APPROACHES TO THE ANALYSIS OF LATIN PROSE, APPLIED TO CATO, SALLUST AND LIVY.

There are many ways of approaching a text using a computer as an intermediary, some of which have been used for many decades even without the assistance of machines. This paper details four such approaches and gives the data resulting from their application to three Latin authors. No attempt has been made to develop a thorough analysis of these results; indeed, it would be premature to do so before material from a number of other authors is available. The intention of this paper is simply to describe the methods used and to provide material for eventual comparisons when similar data becomes available for other authors.

The tests used to provide data are from Cato, Sallust, and Livy. For Cato's *De Agri Cultura*, the entire 1962 Teubner text edited by Antonio Mazzarino was used. The fragments of Cato's orations were taken from the third, 1967, edition of the *Oratorum Romanorum Fragmenta* prepared by Henrica Malcovati; fragment 203, which is itself quoted in several sections in fragment 173, was excluded. Cato's remaining fragments are found in the 1860 Teubner text of Henricus Iordan, *Praeter Librum de Re Rustica quae exstant*; this volume was reprinted in 1967. Only the

ipsissima verba were included, and where material was duplicated in Malcovati's and Iordan's editions, the former was given precedence. All of the texts of Cato were prepared in machine-readable form by the author.

The text from Sallust was copied through the courtesy of David W. Packard, then at Harvard University, from material prepared by Timothy Murphy. Arbitrarily selected samples of 1000 card images were taken; these comprised chapters 17.6 to 51.8 of the *Catilina* and 24.4 to 56.1 of the *Jugurtha;* the 1968 Teubner text of Kurfess was used as a standard. Similarly, arbitrarily selected sections of 1000 lines were taken from Livy, including the Praefatio to 1.27.9 and 3.46.5 to 4.1.5; this text had been prepared by Packard using the 1914 Oxford Classical Text edited by Conway and Walters. This sample proved to be extremely accurate; in the course of extensive work, only four mistakes were discovered, and two of these involved punctuation.

The texts were standardized to a spelling using *j* and *v* for consonantal *i* and *u*, with the result that much later work was greatly simplified. In at most a very few words like *solvit* is there occasional doubt about which spelling should be used. For the purposes of analysis, all section headings were excluded; material supplied in the texts by editors was included, but that deleted by editors was omitted. Finally, each selection was divided into units of 1000 words for the purpose of allowing tests for consistency within authors.

While Cato, Sallust, and Livy can be considered as providing a range of

material from Republican and Augustan Latin prose, no particular philosophical considerations dictated the selection of these authors. The primary consideration was that of availability, since accurate preparation of a text is an extremely time-consuming undertaking. As the texts of other authors become more available, theoretical considerations can play a larger part in the selection of material for comparison.

The first test to be described is that of word length. As early as 1901, Thomas C. Mendenhall published a graph showing the distribution of words of various lengths in a 5,000-word sample of Caesar; words with five and seven letters are most common.(1) In the present study, a word was defined to consist of the alphabetic characters between two successive spaces; diphthongs and combinations like *ph* have been considered as two characters, not one. Enclitics were not treated as separate entities, nor were abbreviations and numerals expanded; thus, III is considered a three-letter word. It is possible to quarrel with this practice, although it does mark out certain passages of the *De Agri Cultura* clearly distinguishable in other ways also. Totally ignored in compiling figures were words partially missing because of a lacuna. The number of words in each selection is given in Table I, where A.C. indicates the *De Agri Cultura* and O.F. the fragments of Cato other than those from the orations. The second line gives the average word length, and it is immediately clear that the word length in the *De Agri Cultura* tends to be shorter not only than that in the selections from Sallust and Livy but also than that in fragments from Cato. The standard deviation in the *De Agri Cultura* is also slightly less than that in the other works. The last two lines present the average word length in the 1000-word section in each author with the shortest

average and the corresponding value for the section with the longest average. Since it is possible to draw only a single 1000-word sample from the fragments of Cato outside those from the orations, this column in the table has been left empty.

Averages alone can be misleading or at least uninformative; Table II presents the frequency per 1000 words of words of each possible length for each selection. In the extreme left column of each line is given the number of letters being considered in that line. In parentheses after the values for all the selections except that for the "Other Fragments" of Cato, two other numbers are given. The first is the number of times that a word of a given length is found in whatever 1000-word section contains fewest words of this type; the second value is the number of occurrences in the section where most such words are found. As is obvious, the sections in question are not necessarily the same ones from one line to another; the figures in parentheses merely serve to give some idea of the range of values observed.(2)

The short average word length in the *De Agri Cultura* can be attributed partly to the abundance of numerals in the text. The number I, however, occurs only 95 times, and this numeral is found far more frequently than others. Even if all the one-letter words were omitted from the text, the average for the *De Agri Cultura* would be 5.50 letters, less than that for any other selection. Clearly, numerals are not the sole cause of the short average word length in the *De Agri Cultura*. A reverse question arises about the longer average word length in the *Origines* and miscellaneous fragments: to what extent do individual odd words for which a fragment

has been quoted distort the totals ? An example might be *praeterpropter* (Iordan, *Incertorum Librorum Reliquiae,* 53). Such words probably have a small over-all effect, for, if all words over twelve letters were omitted from the "Other Fragments," the average word length would remain 5.80 letters, still higher than those for Cato's other works.

Word length distributions can vary considerably from one section of an author to another. If χ^2 texts are run comparing pairs of sections within even the most uniform of the possible divisions, Book I of Livy, two of the thirty-six results fall above the level which would be exceeded on a chance basis only 1% of the trials.(3) Within the *De Agri Cultura,* the proportion is far higher. If the narrative and the "quoted" speeches are separated in the two later authors, the speeches have a consistently shorter average word length. At the same time, there seems to be an avoidance of two-letter words in the speeches; perhaps this occurs because such words, often pronouns or the conjunction *et*, are avoided in an "oral" situation as easily lost or inappropriate. In the narrative sections alone, a greater homogeneity is evident than in the texts as a whole. While the speeches in Sallust and Livy have a shorter average length than the narrative, the *De Agri Cultura* has a shorter average than the oratical fragments of Cato; following what perhaps is the normal pattern, the "Other Fragments" have a longer average word length than the fragments of the orations. Is the apparently anomalous situation of the *De Agri Cultura* due entirely to the presence of numerals and abbreviations ?

When the distributions of word lengths in pairs of selections are compared by χ^2 tests, a difference significant at the 1% level is indicated for every

combination of different authors. A similar difference appears in the two comparisons between the *De Agri Cultura* and the different sets of Cato's fragments. These differences may or may not be meaningful, since there is also considerable variation within authors. It is none the less interesting to note that the only comparison of two selections which does not yield a value significant at the 1% level is that between the fragments from Cato's orations and those from his other works.

Rather more rewarding than data on word lengths is that on sentence lengths, yet sentence lengths are more likely to vary from one edition of an author to another because of editorial practice. Spelling variations like *sit/siet* are comparatively rare and would occur relatively infrequently, but punctuation can vary considerably even between two successive editions of the same work, as a comparison of the opening sections of the *De Agri Cultura* in Goetz's 1922 Teubner edition and Mazzarino's 1962 Teubner edition would indicate. Here, a sentence has been defined to consist of the words between any two major marks of punctuation: included are periods, colons, semicolons, question marks, exclamation points, dashes, opening and closing quotation marks, and even opening and closing parentheses; the last are rare. This definition eliminates apparent differences caused by the use of a period in one edition and a colon in another; there still remain other possibilities for variation, and their effect is uncertain. It might also be possible the quarrel with the inclusion of quotation marks among sentence delimiters, yet this choice leads to informative results. Fragments have been considered as closing with a period where they could reasonably be construed as an independent and complete sentence; any sentence containing a lacuna has been ignored.

Since periods were not put after abbreviations within a sentence when the texts were prepared, the totals are not distorted by the occurrence of abbreviations.

Table III presents the number of sentences in each selection and the average sentence length in words with the standard deviation, following these conventions. The last two lines of the table give the shortest and longest averages found in any single 1000-word section of a selection; here, a sentence has been considered as falling entirely within the section in which it ends. It is immediately obvious that the later authors tend to use longer sentences than are found in the *De Agri Cultura*. Despite the fact that some long sentences might have been truncated by partial quotation, the fragments of Cato also contain longer sentences on the average than the *De Agri Cultura*.

Again, averages alone can be misleading and conceal much information. Table IV gives the number of sentences of each length in the different selections; the figures have been standardized to the number of occurrences per hundred sentences. As in Table II, the figures in parentheses present the minima and maxima for individual sections, also standardized. Part of the explanation for the difference in the average sentence lengths of the authors becomes clear from this table: in Sallust, 20.6% of the sentences contain twenty-five words or more; the corresponding figure for Livy is 12.7%, while the maximum for any selection from Cato is 3.0%. Despite the fact that the longest sentence found, 74 words, occurs in a list in the *De Agri Cultura*, this work contains the fewest long sentences, 2.6%. Although the paucity of long sentences in the fragments may be partially

due to the nature of their transmission, no such explanation is valid for the *De Agri Cultura*.

Interesting also are the large number of one-word sentences in Livy. An inspection of the text shows that Livy normally introduces "quotations" with *inquit*; in these selections, he does so twenty-nine times. The word is much less used by Sallust, and, with the exception of three occurrences in the fragments, all forms of the verb are totally avoided by Cato. When Livy uses *inquit*, it is invariably placed in isolation after the opening of the quotation, many times after the first word, thus accounting for at least one of the single-word sentences each time it is found. Here, an obvious variation in sentence lengths serves to underline the use of a stylistic device.

When sentence-length distributions of pairs of sections within selections are compared by the χ^2 test, the *De Agri Cultura* is shown to be quite uniform with the exception of the section from chapter 10.4 to 18.2, which contains numerous lists and has the longest average sentence length of any section in the work. In the 105 comparisons within the *De Agri Cultura* not involving this section, only one result significant at the 1% level is found. In the ten comparisons between sections of Sallust's *Catilina*, no significant value is found, but in the fifteen comparisons in the *Jugurtha,* two results are significant; both of these involve the section between chapters 45.2 and 51.1. In Livy, the comparisons within Book I yield numerous results significant at the 1% level, while none of those within Book III do. Since the number of sentences involved in each comparison is relatively small, it may be that a few short-term variations

can distort the distributions for one section. When the totals for the different works of Sallust are compared, the result is not significant, just as it is not when the totals for the two books of Livy are considered. The suggestion is that sentence lengths within an author may be consistent, given sufficient text.(4) When comparisons between authors are made, the selections from Cato are grouped together as not significantly different, but every comparison between authors produces a result significant at the 1% level. In Latin, then, sentence lengths appear to be able to fulfill the requirements of a test to discriminate between authors.

A third approach to gathering information is a consideration of the variety of vocabulary. Here, the distributions are based on the number of forms that occur with a given frequency in a specific section of fixed length; the fact that one word may be used a certain number of times in one section and another word the same number of times in another section is not material. As in the treatment of word lengths, enclitics have not been considered as separate words. Also, the forms in the text have not been normalized to lemmas; it is not unreasonable to hold that the connotations of *ager* as the subject of a sentence may be somewhat different from those of *agrum* as an object. At the same time, homonymous forms have not been differentiated; there can be some grounds for doubt on occasion whether a dative or ablative plural was intended with such words as *eripio*.

A useful statistic for measuring vocabulary variety is entropy, which has been described by Etienne Evrard.(5) He presents as a formula $-\Sigma p_i.\log(p_i)$, where p_i is the probability of the occurrence of the ith item; for purposes

of analyzing vocabulary, p_i is found by dividing the number of times that a form occurs by the number of words in the text. It seems useful to refine Evrard's formula somewhat. The result can be divided by the logarithm of the length of the text and then put on a 0 to 100 scale, so that texts of different lengths can be more easily compared. None the less, care must be used if the selections being compared are of markedly different lengths; the absurd case is the comparison of a sample so short that no forms are repeated with a selection of more normal length, for the first text would inevitably have a entropy of 100. To avoid the effects of length on the value of the entropy, only the samples containing a full 1000 words have been considered in treating vocabulary; the short sections remaining at the end of each selection have been discarded. A smaller entropy indicates generally greater repetitiveness, whether it be because only a single form is found very many times or because a number of forms are found relatively frequently.

Table V presents certain figures on the variety of vocabulary in the full 1000-word sections of the five selections. The first four lines are concerned with the number of different forms in the various sections. The first two lines give the average of this number within each selection and the standard deviation from this average; the next two lines present the lowest and highest figures encountered to give some idea of the range. Since there is only one 1000-word section obtainable from the "Other Fragments" of Cato, certain lines have been left blank for this selection. The second group of four lines presents the same kind of results for the number of forms repeated within a single section; regardless of how many times a form may be found elsewhere in a work, if it does not occur at

least twice within a single section, it does not add to the total of repeated forms. The ninth line of the table gives the maximum number of times that some one form is repeated within a single section in a selection. The very high figure for the *De Agri Cultura*, 62, is due to the not surprising repetition of the numeral *I* in the lists between chapter 10.4 and 18.2; in the fragments from Cato's orations, the word most repeated is *atque,* although this conjunction is found only six times in the entire *De Agri Cultura*. The most repeated form in the other fragments of Cato is *in*, a fact more in line with results from later authors.

The last four lines of Table V are concerned with entropy. The low average for the *De Agri Cultura* suggests a greater repetitiveness in the work, a suggestion which is supported by the figures earlier in the table. The section with the highest entropy is the first, from the beginning to chapter 5.3; a reasonable inference is that there may have been more concern for variety shown in composing the opening than in preparing the rest of the *De Agri Cultura*. The lowest entropy is found in the section from chapters 72.1 to 89.1, in which there are many recipes and the frequent use of the phrase ... *sic facito*. Both Sallust and Livy evidence much less variation in the entropy values than is seen in the *De Agri Cultura*, and these values are higher; the latter result is not unreasonable, since these authors were presumably writing with more concern for style. Noticeable also is the high amount of variation in the *Origines* and miscellaneous fragments of Cato and, to a lesser extent, in his oratorical fragments. Here particularly, the preservation of fragments because of the use of odd words or forms may be producing an observable effect. This hypothesis is supported by the fact that fragments 112 to 177 have a

lower entropy (87.48) than fragments 17 to 112 (91.81); the former contain the longest passages to be found in Cato's fragments, those from the speeches "Pro Rhodiensibus" and "De Sumptu Suo." Indeed, it is primarily in these fragments that reasonably continuous text is to be found.

For many kinds of detailed comparison, it often is preferable to consider the whole range of values rather than summarizing statistics. Table VI gives the averages over all the full 1000-word sections in a selection of the number of forms occurring once, twice, and up to twenty times in a single section; the last line includes all the forms occurring more than twenty times. The figures in parenthesis are minima and maxima. Again, in compiling this table, only the number of times that a word occurs in a specific section is considered. The data support the contention that in the *De Agri Cultura* relatively few words are used only once in a section and relatively many are used very often.

The variety in vocabulary remains quite consistent within the *De Agri Cultura* when measured by the application of the χ^2 test to the vocabulary distributions as presented in Table VI. The exception is the material up to chapter 10.4, the first two 1000-word sections. It has already been remarked that the first of these sections could be singled out; the second has the second highest entropy of the *De Agri Cultura*, 87.57, more than a full point higher than the third-ranking value. Within the two later authors, the uniformity of the vocabulary distributions within each author is notable; comparison between all the possible pairs of distributions within each major division, *Catilina, Jugurtha,* Book I of Livy, and Book III, produce no result significant at the 1% level. When, however, the

distributions in the different selections are compared,(6) the *De Agri Cultura* is differentiated from every other selection. The fragments of the orations of Cato are differentiated from the *De Agri Cultura* but from no other selection, while the "Other Fragments" are significantly different from every other selection save the oratorical fragments. No significant difference is shown between the selections from Sallust and Livy.

On this basis, vocabulary distribution as defined here would appear to meet at least partially the criteria for a useful test of authorship; with the exception of an obviously mixed work like the *De Agri Cultura*, it remains uniform within a text while to some extent differentiating between material of different authors. Questions remain. To what extent would different sectioning change the results ? For instance, if all of chapter 10 were included with the section from chapters 10.4 to 18.2, the number of occurrences of *I* would be even higher than it is. Is the difference among the texts assigned to Cato due to the fact that they are composed in different genres? Is it to be traced in part to the fragmentary nature of some of the material? A full study of Cicero's various kinds of writing might suggest an answer to the first question, but the second remains more intractable.

Prose rhythm is the last topic to be treated here. Before information about it could be collected, however, additional preparation of the text was necessary. The closing words of each sentence as defined above formed a natural body of material for study, and these were isolated; if a lacuna was encountered where it would affect the analysis, the sentence was disregarded. For purposes of comparison, the texts were recopied

with as many full words as possible on lines of a fixed length; the last five words of each line were extracted, save where a chapter boundary, fragment division, or lacuna occurred within five words of the line's end. The last word of each line, which would tend to be short, was considered only as it affected the preceding word, and the last syllables of the preceding four words were studied. In both kinds of material, only the last five syllables remaining after allowance for elision were considered in the analysis.(7) When a word group with fewer than five syllables was encountered, it was expanded to include enough words to contain five syllables unless, in the sentence ends, a sentence break was encountered; if any of the conditions was encountered which would have caused the word group to be discarded in the original isolating process, the group was eliminated. Thus, there are no arbitrary word groups with fewer than five syllables treated.

All numerals and abbreviations were then expanded, and the syllables in the revised text were marked through the use of a dictionary stored in the computer(8) where possible and through input into an interactive program otherwise. Lewis and Short's *A Latin Dictionary* was used as a standard; where there was an option before a mute followed by a liquid, the syllable was marked as short. In collecting data, elision and ecthlipsis were allowed in every possible instance unless the first word was at the end of a sentence. A mute and liquid in one word did not make position but divided between two words did; final *s* was treated as lengthening a short final syllable if the next word began with any consonant other than *h*. The final syllable of a sentence was not considered *anceps;* Quintilian *(Institutiones Oratoriae,* IX, 4, 93-94) rather than Cicero (*Orator*, 214) was a guide. Particularly as they have been applied to Cato's texts, some of

these decisions may be reconsidered when the results of work presently being done by the author on Plautine metrics become available; in any event, the choices have been applied consistently and should give some indication of what the classical authors heard as the rhythm of earlier Latin prose.

Table VII presents the results of the analysis under these conventions. The extreme left column shows the syllabic patterns; a . has been used to indicate a short syllable, and a - to denote a long, whether by nature or by position. The first figures in each column present the number of occurrences of a pattern per one hundred sentence ends with five or more syllables; those with fewer than five syllables have not been included in the total used to derive these distributions, although a comparable figure for them has been included at the head of the table. The figures in parentheses present the same data for the patterns in the arbitrarily selected word groups. The total under consideration are :

	Cato D.A.	Cato Orat.	Cato O.F.	Sallust (sel.)	Livy (sel.)
Full clausulae	1734	238	107	661	1136
Under five syllables	37	2	0	4	60
Arbitrary word groupes	1747	242	126	1068	1501

The figures given for full clausulae do not include the sentences with fewer than five syllables. The large number of Livy's sentences falling into the latter category can be traced primarily to his use of *inquit.*

While Table VII presents the full data for each selection, Table VIII gives the ranking of the most common patterns in each selection for both the clausulae and the arbitrary word groups. The figures in parentheses give the actual number of times a pattern occurs; in fact, four patterns could fall in the third place in an ordered ranking of the clausulae of Cato's "Other Fragments" : ...-., -.---, --...., and --.-. each occur five times. In the arbitrary word groups from Cato's oratorical fragments, -.--- occurs just as often as -.---, the pattern listed as third.

The patterns in the arbitrary word groups give an indication of the extent to which clusters of long syllables predominate in Latin: the pattern ----- is among the three most common in each selection. Only comparatively rarely does more than one short syllable occur consecutively in these common patterns from the arbitrary word groups. Since these selections can be taken as an arbitrary but fairly large sample of Latin prose, a reasonable inference is that Plautine spondees, iambs, and trochees are more suited to the natural syllabic rhythm of Latin than Vergilian dactyls.

Short syllables occur somewhat more frequently in the common patterns of the clausulae, partly because the following word is not allowed as lengthening a short syllable at the end of a sentence. While there is no ready explanation for the predominance of particular patterns in the other selections, all three of the most common patterns in the *De Agri Cultura* would be found in sentences ending with a so-called future imperative; the exact pattern would vary with the root and conjugation of the verb used. Inspection of the text shows that this imperative is extremely common in the *De Agri Cultura*, occurring over 800 times in

third conjugation verbs alone; often, it is found at the end of a sentence.

When the distribution of patterns from the ends of sentences is compared with that from the arbitrary word groups in the same selection, in nearly every instance there is differentiation at the 1% level according to the χ^2 value. The one exception is Cato's oratorical fragments; here, there are comparatively few sentences under consideration, and some distinctions may be blurred by the grouping necessary to perform the test. It is also possible is that some fragments may have been erroneously considered as comprising full sentences. The general differentiation between the two distributions should not be surprising after a consideration of the fact that, again with the exception of Cato's oratory, none of the patterns most common in the sentence ends is found among those most frequently occurring in the arbitrary word groups of the same selection. Even in the apparently inartistic prose of the *De Agri Cultura*, certain rhythmic patterns seem to have been considered as more appropriate at the ends of sentences. If the final syllable of a pattern is considered *anceps*, there would still be differentiation at the 1% level between the clausulae and arbitrary word groups of the *De Agri Cultura*, Sallust, and Livy; there would continue to be no significant difference indicated for Cato's oratorical fragments, and there no longer would be a significant difference in his "Other Fragments."

When comparisons are made between the distributions of the arbitrary word groups from the different selections, the only one which produces a result significant at the 1% level is that between the *De Agri Cultura* and the fragments from Cato's orations; this value is significant at the 1% level

but not at the 0.1%. A reasonable conclusion is that the basic nature of the rhythm of Latin prose did not have change materially in the 160 years between Cato and Livy, at least as that rhythm was heard by Livy. A markedly different conclusion is suggested by the results of comparisons of the patterns of sentence ends in the different selections. When the 1% level of probability is used as a criterion, the *De Agri Cultura* is shown as unlike every other selection, including the two sets of Cato's fragments. Perhaps because of the paucity of material, the "Other Fragments" are not shown as significantly different from the selections outside the *De Agri Cultura.* The remains of Cato's oratory are differentiated from the selections from the later authors, which in turn differ significantly from each other. If the final syllable of a sentence were considered *anceps*, these results would be changed only in the comparison between the *De Agri Cultura* and the "Other Fragments" of Cato, which no longer would be shown as significantly different. There is insufficient material within a single 1000-word section to allow meaningful tests for consistency within an author on this basis, but when the distributions from the two works of Sallust are compared, the χ^2 value resulting is not significant at the 1% level; the same statement is true for the distributions from the two books of Livy. Apparently the distribution of the patterns of clausulae in each author may form a part of his stylistic individuality even though the rhythm of the language as a whole remains more uniform.

The approaches to analyzing Latin prose which have been suggested here are by no means the only ones which can be facilitated by the use of computers, nor have they been fully developed. Among many others

which would be equally feasible might be mentioned the study of rhythm at the beginnings of sentences and consideration of sound patterns and stress accents. Conditions of word juncture, which are to an appreciable extent under a writer's control, could be analyzed. Totally ignored here has been any consideration of grammar; syntactic patterns at the beginnings and ends of sentences could provide material from which interesting conclusions might be drawn. Studies of specific conjunctions and other individual lexical items are also possible.

In developing further the analyses on the basis of the approaches used here, it would be possible to use other statistical techniques besides the χ^2 test; discriminatory analysis might suggest what rhythmic patterns vary most significantly between authors. It would be possible to treat fully the differences between the "quoted" material and the narrative in Sallust and Livy and to test whether the texts of Cato show the same differences between oratory and other material. Suggestive entries in tables could be investigated and explanations for them sought, although in some instances explanations might not become evident.

While such extensions could be undertaken, often there would be encountered the problem presently hindering induction of firm conclusions from the data given here : until information is available for other authors, any conclusions should be considered at best tentative. The data now available for Cato, Sallust, and Livy can illuminate some features of the authors' writing, a number of which might equally well be discovered in another fashion. Any suggestions about the differences between authors or about the development of Latin prose must remain hypotheses rather than

conclusions until they can be verified with material from additional writers; the same caution must be applied to inferences drawn from comparisons between works of different genres, even when by the same author. In addition, the very nature of fragments has an unknown effect on results involving this kind of material. None the less, with the present information, some suggestions can be made, and the data itself can be used in comparisons with new information as it becomes available; the material given here also has an inherent value as it is able to draw attention to noteworthy features of the writing of the individual authors.

Dartmouth College Stephen V. F. WAITE

NOTES

1. T. C. Mendenhall : "A Mechanical Solution of a Literary Problem", *The Popular Science Monthly*, LX (December 1901), pp. 97-105. Mendenhall had been making counts of word lengths in English texts nearly twenty years before.

2. These methods and results can be compared with those used by Tore Janson as described in "Word, Syllable, and Letter in Latin", published in *Eranos*, LXV (1967), 1-2, pp. 49-64. For sentences 151 to 250 in the Oxford Classical Text of book XVII of Livy, he gives an average word length of 6.06 letters. Janson has also published "The Problems of Measuring Sentence-Length in Classical Texts" in *Studia Linguistica,* XVIII (1964), pp. 26-36.

3. The actual distributions were used for the χ^2 tests, not the normalized ones given in the tables. Expected values were checked working from the last line as given in the tables to the first. If the expected value for any entry in a line was less than 5, all the entries for that line were grouped with those in the preceding line. If an expected value in the first line was less than 5, the entries were added to those in the following lines until every expected value was at least 5. Because of the large number of small observed values in the tables for syllable patterns, the Yate's correction was applied in the analysis of this feature; it was not used elsewhere. Probabilities for the χ^2 values found were computed by a program prepared by Victor E. McGee of Dartmouth College, who followed a program given on page 131 of Donald J. Veldman:

Fortran Programming for the Behavioral Sciences; New York; Holt, Rinehart and Winston, 1967.

4. Homogeneity in sentence-length distributions has been a basis of work with Greek texts done in collaboration with A.Q. Morton. See, for instance, chapter 5 of A.Q. Morton and James McLeman : *Paul, The Man and the Myth;* London; Hodder and Stoughton; 1966, and M. Levison, A.Q. Morton, and A.D. Winspear: "The Seventh Letter of Plato," *Mind*, LXXVII, 307 (July 1968), pp. 309-325.

5. E. Evrard : "Deux programmes d'ordinateur pour l'étude quantitative du vocabulaire," *Revue*, 1967, 3, pp. 81-95.

6. For these comparisons, the distributions in each section of a selection have been summed; these sums, not the averages, have been used.

7. Although the decision to use patterns of five syllables was reached independently, the number of syllables considered is the same as that used by Leonard Brandwood in studying Greek clausulae, as reported in "Plato's Seventh Letter," *Revue,* 1969, 4, pp. 1-25.

8. The computer used was a General Electric /635 system at Dartmouth College; programming was done in BASIC. The dictionary was prepared as forms were first encountered in analyzing the texts.

TABLE I: AVERAGE WORD LENGTHS IN LETTERS

	CATO A.C.	CATO ORAT.	CATO O.F.	SALLUST (SEL.)	LIVY (SEL.)
NUMBER OF WORDS	15809	2717	1396	11122	15979
AVERAGE WORD LENGTH	5.39	5.74	5.91	5.95	5.89
STANDARD DEVIATION	2.49	2.56	2.62	2.60	2.64
MIN. AVER. FOR 1 SECTION	5.17	5.48	-	5.72	5.59
MAX. AVER. FOR 1 SECTION	5.71	5.96	-	6.08	6.04
	CATO A.C.	CATO ORAT.	CATO O.F.	SALLUST (SEL.)	LIVY (SEL.)

TABLE II: DISTRIBUTION OF WORD LENGTHS IN LETTERS

	CATO DE AGRI CULTURA			CATO ORATIONS			CATO O.F.	SALLUST (SELECTIONS)			LIVY (SELECTIONS)		
1	24.3	(0,	113)	6.6	(4,	8)	3.6	10.6	(0,	27)	6.4	(1,	20)
2	119.9	(96,	177)	79.9	(66,	82)	86.7	96.4	(81,	108)	105.2	(82,	125)
3	107.7	(87,	128)	128.8	(110,	157)	107.4	90.5	(74,	128)	102.1	(76,	118)
4	133.7	(113,	165)	139.1	(132,	153)	139.0	116.8	(91,	144)	107.8	(98,	121)
5	145.2	(109,	175)	164.9	(156,	181)	150.4	145.7	(132,	161)	143.5	(116,	174)
6	154.2	(117,	194)	124.8	(122,	124)	113.2	125.9	(115,	139)	146.3	(128,	173)
7	123.0	(104,	151)	109.7	(90,	125)	133.2	129.9	(106,	156)	123.8	(98,	143)
8	78.0	(53,	109)	97.5	(88,	113)	101.0	113.2	(99,	134)	101.6	(83,	132)
9	52.8	(35,	67)	64.8	(58,	72)	70.2	78.3	(60,	95)	65.8	(49,	83)
10	34.0	(23,	46)	37.2	(27,	44)	45.8	50.4	(35,	67)	47.5	(31,	69)
11	14.5	(7,	30)	26.1	(21,	29)	26.5	22.1	(11,	29)	25.4	(13,	36)
12	8.2	(3,	14)	9.6	(4,	17)	10.0	13.0	(5,	22)	13.5	(6,	24)
13	2.5	(0,	7)	7.0	(5,	6)	8.6	4.8	(2,	10)	6.9	(2,	12)
14	1.0	(0,	3)	2.6	(2,	3)	1.4	1.4	(0,	4)	2.7	(0,	6)
15	0.8	(0,	6)	1.1	(0,	1)	1.4	0.6	(0,	2)	0.8	(0,	4)
16	0.1	(0,	1)	0.4	(0,	0)	1.4	0.4	(0,	1)	0.5	(0,	2)
17	0.0	(0,	0)	0.0	(0,	0)	0.0	0.0	(0,	0)	0.1	(0,	1)
18	0.0	(0,	0)	0.0	(0,	0)	0.0	0.0	(0,	0)	0.0	(0,	0)
19	0.0	(0,	0)	0.0	(0,	0)	0.0	0.0	(0,	0)	0.0	(0,	0)
20	0.0	(0,	0)	0.0	(0,	0)	0.0	0.0	(0,	0)	0.1	(0,	2)

	CATO DE AGRI CULTURA	CATO ORATIONS	CATO O.F.	SALLUST (SELECTIONS)	LIVY (SELECTIONS)

TABLE III: AVERAGE SENTENCE LENGTHS IN WORDS

	CATO A.C.	CATO ORAT.	CATO O.F.	SALLUST (SEL.)	LIVY (SEL.)
NUMBER OF SENTENCES	1781	238	100	665	1196
AVERAGE SENTENCE LENGTH	8.87	9.81	10.23	16.72	13.36
STANDARD DEVIATION	4.20	4.34	3.87	6.08	5.78
MIN. AVER. FOR 1 SECTION	7.44	9.20	-	14.31	9.64
MAX. AVER. FOR 1 SECTION	14.03	10.27	-	20.59	18.87
	CATO A.C.	CATO ORAT.	CATO O.F.	SALLUST (SEL.)	LIVY (SEL.)

TABLE IV: DISTRIBUTION OF SENTENCE LENGTHS IN WORDS

	CATO DE AGRI CULTURA	CATO ORATIONS	CATO O.F.	SALLUST (SELECTIONS)	LIVY (SELECTIONS)
1	1.3 (0, 5)	0.0 (0, 0)	0.0	0.5 (0, 3)	4.3 (0,12)
2	3.7 (1, 8)	4.2 (2, 8)	0.0	0.5 (0, 2)	2.4 (0, 5)
3	7.3 (3,11)	5.0 (2, 8)	1.0	0.6 (0, 3)	3.1 (0, 7)
4	9.1 (6,14)	7.1 (3, 9)	8.0	1.4 (0, 6)	5.4 (2, 8)
5	9.4 (1,14)	10.1 (10,11)	11.0	3.2 (0, 7)	6.0 (2,10)
6	9.9 (3,16)	11.8 (11,17)	7.0	6.3 (0,10)	5.3 (0,13)
7	10.6 (4,16)	6.7 (5, 5)	14.0	5.4 (0,12)	5.8 (2,12)
8	8.0 (5,13)	8.8 (7,10)	7.0	6.3 (2,13)	4.9 (0,11)
9	6.8 (5, 9)	3.4 (0, 7)	11.0	6.8 (2,20)	5.5 (2,12)
10	4.9 (1, 8)	6.3 (2, 7)	8.0	5.0 (1,11)	5.4 (2, 9)
11	4.9 (3, 7)	7.6 (7, 9)	6.0	4.8 (0, 9)	5.9 (3,11)
12	4.8 (2, 8)	4.2 (2, 7)	5.0	3.0 (0, 6)	4.4 (0,10)
13	3.7 (1, 8)	3.4 (3, 5)	3.0	5.1 (0, 9)	4.0 (0,11)
14	3.4 (1, 6)	1.7 (0, 2)	2.0	4.1 (0,10)	3.4 (1, 7)
15	2.4 (0, 5)	3.4 (3, 5)	2.0	2.6 (0, 8)	2.5 (0, 7)
16	1.6 (0, 4)	2.1 (1, 5)	2.0	4.5 (0,11)	2.8 (0, 5)
17	1.3 (0, 5)	2.1 (1, 3)	2.0	2.7 (0, 7)	2.5 (0, 5)
18	1.5 (0, 4)	2.1 (1, 2)	1.0	3.2 (0, 7)	2.6 (1, 5)
19	0.8 (0, 3)	1.3 (0, 3)	1.0	2.9 (0, 7)	3.2 (0, 7)
20	0.6 (0, 3)	0.4 (0, 1)	1.0	1.8 (0, 4)	1.5 (0, 4)
21	0.4 (0, 2)	2.1 (1, 3)	2.0	3.0 (0, 6)	1.8 (0, 5)
22	0.4 (0, 2)	1.3 (0, 2)	1.0	2.7 (0,10)	1.6 (0, 7)
23	0.4 (0, 2)	0.4 (0, 1)	2.0	2.0 (0, 6)	1.8 (0, 5)
24	0.2 (0, 2)	1.3 (0, 2)	0.0	1.8 (0, 8)	1.2 (0, 4)
25	0.4 (0, 2)	0.8 (0, 1)	0.0	1.8 (0, 6)	1.8 (0, 5)
26	0.1 (0, 1)	0.4 (0, 0)	1.0	1.1 (0, 4)	1.3 (0, 7)
27	0.3 (0, 3)	0.0 (0, 0)	0.0	2.0 (0, 6)	0.8 (0, 4)
28	0.1 (0, 1)	0.4 (0, 1)	0.0	1.4 (0, 4)	0.5 (0, 2)
29	0.1 (0, 1)	0.4 (0, 1)	0.0	1.2 (0, 3)	0.6 (0, 2)
30	0.1 (0, 1)	0.8 (0, 2)	1.0	1.5 (0, 5)	1.0 (0, 4)
31	0.0 (0, 0)	0.4 (0, 0)	0.0	1.1 (0, 4)	0.4 (0, 2)
32	0.2 (0, 1)	0.0 (0, 0)	0.0	0.9 (0, 2)	0.4 (0, 3)
33	0.1 (0, 1)	0.0 (0, 0)	0.0	1.2 (0, 3)	0.9 (0, 5)
34	0.0 (0, 0)	0.0 (0, 0)	0.0	0.8 (0, 4)	0.2 (0, 2)
35	0.1 (0, 1)	0.0 (0, 0)	0.0	0.3 (0, 2)	0.6 (0, 5)
36	0.1 (0, 1)	0.0 (0, 0)	0.0	0.8 (0, 4)	0.4 (0, 2)
37	0.1 (0, 1)	0.0 (0, 0)	0.0	0.5 (0, 2)	0.2 (0, 2)
38	0.1 (0, 1)	0.0 (0, 0)	0.0	0.9 (0, 2)	0.5 (0, 3)
39	0.1 (0, 1)	0.0 (0, 0)	0.0	0.8 (0, 2)	0.5 (0, 4)
40	0.0 (0, 0)	0.0 (0, 0)	0.0	0.2 (0, 1)	0.4 (0, 6)
OVER 40	0.7 (0, 8)	0.0 (0, 0)	1.0	4.1 (0,10)	2.2 (0, 9)

| | CATO DE AGRI CULTURA | CATO ORATIONS | CATO O.F. | SALLUST (SELECTIONS) | LIVY (SELECTIONS) |

TABLE V: VOCABULARY VARIETY PER 1000-WORD SECTION

	CATO A.C.	CATO ORAT.	CATO O.F.	SALLUST (SEL.)	LIVY (SEL.)
AVERAGE NUMBER OF FORMS	538.73	661.00	736.00	667.36	684.67
STANDARD DEVIATION	48.01	56.00	-	20.73	18.58
MINIMUM IN 1 SECTION	444	605	-	636	650
MAXIMUM IN 1 SECTION	625	717	-	716	714
AVERAGE NUMBER REPEATED	154.20	136.00	101.00	147.27	135.13
STANDARD DEVIATION	8.29	4.00	-	7.79	9.60
MINIMUM IN 1 SECTION	135	132	-	134	111
MAXIMUM IN 1 SECTION	170	140	-	161	150
MAX. REPETITION OF 1 FORM	62	31	22	31	30
AVERAGE ENTROPY	84.59	89.64	91.82	90.30	90.69
STANDARD DEVIATION	2.18	2.16	-	0.69	0.73
MINIMUM FOR 1 SECTION	80.12	87.48	-	89.11	89.51
MAXIMUM FOR 1 SECTION	87.88	91.81	-	91.92	91.89
	CATO A.C.	CATO ORAT.	CATO O.F.	SALLUST (SEL.)	LIVY (SEL.)

TABLE VI: DISTRIBUTION OF FORMS PER SECTION

	CATO DE AGRI CULTURA	CATO ORATIONS	CATO O.F.	SALLUST (SELECTIONS)	LIVY (SELECTIONS)
1	384.5(291,490)	525.0(465,585)	635	520.1(489,582)	549.5(502,596)
2	77.3(61, 87)	75.5(72, 79)	57	85.1(74, 95)	76.5(51, 92)
3	30.3(22, 36)	26.0(24, 28)	16	28.0(22, 37)	25.2(19, 34)
4	14.1(7, 22)	11.5(11, 12)	10	13.3(6, 25)	13.4(5, 22)
5	9.2(4, 13)	7.5(5, 10)	5	6.1(2, 9)	6.7(4, 11)
6	5.4(2, 10)	3.0(1, 5)	3	4.5(2, 8)	3.5(2, 7)
7	3.3(1, 8)	3.5(3, 4)	1	3.0(1, 6)	2.0(0, 4)
8	3.2(0, 7)	2.5(1, 4)	2	1.5(0, 3)	1.3(0, 4)
9	1.8(0, 3)	0.5(0, 1)	1	1.5(0, 3)	1.1(0, 2)
10	1.4(0, 4)	0.5(0, 1)	0	0.6(0, 2)	0.8(0, 2)
11	1.4(0, 6)	0.0(0, 0)	0	0.5(0, 2)	1.2(0, 2)
12	1.5(0, 4)	0.5(0, 1)	1	0.5(0, 2)	0.7(0, 2)
13	0.7(0, 2)	2.0(0, 4)	2	0.5(0, 2)	0.6(0, 2)
14	0.8(0, 2)	0.0(0, 0)	2	0.1(0, 1)	0.3(0, 2)
15	0.4(0, 2)	1.5(0, 3)	0	0.4(0, 1)	0.1(0, 1)
16	0.1(0, 1)	0.0(0, 0)	0	0.3(0, 2)	0.3(0, 1)
17	0.2(0, 2)	0.5(0, 1)	0	0.2(0, 1)	0.1(0, 1)
18	0.3(0, 1)	0.0(0, 0)	0	0.2(0, 1)	0.1(0, 1)
19	0.1(0, 1)	0.0(0, 0)	0	0.3(0, 1)	0.2(0, 1)
20	0.2(0, 1)	0.0(0, 0)	0	0.0(0, 0)	0.2(0, 1)
OVER 20	2.5(1, 5)	1.0(1, 1)	1	0.9(0, 2)	0.8(0, 2)

	CATO DE AGRI CULTURA	CATO ORATIONS	CATO O.F.	SALLUST (SELECTIONS)	LIVY (SELECTIONS)

TABLE VII: DISTRIBUTION OF CLAUSULA PATTERNS
(WITH OTHER SYLLABLE PATTERNS)

	CATO D.A.	CATO ORATIONS	CATO O.F.	SALLUST (SELECTIONS)	LIVY (SELECTIONS)
UNDER FIVE SYLLABLES	2.1 (0.0)	0.8 (0.0)	0.0 (0.0)	0.6 (0.0)	5.3 (0.0)
.....	1.0 (0.6)	0.4 (0.0)	0.0 (0.0)	2.0 (0.5)	0.5 (0.7)
....-	1.3 (2.1)	0.8 (1.2)	0.0 (0.8)	0.6 (1.8)	0.6 (1.9)
...-.	1.5 (0.8)	3.8 (2.9)	4.7 (0.8)	1.4 (1.3)	0.6 (0.7)
...--	0.8 (2.3)	1.3 (2.5)	0.9 (0.8)	0.9 (2.2)	0.5 (2.1)
..-..	1.8 (1.5)	2.5 (2.5)	0.9 (0.8)	2.7 (1.3)	0.6 (1.0)
..-.-	2.2 (3.8)	0.8 (0.8)	0.9 (2.4)	2.0 (3.0)	0.4 (2.6)
..--.	2.3 (1.7)	1.7 (0.8)	0.9 (0.8)	4.1 (1.1)	3.1 (1.4)
..---	1.7 (3.2)	2.1 (3.7)	1.9 (4.8)	1.7 (3.7)	1.7 (3.8)
.-...	2.7 (1.5)	2.5 (1.7)	1.9 (0.0)	2.3 (1.3)	2.7 (1.0)
.-..-	5.9 (3.8)	1.7 (3.3)	1.9 (0.8)	1.5 (3.3)	2.3 (2.7)
.-.-.	3.1 (1.8)	4.2 (2.5)	2.8 (2.4)	1.2 (0.7)	1.3 (1.5)
.-.--	1.7 (3.9)	2.5 (4.1)	1.9 (4.0)	0.5 (3.7)	1.4 (4.1)
.--..	3.5 (2.5)	5.0 (4.5)	1.9 (3.2)	6.2 (2.7)	3.9 (1.7)
.--.-	4.4 (4.8)	1.3 (2.5)	0.9 (4.8)	3.5 (4.7)	3.5 (4.7)
.---.	3.0 (2.3)	5.0 (1.2)	3.7 (3.2)	7.4 (2.1)	6.5 (2.9)
.----	3.6 (4.6)	3.4 (9.5)	3.7 (4.8)	4.4 (5.8)	4.0 (4.5)
-....	1.7 (1.5)	2.1 (1.7)	2.8 (2.4)	1.1 (1.6)	1.8 (1.1)
-...-	2.0 (2.9)	0.8 (1.2)	1.9 (2.4)	1.1 (2.9)	1.6 (1.8)
-..-.	2.7 (1.7)	4.2 (1.2)	3.7 (2.4)	7.7 (2.2)	4.2 (1.3)
-..--	2.0 (3.1)	2.9 (6.6)	1.9 (1.6)	2.9 (3.9)	2.4 (3.7)
-.-..	2.7 (2.2)	2.1 (1.2)	3.7 (1.6)	2.3 (1.9)	4.8 (1.9)
-.-.-	3.7 (4.7)	2.5 (1.7)	0.9 (4.8)	0.9 (3.7)	3.3 (4.3)
-.--.	3.6 (2.6)	6.7 (1.7)	2.8 (1.6)	2.9 (2.9)	4.9 (3.4)
-.---	2.9 (4.5)	2.5 (6.6)	4.7 (3.2)	1.4 (5.2)	2.8 (5.6)
--...	4.2 (2.1)	2.5 (2.1)	4.7 (1.6)	7.1 (1.7)	6.3 (1.7)
--..-	6.5 (5.7)	2.1 (3.7)	1.9 (6.3)	3.0 (5.1)	4.1 (4.9)
--.-.	4.0 (2.6)	2.9 (1.2)	4.7 (4.0)	3.5 (2.7)	5.0 (3.4)
--.--	3.0 (6.0)	3.8 (4.5)	3.7 (9.5)	1.7 (5.3)	4.9 (7.3)
---..	4.2 (2.3)	7.1 (2.9)	13.1 (5.6)	5.7 (2.7)	1.8 (1.7)
---.-	6.5 (6.9)	2.9 (5.4)	3.7 (9.5)	3.2 (8.5)	1.2 (7.5)
----.	4.7 (3.0)	7.6 (2.9)	14.0 (0.0)	8.9 (3.3)	11.9 (3.4)
-----	5.0 (6.9)	8.0 (11.6)	2.8 (9.5)	4.5 (7.0)	5.3 (9.5)
	CATO D.A.	CATO ORATIONS	CATO O.F.	SALLUST (SELECTIONS)	LIVY (SELECTIONS)

TABLE VIII: MOST COMMON RHYTHMIC PATTERNS

	CATO D.A.	CATO ORATIONS	CATO O.F.	SALLUST (SEL.)	LIVY (SEL.)
CLAUSULAE					
FIRST	--.-(113)	-----(19)	----.(15)	----.(59)	----.(135)
SECOND	---.-(113)	----.(18)	---..(14)	-..-.(51)	.---.(74)
THIRD	.--.-(103)	---..(17)	...-.(5)	.---.(49)	--...(72)
ARBITRARY WORD GROUPS					
FIRST	---.-(120)	-----(28)	--.--(12)	---.-(91)	-----(143)
SECOND	-----(120)	.----(23)	---.-(12)	-----(75)	---.-(113)
THIRD	--.--(105)	-.--(16)	-----(12)	.----(62)	--.--(109)
	CATO D.A.	CATO ORATIONS	CATO O.F.	SALLUST (SEL.)	LIVY (SEL.)