the *U.S. Zip Code Directory*), the *McGraw-Hill Concise Encyclopedia of Science and Technology*, and the *Electronic Holmes Companion*. Academics can make use of the approximately one thousand texts maintained by The Oxford Text Archives, which include documents of scholarly interest such as the *Tibetan Book of the Dead* and *Steppenwolf*. The basic assets of computerization — speed of access and quantity of storage — make almost any form of electronic document useful. However, computerization also presents new opportunities in advanced representations for such documents, both for storage and for display. Hypertext is one such opportunity.

Hypertext is a database technology typically characterized by small fragments of text interconnected by machine-supported links [Conk87][Jone87]. Since most existing texts are not fragmented in this way, the key problem in converting them to hypertext is development of the text fragments and the links. Fragmentation must be carried out carefully if the hypertext is to be a faithful representation of the original. If the document was created decades or centuries ago, its authors or editors may not be available for consultation. As a result, structure must be inferred from a careful study of the text and from consultation with existing experts. The characteristics of pen and paper or the printing press often exert a powerful influence on a document, which should be controlled if possible. Finally, there is always the danger that implicit, unrecognized structure will be lost when converting to a new representation.

This paper reports our experiences in evaluating a potential hypertext representation of the *Oxford English Dictionary*. With the completion of the *Supplement* to the *OED*, the Oxford University Press has turned to consideration of an electronic form for the dictionary [Wein87]. The text of the *OED* has been keyboarded by an independent firm, and we are currently designing suitable computer-based tools for storage, editing, and searching [Berg 88]. Our work on the *OED* provides some insight into the problems of converting existing documents to hypertext.

THE OXFORD ENGLISH DICTIONARY

The *Oxford English Dictionary* [Murr28] is the largest and most scholarly dictionary of written English. Its production spanned the period from 1884 to 1928, with nearly thirty years of preliminary effort in planning and collecting material. In its standard form the *OED* consists of twelve books containing 41.81 million words in 252,259 entries, and 1.86 million quotations. An important subsequent work is the four-volume *Supplement* which was produced from 1958 to 1986. The *Supplement* contains 69,372 entries, 14.5 million words, and 560,000 quotations. 70% of the *Supplement* entries describe words which are not covered in the *OED*; the remaining 30% contain emendations and additions for existing *OED* entries.

The basic unit of the *OED* is the entry; each entry details the historical development of a given word. A typical *OED* entry is that for **abbreviate**, part of which is shown in Figure 1.

Figure 1. **abbreviate**.

An entry consists of a main form (the word being defined), its pronunciation, its part of speech, a list of variant forms, the etymology, and a list of senses. Each sense contains a definition and usually some illustrative quotations chosen from different sources and time periods. Senses often have nested subsenses (e.g., those labelled **b.** and **c.** in Figure 1) in the case of commonly used words, prefixes, or suffixes.

While some of the elements in an entry are indicated by special symbols (e.g., etymologies are surrounded by square brackets), most are indicated only implicitly by position and font. In order to preserve its structure, the text of the *OED* was manually keyboarded, with insertion of typographical tags to capture the most salient characteristics of the entries. These tags were designed for ease of data entry rather than completeness of structure representation, and so extensive work was undertaken to build a parser that could complete the tagging and verify the data [Kazm86]. The tagged and parsed text for **abbreviate**, up to and including the first two quotations, is shown in Figure 2. These tags are currently being exploited in the development of a database index and are crucial in identifying potential components of a hypertext.

```
<entry> <hwgp> <hwlem>abbreviate </hwlem> <pron id=0000041884>a&breve.br
<i>i&mac.</i> &sd.vi&sylab.<i>e</i> <su>i</su> t </pron>, <pos>v.</pos> </hwgp>
<vfl> Also <vd>5&en.7</vd> <vf>abreviate</vf>. </vfl> <etym> f.<xra
id=0000041880><xlem>abbreviate</xlem> <pos>ppl. a.</pos> </xra>; or on the
analogy of vbs. so formed; see <xra id=0000041881> <xlem>-ate</xlem> </xra>.
&es.A direct representative of L. <cf>abbrevia&mac.re</cf>; as <xra
id=0000041882><xlem>abridge</xlem> </xra>, and the obs. <xra
id=0000041883><xlem>abrevy</xlem> </xra>, represent it indirectly, through OFr.
<cf>abregier</cf> and mid.Fr. <cf>abre&acu.vier</cf>. &es.Like the latter,
<cf>abbreviate </cf>, was often spelt <cf>a-breviate</cf> in 5&en.7. </etym>
<sen4> <sen6> To make shorter, shorten, cut short in any way. <qpara> <quot>
<qdat>1530</qdat> <auth>Palsgr.</auth>, <qtxt>I abrevyate: I make a thynge
shorte, <i>Je abrege</i>. </qtxt> </quot> <quot> <qdat>1625</qdat>
<auth>Bacon</auth> <wk>Essays</wk> xxiv. 99 (1862) <qtxt>But it is one Thing to
Abbreviate by Contracting, Another by Cutting off.</qtxt> </quot> </qpara>
</sen6> </sen4>
```

Figure 2. Tagged data for **abbreviate.**

WHY HYPERTEXT FOR THE OED?

The main reason for considering a hypertext representation of the *OED* is to support browsing. The *OED* can be treated as a text database to which formal queries are posed, e.g. *What* *interjections were in common use in the period 1670-1720?* [Benb88][Gonn87a][Raym78]. However, experience with an extremely rapid searching program called PAT [Gonn87b] has shown that browsing the dictionary by means of fixed string searches is an invaluable adjunct to formal querying, and is often more fruitful, serendipitous, and enjoyable. Browsing is a two-stage process: users specify a pattern to be located and then navigate in the vicinity of the pattern. Our users have typically employed PAT to search for interesting phrases or words, and then consulted the relevant entries in the original paper dictionary. This switch of medium is necessary because PAT is only somewhat knowledgeable about the structure of the dictionary, and so does not support a sufficient navigational browsing capability. An obvious extension is a hypertext-like browsing facility as a front end to PAT.

A second reason for hypertext representation is to explore alternative methods for entry display. The *OED* was originally contracted for a specific number of pages, which the editors soon learned would be insufficient. As a result, they employed every means at their disposal to control the size of the *OED*, short of reducing its quality or comprehensiveness [Murr79]. The need to conserve page ‘‘real estate’’ led to very dense typesetting and the extensive use of abbreviations and symbols. Spatial techniques were restricted to paragraphs and different sizes of typeface. While the result is admirable, we need not observe this constraint any longer. A hypertext representation should provide better visual salience and more rapid navigation around large or alphabetically-distant entries. A key mechanism to be employed is dynamic reformatting of entries according to user specification.

A third reason for hypertext representation is to integrate the *OED* with the users’ tasks. Our users typically query and browse the *OED* as part of more extended tasks. At a minimum, users want to save their results and queries for use in future sessions, but they also expect access to annotation facilities, the ability to cut and paste fragments of *OED* text into other documents, routines for sorting and filtering extracted quotations, and tools for statistical analysis of selected variables. Hypertext can facilitate a consistent and simple interface to this wide range of tools.

CONVERTING TEXT TO HYPERTEXT

Hypertext’s main characteristic is commonly held to be its nonlinear network of interconnections. Contemplating the conversion of the *OED* to hypertext has led us to address the problem of defining nodes and links. A document can be broken into many arbitrary networks of nodes, but few of these are suitable representations of the original text. The selection of nodes and links has a significant impact on the usability of the hypertext and its validity as a representation of the original document.

From the point of view of document conversion, hypertext’s main characteristic is *fragmentation*. Fragments are pieces of content or structure which are both discrete and independent. By discrete we mean that the distinction between components is explicit and well-defined; by independent we mean that components are capable of standing in multiple relations to one another. Ideally, a hypertext node is a text fragment which contains a single, independent concept; common examples are footnotes, references, and annotations. Similarly, a hypertext link is a relationship with a specific source and destination node, often made concrete with an arrow and a label. Usually a link must be explicitly selected by the user to be traversed.

Texts often contain ideas which are not confined to discrete fragments. For example, Hamlet’s indecisiveness is a theme based a set of actions, some of which are explicit, some only alluded to, and some conspicuous by their absence. Making these actions explicit is a highly subjective process; moreover, it involves a radical departure from the nature of the text. Hypertext proponents sometimes insist that most texts would be better off as hypertext, and that linear structure is primarily a limitation of the media. We suggest that linear structure is a virtue achieved at some cost by the document’s creator. Furthermore, a subtle treatment of a theme may communicate an idea more artistically (and hence more successfully) than any explicit statement.

Thus for some purposes and documents, hypertext representations are inadequate. Fragmentation is intended to make an implicit structure explicit, so the key question in conversion must be: will an explicit structure be as expressive as the implicit structure? When the answer is yes, the document will gain from conversion; otherwise, conversion will degrade the representation of the document.

CHARACTERISTICS OF THE *OED*

The bulk of the *OED* is quotations which have been extracted from other texts. These quotations exhibit sufficient independence to be detached from their source, yet are still illustrative of a particular sense of a word. Hence, fragmentation is not a problem if only a quotation database is considered. Complete entries are also independent, since they are generally intended to be read as individual entities. However, searches are not often satisfied by exactly one entry. Some entries are extremely large, many have important cross-references, and often homonyms must be examined for related evidence.

The subcomponents of the entries (e.g., senses, etymologies) adhere to a formal structure, since we can describe them with a context-free grammar. However, the Oxford editors have taught us that an entry is not merely a collection of independent subcomponents [Raym86]. Entries are stylistic, creative wholes in which the relative sizes and arrangement of the parts often conveys important (although implicit) information about the development of a word. Hence, fragmentation of an entry may camouflage or destroy this implicit structure.

Attempting to represent each entry as a discrete fragment is complicated by the variance in entry size and structure. Entries in the *OED* range from "**Gig**: see JIG" to the entry for **set** *v.*, which requires almost half a megabyte of storage. Some entries have hundreds of quotations, but many have none. An entry may contain a well-balanced arrangement of senses and quotations, or it may consist largely of possible compounds or formations, e.g., the entry for **un-**. Some entries can be read quickly, while viewing even the structural skeleton of long entries is time-consuming. Clearly, attempting to find a representation for such a wide range will not be a simple task.

The range is wide, but what is the distribution? A histogram of the frequency of entry size is shown in Figure 3. This figure displays the frequency of entries of size less than 5000 bytes (entry size has been rounded to the nearest ten characters).

Figure 3. Distribution of entry size.

The most striking feature of the distribution is that approximately 50,000 entries (one-fifth of the total) contain fewer than 50 characters. The majority of these are entries for little-used variants of words and cross-references to other entries. 95% of all entries are smaller than 4000 characters, with 20% of all entries larger than the mean of 1060 characters. It should be noted that some entries will become larger when the material in the *Supplement* is integrated with that in the *OED*.

It would appear that large entries are not extremely common, and hence the problem of representation may be less severe than it first appeared. However, the probability of access is not equal for every entry. The size of an entry is largely dependent on the number of suitable quotations submitted; thus it should not be surprising that many of the smaller entries are for obsolete or infrequently used words, while commonly used words have larger and more complicated entries. If the commonly used words are also more likely to be browsed, the need for a browseable representation increases. In fact, an argument could be made that the larger entries are more likely to be browsed whether they are common or not, since their structure and content are more difficult to perceive and remember.

The probability of accessing large entries is also increased by the nature of our searching tools. For many search patterns there are too many matches to be returned to the user; in these cases PAT can be instructed to give a random sample of the results. The probability of selecting a large entry in a random probe is proportional to the fraction of the *OED* covered by large entries. Only 5% of entries are larger than 4000 characters, but they account for 47% of the bulk of the *OED*. The probability of matching to a large entry in a random sample of ten matches is $1-(1-0.47)^{10}$, a virtual certainty. Empirical observations confirm this analysis; samples of matches nearly always contain one or more large entries.

The display of larger entries might be facilitated by structural views or abbreviations such as those described by Furnas [Furn86]. Structural information can be extracted from the tags and employed in the construction of a structural view. Figure 4 shows the relationship between the number of tags and the size of entries. The graph shows generally that the larger the entry, the more tags it contains.

Figure 4. Distribution of tags versus size of entry.

However, many tags enclose short pieces of text (e.g., a single part of speech, a variant form, or a quotation date). Such tags designate flat structure which does not facilitate abbreviated display. Because the sense tags constitute the bulk of the nested structural information, restricting our attention to them gives a better appreciation of the substructure. This data is shown in Figure 5.

Clearly the number of sense tags is not well correlated with entry size. Some entries have a large number of quotations or long etymological notes but a very shallow sense structure; a sense-oriented view of such an entry is not indicative of its size or comprehensiveness. On the other hand, some entries have a deeper hierarchy with relatively little leaf content; in this case a sense-oriented skeletal view overemphasizes the content of the entry.

Figure 5. Distribution of sense tags versus size of entry.

From the previous analysis it seems that static fragmentation will not be a satisfactory solution, both because of the possibility of losing implicit information, and because no static fragmentation seems appropriate for all entries. As a result, we have chosen to support dynamic fragmentation, controlled by the user. By permitting various types of elision and formatting in a rapid response display tool, users can employ multiple fragmentations while viewing an entry. The results of a prototype formatter are shown in Figure 6. Here **abbreviate** is displayed as *OED*-style proof, a sense structure skeleton, the sense content without quotations, and quotations only.

Figure 6. Dynamic fragmentation of **abbreviate**.

Turning to the consideration of links, first we must ascertain the characteristics of the links that are already present in the data and tags. The most explicit links are the cross-references, which are pointers to other entries. These can appear in etymological notes or sense text, and are visually designated by printing the main form of the cross-referenced entry in small capitals. In Figure 1 there are cross-references to -ATE, ABRIDGE, and ABREVY. The *OED* contains 391,000 cross-references for an average of 1.55 per entry, making these a substantial source of potential hypertext links.

It is interesting to investigate the distribution of cross-reference destinations. Figure 7 contains four graphs showing the distribution of cross-references for each letter of the alphabet, normalized according to the number of entries per letter. Figure 7a shows cross-references to previous letters, Figure 7b shows cross-references to the same letter, Figure 7c shows cross-references to later letters, and Figure 7d shows cross-references to suffixes. Most cross-references are to words beginning with the same letter of the alphabet, with the only major exception being cross-references to suffixes. As might be expected, the number of cross-references to previous letters increases through the alphabet, while the number of cross-references to later letters decreases. However, note that there are several rather odd exceptions to the general trends. These exceptions may be due to the influence of the editor, the lexicographic position of the letter, or etymological characteristics common to words beginning with that letter.

Figure 7. Distribution of cross-reference destinations.

The general trend of the distribution of cross-references is a result of several factors. First, cross-references in the *OED* usually point to words with similar spelling. Many cross-references in definitional text are prefaced with *erroneous spelling of*, *variant of*, *obs. form of*, *verbal form of* and other such phrases which typically indicate a word with very similar spelling. The major exception to this rule is suffixes, which are referenced more evenly from the whole alphabetic range. Cross-references in etymologies are often of the form *root* + *suffix*, for example **Musal**...[f. MUSE *sb.*^{1} + -AL.]. A second factor is the history of the development of the dictionary. Volumes were generally compiled in alphabetical order, and so the compilers of the later volumes naturally had more information on which to base cross-references, and would be more likely to cross-reference existing entries than as-yet uncompiled ones. A third factor is psychological; since an editor would be most aware of the section of the *OED* currently in progress, he or she would be more likely to insert cross-references to entries in that section.

A more general type of link is known as a *lexical* link. These links result from the simple observation that since the *OED* is a book that defines most English words, every word in the *OED* can be seen as the source of a possible link to its definition. The *OED* and other comprehensive dictionaries are thus unique in containing to some extent their own reference material. Lexical links need not be confined to the *OED*; it would be highly desirable to link words in other texts to their *OED* entry. However, automatically determining the source and destination of lexical links is not simple. For example, consider the definition in sense **1.** of **Fossic** *v.*:

> "To search for gold by digging out crevices with knife and pick, or by working in washing-places and abandoned workings in the hope of finding particles or small nuggets overlooked by others. Also, *to fossick about*."

The first problem is to determine which lexical items should be linked to their definitions. Should we link to *digging*, *out*, the phrase *digging out*, or perhaps all three? To provide for the third alternative is equivalent to isolating so-called "open compounds", in general a difficult problem [Amsl87][Chou88]. Having determined the source, it may need to be reduced in tense or number to its morphological root. For example, *crevices* must be altered to *crevice* and *others* to *other*. On the other hand, because *digging* is explicitly defined in the *OED*, it should probably not be reduced to *dig*. Unlike most other dictionaries, the *OED* contains distinct entries for all derivative forms.

The third step is to determine the target, which could be a complete entry, a subentry for a derivative form, or perhaps a particular sense within an entry. In the above example, **washing-place** is listed as a derivative combination within the entry **washing** *vbl. sb.*, but only its subsense ''(*b*) a place where gold is washed out from sand or earth'' is directly applicable. Choosing the correct target involves determining the part of speech and etymological root to identify an entry, and applying sense disambiguation to identify a particular meaning [Lesk86]. For example, to identify the target of *pick*, one has to determine that it is used as a noun rather than as a verb, next to determine that it is the first of four entries for **Pick** *sb.* that applies, and finally to identify which of the eight senses of this entry is the correct one.

The existing cross-reference links in the *OED* are important but limited in their scope. Furthermore, they can be treated as a special case of lexical links, which present several general difficulties. Because of the evolutionary development of the *OED*, not every cross-reference is specified precisely, and hence resolving cross-references to words appearing in later volumes often involves the same target disambiguation as is required to resolve lexical links in general. Supporting explicit intra-*OED* lexical links would therefore require sophisticated access software.

As in the case of text fragments, the tangle of problems involved in defining explicit links raises a question of their utility. In our experience, when rapid searching and dynamic fragmentation of the resulting text is available, the user can usually determine the appropriate fragment by trying out a few alternatives. The cost of maintaining, updating, and storing explicit links would be considerable, and would not be offset by faster access than is provided by our present tools.

RELATED PROBLEMS

In addition to evaluating hypertext for the *OED* itself, we have observed two related areas in which hypertext work could be done.

The first area is the use of the *OED* as a generator of hypertext links for other documents. The quotations in the *OED* can be interpreted as links from the dictionary to other texts. Links with a common source (i.e., those that start at the same sense of an entry) induce relationships among the referenced texts. Co-citation links could be derived for all texts which have some word sense in common; that is, texts are related if they each supply a quotation for a given sense. Useful links within and between other texts can also be identified from word collocations in the *OED*; for example, a large overlap among words in definitions can serve to identify similarity of topic [Lesk86]. Extracting meaningful pairings remains an interesting problem in computational linguistics.

A second area is the issue of developing and editing entries for the *OED*. Computerized lexicography has a need for powerful hypertext-like structure editing, especially for the creation and maintenance of sense structure [Raym86]. The source material for entries is a large set of quotations obtained over the years from volunteer and directed readers, each handwritten on a 6'' by 4'' slip of paper. In a method still practised today, the editors distill the senses of the word from the set of quotations by arranging and re-arranging the slips into spatial categories on a desktop, looking for the pattern of historical development [Murr79]. This approach seems highly susceptible to solution by a system that employs a desktop metaphor, as does NoteCards [Hala87]. However, existing systems seem designed for tens of slips at a time, whereas *OED* editors can be faced with organizing thousands of slips. A more fundamental problem has been identified in experiments we have conducted on the organization of proverbs [Raym88], which indicate that there is a quantifiable decrease in the quality of semantic categorization when a categorization task is performed in an electronic environment employing a spatial metaphor. It appears that the ability to create temporary, unnamed categories is a key factor in the development of good semantic structures, and that current systems and metaphors interfere significantly with this ability.

CONCLUSIONS

The investigation of cross-references in the *OED* showed significant local interconnectivity, but relatively few links between sections of the database that had been compiled at different times. We suspect that this localization of links to be a general tendency in large documents and possibly even in ''hyper-libraries'', simply because of the cost and difficulty of continually integrating old material with new.

Static fragmentation is common in hypertext systems, especially with data that has been created expressly for hypertext. Our experience with the *OED* indicates that static fragmentation is inappropriate for some converted texts. If there is too much implicit structure in the text or if the variance in structure is too great, then a better approach is to provide the user with tools for quickly developing dynamic fragmentations. Similarly, explicit storage and display of lexical links adds little to the user's ability to navigate the data, yet requires complex semantic analysis for proper resolution.

ACKNOWLEDGEMENTS

We gratefully acknowledge the financial assistance received from the Natural Sciences and Engineering Research Council of Canada through the University-Industry Program under grant 0039063.

REFERENCES

[Amsl87] Amsler, R., ''Research Toward the Development of a Lexical Knowledge Base For Natural Language Processing'', Bellcore Technical Memo TM-ARH-010356, October 1987.

[Benb87] Benbow, T., Carrington, P., Johannesen, G., Tompa, F.W., Weiner, E., ''Report on the *New Oxford English Dictionary* User Survey'', OED-87-05, UW Centre for the New Oxford English Dictionary, Waterloo, Ontario, November 1987.

[Berg88] Berg, D.L., Gonnet, G.H., Tompa, F.W., ''The New Oxford English Dictionary Project at the University of Waterloo'', in *Studies in Honour of Bernard Quemada*, A. Zampolli (ed.), Giardini Editioni, Pisa, 1988.

[Chou88] Choueka, Y., ''Looking for Needles in a Haystack, or Locating Interesting Collocational Expressions in Larga Textual Databases'', *Proceedings of the RIAO '88* (Recherche d'informations Assistée par Ordinateur), March 21-24, 1988, pp. 609-623.

[Conk87] Conklin, J., ''Hypertext: An Introduction and Survey'', *IEEE Computer* (September 1987), pp. 17-41

[Furn86] Furnas, G.W., ''Generalized Fisheye Views'', *Proceedings of the CHI '86 Conference on Human Factors in Computing Systems*, April 13-17, 1986, pp. 16-23

[Gonn87a] Gonnet, G.H., Tompa, F.W., ''Mind Your Grammar: A New Approach to Modelling Text'', *VLDB* '87, Brighton, England, September 1987, pp. 339-346

[Gonn87b] Gonnet, G.H., ''Examples of PAT'', OED-87-02, UW Centre for the New Oxford English Dictionary, Waterloo, Ontario, August 1987

[Hala87] Halasz, F.G., Moran, T.P., Trigg, R.H., ''NoteCards in a Nutshell'', *Proceedings of the CHI + GI Conference on Human Factors in Computer Systems and Graphics Interface* , April 5-9, 1987, pp. 45-52

[Jone87] Jones, W.P., ''How Do We Distinguish the Hyper From the Hype in Non-Linear Text?'', *INTERACT '87*, Stuttgart, September 1-4, 1987, pp. 1107-1113

[Kazm86] Kazman, R., ''Structuring the Text of the Oxford English Dictionary Through Finite State Transduction'', CS-86-20, Department of Computer Science, University of Waterloo, Waterloo, Ontario, June 1986

[Lesk86] Lesk, M.E., ‘‘Automatic Sense Disambiguation Using Machine Readable Dictionaries: How to Tell a Pine Cone From an Ice Cream Cone’’, *Proceedings of the ACM SIGDOC Conference*, Toronto, Ontario, June 8-11, 1986.

[Murr28] Murray, J.A.H, Bradley, H., Craigie, W.A., Onions, C.T., *The Oxford English Dictionary*, Oxford at the Clarendon Press, Oxford, England, 1928

[Murr79] Murray, K.M.E., *Caught in the Web of Words: James A.H. Murray and the Oxford English Dictionary*, Oxford University Press, 1979.

[Raym86] Raymond, D.R., Warburton Y., ‘‘Computerization of Lexicographical Activity on the New Oxford English Dictionary’’, OED-87-03, UW Centre for the New Oxford English Dictionary, Waterloo, Ontario, December 10, 1986

[Raym87] Raymond, D.R., Blake, E.G., ‘‘Solving Queries in a Grammar-Defined *OED*’’, unpublished technical report, UW Centre for the New Oxford English Dictionary, Waterloo, Ontario, February 1987

[Raym88] Raymond, D.R., Cañas, A.J., Tompa, F.W., Safayeni, F.R., ‘‘Measuring the Effectiveness of Personal Database Structures’’, *International Journal of Man-Machine Studies* (in press)

[Wein87] Weiner, E., ‘‘The Electronic English Dictionary’’, *Oxford Magazine*, February 1987, pp. 6-9