

Handmade, Computer-assisted, and Electronic Concordances of Chaucer

Willard McCarty

The wise Plato seith, as ye may rede,
The word moot nede accorde with the dede.
If men shal telle proprely a thyng,
The word moot cosyn be to the werkynge.
(*Canterbury Tales*, MancT 207-10)

The concordance is “a general purpose working tool for the study of literature” whose basic function is “to bring together (‘concord’) passages of a text that illustrate the uses of a word” (Howard-Hill 1979: 4). Since the invention of the concordance in the early 13th century as an exegetical tool to assist preaching and teaching of Scripture, the genre has developed many forms. From a technological point of view, these can be designated handmade, computer-assisted, and electronic. The last may be too recent for us to know its impact on scholarship, but its promise is enough to justify a comparative study.

Since we lack a comprehensive history of concordancing, we do not have all the facts at hand, but it is clear that concordancing has from its origins been intimately bound up with how we think about texts. By disassembling a text, then reassembling it as a series of verbal concords, we get a new perspective on it, perhaps even a new text. The reconstruction might be thought of as a rearrangement or transformation according to a non-linear, discontinuous principle of organization, whether topical, alphabetic, or otherwise. A concordance thus implies a synchronic or “collocative” way of reading, in which separated elements may easily be put together. It provides a view of the text as a rearrangeable structure of simultaneously present elements. Concordancing is, then, a technology for modelling the non-linear, associative way people think about texts they have read, i.e. what they do once they have read a text through — once, that is, all the text is present to the understanding. Such a perspective accords closely with how the inventors of the concordance most likely viewed the Bible, for which it was originally devised.¹

¹I am referring to the traditional Christian idea of typology, or organization of the biblical text by discontinuous correspondences of anticipation and fulfilment connecting the Old to the New Testament (McCarty, forthcoming).

I will here consider two major printed concordances and one example of interactive concordancing software for the works of Chaucer. These illustrate the three technologies of my title: *A Concordance to the Works of Geoffrey Chaucer*, by John S. P. Tatlock and Arthur G. Kennedy (1927), was made by hand; *A Complete Concordance to the Works of Geoffrey Chaucer*, by Akio Oizumi (1991), by computer assistance, though conventionally printed; and the displays of which the program *TACT* is capable arise from interaction between compiler-user and text. By comparing the three examples, I will illustrate what has been gained, and what perhaps endangered, by the movement to electronic concordances.

1. *A taxonomy of concordancing*

The first systematic attempt to describe concordancing and define how it should be done is T. H. Howard-Hill's handbook, *Literary Concordances* (1979), which grew from an earlier essay based on a late 19th-century argument for the interpretative neutrality of concordances (1976: 215-20). In this handbook, he uses the term "computer concordance" to mean something produced by machine but designed to be printed. His argument is important for Chaucerians because Oizumi has followed its recommendations closely.

The welter of historical concordances can be roughly reduced to four basic kinds (cf. Howard-Hill 1979: 3).

1. *Verbal*: an alphabetic list of word-forms (often lemmatized for inflected languages), giving citations to the original text but no context; usually known as an *index verborum*;
2. *Contextual*: an alphabetic list of word-forms with context and citations; Howard-Hill's "concordance" proper, the literary tool with which we are most familiar;
3. *Glossarial*: a list of word-forms, with context, organized by grammatical or orthographical lemmas, thus tending to the lexicon or dictionary;
4. *Conceptual*: a list of word-forms, with context, organized by idea or sense, thus tending to the thesaurus or "concept dictionary".

As helpful as this division may seem, some concordances do not fit. Among these are the 'special' forms Howard-Hill describes (1979: 66-73), including the bilingual concordance. One form he does not mention is Parunak's "pic-

torial" concordance (1981), which succinctly condenses frequency information from the concordance into a histogram or graph.

2. *A brief history of concordancing*

The prehistory of the concordance raises broad matters such as alphabetization, practices of reading, biblical exegesis, and the referential habit of mind in Late Antiquity and the early Middle Ages (Daly 1967, Stock 1983, McArthur 1986). Conceived as a solution to the needs of theologians in the late 12th and early 13th centuries for technical aids to teaching and preaching, the first concordances were the subject of prolonged experimentation. Their immediate predecessor was probably collections of biblical *distinctiones*, that is, groups of distinct senses subsumed under the theologically significant words they explained (Rouse 1974b). *Distinctiones* differ from the concordance by their reference to a conceptual scheme rather than to the lines of a text. The medieval Latin term *concordantia* was similarly first used in the plural to denote groups of passages, or *concordantiae*, assembled to show the harmony or union (*concordia*) of their senses (*OED*); see citations in Du Cange for 1244 and in the *Dictionary of Medieval Latin from British Sources* for 1362. Note the musical metaphor implicit in "concordance": it points to harmonious relationships of words, even harmonious discords.

The first concordance, to the Vulgate, was apparently produced by 1247 at the Dominican house of St. Jacques, Paris, perhaps under the direction of Hugh of St. Cher; some 22 manuscripts survive. The earliest reference to it is that by Tholomaeus of Lucca, *ca* 1315, who says that its putative maker "primas concordantias super biblam cum suis fratribus adinvenit" (Rouse 1974b: 7). Words are listed, without context, which "limited the usefulness and thus the circulation of this concordance" (Rouse and Rouse 1974b: 13). The second, so-called "English Concordance," was produced later, probably also at St. Jacques, by English Dominicans, but there are no known copies. Various abbreviations or abridgements of this "ambitious failure," as Rouse and Rouse note (1974b: 16f), suggest that it had a complex coding scheme for grammatical information and "virtually a full sentence of context for each reference, with the result that [it] ...was not just unnecessary but impossibly long." The third and final concordance of the formative period, attributed to Conrad of Halberstadt, was in existence by 1286 and survives in at least 80 manuscripts. It balanced its two predecessors in amount of context and provided the model for "the verbal concordance to the Bible as it was widely known in the Middle Ages, from the late thirteenth century onward" (Rouse and Rouse 1974b: 17). Thus, if we allow 20 years for the production of a concordance, we may estimate a period of about 60 years for the major fea-

tures of the verbal concordance to be developed.

The history of the concordance during the 650 years from the end of the 13th century until the computer has not been systematically studied. Definitions of the term, for example by John Capgrave (1460), John Marbeck (1550), and Alexander Cruden (1737), refer to a table or index to the Bible, created to locate words (*OED* s.v. 'table' 10b). Cruden's preface to his *Complete Concordance of the Old and New Testament* notes that "the various places where [the words] occur are referred to, to assist us in finding out passages, and comparing the several significations of the same word." Remarkably, these representative citations do not mention context and do not distinguish index, concordance, dictionary, and table. Other applications of the term, such as the "concordantial margin" still found in modern Bibles, indicate topical organization. There one may find, between the double columns of each page, numerous citations of parallel passages, and commentary on the Hebrew or Greek underlying the translation. The parallels are typological rather than just verbal; they interpret, even though the format makes the text primary, and the glossing secondary. In its first definition of "concordance", the *OED* stresses this interpretative principle, contrary to Howard-Hill's practice ("concordance," 5b) but true of some modern literary examples (see Whitman 1918, Patterson and Fogle 1940).

Computer-assisted concordances¹ have won the day since then. As far as we know, the last entirely handmade concordance was to Byron by Ione Dodson Young (1965), who expresses wistful regret for the loss of much pleasure "on the unfeeling machine" after her 25 years' work (ix). The first computer-assisted concordance, the *Index Thomisticus*, was started seven years after Young began, in 1947, by the Jesuit Roberto Busa (the first volumes were issued in 1976, and a CD-ROM in 1992), following four years' study of the preposition *in* as a key to the theme of "inwardness" in St. Thomas (Busa 1980). At the outset, Busa realized that every word must be concorded, no matter how insignificant it might appear to the compiler, and that computerization should do more than imitate manual methods. Early practitioners of computerized concordancing were necessarily preoccupied with methodology, as they had to discover how the most basic operations were done. Thus their published concordances tend to discuss the logical structure of the software; some even give flowcharts. The programmer's name often appears, as with Oizumi's programming, attributed to Kunihiro Miki. Furthermore, computer-assisted concordancing required an organized, teamwork

approach, characterized by substantial budget and a university computing centre. With these new requirements, we can understand the need for standards described in Howard-Hill's guidebook, where he spends much space discussing the base text, copyright, pre-editing, statistical needs, the special treatment of Roman numerals, ampersands, contractions, abbreviations, hyphenated compounds, *i / j* and *u / v* distinctions, and elisions. Consistency and rigour are essential because he has a permanent achievement in mind, a bulwark against time that "... will not be superseded by computer networks, remote access terminals or data storage, and display devices conceivably available to the public in the next millenium" (9).

Recently, interactive concordances — which enable the maker to adjust both text and concordancing format — relax the need for editorial decisions regarding matters such as orthographic variation and lemmatization of headwords; treatment of homographs; ordering of citations by sigla, the designation of sigla, and other labelling matters; the amount, style, and kind of context (unless by syntactic or metrical unit); graphical distinction of the concorded words; the manner in which listings are sorted (e.g., for KWIC, alphabetically by following word or by sequence in the text); indeed, the distinction between the concordance and the edition on which it is based. As well, the new technology has permitted the old exegetical, interpretative purpose to resurface in concordance making. Howard-Hill argued that interpretive apparatus in a concordance could only serve those who thought like the compiler, but electronic concordancing allows us to hide that apparatus unless we need it, to modify it, or to add one of our own. Topicality (see Schmidt 1978, Laffal 1969-70) no longer affects the product's usefulness to a wide audience.

3. Printed concordances to Chaucer

Tatlock and Kennedy base their concordance on the Globe edition of 1898 by Pollard, Heath, Liddell, and McCormick. Although "emphatically a concordance to the Globe edition and not a new study of the text" (iii), the editors altered their base text by silently correcting minor errors (iii), by including "certain probably genuine passages" not in the Globe (iv), and by adopting variant readings (iv-v). In their introduction, Tatlock and Kennedy tell the history of the project, initiated by Frederick James Fumivall in 1871, then handed over to three editors before work began in 1915 (see also Tatlock 1923, Flügel 1911 and 1913, Kurath 1954: ix), and outline their plan under headings such as the base text, works included, and additional passages; variant readings, their sources and manner of presentation; extra-textual features; emendations; arrangement and headwords; homonyms; obsolete, com-

¹See Burton 1981, Parrish 1962 and 1970, Ingram 1974, Oakman 1980, Lancashire 1983 and 1986.

pound, and hyphenated words; extended quotations; cross-references; and specimens for the commonest words. Understanding the constraints imposed on them helps us understand what they could do. Thus the state of their base text required close editorial attention to corrections, emendations, and additions. The lack of a *Middle English Dictionary* or other lexical authority forced the issue of headword spelling; and the manual method of compilation made specimens unavoidable. Given these methods, as well, choosing a phrasal or metrical unit of context was reasonable.

Contemporary reviews (see Anonymous 1927, Brandl 1927, Fischer 1927, Menner 1928, Northrup 1928, Royster 1928, Hoops 1928-29) are listed by Griffith (1955: 3-4). Reviewers attest to the accuracy of Tatlock and Kennedy but criticize their use of modern-English headwords, the failure to separate homonyms (i.e., to lemmatize the text, as Larry Benson has recently done), the inconsistent placement of compounds and hyphenated words, and the limited coverage of common words (such as "a", "all", "be", "in"), a fact that, Royster says, restricts linguistic usage. Yet Northrup speaks for all in saying that numerous problems "have been carefully thought out and a sensible plan has been adopted in every case; this the critic must concede even when he himself might have decided differently" (240).

In his introduction, Akio Oizumi claims that his concordance, based exactly on the Riverside edition, "will supercede" Tatlock and Kennedy (1991: v). This is not entirely correct despite the care Oizumi and Miki have taken, the advantages of modern computing technology, and a new reliable, standard edition. There is unambiguous improvement (e.g., access to every instance of all words in Chaucer's works, with full context), but other changes may benefit some users more than others (e.g., unlemmatized grouping of words). The chief differences between the two concordances are summarized in Table 1.

The editorial choices and decisions of the base texts are not within the scope of this essay, but I should note that Tatlock and Kennedy include words from nine lines of text not found in the Riverside edition, hence not in Oizumi.

Oizumi's concordance is both complete and exhaustive; every instance of all words is listed with a context, slightly more than twice as much as in Tatlock and Kennedy. The work occupies ten large volumes, whose subdivision reflects the separate concording of each work — Chaucer's works are not concorded together. Volumes are allotted to separate works: four for *The Canterbury Tales* (one entirely devoted to word indexes); one for *The Book of the Duchess*, *The House of Fame*, *Anelida and Arcite*, and *The Parliament of the Fowles*; one for *Boece*; one for *Troilus and Criseyde*; one for *The Legend of Good Women*, the Short Poems, "Poems Not Ascribed to Chaucer in the Manuscripts," and *A Treatise on the Astrolabe*; one for *The Romaunt of*

the Rose; and a final volume giving "An Integrated Word List to the Works of Geoffrey Chaucer". Each concordance is supplemented with listings easily generated by the computer from the same data: word indexes (i.e., *indices verborum*), descending word-frequency lists, reverse-alphabetical word-lists, and lists of hyphenated compounds, alphabetically sorted by first and by second elements. The result is an order of magnitude larger.

FEATURE	TATLOCK & KENNEDY	OIZUMI
<i>Base text</i>	Globe	Riverside
<i>Additional text</i>	TC 2.1750a-g; NP 4060a-b	none
<i>Volumes</i>	1 (8" by 10")	10 (~8.5" by 11")
<i>Subdivisions</i>	none	by work; within each by type of listing
<i>Pages</i>	1,110	11,608 (9,040 KWIC; 2,266 word-index; 146 frequency; 146 reverse; 10 hyphenated)
<i>Concordance lines</i>	~217,000	~582,540
<i>Concordance total words</i>	~1,736,000	~9,903,180
<i>Concordance head-words</i>	~14,430	~63,280
<i>Concordance type</i>	KWOC or metrical/phrasal	KWIC
<i>Headword type</i>	lemmatized to standard spelling, ignores y-prefix of past participles, &c.	unlemmatized
<i>Headword form</i>	bold, left	bold, left, with freq. in parentheses
<i>Headword collation sequence</i>	mainly alphabetical	alphabetical, sensitive to case of proper nouns, diacesis
<i>Keyword</i>	not distinguished	centered, spaced
<i>Context</i>	metrical/phrasal; ~8 words per entry	by available space; ~16 words per entry
<i>Line-endings</i>	implicit	marked by virgule

FEATURE	TATLOCK & KENNEDY	OIZUMI
<i>Sigla</i>	Globe, Skeat refs.; A-I CT grouping	Riverside references; A-I grouping for CT
<i>Order of entries</i>	text order	alphabetical by word following keyword
<i>Editorial changes</i>	variants marked; emendations marked or silent	Riverside; bracketed phrases as individually bracketed words
<i>Incipits, explicits</i>	included, where Globe follows MS	excluded
<i>Entry collation sequence</i>	alphabetical, ignoring case	alphanumeric, including contiguous marks of punctuation
<i>Cross-references</i>	from unfamiliar to familiar	none
<i>Multiple keyword-occurrences</i>	single line of context serves for all	each occurrence listed separately
<i>Common words</i>	specimens	exhaustive
<i>Foreign words</i>	separated when possible	unseparated
<i>Hyphenated words</i>	case-by-case	as complete units, hyphen ignored for sorting; separate table provided
<i>Other compounds</i>	as printed, with exceptions	as printed
<i>Homonyms</i>	not distinguished, except when one given by specimens	not distinguished
<i>Obsolete letters</i>	Globe diacritics removed	follows Riverside

Oizumi's major departures from Tatlock and Kennedy in micro-structure are his avoidance of any lemmatization (word-forms are not placed under grammatical, syntactic, or semantic categories), his sorting practice, and the "KWIC" or "keyword-in-context" format, which centers the keyword on the line and gives as much context as will fit on either side. Oizumi's strict ordering of entries by their unlemmatized headwords scatters grammatically-related forms, e.g., past participles with y-prefix from other forms of the verb. He follows Tatlock and Kennedy in not separating homonyms but does so with an absolute consistency his predecessors did not adopt. Thus, in Oizumi, different words are occasionally found together, whereas different forms of the same word are not. They differ more radically, however, with

respect to foreign words, which Tatlock and Kennedy separate from English homonyms if possible (cf. vi) but which Oizumi does not.

Capitalization in Oizumi's concordance is preserved for proper nouns designated in the Riverside edition; as a result, such words have different entries from their minuscule forms (e.g., "Love" and "love"). (Nevertheless, one wonders if, for example, personification is such a secure matter as to have it canonized in a concordance.) Words with diaeresis, such as "degreës" (I.PS 4), and italicized words likewise form separate entries. Within entries, contexts are sorted alphabetically by words following the keyword rather than, as with Tatlock and Kennedy, by the order of occurrence of words in the text. This sorting makes phrasal repetitions quick to find, an advance for those tracing an idea through a work, although since Oizumi has concorded each work separately, this advantage is confined within the individual volume. (The word indexes provided for each work assist here.) Finally, except for quotation marks, punctuation is considered, for sorting purposes, as part of the word it immediately follows. The "alphabetical" order for punctuation is " , ; : ! ? ". The virgule, used to mark line endings, also affects sorting.

The KWIC or keyword-in-context format adopted by Oizumi, invented in August 1959 by H. P. Luhn (1966) for mechanical concording of technical documents, has since become standard for published literary concordances. Unlike a concordance with context based on a syntactic or metrical unit, KWIC concordances are easy to generate by machine; they do not need to be segmented for contextual units. Furthermore, the emphasis on the keyword makes for easier linguistic comparisons and helps one locate approximate repetitions (with, say, inflectional variants or synonyms). The eye is here centered on the target word and its immediate environs, but a concordance with a phrasal context makes the keyword harder to find. The amount of context Oizumi allows is generous — averaging one line of poetry on either side of the target word — and serves those interested in collocational patterns. Virgules marking line divisions in the poetry allow one to study the target word and make for more efficient use of space on the page. Finally, in Tatlock and Kennedy, when the target word occurs more than once in the same unit of context, the occurrences are concorded together; thus, for example, 'love' in *Troilus* 2.392 and 2.788. Oizumi, in contrast, gives them separate (and, because of the sorting order, often widely separated) lines.

In cross-referencing, concordance makers recognize that their adopted plan is not the only one possible, that there are connections it does not make or that it hides. Tatlock and Kennedy provide "upwards of two thousand" cross-references, as a rule going from the familiar to the unfamiliar. Oizumi does not give what might have been helpful links, for example, between scattered forms of the same word. (A kind of cross-referencing is provided, however, for hyphenated expressions, which for each work are given in a

separate table, by listing them as sorted both by the first and by the second word; non-hyphenated compounds are treated as single words.) Here, the basic problem is the inescapable fixity of print, which brings stability at the cost of flexibility.

Neither concordance prints letters obsolete in Modern English.

Oizumi gives listings additional to those in Tatlock and Kennedy because new technology renders such things easy to generate from the textual data. These include word indexes, ranked word-frequency lists, reverse word-lists, and hyphenated word-lists.

As we saw, a word index (or *index verborum*) is a listing of headwords with references to the text(s) from which they are derived, but without context. Oizumi's separate word indexes for every concorded work (and for every work in the *Canterbury Tales*) supply, in parentheses following the headword, the number of occurrences and the number of times the word occurs in rhyming position (references to words in rhyming position are as well marked with an asterisk). The tenth volume of *A Complete Concordance* is 'An Integrated Word Index' to Chaucer's works. Here references to specific lines are omitted when these exceed 100 for a single text; the user is then referred to the word index for that text instead. Headwords in the word index follow those in the KWIC listing exactly, so that substantive capitalization, accent marks, and italics are significant, and marks of punctuation are not attached to the headwords. Oizumi's word indexes compensate for limitations already noted. First, as words are listed in order of occurrence, each index displays the linear development of word-usage through a text. Second, the cumulative word indexes to *The Canterbury Tales* and to Chaucer's works as a whole provide a comparative, "comprehensive bird's-eye view of Chaucer's vocabulary" (1.vii) that one gets from Tatlock and Kennedy.

A 'ranking' frequency list gives the words of a text in descending order of frequency, from most to least. Such lists highlight dominant themes and stylistic features in a text, insofar as these are expressed in single words. In combination with a word index, a frequency list may be used to construct other rankings of word-groups. Oizumi gives each separately concorded work a frequency list (not, however, each of the tales in *The Canterbury Tales* individually nor Chaucer's works as a whole). The frequency lists, three columns per page, give the word, its absolute frequency, and the number of times that word occurs in rhyming position in poetic texts. As in the word index, listed words follow the headwords of the KWIC listing in matters of spelling, capitalization, accents, and italics.

In the reverse-alphabetical list, words appear by endings; thus "bad", "forbad", "had", "yhad", and so forth are listed together in that order. Frequency and the count of rhyming positions appear, as usual. This list overcomes some problems caused by the scattering of unlemmatized forms and

helps studies of rhyming and grammatical form.

Hyphenated compounds are listed for each work, although again not for the individual tales of *The Canterbury Tales* or for Chaucer's works as a whole: one sorted by the first element, one by the second, with frequency and rhyme-position counts specified, as usual.

To compare the two concordances, I chose all forms of the word 'love' in *Troilus and Criseyde* as these are subsumed under the headword "Love" in Tatlock and Kennedy (306 occurrences), namely, 'Love', 'love', and 'loven' in Oizumi, respectively, 41, 255, and 6 occurrences for a total of 302. They differ for the following reasons. Four marked variants are included in Tatlock and Kennedy but not in the Riverside edition, and so not in Oizumi; three double occurrences are given in single lines of Tatlock and Kennedy but in separate lines by Oizumi; four occurrences are in the Globe edition and in Tatlock and Kennedy but in neither the Riverside nor in Oizumi; and one occurrence in Riverside and Oizumi but not in Tatlock and Kennedy. (Hence, $306 - 4 + 3 - 4 + 1 = 302$.)

4. Interactive concording of Chaucer

Oizumi's work typifies printed concording by mainframe, "batch" computing. This contrasts with "interactive" concording, in which partial listings in dynamically adjustable format are generated on demand, often only on screen or as electronic data intended for post-processing by other software. Rather than a printed object, the "concordance," we have "concording," in which the investigator repeatedly turns to the computer. Rather than stable features, interactive concording offers a range of possible features from which the user may select. Interactive concording began with John B. Smith's seminal program *ARRAS*, written for time-sharing mainframe computers about 15 years ago (Lancashire 1986: 55, 58). Although the promised version of *ARRAS* for MS-DOS has to date not been generally released, the interactive concordance has continued through MS-DOS software such as *WordCruncher* and *TACT*.² I will focus on applying *TACT* to Chaucer's *Troilus*. See Bradley (1991) for a manual to *TACT* 1.2.

The concording programs in the *TACT* system work with a compiled form of the text, a textual database, rather than with the text itself. Normally

²Although the MS-DOS version of the sophisticated *Oxford Concordance Program* is not, strictly speaking, interactive, it can be used rapidly to generate selective listings for exploratory analysis and files for post-processing; note also the UNIX concording package *hum*. For Macintosh DOS, *Conc*, produced by the Summer Linguistics Institute, is available.

the text is first prepared for compilation by encoding it with tags signifying various editorial decisions (see Wooldridge 1991). These decisions may be rethought, the tags altered, and the database recompiled at any time. Thus the decisions that the compiler of a printed concordance must make, once and for all, may be made provisionally for its electronic counterpart. Hence the roles of compiler and user tend to merge. In addition, no word-forms need be eliminated or reduced to specimens for reasons of economy, so that one important objective of Oizumi's work, allowing for the complete study of Chaucer's language, may be achieved without the cost or bulk of physical production.

For *Troilus*, textual tags might, for example, include the name of the poem, its division into books, stanzas, lines, and speeches, the identity of each speaker, and an indication of the "text type" to distinguish the poem itself from the incipits and explicits. The user-compiler may then select from these tags, e.g., to produce a KWIC concordance of selected word-forms in which book, line, and speaker are identified for each instance. Interpretative markup that defines thematic units or allusions may be added, although such markup may not be as generally useful. If good editorial practice has been followed, markup of any one type can easily be transformed into another type.

The complete works of Chaucer may be compiled as a single database, or separate databases generated for each work or group of works. Assuming one large database in which separate works have been marked, interactive concordancing can proceed either in the manner of Tatlock and Kennedy, across all works simultaneously, or in the manner of Oizumi, separately for each work, by specifying in each case that only instances of the target word(s) from a given work be displayed. Oizumi's KWIC and supplementary listings may be generated by *TACT*, which has five displays, each of which may be modified dynamically: KWIC, Variable Context, Text, Distribution, and Collocates.³ See Karen Arthur's essay in this volume for examples are taken from an electronic text created and encoded by Ian Lancashire and her from Fisher's edition.

TACT queries and displays provide more than any printed concordance can. For example, the query statement (with which one asks the software to return a collection of words or positions in context) includes a pattern-description language. A query may also ask for collocations of two or more words, either contiguous or separated by a span that the user may adjust. The resulting displays are unlike anything possible in a printed vol-

³Here *TACT* terminology here conforms to version 2.1 and so differs from the names given them in previous versions.

ume. As well, *TACT* queries may be stored externally and used repeatedly, in various text databases, to obtain lemmatized listings of dictionary word-forms. The user may then request selective concordances in the styles both of Tatlock and Kennedy and of Oizumi. Furthermore, since lemmatization is completely under the user's control, any scheme may be adopted.

Headwords for *TACT* KWIC and Variable Context displays are usually unlemmatized, like Oizumi's entries, although with Ian Lancashire's *TACTDct* software, lemmas may be systematically substituted for words. *TACT*, unlike Oizumi, also sorts entries by headword in a case-insensitive way, although the case distinction can be made either by separating majuscules from minuscules in the *TACT* collation sequence, or by exporting a display as a file and sorting it. Distinguishing proper nouns from other capitalized words, e.g., at line-beginnings, would require explicit markup of each one, however. The collation sequence could also be used to determine whether, for example, a word with a vowel marked by a diaeresis constitutes a separate headword, how punctuation is to be treated for sorting purposes, or more generally, what forms are concorded together. The frequency information supplied by Oizumi for each headword is, as we have seen, also given by *TACT*, but not the count of occurrences in rhyming position.

With respect to context, *TACT* can produce the phrasal or metrical unit used by Tatlock and Kennedy, but only after explicit markup. Like Oizumi's software, *TACT* selects context by arithmetic units in the Variable Context display: numbers of words, lines, or whatever the user has defined in markup. Thus the KWIC display gives as much as will fit on the screen, an average of approximately +/- 6 words at maximum. In the KWIC display, context may also be eliminated altogether, to produce a word index such as Oizumi supplies, but with adjustable, user-specifiable reference information. A KWIC listing may also be ordered primarily by word-form, and within each by occurrence, or only by occurrence. *TACT* itself cannot make Oizumi's listing of word-occurrences in alphabetical order of the words following the keyword, although this can be accomplished by exporting a KWIC display and sorting it with other software.

In Tatlock and Kennedy the keyword is not marked; in Oizumi it is centred and separated from surrounding text by spaces. In *TACT* KWIC, Variable Context, and Text displays, the keyword may be highlighted on screen, and through a translation process, denoted by any symbol(s) the user designates.

TACT text references are adjustable within the range defined by textual markup and the parameters set during the compilation of the database. Line-divisions in Tatlock and Kennedy are implicit for the poetry, as full lines are usually given; Oizumi uses the virgule for poetic lines. In *UseBase* all line divisions, whether in poetic text or otherwise, are indicated in the KWIC

display by a vertical bar, and lines are wrapped, but in the Variable Context and Text displays, line endings appear as formatted in the text.

In *TACT* the hyphen can be treated for purposes of sorting and selecting in a variety of ways. Depending on which the user chooses, explicitly hyphenated forms may be selected as such, e.g., to form a separate listing such as Oizumi provides, included in a selection, or excluded. Thus "tendre-herted" could turn up in a list of all hyphenated forms, or in listings of all instances of "tendre" and "herted," or not appear in these listings, as the user wishes. In *UseBase*, non-hyphenated compounds could not be selected as a class of words without explicit markup, but by use of the pattern-description language, a query could easily find compounds based on any specifiable element. Homonyms likewise cannot be distinguished by *TACT* without explicit markup.

TACTFreq, based on Ian Lancashire's and Lidio Presutti's *MTAS* (*Micro Text-Analysis System*), may be used to generate a word index, such as Oizumi provides for each work. The program may also produce ranked frequency and reverse word-lists for all word-forms in the database. *TACTStat*, also derived from *MTAS*, generates type-token statistics, and *CollGen* lists of fixed and unfixed phrases and collocations. The second is described in some detail by Ian Lancashire elsewhere in this volume. Finally, Lancashire's *TACTDct* may be used to mark up texts with lemmatized and part-of-speech tags, which may then themselves be analyzed within *UseBase* in the same way as the base text itself.

5. Conclusion: is the printed concordance obsolescent?

Joe Raben once declared that handmade concordances were a thing of the past. Is this becoming true of the printed concordance?

Beside *TACT*, printed concordances do seem limited and awkward. Yet there are problems with interactive concordancing. First is accessibility. Although the *Riverside Chaucer* is now available from Oxford University Press in electronic form, for example, no software is yet available for all operating systems employed by scholars routinely. Second, there is the issue of longevity. As operating systems and computers change, they make software archaic; and no one yet knows how to preserve electronic data securely in a way that subsequent generations will be able to access. History has not been kind to enthusiasts for new media; it warns us that periods of change in the technology of cultural transmission have as a rule been accompanied by massive destruction of the artifacts produced by the older technology (Rouse 1992: 43). Electronic technology is *essentially* mutable. One solution, perhaps, might be to store all possible concordance listings of all works. Ignor-

ing for a moment the problem of data storage, what such a procedure would not preserve, however, is what the genius of interactive concordancing allows: the ability to play, in the sense of *serio ludere*, for our mutual enlightenment. So, serious problems remain, and the production of printed concordances is likely to continue.

We wait to see if the interactive concordancer lasts and proves fruitful. Some of us have no doubts whatever. Meanwhile, there is much work to be done — both books and software — that draw insight from concordances and transmit it into new scholarship.

Bibliography

- ANONYMOUS (1927). "A Concordance to Chaucer." *Times Literary Supplement*. 10 Nov.: 816.
- BRADLEY, John, Lidio Presutti, Michael Stairs and others (1993). *TACT*. Version 2.0 (in *beta* test). Version 1.2 is available from the Centre for Computing in the Humanities, Robarts Library, 14th Floor, University of Toronto, 130 St. George St., Toronto, Ont. M8X 1R5, Canada. E-mail: CCH@EPAS.UTORONTO.CA.
- BRANDL, A. (1927). *Archiv* 152.
- BURTON, D. M. (1981). "Automated Concordances and Word Indexes: The Fifties." *Computers and the Humanities* 15: 1-14.
- BUSA, R. (1980). "The Annals of Humanities Computing: The Index Thomisticus". *Computers and the Humanities* 14: 83-90.
- CHAUCER, Geoffrey (1989). *The Complete Poetry and Prose of Geoffrey Chaucer*. Ed. John H. Fisher. 2nd. ed. New York: Holt.
- DALY, Lloyd W. (1967). *Contributions to a History of Alphabetization in Antiquity and the Middle Ages*. *Revue d'Études Latines*. Bruxelles: Latomus.
- FISCHER, W. (1927). *Beiblatt zur Anglia* 38: 259-62.
- FLÜGEL, E. (1911). "Prolegomena and Side-Notes of the Chaucer Dictionary." *Anglia* 34: 354-422.
- (1913). "Specimen of the Chaucer Dictionary: Letter E." *Anglia* 37: 497-532.
- GRIFFITH, Dudley David (1955). *Bibliography of Chaucer 1908-1953*. Seattle: Univ. of Washington Press: 3-4.
- HOOPS, Johannes (1928/29). *Englische Studien* 63: 86-9.
- HOWARD-HILL, T. H. (1976). "On Literary Concordances: An Early View." *ALLC Bulletin* 4: 215-20.
- (1979). *Literary Concordances: A Guide to the Preparation of Manual and Computer Concordances*. Oxford: Pergamon Press.

- INGRAM, William (1974). "Concordances in the Seventies." *Computers and the Humanities* 8: 273-77.
- KURATH, Hans (1954). "History and Acknowledgements." In *Middle English Dictionary: Plan and Bibliography*. Ann Arbor: Univ. of Michigan Press.
- LAFFAL, Julius (1969-70). "Toward a Conceptual Grammar and Lexicon." *Computers and the Humanities* 4: 173-86.
- LANCASHIRE, Ian, and others (1983). "Verbal Texts: Concordances and Text Comparison." In *Computer Applications in Literary Studies: A Userbook for Students at Toronto*. Toronto: Department of English.
- (1986). "Concordance Programs for Literary Analysis." *Computing in Higher Education, SIGCUE Outlook* 19.1/2 (Spring/Summer).
- (1990). "Back to the Future: Literary and Linguistic Computing 1968-1988." In Yaacov Choueka, ed. *Computers in Literary and Linguistic Research. Literary and Linguistic Computing 1988*. Proceedings of the 15th Inter. Conf., 5-9 June 1988. Paris-Genève: 36-47.
- LUHN, H. P. (1966). "Keyword-in-Context Index for Technical Literature (KWIC Index)." Rpt. in *Readings in Automatic Language Processing*. Ed. David G. Hays. New York: American Elsevier: 159-67.
- MCARTHUR, Tom (1986). *Worlds of Reference: lexicography, learning and language from the clay tablet to the computer*. Cambridge: Cambridge University Press.
- MCCARTY, Willard (forthcoming). "Discontinuity, Metamorphosis, and Coherence: methodologies for computer-assisted textual analysis, with reference to the *Metamorphoses* of Ovid." In *Origins and Method: Towards a New Understanding of Judaism and Christianity*. Sheffield, U.K.: Sheffield Academic Press.
- MENNER, Robert J. (1928). *Modern Language Notes*. 332-6.
- MTAS (Micro Text-Analysis System), vers. 2.0 (1988). By Ian Lancashire and Lidio Presutti. Toronto: Centre for Computing in the Humanities.
- NORTHROP, Clark S. (1928). *Journal of English and Germanic Philology* 27: 238-40.
- OAKMAN, Robert L. (1980). *Computer Methods for Literary Research*. Columbia, SC: University of South Carolina Press: 69-87.
- OIZUMI, Akio (1991). *A Complete Concordance to the Works of Geoffrey Chaucer*.
- PARRISH, Stephen M. (1962). "Problems in the Making of Computer Concordances." *Studies in Bibliography* 15: 1-14.
- (1970). "Concordance-Making by Computer: Its Past, Future, Techniques, and Applications." In *Proceedings: Computer Applications to Problems in the Humanities, A Conversation in the Disciplines*. Ed. Frederick M. Burelbach, Jr. Brockport, N.Y.: State University of New York: 16-33.

- PARUNAK, H. Van Dyke (1981). "Prolegomena to Pictorial Concordances." *Computers and the Humanities* 15: 15-36.
- PATTERSON, Frank Allen, and French Rowe Fogle (1940). *An Index to the Columbia Edition of the Works of John Milton*. 3 vols. New York: Columbia University Press.
- RABEN, Joseph (1969). "The death of the handmade concordance." *Scholarly Publishing* 1: 61-69.
- ROUSE, R. H. (1992). "The Transmission of the Texts." In *The Legacy of Rome: A New Appraisal*. Ed. Richard Jenkyns. Oxford: Oxford University Press.
- and M. A. Rouse (1974a). "Biblical Distinctions in the Thirteenth Century." *Archives d'Histoire Doctrinale et Littéraire du Moyen Age* 41: 27-37.
- (1974b). "The Verbal Concordance to the Scriptures." *Archivum Fratrum Praedicatorum* 44: 5-30.
- ROYSTER, James F. (1928). "The Chaucer Concordance." *Studies in Philology* 25: 62-9.
- SCHMIDT, Klaus M. (1978). "Conceptual Glossaries: A New Tool for Medievalists." *Computers and the Humanities* 12: 19-26.
- STOCK, Brian (1983). *The Implications of Literacy: Written Language and Models of Interpretation in the Eleventh Century*. Princeton, N.J.: Princeton University Press.
- TATLOCK, John S. P. (1923). "The Chaucer Concordance." *Modern Language Notes* 38: 504-6.
- and Arthur G. Kennedy (1927). *A Concordance to the Works of Geoffrey Chaucer*.
- WHITMAN, Charles Huntington (1918). *A Subject Index to the Poems of Edmund Spenser*. New Haven: Yale University Press.
- WOOLDRIDGE, T. R., ed. (1991). *A TACT Exemplar*. CCH Working Papers 1. Toronto: Centre for Computing in the Humanities, University of Toronto.
- YOUNG, Ione Dodson, ed. (1965). *A Concordance to the Poetry of Byron*. Austin, Texas: Pemberton.