# THE PROBLEM OF THE AUTHENTICITY OF THE *CINQUIESME LIVRE DE PANTAGRUEL* : A QUANTITATIVE STUDY

by George A. Petrossian



## Chapter I
## INTRODUCTION

Controversy over the authenticity of the *Cinquiesme Livre de Pantagruel* has existed almost since its initial publication. It was published in two installments nine years after the assumed date of Rabelais's death.[1]

The date of April 9, 1553, is generally accepted, but it also is disputed. See the following articles. Abel Lefranc, "Conjecture sur la date de la naissance de Rabelais," *Revue des Etudes Rabelaisiennes*, VI (1908), pp. 265-270. By V. L. Saulnier, "Sur la date de naissance de Rabelais," in *Bibliothèque d'Humanisme et Renaissance*, VII (1945), pp. 245-246. And by Floyd Gray, "*Adolescens* and the Date of Rabelais' Birth," *Modern Language Notes*, LXXVI (1961), pp. 733-735.

The first part appeared in 1562 under the title of *L'Isle Sonante* ..., acknowledging Rabelais as the author, but without the name and the place of the publisher.[2]

The full title of the book is *L'Isle Sonante, par M. François Rabelays, qui n'a point encores esté imprimée ne mise en lumière : en laquelle est continuée la navigation faicte par Pantagruel, Panurge et autres ses Officiers.* Imprimé nouvellement. 1562. At the end of the book it is printed : Fin du voyage de l'Isle Sonante.

It had sixteen chapters and was a continuation of the sea theme and search for the "Dive Bouteille," which had begun in the *Quart Livre de Pantagruel*. In 1564, another book was published under the heading of *Le Cinquiesme et dernier livre* ...[3]

The full title of the book is *Le Cinquiesme et dernier livre des faicts et dicts heröiques du bon Pantagruel, composé par M. François Rabelais, Docteur en Medecine. Auquel est contenu la visitation de l'Oracle de la Dive Bacbuc, et le mot de la Bouteille : pour lequel avoir, est entrepris tout ce long voyage. Nouvellement mis en lumière*, 1564.

Once again there was neither name nor place of publication, and its authorship was credited to Rabelais. This new book had incorporated the first fifteen chapters of the *Isle Sonante* version with some slight variations of words and sentences, and added to it 32 new chapters, a Prologue, and, in addition, had a quatrain at the end of the book which began, "Is Rabelais dead ? Here is another book...," signed NATURE QUITE.[4]

A. Dupont suggested that NATURE QUITE is the anagram of a physician friend of Rabelais known as Jean Turquet. "Note sur le Quatrain de Nature Quite," *Revue du Seizième Siècle*, XII (1925), p. 403. Louis Cons in another article claimed that NATURE QUITE was the anagram of the well-known traveler Jean Quentin, whom he thought was the author of the last part of the *Cinquième Livre*, and that Rabelais was the author of *L'Isle Sonante*. *Le Problème du Cinquième Livre, Le Continuateur*, in *Revue Bleue*, April 25, 1914, pp. 524-526. Jean Plattard has rejected entirely Cons' conjecture. See his article in *Revue du Seizième Siècle*, II (1914), pp. 279-282.

This book contained more allegorical descriptions of the voyage, and ended abruptly with the chapters on the "Dive Bouteille." 

Until the Nineteenth Century only these two versions of the *Cinquiesme Livre* were known to exist. In 1840, Paul Lacroix found at the Bibliothèque Nationale a manuscript under the heading of *Cinquiesme Livre de Pantagruel, fragment du Prologue*. It had neither signature nor Rabelais's name, and it was not dated. In this manuscript the Prologue was fragmentary, about one-fourth of the length of the prologue in the book, and it ended in the middle of a sentence. The quatrain by NATURE QUITE, and chapters XXIII and XXIV on the chess tournament entitled respectively "Comment fut, en presence de la Quinte, faict un bal joyeux en forme de tournay" and "Comment les trente deux personnages du bal combatent," were missing. Instead it contained a new chapter entitled "Comment furent les dames Lanternois servies a soupper," and a new, much longer ending. There were also other but minor discrepancies.[5]

For a detailed discussion of these chapters of the manuscript see Lazare Sainean's "Le Cinquième Livre de Rabelais : Son authenticité et ses parties constitutives. In *Problèmes Littéraires du Seizième Siècle* (Paris : Boccard, 1927), pp. 94-97 ; and Arthur Tilley, "Rabelais and the Fifth Book, *Studies in the French Renaissance* (Cambridge : Cambridge University Press, 1922), pp. 85-122.

Paul Lacroix thought that the manuscript was written by Rabelais himself, and considered its existence as sound proof of the authenticity of the *Cinquiesme Livre*[6]

Paul Lacroix, *Journal des Débats*, March 13, 1847.

However, Paulin Paris proved that the penmanship of the manuscript could not have been that of Rabelais ; rather, it was that of a late Sixteenth Century hand.[7]

Paulin Paris, *Journal des Débats*, March 19, 1847.

Paulin Paris's conclusion was again substantiated by Jacques Boulenger in 1905 ; he thought that the manuscript was not original and had been badly transcribed by a scribe.[8]

See the book co-authored by Jacques Boulenger and Abel Lefranc, *L'Isle Sonante* par M. Françoys Rabelais réimprimée pour la première fois par Abel Lefranc, professeur au Collège de France, et Jacques Boulenger, de la Bibliothèque Sainte-Geneviève, Paris, Honoré Champion, 1905, p. iii.

The existence of two published texts and a manuscript, all differing from each other in several specific content areas ; and the partial publication of the *Cinquiesme Livre* on the eve of the Religious Wars had brought about

elements of doubt, and sentiments of suspicion concerning its genuineness. Thus the question raised by many critics throughout the last four centuries has been : Is Rabelais truly the author of the *Cinquiesme Livre de Pantagruel ?*

The need for a satisfactory solution of the authorship of the *Cinquiesme Livre de Pantagruel* in order to understand better Rabelais was voiced by Abel Lefranc at the beginning of this century. While inaugurating the publication of the *Revue des Etudes Rabelaisiennes*, he stated his concern that : "... la question du Ve livre subsiste toujours pendante, jetant dans toutes les études du domaine rabelaisien des éléments de doute et d'incertitude singulièrement embarrassants...."[9]

Abel Lefranc, *Revues des Etudes Rabelaisiennes*, i (June 1903), p. ii.

More than half a century later, Pierre Jourda in his critical edition of the complete works of Rabelais was expressing his doubts as to the possibility of finding a solution for the authenticity of the *Cinquiesme Livre* because : p. 3

On a donné le *Cinquième Livre* à Henri Estienne... et au médecin Turquet ; on a écrit qu'il avait été arrangé et continué par Jean Quentin.... Aucune de ces attributions n'est plus admises.... A moins d'une miraculeuse mais improbable découverte, le problème doit demeurer insoluble — et irrésolu... .[10]

*Rabelais Œuvres Complètes*, Edition de P. Jourda, Paris, Garnier, 1962, II, pp. 271-272

An overall assessment of the hypothesis expounded by sixty previous investigators from 1572-1970 reveals the following :

1. The *Cinquiesme Livre* was not written by Rabelais at all. It is the work of a very skillful imitator, a Huguenot pamphleteer who was attacking the Roman Catholic Church, the Pope, and royal institutions to stir up the emotions of his fellow Huguenots at the outbreak of the Religious Wars. Among adherents of this thesis are F. Brunetière, G. d'Orcet, G. Lanson, P.-P. Plan and others.
2. The *Cinquiesme Livre* could have been written only by Rabelais. In it we find "la griffe du maître" ; no one else was capable of writing like that. This opinion is shared by F. Audiger, P. Lacroix, Le Duchat, Ch. Le Normant, and some others.
3. The *Isle Sonante*, and the major portion of the last 32 chapters of the *Cinquiesme Livre* are definitely from Rabelais's pen, although they have not received final touches, since they show the imprint of his incomparable popular, literary and scientific knowledge. This is the conviction of J. Boulenger, Abel Lefranc, L. Sainéan, W. F. Smith and A. Tilley.
4. Some isolated chapters in the *Cinquiesme Livre* are without doubt Rabelais's, but the major portion of the book lacks the fine satire, and subtle imagination of his previous books. The authenticity of the entire book should remain in doubt until such time as a more objective, concrete and convincing proof has been found. This is the thesis of Jean Plattard, as well as P. Jourda, G. Lote and P. Villey.

The evidence presented here indicates that the problem of the authenticity of the *Cinquiesme Livre de Pantagruel* is still pending, and "tout n'est pas dit."

Although many critics have expressed the problem of the authenticity of the *Cinquiesme Livre* to be somewhat of a gargantuan task ; nevertheless, with modern statistical analysis and high-speed computer calculations, the task does not appear to be impossible. This study has been undertaken for the purpose of satisfactorily finding an objective solution of the verification of the authorship of the *Cinquiesme Livre de Pantagruel*. The investigation will be carried out primarily by a computer assisted stylo-statistical method of analysis. The frequency of a number of selected grammatical categories, special words, and expressions found in the *Pantagruel*, *Gargantua*, *Tiers Livre*, and *Quart Livre* of Rabelais will be compared with the frequency of the same variables found in the *Cinquiesme Livre* as well as three texts written by authors contemporary to Rabelais. These texts will be used as controls. p. 4

The computer assisted method of authentication will be supplemented by other statistical analyses to substantiate further any results claimed by the researcher.

It should be mentioned that none of the critics in the past have ever attempted to make use of the stylo-statistical method to determine the authorship of the *Cinquiesme Livre*. In the next chapter we will first give a summary of historical background of stylo-statistics, and specifically its application to three well known disputed authorships. Afterward we will discuss the method and procedure used in this investigation.

## Chapter II

p. 5

## BACKGROUND OF QUANTITATIVE ANALYSIS

### by George A. Petrossian

Among the research methods being employed in the problem area of disputed authorship is the use of quantitative analysis of writing. The quantitative method is called stylostatistics, and it has been defined by G. Herdan as follows :

The basis for a quantitative characterization of style is the frequency distribution of vocabulary according to frequency of occurrence of individual vocabulary items. The characterization, as a basis of comparison, would conceivably be in terms of the relation between vocabulary and occurrence (text length in terms of number of words), and in terms of either, that is of vocabulary alone or of frequency of occurrence alone.[1]

Gustav Herdan, *"Type-Token Mathematics" ; A Textbook of Mathematical Linguistics* (Mouton and Co., 1960), p. 25.

He further amplifies :

The first type of characterization may be in terms of the whole distribution or in terms of an overall statistical

parameter, such as the Type/Token ratio, the logarithmic Type/Token ratio and the Characteristic K (vm) ; the second may be in terms of the partitioning of vocabulary among the parts of a whole or in terms of probability of occurrence of particular vocabulary items.[2]

*Ibid.*, p. 25.

Hence the basic principle of stylostatistics is the frequency count. The main reason for employing the frequency counts is that they tend to give precision to analytical methods. When writers are being studied, such precision makes stylistic studies specific and clear. Since observations gained by such analysis are also objectively verifiable, they are considered as important evidence in matters of disputed authorships.

Writing may be regarded partly as selecting a sample of word material according to the writer's individual choice. It is this individual element of choice that contributes to the differentiation of one author's style from another.

The author's particular choice could appear under varied forms in his writings.[3]

See Louis T. Milic's article, "Unconscious Ordering in the Prose of Swift," in *The Computer and Literary Style*, ed. Jacob Leed (Kent, Ohio : Kent State University Press, 1966), p. 81.

It may be his choice of frequent use of adverbs in general or some limited use of specific ones ; abundant use of substantives and adjectives ; long enumerations ; constant interventions of some conjunctions such as *and*, *but*, etc. It could be also the considerable use of active verbs, or some peculiar expressions ; his preference for short sentences, and many other grammatical categories may also be indicative of personal style. In more complex forms it could be the excessive use of metaphors, hyperboles, anaphoras, personifications and other rhetorical devices. The relative frequencies of any one or more of these categories can be tabulated within the text of an author. If norms exist, or if they have been established, then with proper statistical methods a quantitative identification of some aspects of an author's style can be verified. Thus, as G. Hardan has stated, "the statistician is fully aware that the frequency distribution characterizes style, that is *form*, not *content*."[4] p. 6

*Op. cit.*, p. 40.

At the same time he also cautions that in statistics one never proves, but only aims at stating the probability with which an event may occur.[5]

*Ibid.*, p. 41.

It is with the same thought in mind that Pierre Guiraud in his *Problèmes et Méthodes de la Statistique Linguistique*, comments,

C'est par la fréquence des mots ou de tout autre élément du langage que l'écrivain agit à la fois sur le lecteur et sur la langue. Un style est un écart qui se définit quantitativement par rapport à une norme.

Certes il ne s'agit pas de substituer une analyse quantitative objective à une appréciation qualitative subjective ; les deux sont inséparables.[6]

Pierre Guiraud, *Problèmes et méthodes de la statistique linguistique* (Paris : Presse Universitaire de France, 1960), p. 19.

Thus, the frequency counts provide us with a numerical value for comparison. Since language is not a precise science *per se*, therefore, it is not so much a specific numerical value that may have significance, but rather the range of these values.

### *The "Frequency Count" Stylistics*

The modern theoretical concept of a quantitative analysis of style to solve the problem of disputed authorship goes back to Augustus de Morgan, Professor of Mathematics of University College of London. In 1851, he wrote a letter to his friend Reverend W. Heald at Cambridge stating :

I wish you would do this ; run your eye over any part of those of St. Paul's Epistles which begin with παῦλος — the (Greek I mean — and without paying any attention to the meaning. Then do the same to the Epistle of the Hebrews, and try to balance in your own mind the question whether that latter does not deal in longer words than the former. It has always run in my head that a little expenditure of money would settle questions of authorship this way. The best mode of explaining what I would try, will be to put down the results I should *expect* as if I had tried them.

Count a large number of words in Herodotus, say all of the first book, and count all the letters ; divide the second number by the first, giving the average number of letters per word *in that book*. Do the same in the second book. I should expect a very close approximation. If Book 1 gave 5.624 letters per word, it would not surprise me if Book 2 gave 5.619. I judge by other things. But I should not wonder if the same result applied to two books of Thucydides gave, say 5.713 and 5.728. That is to say, I should expect the slight difference between one writer and another to be well maintained against each other, and very well agreeing with themselves. If this fact were established, then, if St. Paul's Epistles which began with παῦλος gave 5.428 and the Hebrews gave 5.516 for example, I should feel quite sure that the *Greek* of the Hebrews (passing no verdict on whether or not Paul wrote in Hebrew and others translated) was not from the pen of Paul.... p. 7

I would have Greek, Latin and English tried and I should expect to find that one man writing on two différent subjects agree more closely with himself than two différent men writing on the same subject. Some of these days spurious writings will be detected by this test. Mind, I told you so.[7]

C. B. Williams, *Style and Vocabulary : Numerical Studies* (London : Griffin, 1970), pp. 4-5.

C. B. Williams, from whose book *Style and Vocabulary : Numerical Studies* we have extracted de Morgan's letter, comments that "de Morgan had grasped so many of the basic principles necessary for reliable statistical interpretation : the fact that words could be studied as words only, without reference to their individual meaning ; the necessity of more than one sample for each writer ; the necessity of large samples, and the use of mean values."

One of the first persons to put de Morgan's hypothesis into application was T. C. Mendenhall, Professor of Physics

at the Ohio State University. He thought that the distribution of different word-lengths can be compared to spectrums of light and used the term of the "word spectrum of an author."[8]

T. C. Mendenhall, "The characteristic curves of composition," in *Science*, 11 (supplement, March 1887), pp. 237-249.

For his experiment Mendenhall had selected two political writers, Edward Atkinson and John Stuart Mill. He wanted to study the possible sources of differences between writings by the same author, and between writings by two different authors. In each case two 5,000 word samples from similar subject (economy and labor) were considered. The word spectrum graph for both Atkinson samples had their peak frequency of three-letter words, whereas for Mill it was two-letter words. He then had combined two samples into a 10,000 word sample for Atkinson and one for Mill. Once again their respective peak frequencies of three-letter words, and two-letter words were observed, thus obtaining an experimental proof of de Morgan's conjecture.

Unaware of Mendenhall's work on word-length studies, G. U. Yule in 1939 published a paper suggesting that sentence-length could be a characteristic of authorship.[9]

George Udny Yule, "On sentence-length as a statistical characteristic of style in prose : With application to two cases of disputed authorship," *Biometrika*, XXX (1939), pp. 363-390.

He had selected samples of about 500 sentences each from the works of Bacon, Coleridge, Lamb and Macaulay. Yule showed that the arithmetic means of sentence-lengths were 48.5, 40.3, 26.2 and 22.1 words per sentence respectively. He commented that in view of these great ranges of differences"... we may conclude accordingly that sentence-length *is* a characteristic of an author's style."[10]

p. 8

*Ibid.*, p. 370.

Yule then applied this criterion to a disputed authorship case — that of *De Imitatione Christi*, a Fifteenth Century religious work, attributed to Thomas à Kempis a monk, and Jean Charlier de Gerson, Chancellor of the University of Paris. The favorite candidate had been Thomas à Kempis. This time Yule had selected 1200 sentences from *Imitatio Christi*, the miscellaneous works of à Kempis, and also de Gerson. The results of his statistical analysis were :

*Imitatio Christi* 16.2 words/sentence à Kempis 17.9 Gerson : Selected samples 23.4 Gerson : Random samples 22.7

Thus he concluded that Thomas à Kempis was the author of *De Imitatione Christi*[11]

*Ibid.*, p. 376.

In 1944 Yule approached the problem of a disputed authorship from the point of distribution of vocabulary and relative frequency of the usage of different words.[12]

George Udny Yule, *The Statistical Study of Literary Vocabulary* (Cambridge, England : The Cambridge University Press, 1944).

He had developed a word-distribution, that is richness of vocabulary formula which is called Yule's characteristic *K*[13]

*Ibid.*, p. 53.

He used this characteristic *K* to determine again the authorship of *De Imitatione Christi*. This time Yule had chosen nouns and their distribution as criteria for style. He counted ail the nouns occurring once, twice, three times and so forth in approximately 8,000 word samples from *Imitatione Christi*, the *Miscellanea* of à Kempis, and theological works of de Gerson. He had obtained for the characteristic *K* the following results :

*De Imitatione Christi* 84.2 First samples from à Kempis 73.2 Whole samples from à Kempis 59.7 Whole samples from de Gerson 35.9

Because à Kempis's range of 59.7 to 73.2 is much closer to *Imitatio's* 84.2 than de Gerson's 35.9, Yule concluded that Thomas à Kempis is the most likely author of the *Imitatio*.[14]

*Ibid.*, pp. 237-238.

Yule concluded his study with the following comments :

> We judged that it was best, if fairly trustworthy results were wanted, not to make a sample of much less than some 2000 occurrences of a noun. This would imply a sample of something like 10,000 words, more or less. But many works of authorship of which is disputed are brief tracts far shorter than this.... My impression is that the inclusion of ail words without exception would be a mistake ; that the inclusion of *a* and *the* and *is* and the like, each with a very large number of occurrences in *any author*, would merely tend to obscure differences, and it would be best to limit data to what are in some sense significant words.... Only investigation of a test case by varied methods can throw real light on the matter and suggest the best rules for practice.[15]
> *Ibid.*, p. 281.

p. 9

In 1962 Alvar Ellegard approached the problem of disputed authorship by means of selection of 458 rarely occurring words and expressions. Many of thèse words had a low frequency range of 0.0001 to 0.0002. This was an attempt by stylo-statistics to identify "Junius," a pseudonym signed in politically oriented letters.[16]

Alvar Ellegard, *Who Was Junius ?* (Stockholm : Göteborg-Upsala, 1962).

These letters were printed in a London newspaper, the *Public Advertiser* from 1769 to 1772. In the past Sir Philip Francis had been claimed a strong candidate for the authorship of the Junian letters. The basic principle used by Ellegard is the distinctiveness ratio for plus words and minus words. In the primary Junius material, he had counted approximately 82,000 running words. He had also found that the adjective *uniform* had been used 23 times, which gives a relative frequency of 0.000280. Then in a million word sample from contemporary texts *uniform* occurs 65 times. Hence, its relative frequency is 0.000065. This is about one-fourth of the corresponding Junius figure. Thus *uniform* becomes a Junian plus word with distinctiveness ratio of 280/65 or 4.3. If the occurrence of a word in the

Junius sample is smaller than the one in the control sample, the word becomes a Junian minus word, and its ratio is between + 1 and 0.[17]

*Ibid.*, pp. 104-105.

Ellegard stated that he had grouped together words with very low relative frequencies such as 2 or 3 times in a text of 10,000 running words. His reasoning was :

> … if an author uses each word with a certain characteristic frequency, he will also use any one of a number of words with a certain frequency. If the Junian frequency of *uniform* is 0.00028, and that of *interpose* 0.00016, their combined relative frequency is 0.00044. That is to say, the instances of *uniform* and the instances of *interpose* will together number around 4 or 5 in a Junian text of 10,000 words.[18]
> *Ibid.*, p. 105.

Thus the aggregate frequency of several words combined together could be of great value in frequency counts. However, he cautioned that the grouping of the words should be made carefully, keeping the plus words in one set and minus words in another. He recognized that with low frequency words great fluctuations would occur, but with aggregate frequency the fluctuations would be kept within reasonable limits. These distinctiveness ratios of rare words and expressions obtained from Junian texts were compared with the ones derived from the texts of Sir Francis. He found that in many instances their ranges were very close. Hence, he concluded, linguistic evidence sufficiently indicated that Sir Francis and Junius were one and the same, thus confirming John Taylor's Franciscan hypothesis.[19]

p. 10

*Ibid.*, p. 118.

Ellegard presented his numerical results in diagram forms also. He showed the aggregate frequencies of Junian plus expression, and Junian minus expression, along with those of Francis and the control group. He stated that "… the words avoided by Junius are avoided by Francis, the words favoured by Junius are favoured by Francis."[20]

*Ibid.*, p. 108.

Hence there existed a graphic representation of close identity between Francis and Junius.

Ellegard had elaborated more about his theory of a distinctiveness ratio and the problem of determination of authorship in another book.[21]

Alvard Ellegard, *A Statistical Method for Determining Authorship* (Göteborg, 1962).

He commented that the main assumption in quantitative analysis of style is that :

> Some features, or combinations of features, in a particular writer's style or language, remain reasonably constant, or change in a predictable manner, throughout his production. Moreover, we assume that some at least of these features are sufficiently rare to set the author apart from ail or most of his contemporaries.[22]
> *Ibid.*, p. 8.

He then pointed out three important aspects of style analysis : (1) A text of ten thousand words was long enough to reveal the personal stylistic features of an author. (2) A writer used different styles when composing formai letters and creating fiction. Hence some fluctuations in the style of an author are expected with change of subject matter. (3) The style of an author often underwent changes with the passage of the time.[23]

*Ibid.*, p. 9.

Thus numerical values obtained by frequency counts were *not* absolutely constant. Any interpretation of these frequencies should be done by range concept. He then defined the word "style" in statistical study as being synonymous with

> … constant features or combinations of features in an author's way of writing… therefore, style has none of the connotations of "good style" or "literary style." A bad style, a completely artless and inadequate way of writing may be as constant, and as distinct from ail others, and thus as good for the purpose of identification, as the most admirable stylistic masterpiece… the constant features of an author's style are certainly more likely to be unconscious or subconscious linguistic habits, than the conscious and deliberate rhetorical figures that marked the literary style….[24]
> *Ibid.*, pp. 9-10.

Ellegard thought that for a distinctiveness ratio to be considered useful for identification, it must have a range of above 1.5 and below 0.7 for plus and minus words respectively. Further, the width of each range depended upon the number of words that the group contained.[25]

*Ibid.*, pp. 18-19.

Some of the groups for the Junian texts had attained distinctiveness ratio ranges of 1.5-1.9 ; 2.0-2.4 ; 3-3.9 and higher. Among the examples of words and expressions selected for the testing lists were : *accordingly*, *under auspices*, *frequently* and *in regard to* with total occurrences of 1, 4, 12, 2 respectively in 82,200 words samples.[26]

p. 11

*Ibid* ; Appendix II.

These are extremely low frequency words, in the order of magnitudes of 0,000012 to 0,00015 occurrences ; that is, approximately 1 occurrence in 10,000 words to 1 occurrence in 100,000 words. This was the main reason that Ellegard had to have the one-million word sample control group. The whole concept of his method was based on the rare word phenomenon. This program was computerized.

Our last example of disputed authorship problem is the *Federalist Papers*, investigated in 1964 by Mosteller and Wallace.[27]

Frederick Mosteller and David L. Wallace, *Inference and Disputed Authorship : The Federalist* (Reading, Massachusetts : Addison-Wesley Publishing Company, Inc., 1964).

The *Federalist* papers were written in 1787-1788 by Alexander Hamilton, John Jay and James Madison to persuade the citizens of the state of New York to ratify the Constitution. These were short essays ranging from 900 to 3500 words in length. They were printed in newspapers signed with the nom de plume "Publius." From external

evidence it is generally agreed that of the 77 papers Jay wrote five ; historians agree that 43 belonged to Hamilton, 14 to Madison and 3 were written jointly. The authorship of the remaining 12 papers has been disputed with opinions vascillating between Hamilton and Madison. These are referred to then as “the disputed papers.” The authorship of these 12 papers has gone back and forth to Hamilton or Madison depending on whom the critic was favoring intuitively. For instance, the celebrated “Benson list” made public in 1817 attributed the 12 disputed papers to Hamilton.[28]

*Ibid.*, p. 4.

It was said that John C. Hamilton, the son of A. Hamilton claimed in 1864, that his father had written all the disputed papers, whereas historian Douglass Adlair in 1944 stated that Madison was the author of the disputed 12 papers. In 1941 F. Williams and F. Mosteller attempted to solve the authorship problem of the *Federalist* by means of sentence length distribution. They used the undisputed papers of each author in order to obtain norms of the mean sentence length for comparative purpose. The resuit of their efforts revealed that the average sentence length for Hamilton’s papers was 34.55 words, and for Madison 34.59 words with almost equal standard deviations.[29]

*Ibid.*, p. 7.

Thus it was impossible in this case to use sentence length as a criterion for authorship differentiation. In 1963 Mosteller teamed up with David Wallace and approached the problem with a different method, study of relative frequency of selected words used most frequently. They made use of both the computer and statistics to identify the disputed *Federalist* papers by means of function words such as *all*, *and*, *enough*, *for*, *only*, *upon* and others. From an initial list of 165 different words they finally selected 30 words which were called “markers” either for Hamilton or for Madison. Eighteen undisputed Hamilton papers amounting to 36,577 words, and fourteen Madison papers totaling 37,095 words, plus some exterior texts were used. Among the markers that their frequency counts had generated were *enough*, *while*, *whilst* and *upon*. The rate per 1,000 words for these words were : p. 12

*Words Hamilton Madison* enough 0.59 0 while 0.26 0 whilst 0 0.47 upon 2.93 0.16

Thus the words *enough*, *while* and *upon* became high frequency Hamilton markers, and *whilst* a Madison marker.[30]

*Ibid.*, pp. 12-13.

The most striking was *upon* which had a ratio 18 to 1 in favor of Hamilton. By means of these four function words Mosteller and Wallace were able to identify the twelve disputed papers. They stated that “these data show rather clearly that the disputed papers as a whole are Madisonian, but in this form they cannot settle the papers singly.”[31]

*Ibid.*, p. 12.

To identify each paper separately they used highly complex Bayesian statistics based on prior odds. The resuit of their final findings were that :

> Our data independently supplement those of the historian. On the basis of our data alone, Madison is extremely likely, in the sense of degree of belief, to have written ail the disputed *Federalists*... with the possible exception of No. 55. For No. 55 our evidence is relatively weak because suitably deflated odds are about 90 to 1 in favor of Madison.

And they concluded that :

> In summary, we can say with better foundation than ever before that Madison was the author of the 12 disputed papers.[32]
> *Ibid.*, p. 263.
> thus confirming Adlair’s conjuncture of Madisonian authorship.

Three well known cases of disputed authorship problems have been reviewed. In each instance, a different method of quantitative analysis was used. In *De Imitatione Christi* Yule made use of sentence length and Yule’s characteristic *K* of noun diversity index as authorship attribution criteria. Ellegard in the *Junius Letters* used the distinctiveness ratio of low frequency rare words as authorship determination concept. In the *Federalist Papers*, Mosteller and Wallace had a list of 30 high frequency function words as possible discriminators of authorship.

It should be pointed out that in all three cases, at least one contender was already known to be the likely author of the disputed text. The problem then was to compare quantitatively some stylostatistical features of the disputed text with the style of the contending author. Thus there existed a possible norm to compare the results. p. 13

One of the major difficulties involved in the problem of the authenticity of the *Cinquiesme Livre de Pantagruel* has been that there is no definite candidate of authorship in view, so that we could compare the style of the *Cinquiesme Livre* with that of the contending author. There has been mentioned a skillful “pasticheur” of Rabelais, but who he was, what other books he has written, no one seems to know. The reasonable thing to do then is to consider the stylostatistic contents of four authentic books of Rabelais as the main source of the norm for authentication of the *Cinquiesme Livre*. The following chapter gives the methodology of the study.

## Chapter III

p. 14

## THE METHOD

## by George A. Petrossian

The basic methodology implemented in this investigation was the frequency count of certain words to determine the authenticity of the *Cinquiesme Livre*, by means of random sampling based on the Analysis of Variance. A set of 25 selected function words, compiled from a larger screening list, were used as potential authorship discriminators. First

the mean frequencies of each selected word derived from four authentic books of Rabelais were compared to the ones in the control group. The control consisted of three authors who were contemporaries of Rabelais. From this comparison a new set of words were chosen as possible "Rabelais indicators." These words were selected on the basis of having attained a certain distinctiveness ratio. The significant level of distinctiveness ratio was designated as 2 and above.[1]

Alavar Ellegard in his study of *Junius Letters* had considered the distinctiveness ratio 1.5 and higher as being significant for plus words, c.f. p. 72.

These possible "Rabelais indicators" in turn were grouped ail together to constitute *The Rabelais Indicator Index*.

It was by means of this RII that ail aspects of the authenticity of the *Cinquiesme Livre* were determined. Among them were : (1) the determination of the authenticity of the book as a whole, (2) the *Isle sonante* and the last 32 chapters of the *Cinquiesme Livre* separately, (3) the authenticity of the Prologue to the *Cinquiesme Livre*, (4) the much disputed chapters of "Apedeftes" (Ch. 16), two chapters of "Tournoi de la Quinte" (Ch. 23-24), "Comment furent les dames Lanternes servies a soupper," and the additional ending of the last chapter (Ch. 47), of the book.

A Flow Chart is provided in Figure 1 for the purpose of visualizing the method of investigation. A two-tailed *t*-test at the .05 level of significance was also carried out to further verify the selection of Rabelais Indicators.[2]

The formula and its explanation may be found in Paul A. Games and George R. Klare, *Elementary Statistics and Data Analysis for Behavioral Sciences* (New York : McGraw-Hill, 1967), pp. 298-338. The computerized portion of this investigation was carried out under the supervision of a staff programmer.

As was mentioned these indicators were first selected on the basis of their high distinctiveness ratio criterion. An attempt was also made to substantiate further the results of this random sampling analysis. Some raw data provided by previous investigators of Rabelais's work were evaluated by means of Type-Token ratio calculations. These findings were then compared to the ones obtained by computer analysis. p. 15

## *Analysis Design of This Investigation*

### Texts

For the analysis of Rabelais's work Pierre Jourda's edition of *Rabelais's Œuvres Complètes*, Paris, Garnier, 1962, was used in this study. For chronological purposes, *Pantagruel* is designated as the first book followed by *Gargantua*, *Tiers Livre* and *Quart Livre*. Again for chronological purposes, the *Cinquiesme Livre* was considered as two separate books. The portion of *Isle Sonante* was designated as V-16 and the last 32 chapters as book V-32. The prototype of the *Quart Livre* known as the 1548 edition and *Sciomachie* were used as the calibrating set for authentication of the Rabelais Indicator Index developed in this study.

Three texts of authors contemporary to Rabelais were used as controls. One control was *Les Navigations de Panurge* edited by Marcel Guilbaud.[3]

This text was included in his *Œuvres Complètes de Maître François Rabelais* (Paris : Imprimerie Nationale, 1957).

As has been mentioned, this book was one of the earliest imitations of Rabelais's novel of *Pantagruel*, hence, having a close thematic relation with the texts to be analyzed. The second control text was the *Nouvelles Récréations et Joyeux Devis* of Bonaventure Des Périers edited originally by Paul Lacroix, new edition of Garnier frères, 1925. For sampling, only the first 90 nouvelles were considered because the authenticity of the remaining 39 chapters has been in doubt. As a conteur Des Périers belongs to Rabelais's generation, thus being a good source of stylistic comparison.

The last control was the *Propos Rustiques* of Noël du Fail, edited by Jacques Boulenger, Paris, Editions Bossard, 1921. This is the original authentic version of the 1547 edition. Du Fail was considered as a control because he was one of the best recognized Rabelais imitators.

### Sampling

Since the analysis of the entire work of Rabelais was outside the scope of this investigation, a random sampling method was implemented for the quantitative analysis. In this study each book was considered as an independent sampling frame with the chapter as sampling unit and the word as measuring unit. The chapter was chosen as an appropriate sampling unit for several reasons. First, ail the books were divided into chapters. Hence, there already existed ready and convenient divisions. Second, by keeping the chapter as the principal sampling unit, it was possible to analyze the author's thought pattern in more or less complète form, thus avoiding an unnatural cutting

Fig. 1. — Analysis design components : the flow chart of the *Cinquiesme Livre* authentication p. 16

1 Level of significance 2 and above, computation done by a desk calculator.

2 By *t* statistic, significant at the .05 level in a two-tailed test, computation performed through computer programming.

off of the writer's continuity of thoughts. The fact that each book had a good number of chapters provided an ample source of independent observation of the total population (words). Another advantage of using the chapter was that most of them were long enough to provide a meaningful and effective frequency count of word distributions. p. 17

Most of the investigators of disputed authorship considered a sample of about 10,000 words good enough for a quantitative style analysis of a writer. Since the average length of a chapter in this study is approximately 800-1,000 words, it was decided that a sample of about 16 chapters or 12,000 to 16,000 words sample from each book would provide a sufficient sampling population for this investigation. A total of 11 texts were sampled. Of these, seven were by random sampling and the remaining four were entire text samples. (See Table 1.)

## Word Counts

All the word counts were done by hand on the original pages of the text or on Xeroxed copies. First, the word counts in each line were written on the margins of the page. When all the words in a chapter were counted, they were

## Table 1
## IDENTIFICATION OF CHAPTERS IN ALL SAMPLED TEXTS

*Texts No. of chapters sampled Sampled chapters identified Pantagruel* 16 1, 2, 3, 5, 6, 8, 13, 16, 17, 20, 21, 23, 28, 32, 33, 34 *Gargantua* 16 1, 3, 4, 6, 9, 14, 17, 23, 27, 33, 37, 38, 48, 55, 56, 57 *Tiers Livre* 16 1, 2, 3, 5, 8, 11, 13, 17, 21, 29, 32, 34, 35, 43, 47, 52 *Quart Livre* 16 1, 2, 3, 5, 8, 13, 21, 28, 34, 38, 45, 48, 50, 55, 62, 67 V-16 16 full text V-32 16 16 bis, 17, 18, 20, 21, 23, 24, 29, 32, 33, 37, 39, 40, 42, 45, 47 *48-Quart Livre* 11 full text *Sciomachie* 1 full text *Navigations de Panurge* 31 full text *Nouvelles Récréations* 16 2, 3, 9, 10, 15, 19, 20, 25, 34, 35, 37, 42, 43, 60, 64, 88 *Propos Rustiques* 8 1, 2, 3, 5, 6, 8, 10, 13

totaled by an adding machine, and recorded on a master sheet. All the counting was checked at least three times. If two words were written together at times such as *tresnoble*, it was counted as two words. Similarly, if a word was written sometimes in two separate parts, and sometimes in one, the modem spelling was considered the right one. For example, *ce pendant*, was written also *ce pendant*. The former counted as one word. The rules of contraction and elision were observed. *Au bois* was counted two words, and *à l'église* three words. Words in verses and quotations were not counted, but proper nouns and book titles were included. p. 18

Table 2 provides all the word counts tabulated in this study. Column 1 gives the text ; column 2 the number of chapters counted ; and column 3 the total number of words counted in each book. Column 4 gives the total number of chapters in each book, and column 5 the percentage of the sample.

## Table 2
## RELATIVE SAMPLE SIZES OF ALL THE TEXTS ANALYZED

Table 2 indicates that a total of 55,967 words were counted from four authentic Rabelais texts. This is about 30% of the total word population. In addition, two other Rabelais texts were used as calibration set for RII verification. A total of 15,299 words were counted. These texts represented a 100 % count each. The total count for the *Cinquiesme Livre* was : V-16 equalled 12,513 and V-32 equalled 13,592 to produce a combined total of 26,105 words. This is approximately 66 % of the entire word population. The count for three controls was 38,330 words, about 41 % of the total population. Thus the combined total word count in this study was 135,701 tokens. This number does not include the Prologues and Three Letters written by Rabelais from Rome which were counted separately. Hence, from a purely statistical viewpoint, these were respectable samples, and they should provide a measurable amount of useful and reliable results. p. 19

## Initial Screening of Words

Approximately 120 words, grammatical categories and narrative expressions were considered as possible authorship indicators. Among these were prepositions like *avec, par, pour* and adverbs like *aussi, bien, comme, là* and *plus*. Also considered were adjectives like *autre, grand, gros, petit, tout,* and cardinal numbers like *trois, quatre, cinq, six, sept, vingt, cent, mille*, and others, providing a total of 18 different ones. General large categories such as all adjectives ending in *-ique*, and all adverbs ending in *-ment*. Some other considered examples include the following : large grammatical entities such as verbs ; narrative expressions like *adonc, c'est-à-dire, c'étaient, lors, de sorte que* ; rare words like *basta, fat, gorgias, mirifique, paragon, simpiternel* and many others ; punctuation marks — the period, question mark, and exclamation mark ; special stylistic devices such as parentheses, and ô, ô Noël incantation ; contextual words such as *diable, Dieu, gens de biens, peur* and *vin*.

A preliminary count of all these items mentioned as well as many others were made in sample texts of about 5,000

to 6,000 running words. These sample texts were chapters from *Pantagruel*, *Gargantua*, *Tiers Livre*, *Quart Livre*, the 1548 partial edition of the *Quart Livre* and *Sciomachie*. The elimination of the words was done on the basis of several criteria. For instance, many words had a similar type-token ratio in all samples examined ; among these were most of the prepositions and the use of verbs. Many rare words attained very low frequency counts in the order of 0.0001 to 0.0002, hence, not being suitable for the sample size of this investigation.

## Final Selection of Words

The guiding principle of final selection of words was four-fold : (1) to avoid selecting, if possible, any contextual words and expressions, (2) not to include words and stylistic devices which become obvious to the reader, hence easy to imitate by a "pasticheur",[4]

Noël du Fail is a good example ; his texts are full of ( ), ô, ô, and favorite Rabelais swear words and expressions.

(3) to select as many as possible of the function words, commonly used by most writers. Thus if there appears to p. 20
be a very high frequency count, it could possibly reveal an individual choice and preference ; (4) to select from a large grammatical category only the words that have attained the highest frequencies, by order of descending rank. These criteria were used for my final selection of eight adverbs ending in *-ment* and four cardinal numbers. Table 3 provides the list of the final 25 selected words. Column 1 gives the assigned code number of each word or expression. These code numbers were necessary for the computer programming and provide clarity and brevity to the tables as well. Column 2 describes the selected word and column 3 gives the category division of the words. These words were divided into eight distinct categories. In category A we have only the conjunction *et*, in B the adverb *comme*, in C adjectives ending in *-ique* and in D adverbs ending in *-ment*. These are all individually high frequency items. E is our first group category ; it includes two adjectives and two adverbs *autre*, *gros* and *bien*, *plus* respectively. These are also high frequency words. F contains the eight selected adverbs ending in *-ment*. As has been mentioned

### Table 3
### LIST OF 25 FINAL WORDS

*Code No*. *Word Category* 1 et A 2 comme B 3 adjectives ending in *-ique* C 4 adverbs ending in *-ment* D 5 autre 6 bien (adverb) E 7 gros 8 plus 9 autrement 10 mêmement 11 ordinairement 12 pareillement 13 premièrement F 14 présentement 15 seulement 16 vraiment 17 c'est-à-dire 18 c'était, c'étaient 19 savoir est G 20 toutefois 21 voire 22 cent 23 mille 24 quatre H 25 trois

they were selected by rank order. The first rank was attained by *seulement*, followed by *pareillement*, *autrement* and p. 21
then *vraiment* ; then in descending order *mêmement*, *premièrement*, *ordinairement* and *présentement*. They were selected among approximately 200 different adverbs ending in *-ment* with total occurrences of approximately 500 in a sample population of 75,000 words. Because of these relatively low frequencies, they were grouped together ; this is similar to the method used by Ellegard. Category G contains the narrative expressions which I have identified by the Spitzer method of reading the entire Rabelais books several times, making a mental note of unusual words as frequently appearing in the pages. These are *c'est à dire*, *c'était*, *savoir est*, *toutefois* and *voire*. Category H had four numerals selected by the rank frequency method. These are, in descending order, *trois*, *quatre*, *cent* and *mille*. They were selected from 18 different cardinal numbers which had about 700 total occurrences in a 75,000 word sample.

The means for Categories E, code number 5-8 ; F, code number 17-21 ; and H, code number 22-25 were calculated for possible generation of Rabelais indicators. These groupings were given additional code numbers for computer use. They were 26, 27, 28 and 29 respectively. It should be mentioned that in selecting these 25 final words Vender Beke's *French Word Book* was also consulted for possible norm evaluation ; unfortunately, it was of no important use because the samplings were done from the works of mainly nineteenth and twentieth century writers.[5]

George E. Vender Beke, *French Word Book* (New York : The MacMillan Company, 1929).

Some of our selected adverbs ending in *-ment* did not even appear in their frequency counts.

For the convenience of the reader, a list of the abbreviations used in the rest of this and the following chapters follows. This will also facilitate the reading of the Tables, Figures and Diagrams that are within the context of this work. *Abbreviation Designation Abbreviation Designation* P. *Pantagruel* (Book II) P.R. *Propos Rustiques* G. *Gargantua* (Book I) ANOVA Analysis of Variance III. *Tiers Livre* C Control IV. *Quart Livre* DR Distinctiveness Ratio V. *Cinquiesme Livre*, 1564 M Mean Value V-16 *Isle Sonante* R Rabelais V-32 Last 32 chapters of the RI Rabelais Indicator *Cinquiesme Livre* 48-IV. Partial edition of RII Rabelais Indicator *Quart Livre*, 1548 Index SCIO. *Sciomachie* SD Standard Deviation N.P. *Navigations de Panurge* SE Standard Error N.R. *Nouvelles Récreations* TTR Type-Token Ratio p. 22

## *Procedure*

All the frequency counts of words were done by hand in all the designated chapters (see Table 1). Once the final 25 words were selected, the frequency count of each word was recorded on a master sheet for each book. (See Appendix I for detailed counts.) The summary of the raw scores are given in Table 4. Then, for every sampled

chapter, an IBM punch card was provided carrying the following information :
Columns 1-8 Book number, chapter number and total word count for the chapter, respectively Column 9 Blank
Columns 10-70 The frequency count of the 25 variables (words) Columns 71-80 Program coding

Thus, if a book had 16 sampled chapters, 16 IBM cards were employed. This was a very economical way of using computer time. Hence a total of only 162 IBM punch cards were used for ten books that were analyzed.[6]

*Sciomachie* did not have a punched card because it had only one chapter. The means for this text were computed by an electronic desk calculator.

The means of all 25 words in the ten books were given on a print-out sheet. From this information the Distinctiveness Ratio of every word was calculated by an electronic desk calculator. Since DR is simply a Type-Token Ratio, the differences observed among the variables could be the result of chance. A t-test was programmed to find out whether differences observed in the means were large enough to be regarded as statistically significant. Thus the test of significance was carried out at a .05 level of confidence.

Because the means by themselves are often deceptive and could hide a great amount of variation, Standard Deviations (SD) were provided for each sampled chapter to determine how much fluctuation from the means existed for each word. Similarly, the Standard Error (SE) was also computed so that we could establish an upper and lower limit range for the mean.

Once the Rabelais Indicators (RI) were identified by means of the Distinctiveness Ratio (DR), Analysis of Variance (ANOVA) and the t-test, an additional grouping was designated as Code Number 30. This new grouping represented the Rabelais Indicator Index (RII) and was calculated on the basis of a rate per 1,000 running words. It was then primarily with this RII that all aspects of the *Cinquiesme Livre de Pantagruel* were authenticated. It should be pointed out that the RII is not a constant. It represents a range value between the upper and lower limit of the mean.

The ANOVA portion of this investigation was performed by an IBM 360/50 computer. It was carried out under the supervision of a staff programmer, Mr. Ch. Ramesh, of the Illinois State University Computer Center.

## Table 4

## RELATIVE OCCURRENCES OF THE 25 SELECTED WORDS IN FOUR AUTHENTIC BOOKS OF RABELAIS THE *CINQUIESME LIVRE*, AND THREE CONTROLS

### Raw Scores





## Chapter IV

## RESULTS

### by George A. Petrossian

One of the major findings of this study was that there exists a significance for some words and no significant differences for others insofar as concerns the mean frequencies between the four authentic books of Rabelais and the three control texts. Table 5 shows the summary of the mean frequencies per 1,000 words of the 25 final words. Column 1 gives the code number of the word, columns 3-6 present the means for *Pantagruel*, *Gargantua*, *Tiers Livre* and *Quart Livre* respectively and column 7 the combined means. Similarly columns 8-10 show the means for *Nouvelles Récréations*, *Navigations de Panurge*, and *Propos Rustiques* respectively. The last column gives the means of three controls. (See Appendices II, III, IV, and V for their SD and SE values.) Our next task was to find out which one or ones of these words have the high distinctiveness ratio to become a possible Rabelais Indicator. The combined Rabelais means for each word was divided by the combined control means. These results are shown in Table 6, column 6. We note that only three categories have attained the required significant level of DRs. Hence they were designated as possible Rabelais Indicators. These are adjectives ending in *-ique*, adverbs ending in *-ment* and the grouping of eight selected adverbs — autrement, mêmement, ordinairement, pareillement, premièrement, présentement, seulement, vraiment. They attain DRs of 8.87, 2.26 and 2.92 respectively.

These are all highly significant figures, as DR 2 and above was stipulated in the methodology of this investigation to be an acceptable level of significance. Since DRs by themselves are only arithmetic means without any confidence level attached, I thought it would be better to verify them further by a t-test at the .05 level of confidence.[1]

A significant t-test at the .05 level means that the distribution of vocabulary in the three categories of words according to their relation to the other books is such that we may be wrong 5 times in 100 with our assumption of a real difference.

Column 4 in Table 7 gives the scores of the t-test. All three categories of words attained a level of significance not only at the .05 level, but beyond the .001 level. This means that we would be wrong once in 1,000 times with our assumption of a real difference. Hence, adjectives ending in *-ique*,

## Table 5

## THE MEAN FREQUENCIES OF TWENTY-FIVE SELECTED WORDS IN FOUR AUTHENTIC BOOKS OF RABELAIS AND THREE CONTROL TEXTS (Rate per 1,000 words)



## Table 6

## RELATIVE DISTINCTIVENESS RATIO OF TWENTY-FIVE SELECTED WORDS BETWEEN FOUR AUTHENTIC TEXTS OF RABELAIS AND THREE CONTROL TEXTS

Significant DR is 2 and above for this investigation.

Significant DR is 2 and above for this investigation.

Significant DR is 2 and above for this investigation.

Significant DR is 2 and above for this investigation.

Significant DR is 2 and above for this investigation.

Significant DR is 2 and above for this investigation.

Significant DR is 2 and above for this investigation.

Significant DR is 2 and above for this investigation.

adverbs ending in *-ment* and the eight selected adverbs ending in *-ment* were designated as Rabelais Indicators with added confidence.

To obtain an overall measuring factor for possible Rabelais texts authentication, these three categories were grouped together to form a new grouping. This grouping was designated with Code No. 30, thus the numerical value of No. 30 constitutes the Rabelais Indicator Index. Table 8 gives the means of these RIs for the four authentic texts of Rabelais and three controls. In comparing the means in column 7 of Rabelais's texts for categories C, D and F with  those of the Control in column 11, it is noted that there exists a great amount of difference in all three categories between Rabelais and the controls. For example the M rate of adjectives ending in *—ique* for Rabelais is 2.55 per 1,000 words, whereas the M for the controls is 0.29 per 1,000 words. This results in a very high distinctiveness ratio of 8.87. Great differences exist also

## Table 7

## THE DISTINCTIVENESS RATIO OF RABELAIS INDICATORS, AND THEIR RESPECTIVE t-TEST SCORES BY MEANS OF ANALYSIS OF VARIANCE

$t$ .001,117 = 3.37.

Significant at the .001 level on a two-tailed test.

Significant at the .001 level on a two-tailed test.

Significant at the .001 level on a two-tailed test.

for No. 4 and No. 27. Similarly in comparing the RII between Rabelais and the controls we have the rate of 13 per 1,000 words for the former and 4.69 for the latter. That approaches a 3 to 1 ratio in favor of Rabelais. Whereas, when we compare the controls among themselves for RII rates, the ratios vary slightly from a minimum of 3.88 per 1,000 words for *Navigations de Panurge* to 5.46 and 6.26 for *Nouvelles Récréations* and *Propos Rustiques* respectively. This is almost a 1 :1 or a 1 :1.5 ratio.

We also note that there exists a distinct division within the Rabelais texts for RII means. For example, the RII mean between *Pantagruel* and *Gargantua* on one hand, and *Tiers Livre* and *Quart Livre* together on the other varies from 9

to 18 per 1,000 words respectively. Similarly this divergence continues for words No. 3, N. 4 and No. 27. The ratios are from 1 to 2 all the way to 1 to 3. This examplifies an increase from the P. and G. period to the III and IV period. On the other hand there exist very slight changes between P. and G. for RII ; it is from 9.07 to 9.27 respectively, an almost 1 :1 ratio. The same ratio holds true for word No. 4 ; it is slightly higher for No. 3 — from 1.91 to 1.66, about a 1 : 1.1 ratio. Similarly the same pattern of close ratios for RII and variables No. 4 and No. 27 holds true between the *Tiers Livre* and *Quart Livre*. For example the RIIs for III and IV are 17.97 and 15.67 per 1,000 words respectively. This gives us a ratio of 1 :1.2 for No. 4. Could this possibly imply a change in Rabelais's style of composition, i.e., from P-G period (15321534) to III-IV period (15461552) ?[2]

This trend in change of style from P-G period to III-IV period will be revealed again with other variables such as conj. *et* and use of "noms propres." These will be discussed later.

p. 28

## Table 8

## RELATIVE MEANS OF RABELAIS INDICATORS AND RABELAIS INDICATOR INDEX IN FOUR AUTHENTIC TEXTS OF RABELAIS AND THREE CONTROL TEXTS

## (Rate per 1,000 words)

p. 29

It should be pointed out that previous critics have also noticed this change in Rabelais's style. For example, Arthur Tilley also identified this change with the Third Book and called it Rabelais's later style.[3]

Arthur Tilley, "Rabelais and the Fifth Book," *Studies in French Renaissance* (Cambridge, 1922), p. 112.

The Rabelais Indicator Index is a critical instrument for the authentication of the *Cinquiesme Livre*. It was essential to establish a range of its numerical value at a 99 percent confidence level. This was done by means of the standard error obtained through Analysis of Variance for the RII of seven texts, i.e., four authentic Rabelais texts and three Controls. Table 9 indicates these ranges.

In column 1 the four authentic Rabelais texts are grouped in the *Pantagruel-Gargantua* period and *Tiers Livre-Quart Livre* period. All three Controls are grouped together. In column 2 the RII values of each group are shown at the rate per 1,000 words. Column 3 gives the standard error of each RII, and column 4 indicates the RII range for each grouping. We note that the differences between these groups are again distinctly marked. For example, the range of the upper limit of the control group is still outside the range of

## Table 9

## RELATIVE RANGES OF RABELAIS INDICATOR INDEX IN FOUR AUTHENTIC TEXTS OF RABELAIS AND THREE CONTROL TEXTS

## (Rate per 1,000 words)

1

We have calcultaed the upper limit and the lower limit of each range at a 99 percent confidence level, i.e., RII ± 2 SE.

the lower limit of the Authentic Rabelais group. The difference is approximately 2 units. This indicates that the RII has the possibility to differentiate between non-Rabelais and Rabelais texts. Similarly the upper limit range of the *Pantagruel-Gargantua* period is below the lower limit range of *Tiers Livre- Quart Livre* period, by approximately 3.5 units. Thus DR and ANOVA have generated three distinct RII ranges. For the P-G period, it is between 7.2-11.1 per 1,000 words ; III-IV between 14.619, and for non-Rabelais texts it is between 3.7-5.7.

p. 30

It should be remembered that even with this flexibility of range values, any inferences drawn from a comparison of texts is still in the domain of statistical probability, particularly if the sample texts are not long.

Before we attempt to authenticate the *Cinquiesme Livre* by means of the Rabelais Indicator Index generated by this investigation, we should first apply it to some outside authentic Rabelais texts and see if it is capable of authenticating them. By authentication it is meant that if the numerical values of the RII in these genuine Rabelais texts fall within the overall Rabelais range, it should be good proof. This was established to be between 7.2 to 19 per 1,000 words with the mean value for the P-G period being 9.2 and for the III-IV period 16.8. For our calibrating set we have selected the partial 1548 edition of the *Quart Livre* (which is very closely associated with the *Quart Livre*) and *Sciomachie*.[4]

Written and published in 1549 while Rabelais was in Rome with the retinue of Cardinal Jean du Bellay.

These two texts count 9,868 and 5,431 running words respectively. The 48-IV is definitely a good size sample and the *Sciomachie* falls within the acceptable limit. Table 10 gives the results of the authentication.

## Table 10

*Texts RII Mean SE Range* 48-IV 11.21 0.94 9.33 - 13.09 Sciomachie 14.17 —[1]

There was no SE because it was counted as a single chapter.

14.17

We note that both values of RII for these authentic Rabelais texts remain within the Rabelais range.

We shall now proceed to authenticate the *Cinquiesme Livre*. The means for the Rabelais Indicator Index for both the *Isle Sonante* (V-16), and the last 32 chapters of the *Cinquiesme Livre* (V-32), and the *Cinquiesme Livre* as a single entity is given in Table 11. This table reveals many important and interesting facts about the *Cinquiesme Livre*. First, taken as two separate entities as well as a single unit, all the RII mean values remain within the authentic Rabelais range. On the other hand, the divergence of ranges within the two component parts, i.e., the *Isle Sonante* and the last 32 chapters of the *Cinquiesme Livre* most likely indicates the heterogenous nature of the entire book. There is not even an overlapping between the upper limit range of 14.7 of the *Isle Sonante* and the lower limit range of 16.4 for V-32. In fact, they are approximately 2 units apart. Could a possible apocryphal nature of the book be inferred from this ? From statistical facts such an assumption seems not likely since all three RII means, 13.12, 18.32 and 15.72, for V-16, V-32 and V respectively are solidly within the authentic Rabelais ranges. To understand fully this p. 31
heterogenous nature of V-16 and V-32, we shall present more detailed facts which this investigation has revealed. Before doing this, the relative distributions of the Rabelais Indicator Index for four authentic texts of Rabelais, the three control texts and V-16, V-32 are presented in Table 12. For a visual interpretation of the statistical data see Figure 2.

## Table 11

## THE AUTHENTICATION OF THE *CINQUIESME LIVRE DE PANTAGRUEL* BY MEANS OF RABELAIS INDICATOR INDEX

### (Rate per 1,000 words)

[1] See Appendix IV for standard errors.

## Table 12

## THE RELATIVE MEANS OF RABELAIS INDICATOR INDEX FOR FOUR AUTHENTIC TEXTS OF RABELAIS, THE THREE CONTROL TEXTS, AND THE CINQUIESME LIVRE DE PANTAGRUEL

*Texts RII Pantagruel*. 9.07 *Gargantua*. 9.27 *Tiers Livre* 17.97 *Quart Livre* 15.67 *Nouvelles Récréations*. 5.46 *Navigations de Panurge*. 3.88 *Propos Rustiques*. 6.26 V-16 13.12 V-32 18.32

In examining the Figure 2 we observe at once how clearly the numerical values associated with these texts describe differences visually. First there exists a close RII relationship between *Pantagruel* and *Gargantua*. Similarly

Fig. 2. — Distribution of Rabelais Indicator Index in four authentic Rabelais texts, the *Cinquiesme Livre*, and three controls. p. 32

we note the close relationship between *Tiers Livre* and *Quart Livre*. The controls also reveal some differences p. 33
among themselves, but with a highly significant difference as compared to all four Rabelais texts.

Looking at the bargrams of the *Isle Sonante* and V-32, the heterogeneous aspect of their RII characteristics is evident. Yet they both blend with an overall style profile of the four authentic Rabelais texts. Most particularly one could easily associate them with the bargrams of the *Tiers Livre* and *Quart Livre* period. Thus the inference that may be obtained from these observations is that : considering the adjectives ending in *-ique*, the adverbs ending in *-ment*, and the selected adverbs *autrement, mêmement, ordinairement, pareillement, premièrement, présentement, seulement* and *vraiment*, as quantitative characteristics of one aspect of Rabelais's style, the *Cinquiesme Livre de Pantagruel* ostensibly appears to be that of François Rabelais.

We will now continue to present more statistical facts to further substantiate these claims of authenticity. To get an overall picture of the relative frequencies of 25 selected words and their respective groupings the computer center provided a printout sheet. In it were included the mean frequencies of these 25 variables for four authentic books of Rabelais, and the *Cinquiesme Livre*. These means were computed per 1,000 words. Table 13 provides this information. Checking for a possible developing pattern for any one of these variables in the six texts, we note that the

three Rabelais Indicators — No. 3, No. 4 grouping No. 27, and RII No. 30 all have different established patterns of frequencies. This is true in the four authentic texts of Rabelais as well as in both components of the *Cinquiesme Livre*. Similarly the conjunction *et* shows a great amount of divergence of frequency.

For example the M frequency of *et* in *Pantagruel* is 42.10 per 1,000 words ; it decreases gradually to 39.48, 39.07, 36.32, and 34.37 in *Gargantua*, *Tiers Livre*, *Quart Livre*, and V-16. Then it increases again to 38.04 in V-32. For the use of adverbs ending in *-ment* the pattern is the opposite : it has a M frequency of 6.22 per 1,000 words in *Pantagruel*. It holds even in *Gargantua* with a M value of 6.17, then increases significantly in the *Tiers Livre* with a M value of 10.64. It holds to 10.16 in the *Quart Livre*, then descends to 7.34 in V-16 and goes up to 10.62 in the V-32. For the use of adjectives ending in *-ique*, the pattern is more divergent. It has M frequency of 1.91 per 1,000 words in *Pantagruel* and holds to 1.66 in *Gargantua*. Then it increases sharply in the *Tiers Livre* to 4.43, then goes down in the *Quart Livre* and V-16 with 2.19 and 2.45 respectively. It increases again to 5.04 in V-32. For RII it begins with 9.07 in *Pantagruel* and stays close to 9.27 in *Gargantua*. It increases sharply to 17.97 in the *Tiers Livre*, then goes down to 15.67, 13.12 in the *Quart Livre* and V-16 respectively. And in V-32 it goes up again to 18.32.

It is true that the means of any word show fluctuations in any sampling counts, particularly when we have small samples such as 1,000 to 1,500 words. But these computer figures that we are using are the mean of the means. That is to say, the relative frequency of these words is averaged first in each chapter population, and then the results are added together. This again is divided by the number of the chapters in the sample to obtain a new mean

## Table 13

p. 34

## THE MEAN FREQUENCIES OF TWENTY-FIVE SELECTED WORDS AND THEIR GROUPINGS IN FOUR AUTHENTIC BOOKS OF RABELAIS, AND THE *CINQUIESME LIVRE* (Rate per 1,000 words)

value of all the means. In the case of this investigation, most of the books have sixteen chapters sampled with more than 12,000 words in each instance. Thus with a certain amount of confidence we could use these mean frequencies for comparison purposes. It should be pointed out that we are more interested, in this instance, in the emergence of a pattern or trend connected with some twenty years of Rabelais's writings than the mean values of a single variable. In Figure 3 the curves of these four variables in six texts of Rabelais are plotted.

Once again the heterogeneous nature of the *Cinquiesme Livre* appears. The first thing that strikes the eye is the relative position of V16 and V32 in relation to these four variables. V-16 is located at the lowest point of the curves of all four authentic books of Rabelais, while V-32 projects towards

Fig. 3. — Evolution of the Use of Rabelais Indicators and conj. *et* in four authentic books of Rabelais and the *Cinquiesme Livre de Pantagruel*.

p. 35

the highest points of the curves. In fact, it tends towards the position that *Tiers Livre* has attained. The question at once arises : is this a chance occurrence or a definite pattern of evolution in the style of Rabelais ?

It seems very likely that we are witnessing some aspect of evolution in Rabelais's writing pattern. Now our final question arises : Is it possible to use this statistical evidence to establish the probable dates of the composition of the *Isle Sonante* and the last 32 chapters of the *Cinquiesme Livre* ?

p. 36

If we examine all aspects of these curves, we see that wherever the slopes of the four authentic texts of Rabelais are at their lowest range limit, V-16 continues this range but V-32 diverges toward the upper limit range of the slopes. Thus it appears that V-16 is closely associated with the lower end of the slopes and V-32 the upper end. What inference could we make in the light of these statistical facts ? Considering the adverbs ending in *-ment*, adjectives ending in *-ique*, and the combined RII as our criteria of Rabelais's style analysis, we could state that there is a strong likelihood that V-16 may have been composed during or after the composition of the *Quart Livre*, and that the chapters of V-32 have been composed between the period of the *Tiers Livre* and the *Quart Livre*. Thus the fragments theory advanced previously by several critics can be now objectively and statistically verified. It is thus likely that the last 32 chapters of the *Cinquiesme Livre* are associated with the style profile of Rabelais's *Tiers Livre* and *Quart Livre* period.

The analysis of four authentic texts of Rabelais for the use of adjectives ending in *-ique*, adverbs ending in *-ment*, and eight selected adverbs ending in *-ment* has revealed that there exists two distinct periods. The first one corresponds to the *Pantagruel-Gargantua* period and the second one is the *Tiers Livre-Quart Livre*. The RII ranges for these two periods are 7-11 and 14-19 per 1,000 words respectively. It should be borne in mind that these ranges were established for text samples containing 12,000 or more words each.

This does not mean that we cannot apply these RII ranges for shorter text authentications. It only means that if, for

example, we want to authenticate 500-2,000 word samples, and we have obtained RII values outside both ends of the range, we should not conclude at once that the texts in question do not belong to Rabelais. To authenticate short texts by frequency counts alone is generally very difficult. Mosteller and Wallace in their attempt to authenticate the *Federalist Papers* had to use highly complex Bayesian statistics based on prior odds.[5]

Cf. p. 12.

It is with this reservation in mind that an attempt is made to authenticate by means of Rabelais Indicator Index first, the prologue to the *Cinquiesme Livre* and, second, the authentications of chapters XVI, XII-XIV, XXXII bis and the additional ending of the last chapter (Ch. 47) of the *Cinquiesme Livre* found in the manuscript.

During the literature survey of this study we have noted why some critics thought that the prologue to the *Cinquiesme Livre* did not belong to Rabelais. One of the critics was Sainéan who believed that the entire *Cinquiesme Livre* was authentic except the prologue and Ch. XXXII bis. In order to get a fair evaluation of RII authentication, I have included also the authentications of the prologues to *Pantagruel*, *Gargantua*, *Tiers Livre*, *Quart Livre*, and the Ancien prologue to the 1548 partial edition of the *Quart Livre*. Table 14 gives the results of authentication by RII.

## Table 14

p. 37

## RELATIVE RII VALUES FOR PROLOGUES TO *PANTAGRUEL, GARGANTUA, TIERS LIVRE, QUART LIVRE, 48-QUART LIVRE* AND *CINQUIESME LIVRE* (Rate per 1,000 words)

We note at once that even in the prologues the distinction between P-G period of composition and the III-IV period appears evident.

The RII values of the prologues to *Pantagruel* and *Gargantua* are 7.62 and 8.74 per 1,000 words respectively. The RII range for P-G period is 7-11 per 1,000 words. Hence authentication by RII means has proven to be valid. Similarly the RII value of 14.18 for the prologue to *Tiers Livre* falls within the RII range of III-IV period, i.e., 14-19 per 1,000 words. The prologue to the *Quart Livre* and Ancien prologue to *Quart Livre* have 13.42 and 13.08 per 1,000 words respectively. They both are also close to the III-IV range.

The prologue to the *Cinquiesme Livre* has RII value of 17.71 per 1,000 words. This is solidly within the RII range of III-IV. Thus from the point of view of quantitative analysis it appears that there is no reason to doubt the Rabelaisian authorship to the prologue to the *Cinquiesme Livre*.

On several occasions in the course of our literature survey we have taken notice of the arguments that some critics had for doubting the authenticity of the chapters XVI, XXIII, XXIV, and XXXII bis. In Table 15 the results of RII values of these chapters are shown. We note that chapters XXIV, XXXII bis, and the additional ending of Ch. XLVII have RII values of 14.34, 16.67, and 16.34 per 1,000 words respectively. These values are within the RII range of the *Tiers Livre-Quart Livre* period. We note also that chapters XXIII and XVI have RII values of 10.06 and 3.17 per 1,000 words respectively. The former is still within the overall Rabelais range, but the latter is completely outside RII range. This low RII value could perhaps be the resuit of some interpolation or tampering. As it has been mentioned earlier, the authentica tion of short texts by quantitative analysis is extremely difficult. It requires highly complex mathematics. For the time being, then, the authenticity of Chapter XVI can be neither confirmed nor denied by the method developed in this study.

## Table 15

p. 38

## RELATIVE RII VALUES IN CHAPTERS XVI, XXIII, XXIV, XXXII BIS OF THE *CINQUIESME LIVRE*, AND THE ADDITIONAL ENDING OF CH. XLVII (Rate per 1,000 words)

Outside the range of III-IV period, but within overall Rabelais range.

Outside Rabelais range.

1

This is only a partiel word count of the chapter. The long enumeration of the names of the dances and dishes are excluded.

A final test was made to find out if Rabelais Indicator Index can authen ticate genuine Rabelais texts not connected to the stories of *Pantagruel*. For this analysis Three Letters written by Rabelais while he was in Rome at the service of Cardinal Jean du Bellay were chosen. These letters were written from December 1535 to February 1536, and they

were addressed to his protector, Geoffroy d'Estissac.[6]

There have been some questions raised in the past about the authenticity of these letters. However V.-L. Bourrilly has proven that they were authentic. See his *Lettres ecrites d'Italie par François Rabelais* (Paris : Honoré Champion, 1910).

In Table 16 the RII values of the Three Letters are given.

## Table 16

## RELATIVE RII VALUES FOR THREE LETTERS WRITTEN BY RABELAIS FROM ROME FROM 1535-1536

## (Rate per 1,000 words)

We note that the mean RII value for the Three Letters is 8.22 per 1,000 words. This is well within the RII range of P-G period. Once again the validity of RII authentication of Rabelais texts has been proven. The fact that these letters were written during 1535-1536 closely relates to the established range of 1532-1534, the *Pantagruel-Gargantua* period. p. 39

## An Additional Verification of the Authenticity of the *Cinquiesme Livre*

In our discussion of the methodology it was mentioned that some raw data provided by a previous research will be analyzed by means of the Type Token ratio. The results of this quantitative analysis will be then compared with the results obtained in this study concerning the authenticity of the *Cinquiesme Livre*. For the purpose of this analysis I have chosen the use of some selected proper nouns that appear in *Pantagruel, Gargantua, Tiers Livre,* and *Quart Livre*, and their respective occurrences in the *Cinquiesme Livre*.

The criterion for selecting these proper nouns was based on the rank frequency principle. This is similar to the manner in which the eight selected adverbs ending in *-ment* were chosen as possible Rabelais Indicators. The data for the use of the Proper Nouns category was compiled from Marty Laveaux's "Glossaire des Noms Propres" in his critical edition of *Les Œuvres de Maistre François Rabelais*.[7]

*Op. cit.*, Vol. 6, pp. 253-317.

First a list was made of all nouns that appeared five times or more in all four authentic books of Rabelais. From these nouns, then, the ten most frequently used were selected. These Proper Nouns, in descending order, were : Jupiter, Platon, Homere, Hercules, and Bacchus. Aristoteles and Cicéron attained the sixth place ; the remaining three were Aeneas, Alexandre le Grand and Venus. In Table 17 the respective occurrences of these ten nouns are given in the four authentic books of Rabelais as well as the *Cinquiesme Livre*. The Type-Token ratio per 1,000 words is also pre sented, since these nouns were counted in the complete word populations of each book. In our calculation of rate per 1,000 words, a statistical approxima tion of the total word population of each book was considered. These estimates were calculated from the sampled population of each book counted for this study.

We note that these ten nouns appeared in *Pantagruel* at the rate of 0.78 per 1,000 words. This rate was kept steady for *Gargantua* ; in the *Tiers Livre* it reached 1.72 per 1,000 words, then it decreased to 0.87 in the *Quart Livre*.

The descent continued in V-16 which exhibited the rate of 0.64 per 1,000 words. It increased very sharply to 1.63 per 1,000 words in V-32. We notice how closely this pattern resembles the one generated in this study from Rabelais Indicators. In Figure 4 we see a clear presentation of the familiar drop (as shown in Figure 3) between the *Quart Livre* and V-16, as well as the sharp jump from V-16 to V-32. Hence, the use of the ten most frequently

## Table 17 p. 40

## RELATIVE OCCURRENCES OF THE TEN MOST FREQUENTLY USED PROPER NOUNS IN FOUR AUTHENTIC TEXTS OF RABELAIS AND THE *CINQUIESME LIVRE*

1

Raw scores were compiled from Marty Laveaux "Glossaire des Noms Propres," Vol. 6 *Les Œuvre de Maistre François Rabelais*, pp. 253-317.

2

Based on statistical estimates : *Pantagruel*, 36,000 words, *Gargantua*, 46,000 words ; *Tiers Livre*, 53,000 words ; *Quart Livre*, 55,000 words ; V-16, 12,513 words (actual count) ; V-32, 29,500 words.

Note : It should be pointed out that principal characters such as Pantagruel, Panurge, Frère Jean, and others were excluded for contextual reason.

used “Noms Propres” shows also the evidence of the heterogeneous nature of the *Cinquiesme Livre*. Once again this graph shows unmistakably that the last 32 chapters of the *Cinquiesme Livre* belonged to the period of composition between the *Tiers Livre* and the *Quart Livre*.

Fig. 4. — The Use of Ten Selected Proper Nouns in Four Authentic Books of Rabelais and the *Cinquiesme Livre de Pantagruel*. p. 41

What inference can we make from this additional quantitative analysis ? One fact remains evident : that the *Cinquiesme Livre* is definitely a part of the *Tiers Livre-Quart Livre* period, and Rabelais appears to be quite unmistakably its author.

# Chapter V

p. 42

# CONCLUSIONS

## by George A. Petrossian

### *Summary of the Principal Findings*

The findings of this study were based upon the analyses and evaluations of the data as described in the previous chapters. They are as follows :

1. Considering adjectives ending in *-ique*, adverbs ending in *-meni*, and the adverbs *autrement*, *mêmement*, *ordinairement*, *pareillement*, *première ment*, *presentement*, *seulement* and *vraiment* as characteristic of some aspects of Rabelais’s style, the authenticity of the *Cinquiesme Livre* could not be disputed.
2. The *Cinquiesme Livre* is a heterogeneous work. There exists a marked difference of style characteristics between the *Isle Sonante* and the last 32 chapters of the *Cinquiesme Livre*.
3. The prologue to the *Cinquiesme Livre*, the two chapters of “Tournoi de la Quinte,” chapter XXXII bis, and the additional ending of chapter XLVII — all appear to be genuine Rabelais texts. However, the authenticity of chapter XVI, “Apedeftes,” cannot be determined.
4. This study has proved that there definitely exists a marked quantitative stylistic difference between the two earliest works of Rabelais, *Pantagruel* and *Gargantua*, and his later works of *Tiers Livre* and *Quart Livre*.
5. There is a great likelihood that the *Isle Sonante* was written during or after the composition of the *Quart Livre*. The last 32 chapters of the *Cinquiesme Livre* were written in the period between the compositions of the *Tiers Livre* and the *Quart Livre*.
6. A bonus finding during this investigation was that one of the control texts, the *Navigations de Panurge*, which has been at times attributed to Rabelais, does not contain any elements of the style characteristics associated with Rabelais’s works, either those of his earlier period or those characteristic of his later period.

The two critics in the past who had disputed the Rabelaisian authorship of the *Cinquiesme Livre* had promised to write an elaborate and well documented expose to further substantiate their claims. They were Burgaud des Marets and P.-P. Plan. These substantiations were never published. The reason is obvious ; the more thoroughly and p. 43
extensively the *Cinquiesme Livre* is analyzed the more its authenticity is apparent. This conclusion was revealed by our survey of qualitative analyses of Paul Lacroix, Jacques Boulenger, Abel Lefranc, William F. Smith, Arthur Tilley, Lazare Sainéan and Nan C. Car penter.

This investigation, which is based on statistical analyses, in turn corrob orates the findings of these seven critics that François Rabelais is definitely — the author of the *Cinquiesme Livre de Pantagruel*.

Any quantitative analysis is, of course, subject to the laws of statistical probabilities. However, the strength of the findings of the present investigation lies not in the correlation of single or multiple words, but rather it is based on the correlation of a changing pattern of analysis of twenty years of Rabelais’s writing career. Whichever direction the curves of this changing pattern have followed, the style characteristics of the *Cinquiesme Livre* has faithfully adhered to it.

If there has been a decrease in the use of the conjunction *et* from *Panta-gruel-Gargantua* period to *Tiers Livre-Quart Livre* period, this decrease is also noted in the *Cinquiesme Livre*. On the other hand if there has been a marked increase in the use of adverbs ending in *-ment* from Rabelais’s earlier works to his later works, the *Cinquiesme Livre* has exhibited also this increase — thus associating unmistakably the composition of the *Cinquiesme Livre* with *Tiers Livre-Quart Livre* period.

It should be mentioned that the Rabelais Indicators which have deter mined the authenticity of the *Cinquiesme Livre* all have a discriminatory power beyond the .001 level of significance. This is a very powerful factor of statistical proof. In addition to this, the authenticity of the *Cinquiesme Livre* was also proven by Type-Token Ratio analysis. This was based on the usage of the ten most frequently mentioned famous names from antiquity which have appeared in the four authentic books of Rabelais and the *Cinquiesme Livre*. These “Proper Nouns” were : Aenéas, Alexandre le Grand, Aristoteles, Bacchus, Cicéron, Hercule, Homère, Jupiter, Platon and Venus. Once again there was a

considerable increase in the usage of these ten names from the *Pantagruel Gargantua* period to the *Tiers Livre-Quart Livre* period. The *Cinquiesme Livre* also unfailingly showed the rate of increase associated with Rabelais's later works.

Thus the suggestions of many past critics that the *Cinquiesme Livre* was the work of a very skillful "pasticheur" seems most unlikely in the light of our statistical facts. The authenticity of the *Cinquiesme Livre* was proven by means of six parameters containing twenty-one variables — these are indeed staggering mathematical combinations for an imitation.

## *Recommendation*

It would be most desirable to make use of both the Rabelais Indicator and the Rabelais Indicator Index to verify the authenticity of some texts attributed to Rabelais. Among them are *Les Chroniques Admirables du puissant Roy Gargantua* and *La Bataille Fantastique des Roys Rondilarus et Croacus*. This study has generated an objective authorship determinant with a very high level of confidence. It should be put to further use by other investigators to examine books of questionable authenticity. 

Grenoble.

## ACKNOWLEDGMENTS

I wish to express many thanks to all those who have made this study possible. In particular, I am most grateful to Professor Floyd Gray of The University of Michigan for the numerous and very valuable suggestions he has made and the keen interest he has shown throughout the preparation of this dissertation. My thanks are due also to each member of the dissertation committee for their helpful assistance and responsiveness.

Finally, the writer recognizes his obligation to his wife Natica for her continued encouragement and support that has helped to bring this to a conclusion.