

Leeds Studies in English

Article:

Charles Barber and Nicolas Barber, 'The Versification of The Canterbury Tales A Computer-based Statistical Study', *Leeds Studies in English*, n.s. 22, (1991), 57-83

Permanent URL:

https://ludos.leeds.ac.uk:443/R/-?func=dbin-jump-full&object_id=121847&silos_library=GEN01



Leeds Studies in English
School of English
University of Leeds
<http://www.leeds.ac.uk/lse>

The Versification of *The Canterbury Tales* A Computer-based Statistical Study

Charles Barber and Nicolas Barber

PART II

In which words was E pronounced?

In Part I we showed that there are strong grounds for believing that about half the examples of unelided word-final *-e* in *The Canterbury Tales* were pronounced. But in which words was it pronounced, and in which not? All words containing UE (unelided word-final *-e* within the verse-line) were recorded in Columns 7 and 8 of the data, and an examination of this material may give a partial answer to the question.¹ Statistical material cannot indeed tell us whether UE was pronounced in a particular word in a particular line, but it can provide strong evidence that UE was frequently pronounced (or not pronounced) in certain words. As we shall see later, the material does indeed provide evidence that in some words UE was usually pronounced, while in others it was not usually pronounced; in yet a third category of word, the evidence suggests that UE was sometimes pronounced, but sometimes not. For convenience, we can label these three categories P ('pronounced'), NP ('not pronounced'), and SP ('sometimes pronounced, sometimes not'); and the words in question can be called P-words, NP-words, and SP-words.

It would have been possible for us to draw up a matrix for every UE-word in the material, plotting Basic Line Length (BLL) against number of UE in the line, and entering in each square of the grid the number of examples for the word in question. This is perhaps a task for the future, but it is beyond the scope of the present study: the mere presentation of the matrices, without any interpretation, would occupy more space than the whole of this article. We have therefore adopted a simpler method, but one which we believe to be quite effective.

For each UE-word which we consider, let us set up the hypothesis that UE was always pronounced; we can then test this hypothesis by examining the line-lengths in which the word occurs. We begin from the fact that, in lines containing no examples of UE, the mean Basic Line Length is 10.119 syllables (Table 4 in Part I). We can say, therefore, that a particular line is of average length if it contains ten syllables (excluding line-final unstressed syllables from the count). Let us now, in each line containing UE, add the value of UE to the BLL, and see how many syllables this adds up to. Under our hypothesis that UE was pronounced, a line in which UE + BLL adds up to 10 is one of average length; if it adds up to more than 10, the line is longer than average; if to less than 10, it is shorter than average. Now, if a particular UE-word occurs almost exclusively in lines which are longer than average, and hardly ever occurs in average-length lines or short lines, this is strong evidence that our hypothesis is false, and that its UE was not pronounced, or at any rate was seldom pronounced. On the other hand, if a particular UE-word occurs almost exclusively in average-length and short lines, and very rarely in long lines, this is strong evidence that our hypothesis is correct, and that its UE was pronounced. It is necessary, of course, that the word should occur fairly frequently: if a word occurs only four or five times in the material, no reliable conclusions can be drawn from its distribution.

We have therefore classified all lines in the material according to their length, length being defined as $BLL + UE$. For convenience we can call this their Nominal Line Length (NLL). We have set up five categories of Nominal Line Length: (1) Very Long lines; (2) Long lines; (3) Average lines; (4) Ambiguous lines; (5) Short lines. The criteria for the various categories are as follows:

1. *Very long lines*
 $BLL > 11$
2. *Long lines*
 $BLL = 11$, or
 $BLL = 10$, $UE > 0$
- 3 *Average lines*
 $BLL + UE = 10$

4. *Ambiguous lines*

BLL = 7, UE > 3, *or*

BLL = 8, UE > 2, *or*

BLL = 9, UE > 1

5. *Short lines*

BLL + UE < 10

These criteria embrace every line in the data, including those in which UE does not occur. The reason for including lines without UE is that we are also interested in the 'suspicious' words recorded in Columns 7 and 8, words which lack UE but which we think may have fewer syllables than we have attributed to them (or very occasionally more syllables). We want these words too to be classified according to the Nominal Length of the lines they occur in, so that later we can modify the data as necessary.

The fourth category, Ambiguous lines, calls for explanation. A line with BLL = 9 and UE = 2 has a Nominal Line Length of 11, one more than our average length of 10; but if one of the UE-words in the line is responsible for its extra length, we do not know which of them it is; accordingly we register the line as ambiguous. At a later stage, when we have evidence about the words which frequently have UE unpronounced, we may be able to reclassify the line, and put one of its examples in the 'average' group and the other in the 'long' group.

It will be noticed that the criterion for Type 1 (Very Long lines) makes no reference to the number of UE in the line, but solely to the fact that the Basic Line Length is extremely high. The purpose of this category is to help us to identify those 'suspicious' words to which we have in the original analysis attributed too many syllables. For the categorisation of UE-words, Types 1 and 2 can in fact be amalgamated into a single 'long' category.

For each of these five categories of line-length, the computer has recorded all the words which occur in Columns 7 and 8, and it is the distribution of particular words among these categories which we must now examine. First, however, we need to know how different line-lengths are distributed for lines containing no examples of UE. This will give us a distribution-pattern against which to test the remaining lines. The material contains 9156 lines in which UE = 0, and their length-distribution is as follows:

Basic Line Length					
13	12	11	10	9	8
5	119	1260	7367	388	17
(0.1%)	(1.3%)	(13.8%)	(80.5%)	(4.2%)	(0.2%)

Basic Line Length, it will be remembered, is the number of syllables in the line, excluding all examples of E (word-final unstressed *-e*), and also excluding any unstressed syllables at the end of the line. It will be seen that, among lines containing no example of UE (unelided *-e*), over 80% have a BLL of 10 syllables. A sizeable minority contain 11 syllables, while there are very few lines of other lengths.

Lines containing one P-word (a word in which the UE was normally pronounced) could be expected to show a pattern of Nominal Line Lengths (BLL + UE) very similar to this. Indeed, lines containing more than one P-word could be expected to show this same distribution-pattern, provided that there were no NP-words also present.

But for lines containing one NP-word (a word in which UE was *not* normally pronounced), the figures would be displaced one column to the left, since in the analysis a syllable had been attributed to the line which should not have been. This would give a pattern like the following:

14	13	12	11	10	9
5	119	1260	7367	388	17
(0.1%)	(1.3%)	(13.8%)	(80.5%)	(4.2%)	(0.2%)

The dominant length is now 11 syllables, followed by 12, and the rest very small. This is the distribution to be expected for lines with one example of UE which is not pronounced. And lines containing only one example of UE, it is to be remembered, constitute over 79% of all UE-lines.²

But what about lines with more than one example of UE? Of the 6786 lines containing UE, about 18% (1218) contain two examples. It will be remembered that

The Versification of The Canterbury Tales

the evidence of Part I (p. 101) indicates that almost exactly half the examples of UE in the material were pronounced. So of lines containing two examples of UE, one quarter could be expected to contain two P-words, one quarter could be expected to contain two NP-words, and one half could be expected to contain one P-word and one NP-word. Lines containing two examples of P-words are no problem: they will give the same distribution as lines with one P-word. In lines containing two NP-words, the distribution-pattern will be moved yet one more column to the left, giving an even greater preponderance of long lines; this causes no serious problem of interpretation, and in any case the number of such lines must be small (roughly 4.5% of all UE-lines). As for lines containing one P-word and one NP-word, they will show the same distribution-pattern as lines with just one NP-word; this means that the statistics for the P-word in the line will be slightly distorted, since they will be moved a little towards the Long end of the spectrum. Most of these lines, however, can be expected to fall into the Ambiguous (Type 4) category, and the distortion can be minimised by leaving Type 4 lines out of account. Similar considerations apply to lines which contain either three or four examples of UE; but such lines are in fact rare, accounting for only about 2.5% of all UE-lines. Any distorting effect is therefore likely to be extremely small.

These considerations do suggest, however, that we should exclude Type 4 (Ambiguous) lines from the statistics for the time being, and also that we should use just three categories of Nominal Line Length – Long, Average, and Short. This latter simplification is in any case desirable because of the particular criteria we have used for Types 1 and 2. Accordingly, we simplify the typical patterns given above by amalgamating all line-lengths greater than 10 (Long), and all line-lengths less than 10 (Short), leaving 10 alone to form the Average category. This produces the following type-patterns:

Distribution of line-lengths for UE = 0

	BLL	
Long	Average	Short
1384	7367	405
(15.1%)	(80.5%)	(4.4%)

This type of distribution could also be expected to occur for the Nominal Line

Length of lines containing examples of UE which were pronounced. For Lines in which an example of UE was *not* pronounced, however, the pattern would be as follows:

Expected pattern for lines in which UE not pronounced		
	NLL	
Long	Average	Short
8571 (95.6%)	388 (4.2%)	17 (0.2%)

So, when we examine the distribution of particular UE-words, those are the kinds of pattern we shall be looking for: 15%-81%-4% for words where UE was pronounced, but 96%-4%-0% for words where UE was not pronounced. We have to bear in mind, however, that the pattern for P-words may be slightly distorted towards the Long end by the effects we have discussed above. SP-words (ones where UE was sometimes pronounced and sometimes not) could be expected to provide a pattern somewhere between the two; a word in which UE was pronounced on half its occurrences could be expected to produce a pattern like 55%-42%-2%.

When we examined the computer print-out showing the distribution of all the UE-words in the material, we were slightly disappointed at the number of words which occur only two or three times in the entire corpus, and indeed at the considerable number which occur only once; about such words, obviously, we can come to no conclusions. This situation was aggravated by the degree of detail in the grammatical categorisation of the words: so for example an adjective like *deere* appears in the lists not as one word but as three: as a strong singular adjective, a weak singular adjective, and a plural adjective.³ This breakdown, however, is essential, since it is quite likely that UE would be pronounced in one or two of these categories but not in the other(s). Nevertheless, there were quite a number of words which occurred reasonably frequently: there are 11 words which occur a hundred times or more, a further 15 words which occur fifty times or more, a further 50 words which occur twenty times or more, and over 70 further words which occur ten times or more. Let us begin with the words which occur 100 times or more: *hadde* (past tense singular), *hire* (personal pronoun), *neuere* (adverb), *oure* (pronoun-determiner singular), *seyde* (past tense singular), *sholde* (past tense

The Versification of The Canterbury Tales

singular), *thanne* 'then' (adverb), *thise* (determiner plural), *we(e)re* 'were' (past tense singular), *wo(o)lde* (past tense singular), and *youre* (pronoun-determiner singular). Their distribution among different types of line is as follows:⁴

Word	No. of Examples	Type 1	Type 2	Type 3	Type 4	Type 5
HADDE vb pt sg	212	0 (0%)	137 (64.6%)	35 (16.5%)	38 (17.9%)	2 (0.9%)
HIRE pronoun	155	1 (0.6%)	134 (86.5%)	3 (1.9%)	17 (10.9%)	0 (0%)
NEUERE adverb	116	1 (0.7%)	94 (81.0%)	2 (1.7%)	19 (16.4%)	0 (0%)
OURE det sg	166	1 (0.6%)	133 (80.1%)	6 (3.6%)	26 (15.7%)	0 (0%)
SEYDE vb pt sg	136	1 (0.7%)	37 (27.2%)	64 (47.1%)	32 (23.5%)	2 (1.5%)
SHOLDE vb pt sg	105	0 (0%)	37 (35.2%)	43 (41.0%)	22 (20.9%)	3 (2.9%)
THANNE adverb	137	0 (0%)	99 (72.3%)	11 (8.0%)	27 (19.7%)	0 (0%)
THISE det pl	110	1 (0.9%)	82 (74.5%)	5 (4.5%)	21 (19.1%)	1 (0.9%)
WE(E)RE vb pt sg	141	0 (0%)	104 (73.8%)	10 (7.1%)	27 (19.1%)	0 (0%)

WO(O)LDE vb pt sg	127	0 (0%)	51 (40.2%)	51 (40.2%)	24 (18.9%)	1 (0.8%)
YOURE pron-det sg	278	0 (0%)	210 (75.5%)	5 (1.8%)	63 (22.7%)	0 (0%)

It will be seen that the Ambiguous lines (Type 4) range from 10.9% to 23.5%. If these lines could be reclassified, they would almost all become either Type 2 or Type 3; it would be impossible for them to become Type 1, and very few would become Type 5 (because of the methods of syllable-counting, described in Part I). For the moment, however, let us disregard Type 4, and regroup the statistics under the headings of Long (Type 1 + Type 2), Average (Type 3), and Short (Type 5). Let us also disregard Type 4 in calculating percentages, so that 100% is Type 1 + Type 2 + Type 3 + Type 5. The percentages for these eleven words then come out (to the nearest whole number) as follows:

Word	Long	Average	Short	Category
HADDE vb pt sg	79%	20%	1%	SP
HIRE pronoun	98%	2%	0%	NP
NEUERE adverb	98%	2%	0%	NP*
OURE pron-det sg	96%	4%	0%	NP
SEYDE vb pt sg	37%	62%	2%	SP
SHOLDE vb pt sg	45%	52%	4%	SP
THANNE adverb	90%	10%	0%	NP
THISE det pl	93%	6%	1%	NP

The Versification of The Canterbury Tales

WE(E)RE vb pt sg	91%	9%	0%	NP
WO(O)LDE vb pt sg	50%	50%	1%	SP
YOURE pron-det sg	98%	2%	0%	NP

These figures are to be compared with the expected distribution for words in which UE was pronounced (15-81-4) and for words in which UE was not pronounced (96-4-0). It is immediately apparent that some of these words can with little doubt be classed as ones in which UE was not pronounced, or at any rate rarely pronounced: *hire, oure, thanne, thise, were*, and *youre*; accordingly, they are classified in the final column as NP.⁵

It might be thought that *neuere* was also an obvious candidate for the NP category. It does indeed show exactly the pattern required, but there could be another explanation for this: the pattern would also be produced if the UE of *neuere* was pronounced, but its middle syllable was not, i.e. if it was pronounced /nɛ:vrə/ (which is historically plausible). That this was in fact the case is supported by the pattern found for *neuer-* (i.e. the same word with elided E), and indeed also by that for *neuer* without any E: for both forms the distribution suggests strongly that *neuer* was monosyllabic. This conclusion is indicated in the final column by the use of an asterisk, the line being categorised as 'NP*'.⁶

Of the remaining words in the list, there are three in which it is plausible to believe that UE was frequently pronounced: *seyde, sholde*, and *wolde*; accordingly, they are classified as SP (UE sometimes pronounced, sometimes not). The one remaining word, *hadde*, probably also falls into the SP category, though not as decisively.

So for this first group of words, our conclusion is that UE was not pronounced in *hire, oure, thanne, thise, were*, and *youre*; that it was sometimes pronounced in *hadde, seyde, sholde*, and *wolde*; and that it was probably pronounced in *neuere*, but that this was a dissyllabic word. It will be noticed that none of the words in this first group falls into the P category. This is not really surprising: words of very frequent occurrence, and especially ones which often occur in unstressed position, would be particularly likely to lose the final *-e*.

The second group of words, those which occur fifty times or more, are the following: *alle* adjective plural (93 occurrences), *bothe* co-ordinating conjunction

(55), *euere* adverb (92), *haue* verb infinitive (62), *haue* 1st person singular present tense (96), *herte* substantive (88), *hise* pronoun-determiner plural (64), *koude* verb past tense singular (55), *loue* substantive (51), *myghte* verb past tense singular (69), *owene* adjective weak singular (64), *sire* substantive (60), *thilke* adjective weak singular (74), *tyme* substantive (76), and *we(e)re* verb past tense plural (81). With these words, let us go straight to the Long-Average-Short categorisation:

Word	Long	Average	Short	Category
ALLE adj pl	56%	41%	2%	SP
BOTHE co-ord conj	55%	43%	2%	SP
EUERE adverb	97%	3%	0%	NP*
HAUE verb infin	96%	4%	0%	NP
HAUE 1st sg pres	94%	6%	0%	NP
HERTE sb	15%	78%	4%	P
HISE pron-det pl	98%	2%	0%	NP
KOUDE vb pt sg	33%	67%	0%	SP
LOUE sb	93%	7%	0%	NP
MYGHTE vb pt sg	36%	60%	4%	SP
OWENE adj wk sg	98%	2%	0%	NP*
SIRE sb	93%	7%	0%	NP
THILKE adj wk sg	14%	84%	2%	P

The Versification of The Canterbury Tales

TYME sb	51%	49%	0%	SP
WE(E)RE vb pt pl	87%	13%	0%	NP

The 96-4-0 type of distribution is clearly seen in *haue* infin, *haue* 1st sg pres, *hise* det pl, *loue* sb, and *sire* sb, and less clearly in *we(e)re* pt pl, so these six words are classified as NP. This type of distribution is also seen in *euere* adv and *owene* adj wk sg, but in these two words (as in *neuere*) it is possible that the middle syllable was not pronounced, and that the final *-e* was. The NP type of distribution is found in *sire* both when it acts as an adjunct to another noun (as in 'sire knight') and when it is the head-word of the noun phrase. Two words, *herte* sb and *thilke* adj wk sg, conform well to the 15-81-4 type of distribution, and can clearly be classified as P (i.e. final *-e* normally pronounced). The remaining words, *alle* pl, *bothe* conj (in constructions of the type 'both . . . and . . .'), *koude* pt sg, *myghte* pt sg, and *tyme* sb, fall into the intermediate category SP (final *-e* sometimes pronounced, sometimes not).

The words occurring 20-49 times are the following. In the second column, the first figure is total number of occurrences, while the figure in brackets is this figure minus the number of Type 4 occurrences.

Word	Number of examples	Long	Average	Short	Cat.
ALLE adj str sg	21 (18)	11%	89%	0%	P
ANSWERDE vb pt sg	23 (19)	16%	74%	11%	P
BOTHE adj pl	25 (18)	28%	67%	6%	SP?
CAUSE sb	22 (17)	18%	82%	0%	P
DAME sb	28 (25)	88%	12%	0%	NP
DEERE adj wk sg	37 (25)	8%	92%	0%	P

FAIRE adj wk sg	23 (21)	14%	86%	0%	P
FALSE adj wk sg	30 (29)	3%	97%	0%	P
FELawe sb	27 (24)	96%	4%	0%	NP*
FIRSTE adj wk sg	28 (21)	10%	76%	14%	P
GRETE adj wk sg	33 (26)	12%	85%	4%	P
HADDE vb pt pl	21 (18)	83%	17%	0%	SP?
HAUE vb pres pl	24 (21)	95%	5%	0%	NP
HE(E)RE 'here' adv	38 (32)	100%	0%	0%	NP
HEUENE 'heaven' sb	23 (18)	100%	0%	0%	NP*
HIGHTE vb pt sg	27 (22)	27%	64%	9%	P?
HIRE 'her' pron-det sg	46 (38)	95%	5%	0%	NP
HOUSBONDE sb	25 (19)	95%	5%	0%	NP
MADE vb pt sg	28 (25)	44%	56%	0%	SP
MANERE sb	27 (23)	96%	4%	0%	NP*
MILLERE sb	20 (19)	100%	0%	0%	NP*
MOORE adj str sg	31 (29)	72%	24%	3%	SP
MOORE adv	42 (35)	66%	34%	0%	SP
MOSTE vb pt sg	21 (16)	38%	63%	0%	SP

The Versification of The Canterbury Tales

NAMOORE sb	30 (28)	93%	7%	0%	NP
NOBLE adj wk sg	29 (24)	13%	83%	4%	P
OFTE adv	22 (19)	11%	84%	5%	P
OLDE adj wk sg	42 (33)	21%	79%	0%	P
OLDE adj pl	32 (27)	15%	78%	7%	P
OTHERE adj pl	48 (43)	93%	5%	2%	NP*
OURE pron-det pl	34 (28)	93%	7%	0%	NP
PEPLE sb	36 (30)	17%	80%	3%	P
POORE, POURE adj str sg	22 (22)	27%	59%	14%	P?
SAME adj wk sg	28 (23)	9%	91%	0%	P
SAUE prep	22 (16)	94%	6%	0%	NP
SEYE vb infin	31 (26)	92%	8%	0%	NP
SEYE 1st sg pres	26 (23)	91%	9%	0%	NP
SHOLDE vb pt pl	30 (25)	52%	48%	0%	SP
SOMME 'some' adj pl	31 (25)	84%	12%	4%	SP?
SONE sb	30 (23)	83%	17%	0%	SP?
SOULE sb	29 (23)	17%	83%	0%	P
SWICHE adj pl	26 (21)	81%	19%	0%	SP

TALE sb	41 (31)	13%	87%	0%	P
TELLE vb infin	34 (31)	26%	74%	0%	SP?
THEREFORE adv	43 (38)	97%	3%	0%	NP
THO(U)GHTE vb pt sg	23 (19)	32%	68%	0%	SP
TROUTHE sb	22 (14)	43%	57%	0%	SP
WHICHE rel pl	41 (34)	88%	9%	3%	NP?
YONGE adj wk sg	20 (17)	6%	88%	6%	P
YOURE pron-det pl	29 (27)	96%	4%	0%	NP

The words marked as adjectives include cases where the word is used absolutely, and in the case of *somme* 'some' a large number are of this type, and the word might well be classified instead as a pronoun. It will be noticed that weak singular adjectives regularly fall into the P category, while verb past tenses are usually SP.

Words which occur in our dictionary fewer than twenty times obviously provide less reliable material than the words we have considered so far. Before we examine these less frequent words, therefore, we shall modify the original data on the disks by feeding back some of the conclusions we have so far arrived at. This will result in the reclassification of some Type 4 (Ambiguous) lines as Type 2 or 3 or 5. The following types of change will be made:

(1) *Words in which we have strong grounds for believing that UE was NOT pronounced.* On every occurrence of such a word, the line in question is modified by the reduction of Column 4 (i.e. number of UE in the line) by one. The word is then deleted from Column 7 or Column 8. The words in question are the following: *dame* sb, *euere* adv, *felawe* sb, *haue* vb infin, *haue* vb 1st sg pres, *haue* vb pres pl, *here* adv, *heuene* sb, *hire* pron, *hire* pron-det sg, *hise* pron-det pl, *loue* sb, *manere* sb, *nere* (= 'ne were') vb pt sg and pl, *namoore* sb, *neuere* adv, *othere* adj pl, *oure* pron-det sg and pl, *owene* adj wk sg, *seye* vb infin, *seye* vb 1st sg pres, *sire* sb,

thanne adv, *therefore* adv, *thise* det pl, *were* vb pt sg and pl, *youre* pron-det sg and pl. For simplicity, a single spelling is given for each word, but of course in the computer-programmes all spelling-variants which have been recorded must be included. It will be seen that the list includes all clear examples of the NP category which occur in the material fifty times or more. But words which occur 20-49 times are included only if there are at least 20 occurrences excluding those in Type 4 lines. (An exception to the 20-occurrences rule is made in favour of *heuene* sb, not only because all 18 occurrences are in Type 2 lines, but also because its statistics can reasonably be amalgamated with those of *heuen-*, treated below.) The list also includes words classified as NP*, like *owene* adj wk sg; it is possible that in this word the final *-e* was pronounced, while the medial syllable was not, but this has precisely the same effect on the Nominal Line Length as a normal NP word, so the same procedure is followed.

(2) *Words in which we have strong grounds for believing that UE WAS pronounced.* On every occurrence of such a word, the line in question is modified by the reduction of Column 4 (number of UE) by one, and the increase of Column 2 (Basic Line Length) by one. The word in question is then deleted from Column 7 or Column 8. The words in question are as follows: *deere* adj wk sg, *faire* adj wk sg, *false* adj wk sg, *firste* adj wk sg, *grete* adj wk sg, *herte* sb, *noble* adj wk sg, *olde* adj wk sg, *olde* adj pl, *peple* sb, *same* adj wk sg, *soule* sb, *tale* sb, *thilke* adj wk sg. It will be noticed that these P words are predominantly adjectives (mostly weak singular), with a few substantives. The number of examples used as a criterion for inclusion in the list is the same as for the NP-words in (1) above.

At the same time, it is obviously desirable to feed back information about the 'suspicious' words recorded in Columns 7 and 8 (which we can call 'S-words'); so let us consider the S-words which occur frequently. Our information about such words is, however, less complete than about the UE-words: the latter are recorded exhaustively in Columns 7 and 8, but we cannot be sure that we have recorded all the examples of the S-words which occur in the text (since, for example, we did not begin to record some of them until we were part-way through the analysis). To remedy this deficiency, we have consulted the Tatlock and Kennedy concordance,⁶ and inserted the missing words into our material wherever possible. Unfortunately, however, for words which occur very frequently, the concordance gives only a few specimen examples, and some such words figure among our S-words: *any*, *ellis*, *euerich*, *euery*, *many*, *neuer*. For some of these our own material is extremely

substantial, and moreover was collected right from the beginning, and we include them with confidence despite the absence of support from the concordance; but for others our material is too small for any safe conclusions to be drawn.

The following are the S-words which were recorded twenty times or more, with a break-down into the line-types in which they occurred:

Word	Number	Type 1	Type 2	Type 3	Type 4	Type 5
CREATUR-, CREATURES sb	46	2	4	16	3	21
EUER- adv	72	7	63	2	0	0
EUERICH adj	20	2	18	0	0	0
EUERICHO(O)N, EUERYCHON pron	26	4	22	0	0	0
EUERY adj	212	29	182	0	1	0
HEUEN- sb	18	1	16	1	0	0
IANUARI-	25	4	20	1	0	0
IANUARIE + consonant	8	0	1	7	0	0
IANUARIE line-final	5	0	1	4	0	0
LOUED + vowel, LOUED- vb pt, pp	25	4	20	1	0	0
LOUED + consonant vb pt, pp	14	0	4	10	0	0
MANY A adj + indef art	128	19	106	0	3	0

The Versification of The Canterbury Tales

NAMELY adv	22	1	16	5	0	0
NEUER adv	23	2	21	0	0	0
NEUER- adv	34	4	30	0	0	0
OUER + vowel prep, adv	33	4	23	5	0	1
OUER + consonant	18	0	5	11	1	1
SOUEREYN, -AYN sb	20	0	19	1	0	0
(Y-)CLEPYD, -ID, -ED pp	30	0	15	14	0	1

A hyphen at the end of a word represents an elided E; so for example the transcription CREATUR- means that in the text the word occurred as *creature*, but that the final -e was followed by an initial vowel-phoneme in the next word. Some words behave differently before a vowel and before a consonant, and the statistics are therefore separated. The form *heuen-* is included even though it occurs fewer than 20 times, since its statistics can reasonably be amalgamated with those of *heuene*, with UE.

It will be seen that the S-words are of three types: those where we believe we have attributed too many syllables to the word (e.g. *every*); those where we believe we have attributed too few (e.g. *creatur-*, *creatures*); and those where the evidence is insufficient (e.g. *namely*), or where it suggests that the word varied in its number of syllables (e.g. *yclepyd*). There are thus two further types of modification to be made to the data-disks:

(3) *S-words which we have strong grounds for believing to have had one syllable fewer than attributed.* On every occurrence of such a word, Column 2 (Basic Line Length) is reduced by one, and the word then deleted from Column 7 or Column 8. The words in question are the following: *euer + vowel*, *euer-*, *eueremo*, *eueremoor*, *euerich*, *euerichon*, *euery*, *euerydel*, *heuen-*, *Ianuari-*, *loued-*, *loued + vowel*, *many a*, *neuer*, *neuer-*, *neueremo*, *ouer + vowel*. It will be seen that to the

list have been added two words which do not occur in the table above: *euerydel* and *euer* + vowel. *Euerydel* occurred only fifteen times, but every single occurrence was in a Type 2 line; in any case, the word can obviously be expected to behave in the same way as the simplex *euery*. The word *euer* + vowel occurred only twice, but was included as a subset of *euer-* (the adverb *euere* with final *-e* elided before a vowel).

The word *Ianuarie*, which occurs only in 'The Merchant's Tale', is a curious case. In the original analysis, the *ua* had been treated as two syllables, while the final *-ie* had been taken to represent a single syllable /i:/ (i.e. with no final E). It looks as though both these decisions were mistaken. The statistics suggest that, when the word occurred before a vowel, it had only three syllables (perhaps being pronounced /dʒanwarɪ:/?); but before a consonant it clearly had four syllables, so presumably the final *-ie* represented two syllables, namely /i:/ plus UE (unelided final *-e*). In line-final position the word occurs only five times; it rhymes with *tarye* (three times), *contrarie*, and *Marie*, and in the original analysis the rhyming syllable was taken to be *-ar-*; but the statistics would make sense if, instead, the rhyming syllables were *-i-*.

The word *euery* occurs over 200 times, and the expression *many a* occurs well over 100; these are among the words for which Tatlock and Kennedy give only specimen examples; particularly significant are the large number of occurrences in Type 1 lines, a strong indication of 'overlong' words. It is possible that *many* behaved similarly before all other unstressed vowels as well as the indefinite article, but we collected insufficient evidence on this.

(4) *S-words which we have strong grounds for believing to have had one more syllable than attributed.* On every occurrence of such a word, Column 2 (Basic Line Length) is increased by one, and the word deleted from Column 7 or 8. The words in question are *creatur-* and *creatures*. In these the digraph *ea* had been interpreted as representing one syllable, but clearly it should have been interpreted as representing two. To these two words has been added *creature* (with unelided *-e*), to which the same consideration obviously applies; but in the case of *creature*, the word is not deleted from Column 7 or Column 8, since we still wish to gain information about its final *-e*.

A fifth type of modification was made in respect of one particular word, *benedicite(e)*. This occurs in the material only thirteen times, but six of the

occurrences are in Type 1 lines, and six in Type 2. Moreover, almost all the examples occur in lines of such extraordinary length that we are led to believe that for Chaucer it usually had only three syllables, as against the five which we had attributed to it. In every line in which the word occurs, therefore, Column 2 has been reduced by *two*, and the word deleted from Column 7 or Column 8. It is to be noted that among the fourteenth-century spellings recorded by the OED is one which suggests only *two* syllables, namely *benste*.

These changes were carried out and a revised set of data-disks thus produced, from which a revised dictionary was compiled. These we can call Version 2 of the disks and of the dictionary. From the Version 2 disks it is necessary to recalculate the expected patterns of line-lengths for P-words and NP-words. In the Version 2 data-disks there are 10,942 lines in which UE = 0 (compared with 9156 on the original disks). The length-distribution of these 10,942 lines produces the following patterns: for lines containing no examples of an NP-word, Long 11.14%, Average 84.29%, Short 4.57%; for lines containing one example of an NP-word, Long 95.43%, Average 4.38%, and Short 0.19%. So the two main patterns we shall be looking for are 11%-84%-5% and 95%-4%-0%. As before, Ambiguous (Type 4) lines will be left out of account.⁷

The Version 2 dictionary differs only minimally from the first dictionary in its allocation of those remaining words which occur twenty times or more. A few 20+ words had been left on the data-disks because the number of examples, after the subtraction of Type 4 examples, fell below twenty. In the new dictionary, some of these now get up above the twenty mark: *answerde* vb pt sg (22), *bothe* adj pl (24), *cause* sb (22), and *housbonde* sb (20). In each case, however, the classification of the word remains unaltered, except that for *bothe* the 'SP?' is changed to 'P'. Three words simply have the question-mark removed: *hadde* vb pt pl (now 19 examples), *sone* sb (now 27), and *telle* vb infin (now 33) are all reclassified as 'SP'.

In the new dictionary, the further words which occur at least ten times (excluding Type 4) are the following: *aboute* prep (14) 71%-29%-0%, SP; *Arcite* sb (14) 36%-57%-7%, SP; *brighte* adj wk sg (10) 40%-50%-10%, SP; *chambre* sb (16) 6%-94%-0%, P; *come* vb infin (10) 80%-20%-0%, SP; *come* past part (12) 92%-8%-0%, NP; *Custaunce* sb (14) 64%-34%-0%, SP; *dide* vb pt sg (15) 80%-20%-0%, SP; *dorste* vb pt sg (16) 19%-81%-0%, P; *face* sb (11) 18%-82%-0%, P; *fare (wel)* vb imp sg (10) 100%-0%-0%, NP; *faste* adv (13) 8%-92%-0%, P; *feste* sb (11) 18%-82%-0%, P; *fortune* sb (15) 47%-53%-0%, SP; *frere* sb (13) 23%-

77%-0%, SP?; *fresshe* adj wk sg (19) 11%-89%-0%, P; *goode* adj wk sg (19) 16%-79%-5%, P; *goode* adj pl (13) 46%-54%-0%, SP; *haue* vb imp pl (11) 100%-0%-0%, NP; *heighe* adj wk sg (15) 20%-80%-0%, P?; *herde* vb pt sg (10) 10%-90%-0%, P; *humble* adj str sg (11) 0%-100%-0%, P; *hye* adj wk sg (12) 25%-75%-0%, SP; *ilke* adj wk sg (15) 13%-87%-0%, P; *ioye* sb (12) 50%-50%-0%, SP; *laste* adj wk sg (15) 40%-53%-7%, SP; *leeue* adj wk sg (17) 12%-82%-6%, P; *lettre* sb (12) 8%-92%-0%, P; *longe* adj wk sg (14) 25%-75%-0%, SP; *longe* adv (14) 7%-87%-7%, P; *madame* sb (10) 100%-0%-0%, NP; *make* vb infin (16) 38%-63%-0%, SP; *mayde* sb (18) 11%-83%-6%, P; *morwe* sb (13) 8%-92%-0%, P; *murye* adj str sg (12) 100%-0%-0%, *NP; *myghte* vb pt pl (16) 25%-69%-6%, SP; *namoore* adv (13) 54%-46%-0%, SP; *nekke* sb (10) 0%-90%-10%, P; *nexte* adj wk sg (10) 10%-90%-0%, P; *noble* adj str sg (12) 0%-100%-0%, P; *nolde* vb pt sg (15) 53%-47%-0%, SP; *ofte* adj pl (10) 10%-70%-20%, P; *place* sb (12) 25%-75%-0%, SP; *Rome* sb (15) 20%-73%-7%, P?; *sonne* sb (16) 13%-75%-13%, P; *so(o)re* adv (10) 20%-80%-0%, P?; *sorwe* sb (15) 7%-93%-0%, P; *swe(e)te*, *swote*, adj wk sg (18) 17%-78%-6%, P; *temple* sb (12) 0%-100%-0%, P; *tonge* sb (12) 33%-67%-0%, SP; *trewē* adj str sg (14) 21%-79%-0%, P?; *trowe* vb 1st sg pres (15) 67%-33%-0%, SP; *wende* vb pt sg (10) 70%-30%-0%, SP; *wise* adj wk sg (10) 10%-90%-0%, P; *wiste*, *nyste* vb pt sg (20) 10%-90%-0%, P; *withoutē* prep (10) 0%-90%-10%, P; *wolde* vb pt pl (13) 69%-31%-0%, SP; *yonge* adj pl (13) 23%-77%-0%, SP?.

The figures in brackets show the number of examples found, excluding those in Type 4 lines. As the numbers get down towards ten, the findings obviously become less reliable, since a change in the classification of a small number of lines can produce a large change in the percentage-pattern. Even with only ten examples, however, we can put considerable reliance on the findings when the figures are extreme, as with *fare (wel)* (100% Long) and *withoutē* (0% Long). Indeed, with patterns of this kind we can perhaps have some confidence in words which occur fewer than ten times: the verb infinitive *haue* occurs only nine times, but every single example is in a Long line; and the weak singular adjective *hende* occurs only eight times, but every single example is in an Average-length line.

From the Version 2 disks we also produced a different kind of dictionary, one which recorded line-lengths, not for individual words, but for grammatical categories. In this way we hoped to get some information about the many words which occur in the material only a few times, and which cannot therefore be handled individually. In examining the material, we shall again be looking for the patterns

11%-84%-5% (P) and 95%-4%-0% (NP); a category which was pronounced on exactly half of its occurrences could be expected to give a pattern 53%-44%-2%. In the analysis, as before, examples are disregarded when they occur in Type 4 lines.

Of the 1370 nouns recorded, 764 occurred as subject, object, or complement, and these gave the pattern 43%-56%-1%, SP. A further 543 occurred after prepositions or in positions where they might have been in the dative case, and these gave the pattern 45%-54%-1%, SP. It will be seen that there is no difference between these two groups of nouns, indicating that (as could be expected) no relic of the old dative case remained. Curiously, however, the 63 nouns which occur as attributes (before other nouns) do show a different pattern: 29%-68%-3%. This still puts them in the SP category, but the figures suggest that the UE was pronounced much more often than in other nouns. One possibility is that, when in attributive position, nouns were influenced by the analogy of adjectives; indeed, with some words (e.g. *seinte*) it can be difficult to know whether an example should be classified as an attributive noun or as an adjective.

There were 445 weak singular adjectives, with the pattern 12%-84%-3%, P. It is clear, therefore, that the adjectival inflection for the weak singular was still regularly used. It might have been expected, therefore, that the plural inflection was also regularly used, but the figures do not bear this out: the 356 plural adjectives have the pattern 38%-59%-3%; this does indeed suggest that the UE was pronounced more often than not, but still falls into the SP category. The 292 strong singular adjectives, similarly, give the pattern 29%-69%-2%, SP. When adjectives are used absolutely, they are less likely to have the UE pronounced: the 33 examples give the pattern 64%-30%-2%, SP.

Among verbs, the 382 infinitives give the pattern 40%-60%-0%, SP; the 152 examples of 1st sg pres 51%-48%-1%, SP; the 16 examples of 2nd sg pres 69%-31%-0%, SP; the 88 examples of 3rd sg pres 45%-53%-1%, SP; the 105 examples of pres pl 52%-46%-2%, SP; the 73 examples of imp sg 67%-32%-1%, SP; the 49 examples of imp pl 80%-20%-0%, SP; the 920 examples of pt sg 45%-53%-2%, SP; the 115 examples of pt pl 51%-47%-2%, SP; the 43 examples of past participle 65%-35%-0%, SP; and the 30 examples of present participle 83%-13%-3%, SP?. Most of these are unremarkable and not very helpful, being close to the 50-50 pattern of 53%-44%-2%. Small divergences from this central pattern can be seen in the 2nd sg pres (where all the examples will obviously be subjunctives), the imperative singular and plural, and the past and present participles; the bias in each case is towards the NP end of the scale.

The 211 adverbs give 36%-61%-3%; this is SP, but clearly leans towards the P end of the scale, which is not unexpected in view of the common historical use of *-e* as an adverbial ending. On the other hand, the 62 prepositions lean the other way, with 66%-32%-2%, which is SP biased towards the NP end. The 13 subordinating conjunctions, such as *where*, *wherfore*, and *wherinne*, are quite decisively NP, with 92%-8%-0%. The relative plurals and the co-ordinating conjunctions were each represented by only one form, *bothe* and *whiche*. For interrogatives and for relative singular, there were so few examples that no conclusions could be drawn.

We next went back to the Version 1 disks, and ran a different kind of programme, as a check on our findings. In this we compiled a new dictionary, which we call 'the UE1 dictionary', based solely on those lines in which Column 4 has a value of 1, in other words the lines which contain just one example of unelided E. This has two advantages: first, we avoid the possible distortions caused by some of the lines containing more than one example of UE; and secondly, we eliminate all Type 4 (Ambiguous) lines, since by definition these cannot occur in the 'UE1' lines. The disadvantage of the method is that it uses the evidence of fewer lines than the previous dictionaries; but in fact about 64% of all examples of UE (5389 out of 8375) occur in lines containing only one example, so the numbers are still reasonably large.

In the UE1 dictionary, we did not use the Type 1 to Type 5 style of line classification, but instead recorded the Nominal Line Length (Column 2 plus Column 4) for each example. So, for example, the dictionary records that the word *myghte* vb pt sg occurs twice with NLL 9, 30 times with NLL 10, 7 times with NLL 11, and twice with NLL 12. Such figures are easily converted into the Long-Average-Short mode, in this particular case giving 9 Long, 30 Average, and 2 Short (22%-73%-5%).

When we examined the UE-words which had occurred twenty times or more in the Version 1 dictionary, we found that the UE1 dictionary gave results which were very similar. For only 13 words were the original classifications changed, and in eight of the thirteen cases the change was a marginal borderline one, involving merely the addition or removal of a question mark. These eight were the following: *hadde* vb pt pl (15) 80%-20%-0%, SP (formerly SP?); *highte* vb pt sg (18) 22%-67%-11%, P (formerly P?); *koude* vb pt sg (34) 26%-74%-0%, SP (formerly SP?); *poore*, *poure* adj str sg (18) 22%-61%-17%, P (formerly P?); *somme* adj pl (13) 85%-8%-8%, NP? (formerly SP?); *telle* vb infin (26) 31%-69%-0%, SP (formerly

SP?); *thanne* adv (71) 86%-14%-0%, NP? (formerly NP); *whiche* rel pl (26) 85%-12%-4%, NP? formerly NP). In the case of *hadde* and *telle*, the revised classification is identical with that derived from the Version 2 dictionary.

The five words in which there was a more substantial reclassification were the following: *bothe* adj pl (14) 21%-71%-7%, P (formerly SP?); *dame* sb (19) 84%-16%-0%, SP? (formerly NP); *moste* vb pt sg (12) 25%-75%-0%, P? (formerly SP); *myghte* vb pt sg (41) 22%-73%-5%, P (formerly SP); *were* vb pt pl (45) 82%-18%-0%, SP (formerly NP). In the case of *bothe*, the revised classification is identical with that derived from the Version 2 dictionary.

With only two exceptions, the changes to these thirteen words are all in the same direction: away from the NP end of the spectrum and towards the P end. The two exceptions are *somme* and *telle*, and in the case of the former word the change is minimal (from 84% Long to 85% Long). This favoured direction of change is to be expected: in Part I of this study, we found that the probability of UE being pronounced declined as the number of UE words in the line increased. Since the UE1 dictionary takes into account only those lines with a single occurrence of UE, it follows that the results should be shifted towards the P end of the spectrum compared with our findings from the Version 1 dictionary, in which all UE-lines were considered. The conclusion to be drawn is that such words as *thanne* do not normally have UE pronounced, but that it may occasionally be pronounced if there is no other UE-word in the line; and that in such words as *myghte* the UE is usually pronounced if there is no other UE word in the line, but otherwise usage fluctuates.

In the case of words which in the Version 1 dictionary had occurred fewer than twenty times, the evidence is obviously less reliable. Of the 58 words which occurred at least ten times, 42 achieved the same classification in the UE1 dictionary; seven words had a marginally different classification; and nine showed a more substantial change. Once again, the majority of changes (10 out of 16) were toward the P end of the spectrum. In general, the UE1 dictionary suggests that the findings of the Version 1 and Version 2 dictionaries are in the main reliable, and that any distortions produced by the methods used are small.

At this point, we decided against any further attempts to refine the data. We could have carried out yet another revision of the data-disks, and produced a Version 3 dictionary, but the numbers of examples involved would have been small, and the additions to our findings probably minimal. Our material does indeed suggest that there are a number of lines of enquiry which could profitably be followed up, but our very modest computer-power puts them beyond our scope. If the whole text of

The Canterbury Tales, and indeed the whole of the Chaucer canon, were to be entered into a more powerful computer, it would be possible to produce a complete concordance, and also to pursue numerous linguistic points: for example, the extent to which the *-es* plural morpheme in nouns constituted a syllable, and whether there was any difference in this respect between monosyllabic, dissyllabic, and polysyllabic nouns; and whether, as our material hints, final unstressed *-y* tended to be non-syllabic when it occurred before an unstressed vowel.

Finally, we give a summary of our findings about the pronunciation of unelided word-final unstressed *-e* in various words and word-categories. For this purpose, we have divided the SP category into three: (a) the part of the spectrum abutting on the P category, where it seems that UE was pronounced more often than not; (b) the central part of the SP range, where we would expect UE to be pronounced on roughly half its occurrences; and (c) the part of the spectrum abutting on the NP category, where it seems that UE was most often not pronounced.

Summary

1. *Words in which UE was pronounced*

Weak singular adjectives.

Individual words: *alle* adj str sg; *answerde* vb pt sg; *bothe* adj pl; *cause* sb; *chambre* sb; *deere* adj wk sg; *dorste* vb pt sg; *face* sb; *faire* adj wk sg; *false* adj wk sg; *faste* adv; *feste* sb; *firste* adj wk sg; *fresshe* adj wk sg; *goode* adj wk sg; *grete* adj wk sg; *hende* adj wk sg; *herde* vb pt sg; *herte* sb; *humble* adj str sg; *ilke* adj wk sg; *leeue* adj wk sg; *lettre* sb; *longe* adv; *mayde* sb; *morwe* sb; *nekke* sb; *nexte* adj wk sg; *noble* adj str sg; *noble* adj wk sg; *ofte* adj pl; *ofte* adv; *olde* adj wk sg; *olde* adj pl; *peple* sb; *same* adj wk sg; *sonne* sb; *sorwe* sb; *soule* sb; *swe(e)te*, *swote* adj wk sg; *tale* sb; *temple* sb; *thilke* adj wk sg; *wise* adj wk sg; *wiste*, *nyste* vb pt sg; *withoute* prep; *yonge* adj wk sg.

2. *Words in which UE was sometimes pronounced, sometimes not*

(a) *Words in which UE was more often pronounced than not*

Plural adjectives; strong singular adjectives; adverbs; verb infinitives.

Individual words: *Arcite* sb; *brighte* adj wk sg; *frere* sb; *heighe* adj wk sg; *highte* vb pt sg; *hye* adj wk sg; *koude* vb pt sg; *laste* adj wk sg; *longe* adj wk sg;

made vb pt sg; *make* vb infin; *moste* vb pt sg; *myghte* vb pt sg; *myghte* vb pt pl; *poore*, *poure* adj str sg; *place* sb; *Rome* sb; *seyde* vb pt sg; *sholde* vb pt sg; *so(o)re* adv; *telle* vb infin; *tho(u)ghte* vb pt sg; *tonge* sb; *trewe* adj str sg; *trouthe* sb; *yonge* adj pl.

(b) *Words in which UE was pronounced on about half its occurrences*

Nouns; verbs – 1st sg present, 3rd sg present, present plural, preterite singular, preterite plural.

Individual words: *alle* adj pl; *bothe* co-ord conj; *fortune* sb; *goode* adj pl; *ioye* sb; *namoore* adv; *noide* vb pt sg; *sholde* vb pt pl; *tyme* sb; *wo(o)lde* vb pt sg.

(c) *Words in which UE was most often not pronounced*

Prepositions; present participles; past participles; verbs – 2nd sg present, imperative singular, imperative plural.

Individual words: *aboute*, prep; *come* vb infin; *Custaunce* sb; *dide* vb pt sg; *hadde* vb pt sg; *hadde* vb pt pl; *moore* adj str sg; *moore* adv; *somme* adj pl; *sonne* sb; *swiche* adj pl; *trowe* vb 1st sg pres; *wende* vb pt sg; *wolde* vb pt pl.

3. *Words in which UE was not pronounced*

Pronouns; pronoun-determiners; subordinating conjunctions.

Individual words: *come* past part; *dame* sb; *fare (wel)* vb imp sg; *haue* vb imp pl; *haue* vb infin; *haue* vb 1st sg pres; *haue* vb pres pl; *he(e)re* adv; *hire* 'her' pron; *hire* 'her' pron-det sg; *hise* pron-det pl; *housbonde* sb; *loue* sb; *madame* sb; *namoore* sb; *oure* pron-det sg; *oure* pron-det pl; *saue* prep; *seye* vb infin; *seye* vb 1st sg pres; *sire* sb; *thanne* adv; *therefore* adv; *thise* det pl; *we(e)re* vb pt sg; *we(e)re* vb pt pl; *whiche* rel pl; *youre* pron-det sg; *youre* pron-det pl.

The following words produced the same pattern as ones in which UE was not pronounced, but it is possible that in these cases the UE was pronounced, and a medial syllable not pronounced: *euere* adv; *felawe* sb; *heuene* sb; *manere* sb; *millere* sb; *murye* adj str sg; *neuere* adv; *othere* adj pl; *owene* adj wk sg.

It will be seen that there are occasional contradictions between the findings for individual words and those for word-categories. Some of these may be due to the unreliability of the findings for individual words, in cases where the number of examples is very small. But others must reflect genuine exceptions: adverbs fall into

the SP(a) category (UE more often pronounced than not), whereas the adverb *here* is quite definitely in the NP category. In such cases it must also be remembered that the findings for grammatical categories (except those for pronouns and pronoun-determiners) are based on the Version 2 disks, from which many very common words like *here* had been removed. It is not surprising that some very common words turn out to be exceptions to the usual rule for their grammatical category. In the particular case of *here*, it is to be noted that historically this word did not have the final *-e* commonly used to form adverbs, but goes back to an Old English form *hēr*.

NOTES

¹ Part I of this article appeared in *Leeds Studies in English*, n.s. 21 (1990), 81-103. The abbreviations used were explained and defined in Part I, pp. 83-84.

² See Part I, p. 94.

³ The grammatical categories assigned to UE-words were as follows: 1. Substantive as subject, object or complement (1A substantive used attributively). 2. Substantive after a preposition, or in a position where it could have been in the dative. 3. Strong singular adjective (3A used substantively); singular determiner; singular present participle used attributively. 4. Weak singular adjective (4A used substantively). 5. Plural adjective (5A used substantively); plural determiner; plural present participle used attributively. 6. Verb infinitive. 7. Finite verb: 7A 1st singular present; 7B 2nd singular present; 7C 3rd singular present; 7D present plural; 7E imperative singular; 7F imperative plural; 7G past tense singular; 7H past tense plural. 8. Past participle. 9. Adverb. 10. Preposition. 11. Present participle (not noun-like or adjective-like). 12. Relative plural. 13. Co-ordinating conjunction. 14. Subordinating conjunction. 15. Relative singular. 16. Interrogative. Pronouns (like *hire*) were the only UE-words not given a category-number, and were identified by this fact. No distinction was made between main verbs and auxiliaries, since it is not clear on what criteria these can be distinguished in fourteenth-century English.

⁴ The following abbreviations have been used for grammatical categories: adj 'adjective'; adv 'adverb'; art 'article'; conj 'conjunction'; co-ord 'co-ordinating'; det 'determiner'; imp 'imperative'; indef 'indefinite'; infin 'infinitive'; part 'participle'; pl 'plural'; pp 'past participle'; prep 'preposition'; pres 'present'; pron 'pronoun'; pt 'past tense'; sg 'singular'; rel 'relative'; sb 'substantive'; sg 'singular'; str 'strong'; sub 'subordinating'; vb 'verb'; wk 'weak'.

⁵ In classifying words, we have made an initial assessment (admittedly somewhat arbitrary) based on the percentage of examples occurring in Long Lines, as follows: 0%-23% P; 24%-25% P?; 26%-27% SP?; 28%-82% SP; 83%-84% SP?; 85%-86% NP?; 87%-100% NP. This, however, has not been an absolute rule: in words near the borderlines we have also taken into account the percentage of examples occurring in Short Lines.

⁶ J. S. P. Tatlock and A. G. Kennedy, *A Concordance to the Complete Works of Chaucer* (Washington, 1927).

⁷ In classifying words in the Version 2 Dictionary, we have again made an initial assessment based on the percentage occurring in Long Lines, as follows: 0%-19% P; 20%-21% P?; 22%-23% SP?; 24%-81% SP; 82%-83% SP?; 84%-85% NP?; 86%-100% NP.