

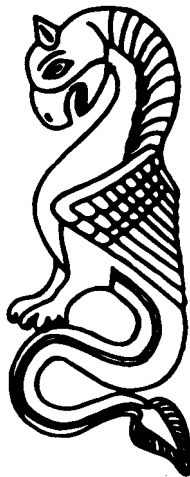
Leeds Studies in English

Article:

Charles Barber and Nicolas Barber, 'The Versification of The Canterbury Tales: A Computer-based Statistical Study', *Leeds Studies in English*, n.s. 21 (1990), 81-103

Permanent URL:

https://ludos.leeds.ac.uk:443/R/-?func=dbin-jump-full&object_id=122548&silo_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 I

Testing Southworth's Hypothesis

During the first half of the present century, there was no great difference of opinion about the pronunciation of word-final unstressed *-e* in Chaucer's verse. It was accepted that word-final *-e* was frequently pronounced (as /ə/), but that it was regularly elided before vowels and before many words whose spelling began with *h*. It was also accepted that in many words the final *-e* was not pronounced, but was merely orthographical. This traditional view is found in the Introduction to F. N. Robinson's edition of Chaucer, which for several generations of students has been the standard text. Robinson says, for example:

The most important difference between Chaucer's English and modern English, for the purpose of versification, lies in the numerous final *-e*'s and other light inflectional endings described in the preceding pages. These endings are ordinarily pronounced in the verse, and indeed are essential to the rhythm. They are also pronounced in rime, and Chaucer with almost complete consistency avoided riming words in *-e* with words not etymologically or grammatically entitled to that ending. But within the verse final *-e* is regularly elided before an initial vowel or before an *h* which is either silent (as in *honour*) or slightly pronounced (as in *he*, *his*, *her*, *him*, *hem*, *hadde*, and a few other words). Before initial consonants *-e* is ordinarily sounded, though there are cases on almost every page where it must have been either slurred or entirely apocopated.¹

In the 1950s, however, this traditional view of Chaucer's versification was challenged in a series of works by J. G. Southworth.² Southworth argued that word-final *-e* had disappeared from speech by Chaucer's time, and was not pronounced in his verse, which had an irregular movement very different from modern iambic verse. Moreover, Southworth vigorously attacked what he saw as a circularity in the procedures of the traditionalists, who devised a prosody by doctoring the text (choosing readings from different manuscripts) to make it 'metrical'.

Subsequent debate on the subject has tended to deal with the linguistic evidence about the date of the loss of *-e* in various sections of the speech-community, the possible relationship between speech and the usage in poetry, and the significance of certain rhymes (such as *Rome/to me*).³ Such discussion, however, can by its nature hardly be conclusive, and Southworth's theory has been neither definitively proved nor definitively disproved. Yet it has surely been obvious from the beginning that, in principle at any rate, the question can be decided beyond any reasonable doubt by statistical methods, and that is what we attempt to do in the present study.

Let us take a simplified example. Suppose we take the first thousand lines of *The Canterbury Tales* in which word-final *-e* does not occur at all inside the line, and count the number of syllables per line. Let us say that the average (i.e. the arithmetic mean) comes out at 10.2 syllables per line. Then let us take the first thousand lines of the same work in which two examples of word-final *-e* occur inside the line, and count the number of syllables per line in this second group (not counting *-e* as a syllable). If Southworth is right, the arithmetic mean for this group too must come out at something near 10.2 syllables per line. If, on the other hand, all the final *-e*'s were pronounced, this second average must come out at something near 8.2; whereas if only half of them were pronounced, the figure must come out at about 9.2. Whether the differences obtained between the different types of line are statistically significant can be calculated by standard methods. The larger the sample, of course, the more reliable the results.

A more elaborate version of this procedure provides the basis for Part I of this study. It is to be noticed that the method makes no prosodic assumptions at all, and would work with verse of any kind, on whatever principles it was written. It is obviously desirable, however, that all the verse in the sample should make use of the same type of metre, though it does not matter at all whether the lines are arranged in couplets or in stanzas of different kinds. It would be generally accepted that, with the sole exception of Chaucer's 'Tale of Sir Thopas', the verse passages of *The*

Canterbury Tales do indeed provide a body of material written in the same metre. The material chosen for our statistical study is accordingly the entire verse of *The Canterbury Tales*, 'Sir Thopas' alone being excluded.

It is necessary, however, to pay heed to Southworth's warnings about the doctoring of texts to fit a metrical theory. For this reason, it is necessary to carry out the analysis, not on an edited text, but on a manuscript. This should be an early and reliable one, and we have chosen to use the Hengwrt manuscript, in the admirable facsimile edition edited by Paul G. Ruggiers.⁴ We have permitted ourselves no departures from this text. If the reading of a line is doubtful, that line is omitted from the study, not emended. Thus some lines are omitted because the manuscript is damaged and part of the line cannot be read. Some are omitted because they are plainly nonsensical; other manuscripts may have a reading which makes sense, but this is disregarded. Some are omitted because they contain a Roman numeral. Some are omitted because they are written in a later hand, such as the one categorised by the Hengwrt editors as 'Hand F', and cannot be considered authentic. But a line is never omitted on metrical grounds – because it is very long, for example, or very short, or reads roughly – for it is essential to avoid making prosodic presuppositions.

After a pilot study carried out manually, we set up computer-programmes for the main study.⁵ One of us (CB) devised the whole research-programme and carried out the analysis of the text. The other (NB) wrote the computer-programmes and did the necessary statistical calculations. To simplify the exposition, and avoid frequent repetition of 'word-final unstressed -e', we shall define a few expressions which we shall use a good deal, and introduce abbreviations for some of them:

Basic line-length (BLL): the number of syllables in a line, excluding all examples of word-final unstressed -e, and also excluding any unstressed syllables at the end of the line.

E: unstressed word-final -e, excluding those in line-final position.

UE: unelided E.

Line-final E (LFE): unstressed word-final -e occurring at the end of a line of verse.

UF: unstressed syllable (excluding LFE) occurring after the last stressed syllable in a line.

Where appropriate, the abbreviations can also mean 'Number of . . . ' or 'Number of examples of . . . '.

It will be noticed that a distinction is made between syllables occurring inside a line, and those occurring at the end. This is because it may well have been the case that final unstressed syllables in a line were extra-metrical, and so should be disregarded in the statistical comparisons; and our pilot study rather suggested that this was in fact so.

We have not, however, disregarded the line-endings; instead, we have broken down the lines into different categories, and analysed each category separately. The first category is that which contains neither LFE nor UF, i.e. in which there are no unstressed syllables at the end of the line; this we can call the 0-0 type of line. The following is an example:⁶

Ex. 1 Of which vertu engendred is the flour
(A0004).

The second category is of lines which contain LFE, but do not contain UF; this is the 0-1 type of line, like the following:

Ex. 2 Whan that Aueryll with his shoures soote
(A0001).

The third category contains UF, but not LFE; this is the 1-0 type of line:

Ex. 3 So priketh hem nature / in hir corages
(A0011).

The fourth category contains both UF and LFE; this is the 1-1 type of line:

Ex. 4 Ire is a synne / oon of the grete of seuene
(B2005).

And finally, there are a few lines which contain *two* UF but no LFE; this is the 2-0 type of line:

Ex. 5 Thomas / of me / thow shalt nat been yflatered
(B1970).

In the analysis, each of these five categories has been handled separately. We have found no examples of a 2-1 type of line, i.e. one containing two UF and one LFE.

This categorisation does, however, raise a question of methodology. We are categorising the lines according to what occurs, or does not occur, after the final stressed syllable of the line. But how can we identify this final stressed syllable without making prosodic assumptions? The criterion we have used for the final stressed syllable is that it is the rhyming syllable. So in Example 1 above, *flour* rhymes with *lycour*, and the rhyming syllables *-cour/flour* are taken to be the final stresses in the line. In Example 2, similarly, *soote* rhymes with *roote*, and (whether or not LFE was pronounced) the rhyming syllables are *soot-/root-*. In Example 3, *corages* rhymes with *pilgrimages*; the rhyming syllables are *-rag-/mag-*, which are therefore taken to be the final stresses. In Example 4, *seuene* rhymes with *heuene*, and the rhyming syllables are *seu-/heu-*. And in Example 5, finally, *yflatered* rhymes with *scatered*, and the rhyming syllables are *flat-/scat-*. It could be argued that this method involves a prosodic assumption, since it takes the rhyming syllable to be the last stressed one in the line; but if so it is a minimal assumption.

In the vast majority of lines there is no difficulty in identifying the rhyming syllable. There are occasional problems, however, which to some extent are caused by Chaucer's habit of sometimes using identical syllables instead of full rhymes. He sometimes rhymes homophones, such as *style* 'style' with *style* 'stile' (E0105-06), or rhymes final identical syllables such as *-ly*, as in *thriftily/yemanly* (A0105-06). In these particular cases there is no problem in identifying the rhyming syllable, nor is there any difficulty in taking the final *-ly* to be stressed, since there is ample evidence that such syllables could be stressed, and indeed that forms with the stressed syllable persisted in English as late as 1600.⁷ Problems arise, however, when there is more than one possible candidate in a line for the honour of being rhyming syllable. Such is the case for example with *priuely/softely* (A4057-58) and with *hardily/holily* (B2285-86). A purely mechanical analysis would take the rhyming syllables to be *-uel-/tel-* and *-dil-/lil-*. This is highly implausible, however, on historical-linguistic grounds, and it is more reasonable to classify the lines as rhyming on the final syllable. Similarly, if the ending *-ioun* is taken to constitute two syllables (*-i-oun*), as it is in our analysis, a problem arises with rhymes like *Scorpioun/confusioun*. Here a mechanical analysis would give the rhyming syllables as *-pi-/si-*, but this is implausible, and we in fact analyse the rhyme as occurring on the final syllable. This analysis is supported by rhymes in stanzaic portions of the poem, where there are often three rhyming words. For

example, there is a rhyme *Abusioun/conclusioun* (D0214-15), which on a mechanical analysis might be taken to rhyme on *-bus-/-clus-*; but these words also rhyme with *doun* (D0212), which shows conclusively that the rhyme is on the final syllable. Other rhymes where similar decisions have been taken include *Citee/solempnytee* (A2701-02), *haliday/any day* (A3309-10) (cf. the rhyme *haliday/gay* at A3339-40), *receyue/deceyue* (I0917-18), and *affiance/alliance* (I1329-30). The total number of such debatable rhymes, however, is very small. Notice that in such cases the decision to treat the final syllable as stressed is not taken on metrical or prosodic grounds: it has nothing to do with the length or the rhythm of the line, and is made solely on the evidence of historical linguistics.⁸

The programme of analysis involved making eight entries for each line of verse. These were recorded on the computer-disk, and also given as a print-out in eight columns. The eight entries were as follows:

Col. 1	Line-number
Col. 2	Basic line-length (BLL)
Col. 3	Value of E
Col. 4	Value of UE
Col. 5	Value of UF
Col. 6	Value of LFE
Col. 7	} A record of all words in the line containing UE,
Col. 8	

and of any other words considered significant.

The line-number (Column 1) consists of one letter and four figures, beginning at A0001 and ending at I2156. The figures are those given to the lines in the Ruggiers edition of the Hengwrt manuscript. The initial upper-case letter is necessary because the Hengwrt editors do not number the verse-lines of the manuscript continuously from 1 to (about) 16,000, but retain the traditional line-numbering of the various sections, thus constantly going back and starting again at 1 (or sometimes at some other number). In our numbering system, Section A begins at the beginning of the General Prologue and ends with A4422 at the end of the incomplete 'Cook's Tale' (p. 225 in the Ruggiers edition). Section B begins with 'The Wife of Bath's Prologue', and ends with B4636 at the end of 'The Nun's Priest's Tale' (p. 422). Section C begins with the Prologue to 'The Manciple's Tale' and ends at C362 at the end of that tale (p. 441). Section D begins with the Prologue to 'The Man of Law's Tale', and ends at D1162 at the end of that tale

(p. 506). Section E begins at E0009 with the beginning of 'The Squire's Tale', and ends at E2440 after the Host's comments on 'The Merchant's Tale' (p. 606). Section F begins with the Host's words to the Franklin (spoken to the Squire in the Ellesmere manuscript), and after eight lines leaps on to F0709 at the beginning of 'The Franklin's Tale'; it ends at F1624 at the end of that tale (p. 654). Section G begins with the Prologue to 'The (Second) Nun's Tale', and ends at G0553 at the end of that tale (p. 686). Section H begins with the Prologue to 'The Clerk's Tale', and ends at H1219 at the end of that tale (p. 758). And finally Section I begins with 'The Physician's Tale' and ends at I2156 immediately before Chaucer's 'Tale of Melibeus' (p. 858); Chaucer's 'Tale of Sir Thopas' occurs in this section (I1902-2108), but is not included in the study. This division into sections is obviously different from the traditional division into Groups or Fragments, but is the most convenient for our purposes.

Column 2 records the Basic Line Length of the line, that is, the number of syllables it contains, excluding all examples of E and also excluding any unstressed syllables at the end of the line (UF and LFE). It is here that we encounter the main practical difficulty of the analysis. For what are we to count as a syllable? Is the word *euery* two syllables, or three? and what about *beautee*, and *bodyes*, and *boldely*, and *creature*, and so on through the alphabet? It is to be noted, however, that (at any rate for Part I of this study) the essential virtue is consistency. It does not really matter whether we analyse *euery* as two syllables or as three, provided we always analyse it the same way. For it is as likely to occur in one kind of line as in another, so that any error in syllable-counting will cancel itself out statistically speaking: all types of line may have too many syllables attributed to them, but this will not affect the *differences* between lines containing different numbers of E. We have therefore aimed to have clear rules for counting syllables, and where we have had to make more-or-less arbitrary decisions about syllable-number we have kept a record of each decision and stuck to it in later occurrences.

The general rule adopted for counting syllables is to reckon everything as a syllable which is spelt as one. So *euery* is analysed as three syllables, one for each of the vowel-symbols in its spelling; and *Atthenes*, *bokeler*, *boldely*, *exiled*, *foweles*, *huntery*s, *lowely*, *remenant*, and *wodecraft* are all similarly analysed as having three syllables. The digraphs *ai*, *ay*, *au*, *ee*, *ei*, *eu*, *ey*, *oi*, *oo*, and *ou* are normally taken to represent only one syllable, as in *batailles* (3 syllables), *bargaynes* (3 syllables), *sauc(e)* (one syllable), *threed* (one syllable), *seith* 'says' (one syllable), *seuretee* 'surety' (3 syllables), *frankeleyn* (3 syllables), *boill(e)* (one syllable), *hoot* 'hot' (one syllable), and *koud(e)* 'could' (one syllable). But in

Classical proper names, the ending *-eus* is on the contrary analysed as two syllables, so that *Theseus* has three syllables. The digraphs *ea*, *ie*, and *ye* cause some difficulty, since each of them can represent either a single vowel or a sequence of two vowels, and decisions have to be made in individual cases. So *diet* 'diet' is analysed as two syllables, but *hierd(e)* 'shepherd' as one; *geant* 'giant' as two, but *-geant* in *sergeant* as one. In some cases arbitrary decisions had to be made: thus forms like *ladyes* were analysed as having three syllables, whereas forms like *ladies* were taken to have only two, and *ladye* and *ladie* were treated similarly. This is hardly likely to represent the scribe's usage (or Chaucer's), but has the advantage of consistency, and also of dividing this group of doubtful cases more or less equally between the two possible types of category.

Other vowel-digraphs are normally taken to represent two syllables, as in *Dian* 'Diana' (2 syllables), *poet* (2 syllables), *Iuels* 'jewels' (2 syllables), *ryot* 'riot' (2 syllables), *ydiot* 'idiot' (3 syllables). Endings such as *-ial*, *-ian*, *-ienc(e)*, *-ient*, *-ion*, *-ioun*, and *-ious* are taken to have two syllables each.

Departures are made from these general rules when they seem called for on historical-linguistic grounds. So in some words, such as *proeued* 'proved, tested' and *moeuer* 'mover', the digraph *oe* is taken to represent only one syllable; and contrariwise in a few words such as *deitee* and *obeisance*, the digraph *ei* is taken to represent two syllables.

In cases where we have had to make decisions about the number of syllables in a word, or class of words, these have always been made on historical-linguistic grounds, never on metrical grounds. And, since consistency is essential, once we have made such a decision we stick to it for the rest of the analysis.

Because of the spelling-criterion used in counting syllables, any departure in the count from Chaucer's own intentions will almost invariably be in the same direction: more syllables will be recorded than Chaucer himself pronounced in reading the verse, not fewer. Occasionally there may be a deviation in the opposite direction: for example, a vowel-digraph may be counted as a single syllable when in fact for Chaucer it was two syllables. But in the vast majority of cases the Basic Line Length allocated will be too long rather than too short (though of course in most cases we hope that it will be exactly right). This bias towards over-long lines in the count is of no consequence in Part I of the study: as we have said, the errors will occur in equal proportions in all the categories analysed, and so will not affect the outcome. In Part II of the study, however, it will be necessary for us to attempt to eliminate this bias; and then, as we shall see, it will be a positive advantage that practically all the deviations are in the same direction.

Column 3 records the value of E for the line, in other words the number of examples within the line of word-final unstressed *-e*. This is merely a matter of recording the number of words (excluding monosyllables) which end with the spelling *-e*. There are occasional decisions to be made: for example, the word *meyne* 'meinie, retinue' (C0228) is taken *not* to have E, an analysis supported both by the modern pronunciation and by the spelling *meynee* elsewhere in the manuscript (e.g. C0231). On the other hand, *pouerte* 'poverty' is analysed as having E, despite the modern pronunciation, in view of the rhyme *pouerte/sherte* (B1185-86).

Monosyllabic words – *be, he, me, she, the, we, ye* – are interpreted as having been pronounced with a vowel other than /ə/, and so as *not* having E. It is indeed possible that occasionally some of them may have lost their vowel when unstressed, especially *the*, but this is commonly shown in the spelling: the manuscript has such forms as *thapostle* 'the apostle' (B0124), and *mendite* (G0033) (= *me endite*). Since *the* is analysed as not containing E, the form *atte* 'at the' is analysed in the same way: it is counted as two syllables in Column 2, but not counted for Column 3.

Column 4 gives the value for the line of UE, unelided E. According to the traditional view, E was commonly pronounced, but not when subject to elision. So Column 4 gives the figure in Column 3 minus the number of E in the line which occur in a position where E would be elided. This position is taken to be before a vowel at the beginning of the following word. It is to be noted, however, that 'vowel' here means 'vowel-phoneme', not 'vowel-symbol'; it refers to pronunciation, not spelling. Many words whose spelling began with *h-* had an initial vowel in pronunciation, since the *h* was not pronounced. One category of word to which this applied was French loan-words, such as *harlot, hazard, heritage, homicide, host, humble*; these had already lost their initial /h/ before they were borrowed into English, and the modern pronunciations are due to the spelling and to Latin influence.⁹ It is possible that some such spelling-pronunciations had arisen by Chaucer's time, but the majority of them date from the Modern English period, and we have therefore assumed that all French words of this type caused elision. There is one group of French loans, however, in which initial /h/ was pronounced; these are words which French itself had earlier borrowed from Germanic, such as *hardy, haste, haunch, herald, and heron*.¹⁰ We have assumed that such words do not cause elision.

In native English words, and in words borrowed from Old Norse, initial /h/ was normally pronounced, and we have assumed that no elision took place before

such words as *hang*, *hap*, *harm*, *hest*, *hide*, and *hurt*. In the thirteenth century, however, /h/ was lost at the beginning of an unstressed syllable. In consequence, words which occurred frequently in both stressed and unstressed positions, like the pronouns *hit* 'it' and *hire* 'her', developed double forms: a strong form with initial /h-/, and a weak form without it.¹¹ Clearly, the weak form would cause elision, but the strong form would not.

Our difficulty here is that we wish to avoid basing any of our decisions on metrical grounds: and if we assert that, in a given line of verse, the word *hire* is stressed (or unstressed), we shall be doing precisely that. To avoid this consequence, we have treated certain words as always occurring in their weak forms, and therefore without the initial /h-/. The words in question are the pronouns *he*, *hem*, *him*, and *hir(e)*, the pronoun-determiners *hir(e)* and *his(e)*, the adverb and conjunction *how*, and all parts of the verb 'to have' (*haue*, *hast*, *hadde*, etc.). These words are taken to cause elision of a preceding E. Just one exception is made: if the words occur in rhyme-position, they must surely be treated as stressed, so in such cases we take it that the strong form was used and no elision occurred, as in the following:

Ex. 6 In a gowne of faldyng to the knee
A daggere hangynge on a laas / hadde he
(A0391-92).

Here it is clear that *he*, being the rhyming syllable, is stressed, and so does not elide the E of *hadde*. No doubt the words in question were sometimes stressed when not in rhyme-position, but probably not very often; and it is better to tolerate a small amount of error rather than sacrifice consistency. It would be dangerous to launch into the extremely subjective business of analysing the stress-pattern of every line of verse, and would introduce into the analysis exactly the kind of circularity which Southworth objects to.

In the case of words where there was no initial *h* in the spelling, we have to remember that the symbols *i*, *u*, and *v* could all represent either a vowel-phoneme or a consonant-phoneme. So for example the word *iuel* 'jewel' begins with a consonant, and will not cause elision. Notice, too, that words like *use* began with a vowel, and not (as today) with the semivowel /j/ (which would not have caused elision). In Chaucer's time, the verb *use* was pronounced /iuz/ or /y:z/, and so would cause elision. The modern pronunciation /ju:z/ was not normal in educated speech until the seventeenth century.¹²

In columns 3 and 4, words which end in Consonant + /l/ or /r/ + -e (like *bette*, *noble*, *propre*, and *tendre*) are treated in exactly the same way as other words. That is, the final -e is interpreted as E, and is elided before a vowel. When the E is elided, a word like *noble* becomes a monosyllable. This may seem odd to present-day readers, because today the word is dissyllabic, and the final consonant-cluster /-bl/ is not permitted in Present-day English unless the /l/ is syllabic. But this was not necessarily the case in Chaucer's time. The E was certainly lost in words like *noble*, for it is not pronounced today. What has happened is that, at some period after the loss of E, a svarabhakti vowel has developed before the /r/ or /l/, or (in the case of /l/) the final consonant has become syllabic. But the exact date at which this happened is not known, and it is quite likely that in the late fourteenth century *noble* was a monosyllable (as the corresponding word is in Present-day French). Similar considerations (with some differences of detail) apply to words ending in Consonant + /w/ + -e, like *narwe*, *swalwe*, *widwe*.

Column 5 records the value of UF for the line, that is, the number of unstressed syllables (other than -e) which occur at the end of the line; this value can be 0, 1, or 2. Column 6 records whether or not LFE (line-final -e) occurs in the line; if it occurs, the value shown is 1; if it does not occur, 0. The use of the rhyming-syllable as the criterion for the last stressed syllable in the line has already been discussed above.

Columns 7 and 8 are Comment columns. In them are recorded all the words in the line which contain UE (unelided E); these words will be analysed in Part II of the study. It is possible to enter two words into each of the Comment Columns, by joining them with a + sign. The material contains no lines with more than four examples of UE. When these words are recorded, it is obviously necessary to record also their grammatical status; at the end of each word, therefore, is added a number which shows its word-type (noun, adjective, verb, etc.), often with sub-classification (adjectives, for example, being shown as strong singular, weak singular, or plural).

In the initial pilot-study, it was seen that some words occurred with great frequency in lines of more than average length, and rarely in lines of average or less than average length, giving rise to the suspicion that they were pronounced by Chaucer with fewer syllables than we had attributed to them in the analysis. Striking examples are the word *euery* and the expression *many a*. Such words are recorded in Columns 7 and 8. There are, on the other hand, a few words (very few) which appear to occur predominantly in lines of less than average length, leading to the suspicion that our analysis has attributed too few syllables to them; an example is the

word *creature*, where we took the digraph *ea* to represent only one syllable; the distribution of the word suggests that we should, on the contrary, have analysed *ea* as representing two syllables (as indeed we did in such words as *create* and *creation*). Such words too are entered in Columns 7 and 8. These 'overlong' and 'overshort' words will be used in Part II of the study, where we shall attempt to produce a revised version of the data taking such evidence into account. It must be emphasised once again that misinterpretations of this kind have no effect on the findings of Part I of the study.

Occasionally, there is no room in Columns 7 and 8 for all the material which we wish to enter, since our disk-space is somewhat limited. In such cases the omitted material is recorded manually, and entered into the data at a later stage.

As an actual example of the analysis, let us take the opening lines of 'The Wife of Bath's Prologue', which in the Hengwrt manuscript appear as follows:

- Experience / thogh noon Auctoritee
Were in this world / is right ynogh for me
To speke of wo / that is in mariage
For lordynges / sith þat I twelf yeer was of age
5 Thonked be god / that is eterne on lyue
Housbondes atte chirche dore / I haue had fyue
If I so ofte / myghte han wedded be
And alle were worthy men / in hir degree
But me was told certeyn / noght longe agon is
10 That sith þat Crist ne wente neuere but onys
To weddyng in the Cane of Galilee
That by the same ensample / taughte he me
That I ne sholde / wedded be but ones
Herke eek / lo / which a sharp word for the nones
15 Bisyde a welle / Iesus / god and man
Spak / in repreeue of the Samaritan
Thow hast yhad / fyue housbondes quod he
And that ilke man / which that now hath thee
Is nat thyn housbonde / thus he seyde certeyn
20 What that he mente ther by / I kan nat seyn
(B0001-0020)

The analysis is as follows:

The Versification of The Canterbury Tales

Column							
1.	2.	3.	4.	5.	6.	7.	8.
(Line No.)	(BLL)	(E)	(UE)	(UF)	(LFE)	(Comment)	(Comment)
B0001	10	1	1	0	0	EXPERIENCE1	
B0002	10	1	0	0	0		
B0003	10	1	0	0	1		
B0004	12	0	0	0	1	LORDYNGES	
B0005	10	1	0	0	1		
B0006	11	3	1	0	1	CHIRCHE1A	HOUSBONDES
B0007	9	2	1	0	0	OFTE9	
B0008	10	2	2	0	0	ALLE5A	WERE7H
B0009	10	1	0	1	0		
B0010	10	2	2	1	0	WENTE7G	NEUERE9
B0011	10	1	0	0	0		
B0012	9	3	1	0	0	ENSAMPLE2	
B0013	9	1	1	1	0	SHOLDE7G	
B0014	10	1	0	1	0		
B0015	9	2	1	0	0	WELLE2	
B0016	10	1	0	0	0		
B0017	10	1	1	0	0	FYUE5	HOUSBONDES
B0018	9	1	1	0	0	ILKE4	
B0019	10	2	2	0	0	HOUSBONDE1	SEYDE7G
B0020	10	1	1	0	0	MENTE7G	

The syllable-counting rules already described are illustrated by the fact that, for example, *lordynges* (4) and *Housbondes* (6) are counted as three syllables, and *Thonked* and *atte* as two. Notice that there is elision before *had* (6), *han* (7) and *he* (12), but not before *housbondes* (17). All words with UE are recorded in Column 7 or Column 8, and their grammatical status marked: for example, the marking 7G after *wente* (10) records that the word is a verb, past tense singular. The non-E words recorded in Columns 7 and 8 are all ones where it is plausible to imagine that the number of syllables in pronunciation may have been different from that attributed by our rules: for example, the ending *-es* in *lordynges* (4) could well have been non-syllabic. Columns 5 and 6 record the kind of line-ending, and it will be seen that in this passage the 0-0 type of ending is predominant.

In all, we analysed 15,942 lines of verse. The distribution of different types of line ending was as follows:

0-0 endings: 7130

0-1 endings: 7927

1-0 endings: 821

1-1 endings: 51

2-0 endings: 13

In these 15,942 lines, we found 16,460 examples of E. The number ranged from zero to five in individual lines, distributed as follows:

Lines with 0 examples of E: 4811

Lines with 1 example of E: 6822

Lines with 2 examples of E: 3393

Lines with 3 examples of E: 819

Lines with 4 examples of E: 90

Lines with 5 examples of E: 7

Very nearly half of these examples of E were in positions where they would be subject to elision, leaving a total of 8375 examples of UE. No line contained more than four examples of UE, and the distribution was as follows:

Lines with 0 examples of UE: 9156

Lines with 1 example of UE: 5389

Lines with 2 examples of UE: 1218

Lines with 3 examples of UE: 166

Lines with 4 examples of UE: 13

In the statistical analysis, lines with different line-endings were treated as separate groups. In each group, the mean Basic Line Length was calculated for each different value of E (Type 1 tests) and for each different value of UE (Type 2 tests). The complete results of these tests are given in Tables 1 and 3. The computer calculated the mean value of BLL to five decimal places, but for simplicity the tables give the results only to three significant decimal places.

Table 1 gives the findings of the Type 1 tests (Tests 1-30), in which the variable is E. In the first group of tests (Tests 1 to 6), the line-ending is 0-0 (i.e. there are no unstressed syllables at all at the end of the line). It will be seen that, for lines with no examples of E (Test 1), the mean value of BLL is 10.170 syllables. If Southworth were right, and E were never pronounced in Chaucer's verse, the figure

for BLL should remain at about this figure in Tests 2 to 6. This is not so however, for the Basic Line Length in fact goes down as the value of E increases. (The figure for Test No. 6 can be disregarded, since the number of examples is so small: in general, samples of less than 10 must be regarded as somewhat unreliable, and samples of less than 5 as extremely unreliable.) The reduction in BLL, however, is a great deal smaller than could be expected if *all* the examples of E were pronounced: for example, the value of BLL for Test No. 2 is about 0.32 lower than that for Test 1, suggesting that nearly one-third of E were pronounced; while that for Test 3 is about 0.56 lower than that for Test 1, which suggests a somewhat lower figure for the proportion of E pronounced, about 28%. It is striking, however, that (leaving aside Test No. 6) the trend in the extreme right-hand column is consistently downward as E increases. It is curious, however, that, as we go down the table, the differences become progressively smaller: that between Test 1 and Test 2 is 0.320, but the difference between Test 4 and Test 5 is only 0.263. This flattening of the curve is something we shall find in other tables, and a possible explanation will be offered later.

The second group of tests (Tests 7 to 12) gives the findings for lines in which the line-ending is 0-1 (that is, lines ending in unstressed final *-e*). The figures are strikingly similar to those for Tests 1 to 6: in both groups of tests the value of BLL is just over 10 when E = 0, and declines gradually as E increases. And in this group, as in the first one, the curve flattens as we go down the table, the decrements in the value of BLL becoming less and less. (Test 12 must be left out of account, since the number of examples is so small.)

The third group of tests (13 to 18) gives the figures for lines with the ending 1-0 (that is, lines with a final unstressed syllable other than *-e*). Here we have the same pattern yet again, with the value of BLL beginning just above 10, and decreasing gradually as E increases. The values are indeed lower than in the previous two tables (except for E = 2), but the difference is extremely small, and the evidence suggests strongly that Basic Line Length is unaffected by the presence or absence of an unstressed syllable at the end of the line. There is a slight difference from the previous two tables in that the value of BLL decreases less between Tests 13 and 14 than between Tests 14 and 15; but the flattening of the curve then occurs between Tests 15 and 16. The sharp drop in Test 17 is to be disregarded, since the material is so small.

The fourth group of tests (19 to 24) gives the results for lines with the ending 1-1, i.e. lines which have both line-final *-e* and another unstressed syllable after the rhyme-syllable. The material is here very small, but at any rate there are enough

Table 1
Type 1 Tests

Test No.	Conditions			Occurrences	Average of Col. 2 (BLI)
	Col. 3 (E)	Col. 5 (UF)	Col. 6 (LFE)		
1.	0	0	0	2150	10.170
2.	1	0	0	3041	9.850
3.	2	0	0	1529	9.162
4.	3	0	0	367	9.493
5.	4	0	0	39	9.230
6.	5	0	0	4	10.000
7.	0	0	1	2393	10.111
8.	1	0	1	3414	9.842
9.	2	0	1	1681	9.577
10.	3	0	1	393	9.453
11.	4	0	1	43	9.442
12.	5	0	1	3	10.000
13.	0	1	0	247	10.077
14.	1	1	0	343	9.834
15.	2	1	0	169	9.586
16.	3	1	0	54	9.352
17.	4	1	0	8	8.750
18.	5	1	0	0	—
19.	0	1	1	16	10.188
20.	1	1	1	21	9.667
21.	2	1	1	10	9.600
22.	3	1	1	4	8.750
23.	4	1	1	0	—
24.	5	1	1	0	—
25.	0	2	0	5	10.400
26.	1	2	0	3	9.333
27.	2	2	0	4	9.500
28.	3	2	0	1	10.000
29.	4	2	0	0	—
30.	5	2	0	0	—

The Versification of The Canterbury Tales

examples in the first three lines for them not to be negligible, and the pattern is precisely the same as in the previous three tests, with the value of BLL above 10 when $E = 0$, and decreasing gradually as E increases.

The final group of Type 1 tests, Numbers 25 to 30, gives the findings for lines with the ending 2-0, i.e. lines with two unstressed syllables (excluding *-e*) after the rhyme-syllable. Here the number of examples is much too small for any conclusions to be drawn, except that we can perhaps say that the figures offer no evidence for anything wildly different from the previous four tables; given the smallness of the material, it is at any rate consistent with the earlier findings.

The evidence of Table 1 surely gives us strong grounds for believing that Southworth's hypothesis is false. The figures clearly suggest that, in the verse of *The Canterbury Tales*, some of the word-final *-e*'s must certainly have been pronounced by Chaucer when he wrote and read it. The statistics suggest that the number pronounced was something between one-quarter and one-third of those written in the manuscript.

Table 1 also shows beyond any reasonable doubt that line-final *-e*, and line-final unstressed syllables other than *-e*, were extra-metrical. By this we mean that the number of syllables in a line, up to and including the rhyme-syllable, is independent of the existence or non-existence of unstressed syllables after this rhyme-syllable. This can be seen clearly if we take the figures for Basic Line Length in all lines in which $E = 0$, broken down according to line ending:

Type of Ending	Average BLL
0-0	10.170
0-1	10.111
1-0	10.077
1-1	10.188
2-0	10.400

The only figure markedly different from the others is that for 2-0, which is based on only 5 examples; and in any case this figure of 10.4 hardly contradicts our view, since it is in the 'wrong' direction: if the existence of two unstressed syllables at the end of the line were to have any effect on BLL, it could surely be expected to make it shorter, not longer. All the other figures for BLL, ranging from 10.077 to 10.188 differ from one another insignificantly, and moreover show no consistent trend either upwards or downwards as the number of line-final syllables increases.

Since this is so, we can confidently amalgamate the statistics from the five

different groups of tests in Table 1, and produce an overall table in which line-ending is disregarded:

Table 2		
Lines with all types of ending		
Col. 3	Occurrences	Average of Col. 2
(E)		(BLL)
0	4811	10.136
1	6822	9.845
2	3393	9.593
3	819	9.462
4	90	9.289
5	7	10.000

Even in this consolidated table, the number of lines containing five examples of E amounts to only seven, and the bottom line of the table must therefore be left out of account as statistically unreliable. The first five lines of the table, however, show very clearly the decrease of BLL as the value of E rises, from over 10.1 when E is zero to just under 9.3 when E = 4. Moreover, the differences in the value of BLL are statistically significant at a confidence-level greater than 99.9%; in other words, the odds against these differences being due to chance alone are greater than a thousand to one.¹³ A simple calculation shows that, according to the figures in this consolidated table, about 27% of the examples of E were pronounced.

Table 2 also shows clearly, however, the way in which the decrements in the value of BLL become smaller as E increases. Between E = 0 and E = 1, the average value of BLL decreases by 0.29; between E = 1 and E = 2, by 0.25; between E = 2 and E = 3, by 0.13; and between E = 3 and E = 4, by 0.17. Clearly the curve flattens out as E increases, though with a slight reversal at the end. This is a puzzling phenomenon, but we should like to suggest a possible explanation.

The evidence we have adduced gives us good grounds for believing that in a considerable number of words the E was pronounced. In some words or classes of word it may perhaps have always been pronounced. In others, it was perhaps never pronounced. But there may well have been a third category of words or word-types where the pronunciation of E was optional, the poet choosing to pronounce it or to omit it according to metrical convenience; and in Part II of this study evidence will be given which supports this view. But if this were so, there would surely be a tendency for the poet to choose the E-silent form in lines where there were already

The Versification of The Canterbury Tales

Table 3
Type 2 Tests

Test No.	Conditions			Occurrences	Mean Value of Col. 2 (BLL)
	Col. 4 (UE)	Col. 5 (UF)	Col. 6 (LFE)		
31.	0	0	0	4050	10.144
32.	1	0	0	2444	9.603
33.	2	0	0	551	9.232
34.	3	0	0	80	9.038
35.	4	0	0	5	8.200
36.	0	0	1	4622	10.102
37.	1	0	1	2655	9.575
38.	2	0	1	578	9.138
39.	3	0	1	65	9.046
40.	4	0	1	7	9.286
41.	0	1	0	444	10.074
42.	1	1	0	272	9.614
43.	2	1	0	84	9.238
44.	3	1	0	20	9.200
45.	4	1	0	1	9.000
46.	0	1	1	33	10.121
47.	1	1	1	14	9.286
48.	2	1	1	3	8.667
49.	3	1	1	1	7.000
50.	4	1	1	0	—
51.	0	2	0	7	10.143
52.	1	2	0	4	9.500
53.	2	2	0	2	9.500
54.	3	2	0	0	—
55.	4	2	0	0	—

a number of E¹ words, partly for euphony, and partly to keep the line down to length. The result of this, clearly, would be that the more examples of E there were in a line, the fewer examples of E would be pronounced; and this would produce the kind of pattern for BLL seen in our tables. Plainly we can produce no supporting evidence for this theory, but it is the only one we can think of at present which explains the facts.

Since about 73% of the examples of E appearing in the manuscript seem not to have been pronounced, the traditional theory of the elision of E before vowels is a plausible one. The tests have therefore been rerun with UE substituted for E. These are Type 2 tests, Numbers 31 to 55, and the results are given in Table 3. It will be seen that the patterns found are extremely similar to those produced by the Type 1 tests in Table 1: in each of the five groups of tests in Table 3, the value of BLL is about 10.1 when UE = 0, and decreases gradually as UE increases from 0 to 4. The few aberrant figures, such as that for Test 40, can clearly be attributed to the very small number of examples in the tests in question. Table 3 also confirms the finding of Table 1, that line-ending has no significant effect on the value of BLL. Here again, therefore, we can look at a consolidated table, showing the outcome of Type 2 tests for all types of line-ending:

Table 4
Lines with all types of ending

Col. 4 (UE)	Occurrences	Average of Col. 2 (BLL)
0	9156	10.119
1	5389	9.589
2	1218	9.187
3	166	9.048
4	13	8.846

The value of BLL is just over 10.1 when UE is zero, and decreases as the value of UE increases. As is to be expected, the decrease in BLL is much greater than in Table 2: for example, between UE = 0 and UE = 3, BLL decreases by about 1.07, whereas the corresponding figure for E (in Table 2) is about 0.67. The differences between the various values of BLL in Table 4 are statistically significant at the 99.9% confidence level.¹³

As in Table 2, the decrements of BLL become smaller as we go down the table, with just a small reversal at the very end: the successive decrements,

beginning at the top of the table and working downwards, are 0.530, 0.402, 0.139, and 0.202. These figures suggest that, in lines containing one example of UE, about 53% of these examples were pronounced; in lines containing two examples, about 47% were pronounced; in lines containing three examples, about 36% were pronounced; and in lines containing four examples, only about 32% were pronounced. Overall, it can be calculated from the figures that almost exactly 50% of all examples of UE were pronounced. So our conclusion is that, in the verse of *The Canterbury Tales*, almost exactly half the examples of unelided word-final *-e* occurring inside the line were pronounced by the author when he wrote and read the poem.

The tables give us no information about the pronunciation or non-pronunciation of line-final *-e*, but it would seem reasonable to assume that words in which E was pronounced inside the line also had LFE pronounced at the end of the line. It is to be noted that, in both Middle English and Modern English phonology, a following pause usually has the same effect as a following consonant, not as a following vowel;¹⁴ one would therefore expect LFE to survive in pronunciation as long as UE.

In Part II of the study, we shall try to establish which words in the poem commonly had E pronounced, and which did not, by examining their distribution among lines of different lengths. Reliable information can only be obtained, obviously, about words which occur in the poem reasonably frequently. At the same time we shall try to detect words to which, in our analysis, we have attributed too many or too few syllables. The information obtained will then be fed back into the computer-disks, producing a revised version of the data. From the revised data it may be possible to obtain more refined results about the pronunciation of various E-words. It is proposed that Part II of the study shall be published in *Leeds Studies in English* in 1991.

NOTES

¹ *The Works of Geoffrey Chaucer*, edited by F. N. Robinson, second edition (London, 1957), p. xxxv. This work has now been superseded by the admirable Riverside Chaucer, in which the section on Language and Versification (by the late Norman Davis) gives an account very similar to Robinson's: see *The Riverside Chaucer*, general editor Larry D. Benson, third edition (Boston, 1987), pp. xxix-xlv.

² J. G. Southworth, *Verses of Cadence* (Oxford, 1954); *The Prosody of Chaucer and his Followers* (Oxford, 1962); 'Chaucer: a Plea for a Reliable Text', *College English*, 26 (1964), 173-79 (reprinted, with revisions, as 'Chaucer's Prosody, a Plea for a Reliable Text', in *Chaucer's Mind and Art*, edited by A. C. Cawley [Edinburgh and London, 1969], pp. 86-96).

³ See for example Ian Robinson, *Chaucer's Prosody* (Cambridge, 1971).

⁴ *The Canterbury Tales. A Facsimile and Transcription of the Hengwrt Manuscript*, edited by Paul G. Ruggiers (Oklahoma and Folkestone, 1979).

⁵ The computer-programmes are mostly written in Crystal Basic 4.12, except for the sorting programmes in Part II, which are written in BBC Basic Version 2.31. The machine used was a Tatung Einstein with two disk-drives.

⁶ For the method of indicating line-numbers, see pp. 86-87 above.

⁷ See for example E. J. Dobson, *English Pronunciation 1500-1700*, second edition, 2 vols (Oxford, 1968), II, pp. 445-46, 827-38, 842-46.

⁸ The existence of the forms in question is attested by their survival into much later times. See Dobson, *English Pronunciation 1500-1700*, pp. 827-55; C. Barber, *Early Modern English* (London, 1976), pp. 322-27.

⁹ See M. S. Serjeantson, *A History of Foreign Words in English* (London, 1935), p. 300.

¹⁰ See Richard Jordan, *Handbuch der mittelhochdeutschen Grammatik* (Heidelberg, 1934), pp. 228-29.

¹¹ See B. M. H. Strang, *A History of English* (London, 1970), p. 250; Dobson, *English Pronunciation*, pp. 450-64.

¹² Dobson, *English Pronunciation*, pp. 699-713.

¹³ The significance has been calculated by variance-analysis: the variance within groups has been compared with the variance between groups. For this purpose it is necessary to break down the entire material according to Basic Line Length (Column 2). When the variable is E (Column 3), the breakdown comes out as follows:

The Versification of The Canterbury Tales

Value of E	BLL=7 ¹	BLL=8	BLL=9	BLL=10	BLL=11	BLL=12	BLL=13	Total
0	0	8	183	3844	700	73	3	4811
1	2	61	1704	4336	667	50	2	6822
2	0	219	1185	1760	215	14	0	3393
3	11	87	287	385	46	2	1	819
4	2	14	32	40	2	0	0	90
5	0	0	1	5	1	0	0	7

From this table, and Table 2, it can be calculated that lines with different values of E constitute different populations: the odds against the differences in mean BLL being due solely to chance are greater than a thousand to one. A similar distribution-table can be drawn up when the variable is UE, and from this, combined with Table 4, it can be shown that lines with different values of UE similarly constitute different populations. For statistical methods, we have consulted G. Udny Yule and M. G. Kendall, *An Introduction to the Theory of Statistics*, fourteenth edition (London, 1958).

¹⁴ For example, in Middle English, word-final /n/ is lost in unstressed syllables before a consonant or a pause, but retained before a vowel; and in many varieties of Modern English, word-final /r/ is lost before a consonant or a pause, but retained before a vowel. Presumably, therefore, a pause (such as can be expected at the end of a line of verse) has the same kind of phonological effect as a following consonant.