and others. 1st ed. 12 vols. Oxford: Clarendon.
TOMBEUR, Paul (1991). *Cetedoc Library of Christian Latin Texts.* Computer software. Louvain-la-Neuve: Brepol.
TRIVET, Nicholas (n.d.). *Exposicio Fratris Nicoli Trevethi Anglici* Ordinis Praedicatorum Super Boecio De Consolatione. Ts. Ed. Edmund Taite Silk. Pontifical Institute of Mediaeval Studies, Toronto.

Phrasal Repetends and "The Manciple's Prologue and Tale"

Ian Lancashire

My previous study of *The General Prologue* to *The Canterbury Tales* showed that phrases repeat in Chaucer's poetic language far more than expected (Lancashire 1993). Over 460 different collocations or fixed phrases from *The General Prologue* occur at least twice either in those 858 lines alone or in them and the rest of the tales. These phrases consist of at least two open-class words, or a sequence of any four or more words, or an idiom. Excluded from this count are most fixed phrases with two or more function words. Results for "The Manciple's Prologue and Tale" are similar. 377 different fixed phrases or collocations occur more than once in its 362 lines alone or in them and the rest of the tales. There is every reason to think that the other tales have comparable numbers. The *CollGen*-generated file of maximal repeating phrases of two or more words for the 182,000 words of the full unlemmatized *Canterbury Tales* has almost 22,000 different entries, which have over 3,400 different words in their phrasal lexicon. The ratio of maximal phrases to the tales' vocabulary, 12,047 different words, is 11:6. 60 percent of these entries have a frequency of 3 or more occurrences in the tales. This covers more text, in fact, than are in the tales themselves, even if we ignore persistent collocations among the repetends. Assuming that the phrasal repetends in the "Manciple's Prologue and Tale" are typical, with an average frequency of 4.59 occurrences, and an average length of 3.09 words (see Table I), 22,000 phrasal repetends will amount to 310,000 words, 170 percent of the tales' complete length in words. (This duplication takes place because the same words often participate in several different phrasal repetends at the same time.) Phrasal repetition, then, is a crucial aspect of Chaucer's writing. Among the 22,000 fixed phrases are 34 instances in which Chaucer repeats, verbatim or with minor variations, a complete line of verse (see Lancashire 1993). Under one-third of these have been noticed to date.

We knew that Chaucer had a large vocabulary, but these figures show that it was substantially phrasal and that we have to reassess what his poetic language was. Because phrasal repetition appears in everyone's English, no one should jump to conclude that Chaucer was an oral-formulaic poet. Consider three examples. Shakespeare's collection of poems in 1609 (4,321 different words in a word-count of 20,654) has 1,680 different phrases. Ben Jonson's *Volpone* (4,636 different words, with a word-count of 29,089) has

3,066 different phrases. Emily Bronte's *Wuthering Heights* (9,815 different words, with a word-count of 116,592) has 15,824 maximal fixed phrases of between two and ten words long. As texts become shorter, their vocabularies gradually exceed the number of their maximal phrases. Chaucer's repeated phrases must be viewed in the context of language itself for what they can tell us about Chaucer's style and mind.

1. *Phrasal repetends*

By phrasal repetends I mean (a) repeating fixed phrases and (b) repeating collocations, in which words co-occur, although in varying order, sometimes with words intervening. Fixed phrases are in fact a sub-set of collocations — that is, combinations of words or word-groups — because every word in a fixed phrase co-occurs with every other word in that phrase, whether or not the order is unchanging. Persistent association is logically prior to word-order. In practice, as well, most popular fixed phrases come unfixed easily. For example, the idiom, "for better, for worse," looks like a fixed phrase until paired with its regular associates, "for richer, for poorer," and "in sickness and in health." Because people do quote these three in different order, and in fragments, this phrasal repetend really consists of three collocating fixed phrases, any two of which may be absent. For this reason, phrasal repetends are fuzzy things. Their constituents combine unpredictably, and the boundaries for repeating collocation are never certain.

Some language researchers restrict phrasal repetition to lexically or semantically significant combinations, that is, fixed or unfixed phrases that mean something different from the senses their constituent words have by themselves. The phrases "government-binding theory" and "g-man," thus, are called linguistic collocations, but not "down with the government" or "her majesty's government." Defining phrasal repetends semantically also excludes syntactic structures such as "it is this that can be" or "might have been in the" as well as semantically-impoverished paralanguage like "well you know I mean," and associational clusters that extend beyond sentence boundaries.

Phrasal repetends are best treated as a fundamental characteristic of language, far too general to be restricted to linguistic collocations. Sociolinguist Deborah Tannen and corpus linguist Bengt Altenberg agree that speech unfolds chunkily as sequences of phrases that pile up, one after another, rather than fit into a prior, over-governing sentence structure, and that these phrases recur, uncomfortably for most writers whose process of composition begins, at least, as inner speech stored on paper or in computer memory. Most of us rely on a limited stock of phrases for everyday use.

For this reason, interviewers prune taped conversations of awkward repetitions before broadcasting them, and writers edit them away during revision.

The process of speech, outer or inner, uses repetition for the same reason that redundancy, within information theory, is a basic feature in the transmission of messages along a channel. Redundancy, which pads messages with repeated codes, ensures that noise does not distort communication. Something might be missed or destroyed once in transmission, but not twice or three times. Yet it would be misleading to compare human language process only to a channel, because experimental work in cognitive psychology and neuropsychology has given us a realistic paradigm for understanding why everything we utter is characterized by phrases. Working memory, which used to be called short-term memory, shows that our everyday phonological store, which we use to process most of what we hear, read, speak and write, holds no more than the words we can speak in under two seconds. This — our internal holding area for language, which is 'written over' every two seconds — explains the phrasal pulse of conversation and writing. Phrase length seldom extends beyond the boundaries of this phonological store. The combinatorial aspect of the phrase, linking two or more co-occurring words or word groups, is explained by the associative nature of long-term memory, not working memory. The more paths into a concept, the easier it is to recall; and it is more likely that that concept will be remembered with one or more associated ideas, that is, as a collocation, than just by itself. Neurons, connected to one another across synapses found on dendrites attached to their axons, provide the physical basis for associative memory.

Phrasal repetends, among the most banal of language features, may well then hold the key to seeing aspects of Chaucer's long-term memory and thus to his style, the way of writing that is distinctively his.

2. *Finding Phrasal Repetends*

Until *CollGen* was added to the Toronto text-retrieval and analysis system *TACT*, fixed phrases and collocating word-pairs in any text had to be discovered manually by combing through concordances. By running an eye down the keywords in contexts under each headword and checking for repeated words in front of and after the keywords, a reader might compile a list of fixed phrasal repetends. Because a concordance is many times the size of the original text, however, this process might take some time; and even then, at the end, very few collocations — recurring unfixed word-combinations — will have been recovered. Computer concording and text-retrieval programs like *Micro-OCP*, *WordCruncher*, Alan Reed's early *CLOC*, and recent Macintosh software like *Conc* enabled researchers to search for specified colloc-

ations or fixed phrases, but not to obtain them all at once without having to specify some template (see Lancashire 1991: 438-39, 486-88).

CollGen works off a *TACT* text database file previously created by *MakeBase*, *TACT*'s indexing program, but has its own set-up screen that allows users to select options. The first three concern global variables: length, frequency, and boundaries of the repetends. The screen first asks for the "span context" or the length of the passage in which a repetend may be said to occur, defined in words. If this is set from 2 to 20 words, then *CollGen* will catch fixed phrases as short as two words, and as long as 20 words, and will list all collocates that occur up to 19 words away from any "node" word. The screen next asks for the number of repetitions that have to occur in order for a phrase or collocation to be retrieved. The default setting is from 2 to just over 65,000. Finally, *CollGen* permits one to set boundaries for a collocation, beyond its simple length, by listing a tag variable. If a sentence boundary were set, for example, a collocation crossing into a bordering sentence would not be retrieved, even if the span context extended many words into that other sentence.

The second group of *CollGen* options affects the form of the output files. Collocate pairs appear in one file, each with a field indicating the strength of the association in a z-score (as in the *TACT* collocate display), but fixed phrases have two forms, the first a list formatted with tabs or spaces, and the second as "queries" that may be imported into *TACT* by the user. Both are ASCII files, but the first of the two is especially useful because it may be imported into standard database and spreadsheet programs, and because it implements a very important formal distinction among phrasal repetends. If fixed phrases are permuted, they are listed redundantly alphabetically under each word found in them; thus a five-word fixed phrase would be listed five times in that file. By permuting the phrases, an exhaustive index is created, but there would also be no easy way of determining exactly how many different phrases existed.

```
3  a bowe
2  a bowe he
2  a bowe he bar
2  a bowe he bar and
2  a bowe he bar and arwes
2  a bowe he bar and arwes brighte
2  a bowe he bar and arwes brighte and
2  a bowe he bar and arwes brighte and kene
```

Each substring (with two words or more) of each phrase is listed separately in this file. Thus the five-word phrase "a bowe he bar and arwes brighte and kene" appears eight times in this permuted index. It is clear that there are only two phrasal repetends in this list: one is the longest, or maximal "a bowe he bar and arwes brighte and kene"; and the other is its shorter but more frequent sub-string "a bowe". The rest are just fragments of the maximal repetend. The second fixed-phrase output file yields only maximal and the shorter, higher-frequency phrases that are its sub-strings. This list holds the phrasal pulse of a writer in his writing.

3. *Phrasal repetends in the "Manciple's Prologue and Tale"*

CollGen gave me the list of all maximal phrasal repetends in all the tales, but discovering which ones occurred in the Manciple's sections required me to search this list for all successive word pairs in the poem. This search netted just the fixed phrases. To capture the collocations, I used *UseBase* (the *TACT* program for interactive use) interactively, checking for co-occurrences of successive pairs and triplets of open-class (not function) words. At this stage I added citations to the basic list of fixed and collocating word phrases. About 1,800 maximal fixed phrases appear in the prologue and tale, approximately five for each line, but the majority of these are short strings of function words, or single open-class words with function words. At this point I decided to look only at the subset of repeating phrases that have two or more open-class words (nouns, adjectives, lexical verbs and adverbs, etc.), or that are at least four words in length, or that are prepositional phrases. This subset holds the phrases with important semantic or thematic content, the part of language that varies most, and I used it previously in studying the "General Prologue."

The 377 phrasal repetends in this subset have certain quantitative features. This information appears in a *Quattro Pro* spreadsheet file and, selectively, in Table I. I first imported an ASCII or DOS file of the repetends and their total frequencies into *Quattro Pro*. Each row contains one phrase, and there are columns for its length, number of function words, number of content words, first line number in the "Manciple's Prologue or Tale" in which it occurs, additional line numbers for further occurrences, total number of occurrences of the phrase in the *Canterbury Tales*, occurrences in each of 57 sections of the tales, and the spread of the phrase across these sections (that is, the number of sections in which it occurs). Once entered, the data was analyzed. I used standard spreadsheet formulas to calculate totals, averages, and frequencies for each variable or piece of information. Like most spreadsheet software, *Quattro Pro* also does database analysis, e.g.,

Word-count in all phrase types	1165
Open-class (content) word tokens	597
Closed-class (function) word tokens	568
Total occurrences of phrases (all tales)	1729
Average length	3.1
Average frequency	4.6
Phrases with only content words	104
Phrases with only function words	24
Phrases in *ManchPro&T* only	18

Table I. Profile of phrasal repetends in ManchPro and ManchT

ascending or descending sorts of the rows on numbers in each column. This feature enabled me to print out lists of repeated phrases by section. Finally, the spreadsheet graphed distributions of these phrases by section in the tales, and by line numbers in the "Manciple's Prologue and Tale."

It is unusual in critical analysis to do so much preparatory work and still have no idea of what results to expect, but patterns in data of this magnitude do not pop out at one. With both the "General Prologue" and the "Manciple's Prologue and Tale," I was surprised by the results. In any computer-related work, an open mind is an advantage. Let me review what I found out, with three caveats. It would be advantage to work with Chaucer's works as a whole, not just the *Canterbury Tales*, and with a lemmatized version of that text. First, lemmatized text, by collapsing inflectional and spelling variants of a word into one form, will increase the frequency of repeated phrases, and perhaps reduce their number. Second, patterns in the data would change if I included all phrases, not just ones with "content," although possibly not critically, to judge from my work on *Hamlet*, III.1 (Lancashire forthcoming). Third, interpreting data sensitively asks for more time than I have had with this data-set. This is a working paper in the spirit of the CCH series in which these conference papers will be published.

Table II lists the phrasal repetends after sorting on length in descending order. The two longest phrases both exceed a line; each begins and ends with a rhyme word. Chaucer clearly repeated, over long stretches of the tales, substantial clauses, most of which have no obvious connection to a

tale's subject (only one phrase, "now hadde this Phebus in his hous a", is localized). That alone is surprising, but the low frequency of these repetends is expected. The number of content words in a phrase, and its length, vary proportionally with one another and inversely with frequency, for which see Table III, which sorts phrases in descending order of number of content words in them. All are infrequent, except for a remark about faithlessness, men, and women, and three invocations of God. The shorter the phrase, the more likely it will have a high frequency. Table IV lists phrases in descending order of frequency, with few surprises, "every man" (34 instances) excepted. A comparison of the 18 phrasal repetends that never get beyond the "Manciple's Prologue and Tale" in the tales as a whole (see Table V) shows that, on average, they are longer (4.1:3.09), less frequent (2.6:4.58) and more favouring function words (57% to 51%) than the data-set as a whole. The longer a phrase gets, the more function words it has, the harder it is to recall from memory.

Of the 377 phrases as a whole, we have to ask, do they all belong to a combinatorial lexicon of poetic idioms or stock phrases? Table VI lists phrasal repetends for the first 20 lines of the "Manciple's Prologue." Many of these combinations must reflect Middle English usage, rather than Chaucer the poet. For instance, "jape (or, and) pleye", "(theef, theves) & (robbe, robbours)" and "for cokkes bones" look like popular idioms, absorbed by Chaucer as linguistic forms of his age. One way to find out would be to analyze, for phrases, a representative corpus of Middle English texts, such as David Burnley is setting up, but before that happens a method would have to be found of filtering the phrases to include only ones that occur often and widely enough to qualify as possible idioms. Phrases like "kn[eo]we*t*h* & his penaunce", for instance, occur too infrequently to be called habitual.

Table VII gives a three-way sort of the phrases that, without asking a critic to decide on semantic grounds, may isolate phrases that Chaucer re-used widely and that, because of their length, took some effort to recall. I derived this table as follows. All 377 repetends are sorted first by descending frequency, and only the 117 that occur between 5 and 62 times are culled out for further processing. These 117 are next sorted by descending length (in words), and only the 57 longest (between 5 and 3 words) are carried forward. Last, these 57 are sorted in descending order of spread among the sections, so that phrases appearing in fewer sections move to the bottom. 34 phrases survive if we restrict them to a spread between 6 and 16 sections. I suggest that this method produces a reasonable slice of phrasal repetends characteristic of Chaucer's writing. Of course this pool of candidates excludes repeated phrases that do not occur in this prologue and tale, just two of the 57 sections. Only by examining the entire corpus could we recover a broadly representative set of Chaucer's idioms.

How many of the phrases in Table VII, then, reflect 14th-century English, and how many reflect Chaucer's personal contribution to it? Anyone can guess, but only a computer textbase will give probable cause for deciding.

Fig. 1 shows the distribution of the 377 phrasal repetends in the "Manciple's Prologue and Tale" through that poem. Note the sudden increase of repeating phrases at the very end of the prologue and the start of the tale, a period lasting about 60 lines. The second peak, at about lines 200-20, just after a ten-line section where there is little repetition, occurs when Chaucer meditates on the unfaithfulness of Phoebus' wife. Note the wave-like rhythm of the peaks, about every hundred lines (at ten-line sections ending at lines 20, 110, 210 and 310. Each of these peaks is preceded by a short passage without much repetition (at ten-line sections ending at lines 10, 100, 200 and 290). Is this a periodic reflection of Chaucer's creativity, alternating between moments where he creates new word-combinations, and more extended passages when he re-uses material? Is that Chaucer's "phrasal pulse"? A more detailed study would be needed to find out.

Chaucer clearly relies on stock poetic phrases when he opens and closes sections. That effect occurred in "The General Prologue" and reappears here around lines 104-5, when the prologue shifts into the tale. Table VIII lists lines 90-120 and marks all maximal repeated fixed phrases in bold-face, and collocating terms in italics: it restates the information in the distribution graph. Lines 99-104, the end of the prologue, have many more content-specific repetends than lines 90-98. The portrait of Phoebus that follows is also rich in repeated materials. I think the explanation for this effect — the book-ends, so to speak, of a poem — must relate to a special set of conventions that Chaucer used when beginning and ending something. This is a definable poetic procedure. Note how phrases that are unrepeated, such as "laughen wonder", "good drynke", and "Worship and thank", stand out. Note that eleven lines of 31 consist completely of phrases re-used elsewhere. In order to get a distribution of the 377 repeating phrases from the "Manciple's Prologue and Tale" through all 57 sections of *Canterbury Tales*, we cannot use the frequencies of phrases in them without taking the length of those sections into account. The procedure I use for this normalization is simple. I calculate an "expected" number of phrases for each section (prologue, tale, or connecting passage) by multiplying the total number of occurrences of the phrases by the length of that section in words over the length of the *Canterbury Tales* in words. Figures for the "Manciple's Prologue and Tale" itself are excluded because they are the source of the phrases in whose distribution we are interested. Then I compare the difference between the actual and expected numbers in a graph (Fig. 2).

The distribution is consistent with the one that I produced for the

"General Prologue" (Lancashire), phrases from which remarkably overlapped ones in the "Manciple's Prologue and Tale." In both, "Melibee" and "Parson's Tale" scored much lower than expected, presumably because they are prose, not verse, and the resemblance between the two distributions extends to other tales, also scoring lower in each graph: "Miller's Tale," "Squire's Tale," "Monk's Tale," and "Second Nun's Tale." Tales that scored higher than expected in each include "Cook's Tale," "Wife of Bath's Prologue," "Summoner's Tale," "Merchant's Tale" (although less so in the Manciple's instance), "Physician's Tale," "Pardoner's Tale," and "Nun's Priest's Tale." Graphs for the "General Prologue" and for "Manciple's Prologue and Tale" differ in two respects. Each section overlaps more with its near neighbours, the "General Prologue" with "Knight's Tale" (which has a very low score in "Manciple's") and the "Manciple's" with "Canon Yeoman's"; and "Franklin's Tale," which scores low for the "General Prologue," scores highest, along with "Canon Yeoman's," for the "Manciple's."

How should we interpret this distribution graph? What causes variation in the way phrasal repetends from the "Manciple's Prologue and Tale" appear in other tales? Several explanations suggest themselves: similarity in the topic, shared genre (e.g., fabliau, saint's life, etc.), deliberate echoing, or just proximity in the time of composition.

No confident answer is yet possible, but if we look at phrasal repetends found in the "Canon Yeoman's Tale" — which scores highest — we see that not one of the 52 phrases it shares recalls the subject matter of the two poems (Table IX), that the two poems belong to different genres (one an anti-clerical tale satirizing alchemy, the other a beast fable), and that, so general are the phrases' content, it is hard to imagine any echoes could be recognized as rhetorical features. This leaves proximity in time. If the contents of Chaucer's memory changed over time, as our memories certainly do today, so that some early associational branches revealed in these word combinations weakened, while other ones, laid down later, were recollected more readily, the unevenness in their distribution through the tales might best be explained as the result of their being written at different times. The more phrasal repetends two tales share, other factors not appearing, the closer in time are their dates of composition. Of all the verse tales, the Knight's shares many phrases fewer than expected, and it was probably based on the earlier "Palamon and Arcite." Other tales regarded as having been written before Chaucer started work on the *Canterbury Tales* — the tales of the Squire, the Monk and the Second Nun — also show fewer phrases than expected. Curiously, another tale of marital infidelity, the Miller's tale, appears in this group too.

4. *Chaucer's Cognitive Style*

If proximity in time explains systemic phrasal overlap, Chaucer could not have been consciously aware of why he was using many of the phrases he did. They must have seemed `right' to him, and he put them in. Some reflection on what we ourselves do when we utter inner speech and express it in oral or written form will convince most of us that we speak or write, unconscious of how we are doing so. We do not create a sentence, review it mentally, and then utter it. Knowing how to use language involves cognitive procedural memory (Squire 1987: ; Kosslyn and Koenig 1992: 373). Unlike facts and episodes in one's life, which are part of semantic memory, procedural knowledge of cognitive skills, including how to use language, turns out not to be subject to general loss during amnesia, although we may in fact `forget' language facts, such as rules of good writing, names for rhetorical devices, etc. We can neither recall nor forget how to utter speech. Procedural knowledge is only and always expressed in performance. It must change over time, but how it does so is a mystery.

If this view is right, and it is accepted by most cognitive psychologists today, then language process is far stranger than we think, and the methods of finding out how that process works in the texts it leaves behind may well seem alien to the experience we are trying to describe. Computational text analysis appears just such an alien methodology to most literary critics. If it produces interesting results, however, the discipline should ask why it does so.

One of the strangest results in computer-based text study is the phrasal repetend graphs that materialize when we group the phrases, not by any classification we think of (e.g., metrical tags, proverbs, etc.), but simply by their constituent words. Figs. 3 and 4 are graphs for two words among the 39 most common open-class words employed in the 22,000 maximal phrasal repetends in the *Canterbury Tales*: "wo" (53 occurrences) and "love" (114 occurrences). Each graph displays three kinds of phrases: ones ending with the keyword (these are listed at the top left), ones centered on the keyword (these appear directly to left and right of the keyword), and ones beginning with the keyword (these are listed at the bottom right). The graphs for "wo" and "love" are associative networks by which, in unconscious recall, Chaucer might have found his way to or from those two concepts. Note that at the bottom left of the "love" graph we have the phrase, "love ne wyf". The second graph, for "wo," rather more explicitly shows how a network of "care", "wo", "peyne" and "allas" grows. There is no ambiguity here. These are associative networks, for several decades known to be the basis of our long-term memory, and they emerged without being nudged into place by human hand.

The research described in this paper asks us to re-think Chaucer's style. Repeated phrases and collocations are an essential component of his poetic idiom, and differ in his prose. Phrasal density in the poems exhibits sometimes a pulse-like wave, as if Chaucer alternated innovative use of language with stretches in which he relied on a familiar phrasal store, although openings and closings of sections show particularly heavy clusters of these repeated phrases. They mark conventional passages. The patterns of distribution that these repeated phrases show over many years of composition, on the other hand, may mirror Chaucer's associative long-term memory as it changes over time. Where two poems share many more phrases than we might expect, they may have been written about the same period. A text's profile of phrasal repetends, then, may `date-stamp' a text, locating it chronologically with other texts by the same author. Genre does not yet appear to be a factor in distinguishing repeated phrases into groups. Finally, phrasal repetends coalesce into associative networks that may capture fragments of Chaucer's long-term memory — his semantic store — through a long career.

The next step is to test these observations through Chaucer's entire poetic canon. Fixed phrasal repetends will have to be listed for each poem. The new electronic *Riverside Chaucer* by Larry Benson will be invaluable in making this research possible.

PHRASE	LEN-GHT	FUNCT. WORDS	CONT.	1st OCCUR.	2nd OCCUR.	TOTAL FREQ.
man that is or was sith that the world bigan	10	7	3	119		2
a wyf which that he lovede moore than his lyf	10	7	3	139		2
of litel reputacioun (nat, noght) worth to I (as in, in) comparisoun	9	5	4	200	254	2
now hadde this Phebus in his hous a	8	6	2	130	139	2
(accorde & dede & (speche, word)) & (and that I (thee, yow) biseche)	8	4	4	208		2
the wordes moote be cosyn to the dede, the word moost cosyn be to the werkyn	8	5	3	210		2
this is th'effect ther is namoore to	8	7	1	267		2
(in his hand) & (baar, beren) & ((a myghty bowe), (a bowe))	7	3	4	129		2
mete and drynke of alle deyntes that	7	4	3	165		2
hir thoughte (hir, his) (cursed, sorweful) herte brast atwo	7	3	4	263		2
was sith that the world bigan	6	4	2	120		3
fulfild of & honour and of & worthynesse	6	3	3	123		2
wyf & (had, hath) & sent & for hir lemman	6	3	3	204	238	2
right as a swerd & right so	6	3	3	340		2
for preyere ne for (hyre, meede, yifte)	5	3	2	6		2
(swonken, swynke, yswonke) & ((al the longe nyght), a-nyght, (al nyght))	5	2	3	18		3
(ful pale) & (and (nat, nothyng) reed)	5	3	2	20		2
(fordronken, dronke) & speketh & nose & the pose	5	1	4	60		2
(so, als ever, also) moote* I thryve	5	4	1	80	255	6
I seyde it in my	5	4	1	81		2

Table II. Longest phrasal repetends

Table III. Phrasal repetends with most content words

PHRASE	LEN-GHT	FUNCT. WORDS	CONT.	1st OCCUR.	2nd OCCUR.	TOTAL FREQ.
(fordronken, dronke) & speketh & nose & the pose	5	1	4	60		2
(in his hand) & (baar, beren) & ((a myghty bowe), (a bowe))	7	3	4	129		2
of litel reputacioun (nat, noght) worth to I (as in, in) comparisoun	9	5	4	200	254	2
(accorde & dede & (speche, word)) & (and that I (thee, yow) biseche)	8	4	4	208		2
hir thoughte (hir, his) (cursed, sorweful) herte brast atwo	7	3	4	263		2
God & (yeve me sorwe)	4	1	3	15		4
(swonken, swynke, yswonke) & ((al the longe nyght), a-nyght, (al nyght))	5	2	3	18		3
God & blesse & soule	3		3	21		5
God (his, my) soule blesse	4	1	3	21		5
as olde bookes maken	4	1	3	106		2
man that is or was sith that the world bigan	10	7	3	119		2
fulfild of & honour and of & worthynesse	6	3	3	123		2
a wyf which that he lovede moore than his lyf	10	7	3	139		2
mete and drynke of alle deyntes that	7	4	3	165		2
men & been & untrewe and & women	5	2	3	188		2
wyf & (had, hath) & sent & for hir lemman	6	3	3	204	238	2
the wordes moote be cosyn to the dede, the word moost cosyn be to the werkyn	8	5	3	210		2
(sadd*e* and) & trewe & (of hewe) [rhymes]	5	2	3	275		2
face pale of hewe	4	1	3	276		2
synge ywis as any nyghtyngale, songe whilom lyk a nyghtyngale	5	2	3	294		2

PHRASE	LEN-GHT	FUNCT. WORDS	CONT.	1st OCCUR.	2nd OCCUR.	TOTAL FREQ.
worde*s* & (seye*, seyde*, seith, seyn, seyden)	2		2	207	241,	62
every man	2	1	1	134	283	34
and after that	3	3		270		32
by my (fey, fay, fayth, feith)	3	2	1	13		23
in ‖ (lliylf, lyves*, lyvynge*)	2	1	1	154	311	23
God yeve	2		2	15		18
to and fro	3	3		53		17
no fors	2	1	1	68		17
right anon	2		2	84		17
in ‖lyf	2	1	1	297		17
many a man	3	2	1	326		17
of deeth	2	1	1	86		15
upon a day	3	2	1	110		15
God it woot	3	1	2	160	221	15
(do, dooth) & harm	2		2	232		15
at ‖ tymes*	2	1	1	330		15
and right anon	3	1	2	84		14
lady & (loves*t*, loveres*)	2		2	218		14
men clepen*	2		2	234		14
I yow preye	3	2	1	309		14

Table IV. Phrasal repetends with highest frequency

PHRASE	LEN-GHT	FUNCT. WORDS	CONT.	1st OCCUR.	2nd OCCUR.	TOTAL FREQ.
kepen his tonge weel, ke*pe* wel thy tonge	4	2	2	315	319,	4
avysel*y* & (speeke, speche, spekynge)	2		2	324	327,	3
see how he	3	2	1	9	35	2
breeth & (stynketh, stynkyng)	2		2	32	39	2
gentill*esse & worthynesse [as rhymes]	2		2	124	249	2
now hadde this Phebus in his hous a	8	6	2	130	139	2
of litel reputacioun (nat, noght) worth to I (as in, in) comparisoun	9	5	4	200	254	2
come*th muchel harm	3	1	2	202	337	2
for hir lemman	3	2	1	204	238	2
wyf & (had, hath) & sent & for hir lemman	6	3	3	204	238	2
(ther (is, nys) no difference) & bitwixe	5	4	1	212	223	2
she shal be cleped his	5	4	1	218	220	2
an outlawe or a theef	5	3	2	224	234	2
noght textueel	2	1	1	235	316	2
and to the crowe	4	3	1	292	303	2
and ‖ white fetheres everliylchon	4	2	2	296	304	2
thenk (on, upon) the crowe	4	2	2	318	362	2
to the crowe	3	2	1	270	292,	3

Table V. Phrasal repetends only in "The Manciple's Tale"

PHRASE	LEN-GHT	FUNCT. WORDS	CONT.	1st OCCUR	2nd OCCUR	TOTAL FREQ.
jape (or, and) pleye	3	1	2	4		2
gan & (oure Hooste*)	3	2	1	4		3
in the m[iy]re	3	2	1	5	290	5
for preyere ne for (hyre, meede, yifte)	5	3	2	6		2
and s[ae]lyde sires	3	1	2	6		2
(theef, theves) & (robbe, robbours)	2		2	8		2
see how he	3	2	1	9	35	2
for cokkes bones	3	1	2	9		2
(falle, fil) & atones	2	1	1	10		2
with meschau*nce	2	1	1	11	193	5
cook of londoun	3	1	2	11		3
kn[eo]we*t*h* & his penaunce	3	1	2	12		2
(he, hym) & come forth	3	1	2	12		4
shal telle a tale	4	2	2	13		3
by my (fey, fay, fayth, feith)	3	2	1	13		23
telle a tale	3	1	2	13		6
God & (yeve me sorwe)	4	1	3	15		4
God yeve	2		2	15		18
God yeve thee	3	1	2	15		3
quod he God	3	1	2	15		5
by the morwe	3	2	1	16		7
what eyleth (thee, the, yow)	3	2	1	16		8
(swonken, swynke, yswonke) & ((al the longe nyght), a-nyght, (al nyght))	5	2	3	18		3
(ful pale) & (and (nat, nothyng) reed)	5	3	2	20		2
ful pale	2	1	1	20	30	5

Table VI. Phrasal repetends for "The Manciple's Prologue," lines 1-20

PHRASE	LEN-GHT	FUNCT. WORDS	CONT.	TOTAL FREQ.	SPREAD
by my (fey, fay, fayth, feith)	3	2	1	23	16
and after that	3	3		32	14
upon a day	3	2	1	15	13
to and fro	3	3		17	12
and right anon	3	1	2	14	12
right thus & (seyde, seyn, seye*, seith)	3	1	2	12	11
God it woot	3	1	2	15	10
many a man	3	2	1	17	10
if (he, I) may	3	3		11	9
I yow preye	3	2	1	14	9
what eyleth (thee, the, yow)	3	2	1	8	8
in al this world	4	3	1	9	8
ther may no man	4	3	1	9	8
mete and dr[iy]ke	3	1	2	10	8
so greet (a, an)	3	2	1	13	8
and that anon	3	2	1	9	8
as & clerkes & (s[ae]lyn, writen)	3	1	2	8	8
I dar wel seyn	4	2	2	9	8
to no wight	3	2	1	10	7
if I may	3	3		7	7
(hym in, in) swich (wise, wise as)	5	4	1	7	7
I see* wel	3	2	1	10	7
(in (this the) erthe), (in erthe)	3	2	1	13	7
in al (thy, his, my) lyf	4	3	1	9	7
telle a tale	3	1	2	6	6
(so, als ever, also) moote* I thryve	5	4	1	6	6
sire quod he	3	1	2	8	6
for in this world	4	3	1	6	6
many a day	3	2	1	6	6
nyght and day	3	1	2	10	6
that no man sholde	4	3	1	6	6
as ye may	3	3		6	6
never a deel	3	2	1	8	6
thyng that is	3	3		6	6

Table VII. Most frequent, long, and widely spread phrasal repetends

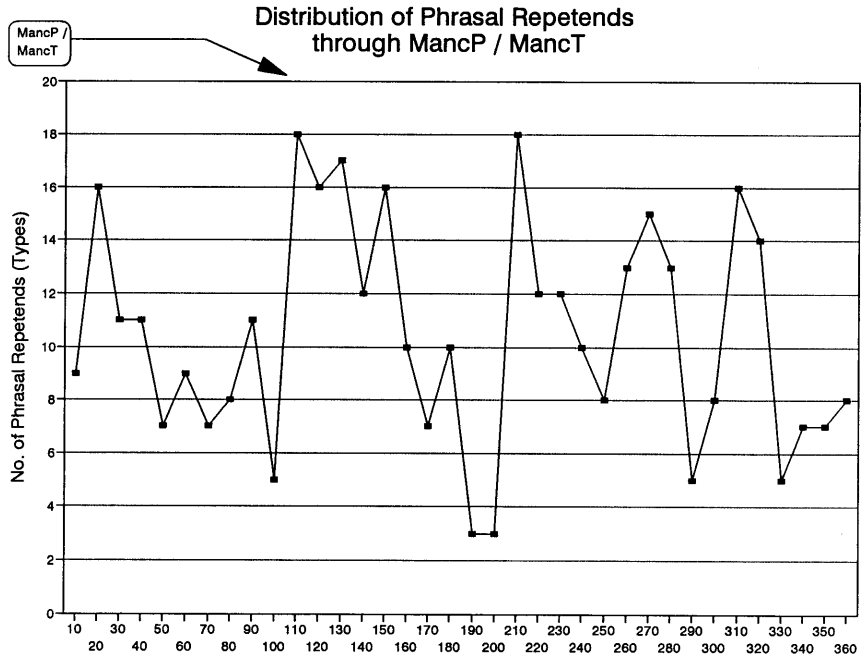

Fig. 1. Distribution of phrasal repetends through "The Manciple's Prologue and Tale"

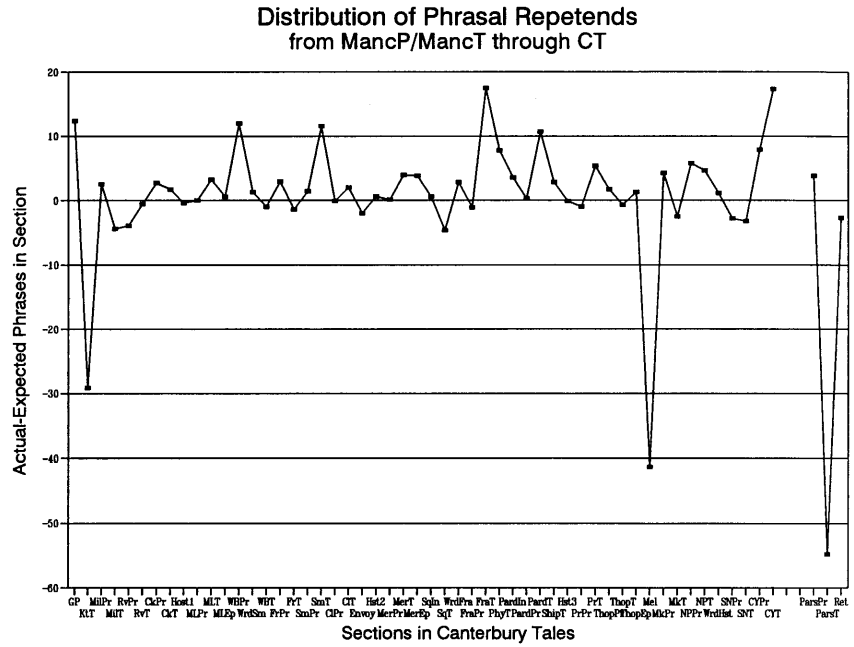

Fig. 2. Distribution of phrasal repetends from "The Manciple's Prologue and Tale" through *The Canterbury Tales*

<L 90> **And whan he hadde** pouped **in this** horn
<L 91> **To the Manciple he took the** gourde **agayn.**
<L 92> **And of that** drynke the Cook **was wonder fayn,**
<L 93> **And thanked hym in swich wise as he koude.**
<L 94> Thanne **gan oure Hoost to** laughen wonder **loude,**
<L 95> **And seyde,** "**I se wel it is necessarie,**
<L 96> **Where that we goon,** good drynke we **with us** carie.
<L 97> For **that wol** turne **rancour and disese**
<L 98> T'acord **and love, and many a** wrong apese.
<L 99> O Bacus, **yblessed be thy name,**
<L 100> **That so** kanst turnen *ernest* into *game*!
<L 101> Worship and thank be **to thy** deitee!
<L 102> **Of that** mateere **ye gete namoore of me.**
<L 103> **Telle on thy tale, Manciple, I thee preye.**"
<L 104> "**Wel, sire,**" **quod he, "now herkneth what I seye.**"

Heere bigynneth the Maunciples Tale of the Crowe.

<L 105> Whan Phebus dwelled **heere in this erthe** adoun,
<L 106> **As olde bookes maken mencioun,**
<L 107> **He was the mooste lusty bachiler**
<L 108> **In al this world, and eek the beste** archer.
<L 109> *He slow* Phitoun *the serpent* **as he lay**
<L 110> Slepynge **agayn the sonne upon a day,**
<L 111> **And many another** *noble worthy* dede
<L 112> **He with his bowe** wroghte, **as men may rede.**
<L 113> *Pleyen* **he koude on every *myrstralcie,***
<L 114> **And** *syngen* **that it was a melodie**
<L 115> **To *heeren* of his** cleere *voys* **the soun.**
<L 116> **Certes the *kyng* of *Thebes* Amphioun,**
<L 117> **That with his** syngyng walled **that citee,**
<L 118> Koude nevere syngen **half so wel as** hee.
<L 119> **Therto he was the semelieste man**
<L 120> **That is or was sith that the world bigan.**

Table VIII. Phrasal repetends in the juncture of the "Manciple's Prologue and Tale"

PHRASE	LEN-GHT	TOTAL FREQ.	CYT SHARED	SPREAD
every man	2	34	5	17
(taken*, taketh) heede	2	7	2	6
no fors	2	17	2	13
another day	2	5	2	4
right anon	2	17	2	11
sire quod he	3	8	2	6
in his hand	3	5	2	4
I yow preye	3	14	2	9
yvele apayd	2	8	2	6
(he, hym) & come forth	3	4	1	4
by my (fey, fay, fayth, feith)	3	23	1	16
in good (feith, fey)	3	2	1	2
good feith	2	3	1	3
fader kyn	2	6	1	6
devel of helle	3	4	1	3
to and fro	3	17	1	12
care and wo	3	3	1	2
by my savacioun	3	3	1	3
in feith	2	10	1	7
(cometh, cam) to & preef	3	2	1	2
yet hadde I levere	4	3	1	3
and wite ye	3	2	1	2
if I may	3	7	1	7
wel it is	3	5	1	4
I see* wel	3	10	1	7
I thee* preye	3	2	1	2
sire quod he now	4	2	1	2
as I (maken, maketh) mencioun he	4	2	1	2
in al this world	4	9	1	8
upon a day	3	15	1	13
(syngeth, syngen*, songe) & nyghtyngale	2	5	1	4
nyght and day	3	10	1	6
save oo*nly	2	10	1	7
(sooth, soothly, sothe) & sfaelyn	2	8	1	7
werke*s* and thoghte*s*	3	6	1	5
labour is in vfaelyn	4	2	1	2
is in vfaelyn	3	3	1	3
al that he kan	4	4	1	4
ther may no man	4	9	1	8
if (he, I) may	3	11	1	9
it happeth & ofte & so	4	3	1	3
com.* & harme*s*	2	7	1	5
worde*s* & (seye*, seyde*, seith, seyn, seyden)	2	62	1	24
right thus & (seyde, seyn, seye*, seith)	3	12	1	11
(I seye) & (the same)	4	4	1	4
blered is (myn, thyn) eye	4	2	1	2
what wol ye	3	6	1	5
for sorwe of which	4	3	1	3
(y*broghte*) & to confusioun	3	3	1	3
many a man	3	17	1	10
at I tymes*	2	15	1	9
thyng that is	3	6	1	6

Table IX. Phrasal repetends from "The Manciple's Prologue and Tale" found in "The Canon's Yeoman's Tale"

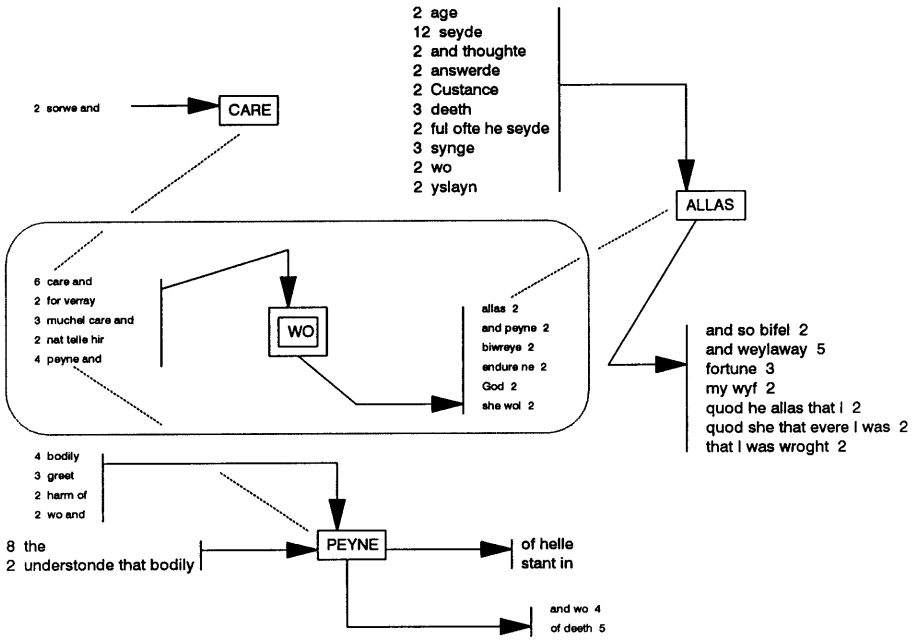

Fig. 3. Fixed phrasal repetend graph for "Wo" in *The Canterbury Tales*

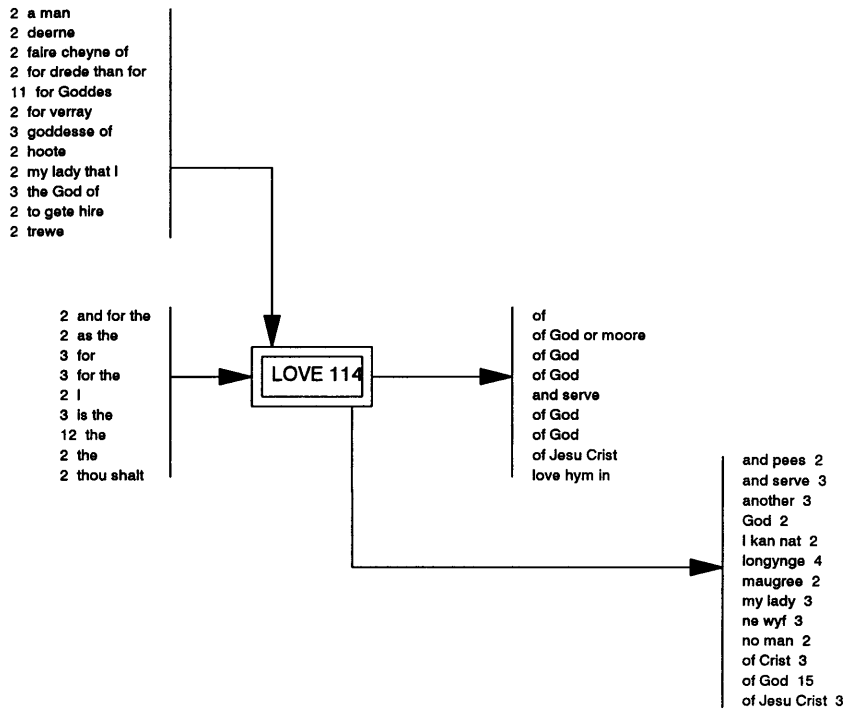

Fig. 4. Fixed phrasal repetend graph for "Love" in *The Canterbury Tales*

Bibliography

BADDELEY, Alan (1990). *Human Memory: Theory and Practice.* Hove and London: Lawrence Erlbaum.

BRADLEY, John, Lidio Presutti, Michael Stairs and others (1993). *TACT.* Version 2.1 (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.

KOSSLYN, Stephen M., and Olivier Koenig (1992). *Wet Mind: The New Cognitive Neuroscience.* New York: The Free Press.

LANCASHIRE, Ian (1991), ed. *The Humanities Computing Yearbook 1989-1990: A Comprehensive Guide to Software and Other Resources.* Oxford: Clarendon Press.

— (1993). `Chaucer's Repetends from The General Prologue of The Canterbury Tales.' 32 pp. Forthcoming in *The Centre and Its Compass: Perspectives on Medieval Literature in Honour of Professor John Leyerle*, ed. J. Burke and others. Western Michigan University Press.

— (forthcoming). "Phrasal Repetends in Literary Stylistics: *Hamlet* III.1." *Research in Humanities Computing*, 4. Oxford: Clarendon Press.

QUATTRO Pro (1992). Version 4.0. Borland International, 1800 Green Hills Road, P.O. Box 66001, Scotts Valley, CA 95067-0001.

SQUIRE, Larry R. (1987). *Memory and Brain.* New York and London: Oxford University Press.

The Sheffield Chaucer Textbase: its Compilation and Uses

David Burnley

Some time ago I published a book on Chaucer's ideas and language in which I sought to introduce as a technical term the word "architecture" to refer to a feature of language use (Burnley 1979; and further 1991). In the present context it would be natural to assume that in selecting that term for what I had in mind I was influenced by the language of computer technology. This would be a quite incorrect assumption. At the time I was not aware of having heard the word mentioned in connection with computers but was rather using a suggestion made as early as 1951 by a Norwegian linguist (Flydal 1951) to describe something like the feature of linguistic usage that I wanted to discuss.

This fact is worth mentioning — apart from its value in explaining the origins of this paper — because it also serves as an illustration of this very aspect of linguistic meaning. This is meaning which lies outside the ordinary scope of glossaries, dictionaries, concordances and grammars, but which is nevertheless crucial to the competent communicative use of a language, and therefore, by extension, to the subtle interpretation of utterances in that language. It is a question of context. Much of the meaning of words is contributed by the context in which they are used. It may be as fundamental a matter as the disambiguation of homophones by immediate linguistic context, or it may simply be that the association with contexts of use and common preoccupations gives resonance to the words of any language. Thus, we probably made an erroneous assumption about my use of the word "architecture" because a `computer sense' is salient in the present circumstances. Words gain their associative resonance not from our knowing the language as much as from our knowing the cultural contexts in which the words are used, and also from responding to the promptings of immediate context.

In order to use and interpret a language with fullest competence, it is necessary to be aware of the consequences of the choice of certain linguistic items in preference to alternatives which might seem linguistically possible in the same context and circumstances. This is a question of appropriateness and acceptability, which are stylistic notions different from the linguistic ones of ungrammaticality or the use of malapropisms. For a remote language, they are much less easy to catalogue exhaustively, but must once have been equally essential to full communicative competence (Hymes 1971). Skill in

CCH WORKING PAPERS 3, 1993, 123-40