

ABSTRACT

BAYLE AND L'AVIS AUX RÉFUGIÉS, A NEW APPROACH

By

Richard Vern Wall

The principal aim of this dissertation was to discover, through a computational stylistic analysis, whether or not Pierre Bayle wrote l'Avis aux réfugiés. To this end, thirty-seven computer programs were developed which combined a series of variables and procedures, some of which had previously been tested individually in other attribution research. New elements were indices of French function words and word-patterns (fixed and variable expressions) and analyses of root-and-variation vocabularies found in works by Bayle and Daniel de Larroque--the most frequently mentioned candidates for the attribution of l'Avis--and in the disputed text. In addition, total texts, rather than samplings, were used. The great amount of data to be analyzed made nearly complete automation the most feasible approach.

Because no method to count them systematically has yet been devised, some aspects of style remain in the realm of subjective analysis. Others, however, when structure allows their incidence to be readily identified and counted, have lent themselves to a quantitative, and therefore more objective, analysis, than that provided by earlier methods. The advent of the electronic computer has provided the literary

scholar a means of processing quantities of information such as a researcher might not previously have assessed during a whole professional career.

In my study, some eight hundred test items were grouped into five major areas represented by the basic computer program which generated their quantitative stylistic measures. These five categories and their corresponding mnemonic program names consist of:

1. Sentence level measures (SENWOL)
2. Sentence beginnings and endings (STYLBEND)
3. Function word frequencies and usage (FREQFUN)
4. Expression frequencies and usage (EXSOR)
5. Vocabulary analysis (ENROOT)

The quantitative use of each variable was computed from works known to have been written by Bayle and Larroque. Figures thus derived were then compared to like values in the disputed text. The results were conclusive. Thirty-six percent of the postulated stylistic discriminants examined showed an appreciable discriminative ability in the works tested. Approximately eight percent of these variables suggested that Bayle authored the disputed text; less than one percent implied that Larroque wrote it. On the other hand, nearly one fourth of all variables tested revealed statistically significant differences of use by Bayle and the unknown author. Likewise, more than twenty-five percent of the variables separated Larroque from the author of the controversial text. Although procedures employed in my study do not represent a foolproof method for authorship attribution, they are, nonetheless, a viable tool for analysis, a means of obtaining objective internal evidence correlative to that obtained through other approaches.

Earlier qualitative and quantitative findings are evaluated and compared to the stylostatistical methods and results produced by my

research. Emphasis was placed on procedures used in attribution studies, particularly those giving substance to the five categories of stylistic indices utilized. Technical sections outline the specific procedures followed and report statistical results of comparative quantitative-use of the variables. Statistical test results led to the conclusion that neither Bayle nor Larroque wrote l'Avis aux réfugiés.

BAYLE AND L'AVIS AUX RÉFUGIÉS

A NEW APPROACH

by

Richard Vern Wall

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Romance Languages

1971

PLEASE NOTE:

**Some Pages have indistinct
print. Filmed as received.**

UNIVERSITY MICROFILMS

© Copyright by
RICHARD VERN WALL
1971

TO MY WIFE
AND CHILDREN FOR THEIR LOVE
AND PATIENCE

ACKNOWLEDGMENTS

Sincere gratitude is extended to all those who contributed to the realization of this project. I would like to thank the Department of Romance Languages for permitting me to develop the subject of this dissertation. I am especially grateful to my committee chairman, Professor Kenneth R. Scholberg, members of my guidance committee, friends and colleagues for their counsel, assistance and encouragement. An especial note of gratitude is acknowledged to the Computer Laboratory of Michigan State University and to Indiana University whose grants-in-aid provided funds for program development and final data processing.

To the Computer Institute for Social Science Research and to the Learning Systems Institute of M.S.U., I express my gratitude for computer time and programming support. This project could not have been realized without special concessions in the use of computing facilities by M.S.U. and Indiana-Purdue computer centers.

A special debt of gratitude is due Paul Gabriel and Gordon Wakefield whose assistance in computer programming and data analysis was invaluable.

Finally, to Ruth Harrod and Drs. Maurice Crane and Helen Lee I express unmeasured thanks.

TABLE OF CONTENTS

CHAPTER	PAGE
LIST OF TABLES	vii
I. INTRODUCTION: THE NEED FOR AN ATTRIBUTION STUDY OF <u>L'AVIS AUX REFUGIES</u>	1
II. THE APPROACH	19
III. PROCEDURES AND PROGRAM DESCRIPTIONS	42
A. Section 1 - Data Description	44
B. Section 2 - Program Descriptions	47
1. SENWOL	47
2. SYLAN	48
3. STYLBEND	50
4. EXSOR	51
5. FREQFUN	55
C. Section 3 - Miscellaneous Programs	58
1. CANAL	58
2. EDITING	59
3. OUTPUT CONVERSION	60
4. FAP	60
5. PIP	61
6. ENROOT	62
7. ANOVAR	65
IV. TEST RESULTS AND ANALYSES	72
A. Sentence-Level Measures	76
1. Summary and Conclusions of SENWOL Results . .	92

CHAPTER	PAGE
B. Sentence Beginnings and Endings	95
1. Statistical Summary of Sentence Beginning Variables	129
2. Conclusions on Sentence Beginnings	131
3. Statistical Summary of Sentence Ending Variables	132
4. Conclusions on Sentence Endings	134
C. Function Word Frequencies and Usage	135
1. FREQFUN Individual Word Analysis	135
a. Verbal Summary of Individual Function Words	147
b. Statistical Summary of Individual Function Words	150
2. FREQFUN Grouped Data	150
a. Statistical Summary of Grouped Function Words	160
3. Function Word Conclusions	162
D. Expression Frequencies and Usage	162
1. EXSOR Individual Expression Analysis	162
2. Verbal Summary of Individual Expressions	177
3. Statistical Summary of Individual Expressions	179
4. EXSOR-Grouped Data	182
5. Statistical Summary of Grouped Expressions	191
6. Conclusions for EXSOR Data	193
E. Vocabulary Analysis	193
F. Summary of Test Results	203
V. SUMMARY AND CONCLUSIONS	214
LIST OF REFERENCES	225

CHAPTER	PAGE
APPENDIX A. LIST OF ARTICLES TESTED	231
APPENDIX B. INPUT DATA - TEXT SAMPLE	232
APPENDIX C. ANNOTATED PROGRAM LIST	233
APPENDIX D. <u>SEnwOL</u> FLOWCHART	237
APPENDIX E. SAMPLE <u>SEnwOL</u> OUTPUT	238
APPENDIX F. TABLE 4:6A. AVERAGE NUMBER OF LETTERS AND SYLLABLES PER WORD AND LETTERS PER SYLLABLE, LESS ONE AND TWO LETTER WORDS	239
APPENDIX G. <u>SYLAN-SUBROUTINE BREAK</u> FLOWCHART	241
APPENDIX H. <u>STYLBEND</u> FLOWCHART	242
APPENDIX I. SAMPLE <u>STYLBEND</u> OUTPUT	243
APPENDIX J. <u>FREQFUN</u> FLOWCHART	244
APPENDIX K. FUNCTION WORD ABSOLUTE FREQUENCIES	245
APPENDIX L. TABLE 4:34A. <u>FREQFUN</u> STATISTICAL COMPARISON, X:L1 88 WORDS	249
APPENDIX M. TABLE 4:34B. <u>FREQFUN</u> STATISTICAL COMPARISON, P:B1 31 WORDS	251
APPENDIX N. TABLE 4:41A. REFINED FUNCTION WORD LIST RESULTS, X = L1 9 WORDS	252
APPENDIX O. TABLE 4:41B REFINED FUNCTION WORD LIST RESULTS, X = B1 15 WORDS	253
APPENDIX P. TABLE 4:41C. REFINED FUNCTION WORD LIST RESULTS, X \neq B OR L1 26 WORDS	254
APPENDIX Q. FUNCTION WORDS IN GROUPS	255
APPENDIX R. TABLE 4:41D. <u>FREQFUN</u> - GROUPED DATA, ABSOLUTE FREQUENCIES	258
APPENDIX S. <u>EXSOR</u> FLOWCHART	259
APPENDIX T. LIST OF <u>EXSOR</u> EXPRESSIONS	262
APPENDIX U. EXPRESSIONS IN GROUPS	269

CHAPTER	PAGE
APPENDIX V. TABLE 4:54A. <u>EXSOR</u> GROUPED DATA, ABSOLUTE FREQUENCIES	215
APPENDIX W. <u>ENROOT</u> FLOWCHART	276
APPENDIX X. <u>ENROOT</u> SYNONYM LIST	277
APPENDIX Y. GLOSSARY OF WORD LISTS FOR <u>ENROOT</u>	290
APPENDIX Z. <u>ENROOT</u> SAMPLE OUTPUT	291
APPENDIX AA. <u>PIP</u> FLOWCHART	292

LIST OF TABLES

TABLE	PAGE
4:1. Average Number of Words per Sentence in One-Hundred-Sentence Units	71
4:2. Overview of Absolute Frequencies	79
4:3. Average Number of Words per Sentence	81
4:4. Average Number of Syllables per Sentence	83
4:5. Average Number of Letters per Sentence	85
4:6. Average Number of Letters per Word, Syllables per Word, and Letters per Syllable	87
4:6A. Average Number of Letters and Syllables per Word and Letters per Syllable, less one and two Letter Words	239
4:7. Regrouping of Test Articles	89
4:8. Average Number of Words, Syllables, and Alphabetic Characters per Sentence; Average Number of Letters per Word, Syllables per Word, and Letters per Syllable	91
4:9. SENWOL Summary	93
4:10. Adjectives as Sentence Beginnings and Endings	97
4:11. Adverbs as Sentence Beginnings and Endings	99
4:12. Articles as Sentence Beginnings and Endings	101
4:13. Conjunctions as Sentence Beginnings and Endings	103
4:14. Exclamations as Sentence Beginnings and Endings	105
4:15. Gerunds as Sentence Beginnings and Endings	107
4:16. Infinitives as Sentence Beginnings and Endings	109
4:17. Interrogatives as Sentence Beginnings and Endings	111
4:18. Latin Words as Sentence Beginnings and Endings	113

TABLE	PAGE
4:19. Negatives as Sentence Beginnings and Endings	115
4:20. Nouns as Sentence Beginnings and Endings	117
4:21. Numbers as Sentence Beginnings and Endings	119
4:22. Prepositions as Sentence Beginnings and Endings	121
4:23. Pronouns as Sentence Beginnings and Endings	123
4:24. Conjugated Verbs as Sentence Beginnings and Endings	125
4:25. Percentage of Use Summary for Sentence Beginnings	127
4:26. Percentage of Use Summary for Sentence Endings	128
4:27. Summary of Sentence Beginnings	130
4:28. Summary of Sentence Endings	133
4:29. Function Words, X = L1	137
4:30. Function Words, X = B1	139
4:31. Function Words, X \neq B1	141
4:32. Function Words, B1 \neq L1	143
4:33. Function Words, X \neq B1 or L1	145
4:34. Function Words, X \neq P	147
4:34A. FREQFUN Statistical Comparison X:L1	249
4:34B. FREQFUN Statistical Comparison P:B1	251
4:35. Summary of Significant Function Words	148
4:36. FREQFUN Individual Words--Statistical Summary	151
4:37. FREQFUN Grouped Data Results	155
4:38. Function-Word Groups Which Suggest Bayle Wrote <u>l'Avis aux réfugiés</u>	157
4:39. Function-Word Groups Which Suggest Larroque Wrote <u>l'Avis aux réfugiés</u>	158
4:40. Function-Word Groups Which Suggest neither Bayle nor Larroque Wrote <u>l'Avis aux réfugiés</u>	160
4:41. FREQFUN Grouped Words Probable Author Summary	161

TABLE	PAGE
4:41A. Refined Function Word List Results, X = L1	252
4:41B. Refined Function Word List Results, X = B1	253
4:41C. Refined Function Word List Results, X = B1 or L1	254
4:41D. FREQFUN Grouped Data, Absolute Frequencies	258
4:42. Expressions, X = L1	165
4:43. Expressions, X = B1	167
4:44. Expressions, X \neq B1	169
4:45. Expressions, B1 \neq L1	172
4:46. Expressions, X \neq B1 or L1	175
4:47. Expressions, X \neq P	176
4:48. Summary of Significant Expressions	178
4:49. EXSOR Individual Expressions--Statistical Summary	180
4:50. EXSOR Grouped Data Results	187
4:51. Expression Groups which Suggest Bayle as Author of <u>l'Avis aux réfugiés</u>	189
4:52. Expression Groups which Suggest Larroque as Author of <u>l'Avis aux réfugiés</u>	190
4:53. Expression Groups which Suggest neither Bayle nor Larroque wrote <u>l'Avis aux réfugiés</u>	191
4:54. EXSOR Grouped Expressions Probable Author Summary	192
4:54A. EXSOR Grouped Data, Absolute Frequencies	266
4:55. Summary of ENROOT Comparative Data	195
4:56. Vocabulary Richness of Five Test Articles	198
4:57. Logarithmic Type/Token Projected Values and Unique Roots . .	200
4:58. "K" Characteristics of ENROOT Data	201
4:59. Results of Verbal to Non-Verbal and Content-Word to Function-Word Comparisons	203
4:60. Summary of Test Results	204
4:61. Percentage of Discriminants which reject Bayle and Larroque as Probable Authors of <u>l'Avis aux réfugiés</u>	206

CHAPTER I

INTRODUCTION

THE NEED FOR AN ATTRIBUTION STUDY OF L'AVIS AUX RÉFUGIÉS

The reasons for assigning certain pieces of literature of dubious authorship to a particular author's canon are diverse. Gerald Bentley once pointed out that theater managers and publishers might attribute plays to a famous playwright merely for advertising purposes.¹ Like plays and other popular non-dramatic literature, polemical pamphlets on religion and political topics may suffer from faulty attribution. Reviewing a letter from M. Arnold in the Nouvelles de la république des lettres, Pierre Bayle examines some weaknesses in the argument for attributing a questionable work to Saint Athanasius. Bayle's arguments against Arnold suggest that he was aware of such publication practices as Bentley mentioned. He writes, for example, that although the title page bore Athanasius' name, it was the copyist who put it there, probably because Athanasius was "plus célèbre et moins odieux" than its real author.² Even though it is frequently possible for human judgment to discern individual writing styles and to attribute passages, or even complete works to certain authors, literary history is full of cases of disputed authorship and canonical uncertainties. In the seventeenth and eighteenth centuries, for example, publication of anonymous and pseudonymous polemical works was a common practice. Many of these works passed into obscurity, while others, such as l'Avis aux réfugiés,³ found their way into the canon of celebrated writers, even though

their literary value and authorship remained in question. The fact that so many disputed works remain suggests two possibilities: either scholars have seen no reason to resurrect an old problem (that of attribution), because stylistically or ideologically the work in question offers nothing new to current criticism, or researchers using human judgment and traditional methods have not been able to agree on definitive attribution.

Elisabeth Labrousse points out that the truth about l'Avis aux réfugiés is so elusive that "l'attribution à Bayle de ce pamphlet est une pomme de discorde parmi les spécialistes et elle a fait couler beaucoup d'encre."⁴ Because qualitatively the style of l'Avis is considered to be "poor," "careless," and "lacking in imaginative power,"⁵ its stylistic qualities would not seem to be justification for its survival. This being the case, there must be other reasons for its continued importance. The appearance of l'Avis was most opportune. Jurieu's Lettres Pastorales⁶ had stirred and offended members of both Protestant and Catholic factions; l'Avis aux réfugiés constituted a timely reply. Shortly after the work appeared, it was unobtrusively refuted⁷ and would likely have been forgotten had not Jurieu, by publicly attributing it to Bayle, revived interest in it, an interest which persists to the present. Therefore, the task remains for modern-day scholars to grapple with the difficulties of definitive attribution of l'Avis aux réfugiés, for if, in reality, Bayle did write l'Avis, it has earned a place in his canon and should be definitively assigned only there. Moreover, since the central theme of l'Avis is strongly political, and since it is the only politically oriented work in his published corpus, proof of his authorship would justify interpretations that Bayle's

works may be infused with political overtones or intent. If, however, Bayle did not write l'Avis, this area of future scholarship should be foreclosed and such interpretations of particular works as Walter Rex's treatment of Bayle's article on David⁸ would have to be judged on inherent evidence alone.

Definitive attribution of l'Avis aux réfugiés is important, then, because 1) it is "one of the few occasional pieces of the period [1685-1715] to survive in French literature,"⁹ 2) if Bayle wrote l'Avis, it has place only in his bibliography, and 3) definite knowledge of authorship will either open or foreclose additional areas of Baylian scholarship.

Studies of l'Avis made since Bayle's death make it clear that scholars of the Bayle canon are concerned about proper attribution of the controversial text, even though their conclusions have differed. Some of the names which have been associated with l'Avis are Péliisson, Larroque, Brueys, Coquelart, Chardon, and Bayle.

Very little is known of Brueys, Coquelart and Chardon¹⁰ whose names have been left out of the most recent studies on l'Avis.

Paul Péliisson-Fontanier was a Protestant courtier and French Academician converted to Catholicism, who had become well known for his polemical works directed against his former "co-religionnaires." A personal friend of Péliisson's and a former diplomat residing in England, Marc-Antoine de Crosat de la Bastide, named him as the author in "l'Auteur de l'Avis aux Réfugiés déchiffré," published in 1716, by l'abbé du Revest.¹¹ In 1907, Charles Bastide reported that "on avait vu chez un libraire de Paris le manuscrit de l'Avis de l'écriture de Péliisson"¹²

Daniel de Larroque, the gifted son of a famous Protestant theologian, at the height of the dispute, just a few weeks after l'Avis appeared, returned to France and converted to Catholicism. Larroque had served as Bayle's "secrétaire" and is reputed to have laid claim to l'Avis.¹³ A young and inexperienced writer, he gained the confidence of the older Bayle and is said to have supplied him with current information on British intellectual life.¹⁴

Many of the exiled Protestants held Bayle, a well-known polemicist among the moderates, responsible for the publication of the controversial pamphlet. Bayle was a master of polemical writing. He had gained a reputation as a literary and religious critic while editing Les Nouvelles de la République des Lettres, and for having written Les Pensées diverses sur la comète (1682), La Critique Générale de l'Histoire du calvinisme de M. Maimbourg (1682), and Le Commentaire philosophique sur ces paroles de Jésus-Christ 'Contrains-les d'entrer' ... (1686).

No one has denied that ideas expressed in l'Avis were currently Bayle's. It has even been suggested that he wrote the preface, having received the manuscript from Larroque.¹⁵ Even so, Bayle repeatedly denied having written l'Avis. However, some of his closest friends, fearing that authorship of the work would be grounds to exile the philosopher who had gone through a conversion to Catholicism and a relapse to Protestantism, who had been exiled from France and who had seen his Chair of Philosophy taken from him both by Louis XIV and by the Dutch Consistory, saw Bayle's hand in the controversial document.¹⁶

In April, 1691 (over a year after l'Avis publication), the first published accusation was made, with Jurieu's Examen d'un libelle contre la religion, contre l'Etat et contre la Révolution

d'Angleterre.... In this pamphlet Jurieu accused Bayle of plotting the overthrow of the Dutch government; stating, then, that Bayle wrote l'Avis as part of a vast "cabale." Thus began "la guerre de pamphlets"¹⁷ which saw the ultimate disintegration of the Bayle-Jurieu friendship.

The "philosophe de Rotterdam" sought to defend himself. He offered "de paraître devant les magistrats de Rotterdam pour être entendu **contradictoirement** avec son adversaire,"¹⁸ but since the consistory appeared to be taking no action, he published his Cabale Chimérique.¹⁹ The entire scandal was so embarrassing to the magistrates, who evidently saw in the "guerre de pamphlets" a personal quarrel rather than a political uprising, that they barred any further dialogue between the two protagonists.²⁰

The first major attribution study of l'Avis after Bayle's death was made by Desmaizeaux as he prepared his Vie de Bayle which precedes the 1730 edition of the Dictionnaire historique et critique. Desmaizeaux recognized the similarities between Bayle's political views and those expressed in l'Avis. His conclusion was, however, that his respected friend did not write the pamphlet. Nevertheless, popular consensus caused l'Avis to be placed among Bayle's Oeuvres Diverses in 1737.

In 1878, M. Deschamps, in his book Genèse du scepticisme érudit de Bayle, refused to recognize Bayle as author of l'Avis. According to Georges Ascoli,²¹ M. Deschamps's refusal was for sentimental, not factual reasons. Then, in 1906, Jean Delvolvé published his Religion, critique et philosophie positive chez Pierre Bayle, in which he associated ideas expressed in l'Avis with those expressed in earlier works known to be by Bayle. His conclusion was that Bayle wrote the work.²² Nevertheless, in 1907, in the Bulletin de la Société d'Histoire du Protestantisme,

Charles Bastide took the opposite stand.²³ He admitted that the internal evidences, that is, the ideas, were attributable to Bayle, but claimed that the external evidences, word of mouth and correspondence of the period, failed to justify the hypothesis of Bayle's authorship. In 1913, Georges Ascoli produced perhaps the most exhaustive study of the external evidence yet made on l'Avis.²⁴ He took as a guide many of the documents referred to by Bastide, using much greater care and perceptive analysis in correlation of the manuscripts he found. As indicated by Mme Labrousse, his arguments are, however, more ingenious than convincing.²⁵ He conclusively eliminates Péliisson from the list of possible writers and through circumstantial, external evidence concludes that Bayle must have been the author.

The most recent study of l'Avis was made by Elizabeth Labrousse in 1965. In true Baylian fashion, she has provided those interested in Bayle with a wealth of documented information. However, her treatment of l'Avis, rather than producing substantial evidence for its proper attribution, offers only a very plausible explanation of why it was written. With regard to the attribution, she discounts most of Ascoli's arguments and agrees with Jacques Basnage's postulation, found in his letter to Desmaizeaux: "Je n'ai point encore abandonné ma première conjecture: c'est que le manuscrit lui en avait été confié. Il le fit imprimer, y ajouta une préface et quelques traits de sa main, [sic] M. Hartsoeker m'a confirmé dans ma conjecture, parce qu'il m'a assuré que M. Larroque, étant prisonnier à Paris, citait souvent cet ouvrage comme une production qui lui appartenait."²⁶ Nevertheless, Mme Labrousse feels that Bayle is to be held totally responsible for the work, even though he might have received the original manuscript from Larroque.²⁷

With the exception of Delvolvé, all of the studies mentioned dealt with bibliographical and external evidence in their attempts to determine authorship of l'Avis aux réfugiés. One type of evidence remained to be examined: ideas, structure and language in the text itself. This is internal evidence. Delvolvé, Howard Robinson and, to a certain extent, Mme Labrousse approached the attribution of l'Avis through the text itself by comparing ideas in the pamphlet to those espoused by the philosopher of Rotterdam.

Delvolvé begins his argument concerning attribution of l'Avis by admitting that "Bayle ne s'est jamais avoué l'auteur de cet ouvrage, que l'on ne songea pas tout d'abord à lui attribuer, mais qui a figuré par la suite dans toutes les éditions de ses oeuvres."²⁸ Nevertheless, he goes on to state that the "idée de la domination du pouvoir de l'Etat au-dessus de toutes prétentions religieuses est présentée dans l'Avis sous diverses formes"²⁹ One of these forms is the doctrine embraced by ancient pagan moralists which centers on "devoirs absolus et sacrés, revêtus d'un caractère quasi religieux ... qu'on doit à la patrie."³⁰ Delvolvé sees a direct relationship between this quasi religious view of state supremacy and Bayle's doctrine of toleration. To elucidate his interpretation, he draws a statement from Bayle's Commentaire philosophique³¹ wherein the philosopher criticizes Catholics whose intolerant attitudes toward the Protestants were not for the good of the public.³² He then points out that reproaches made against the Catholics have a striking similarity to those directed against Protestants by the author of l'Avis aux réfugiés. The link, then, that Delvolvé sees tying together le Commentaire and l'Avis is "le bien public," characterized in the former work by an appeal for toleration and in the latter

by rejection of rebellion. Implementation of these two principles by the exiles would fulfill the most ardent desires of both moderates and zealots. Delvolvé concludes this portion of his argument by suggesting that because orthodox Calvinists (led by Jurieu) exhibited great opposition to the doctrine of tolerance presented in the Commentaire, Bayle wrote l'Avis as "une attaque directe contre des principes et un esprit dont il a maintenant éprouvé l'hostilité."³³

The remaining arguments Delvolvé proposes, dealing less with literary parallelisms and more with logical reasoning, may be summed up as follows:

1. The author of l'Avis encouraged the return of exiles to France: Bayle would have preferred living in Paris;³⁴
2. Bayle, having become strongly embittered toward Jurieu, sought to "atteindre au vif Jurieu et l'orthodoxie, qu'il commençait à haïr ...;"³⁵
3. there was cause to conjecture that the French government "envisagerait peut-être ... un accord avec les plus modérés des réfugiés, qu'on ferait servir à calmer les irritations: en usant à leur égard de clémence, on désorganiserait le parti, rappelant les uns, laissant dehors les éléments les plus dangereux." Such conjectures, according to Delvolvé, prompted Bayle to profit from the situation by raising "au-dessus des intérêts religieux la suprématie de l'intérêt de tolérance ...;"³⁶
4. and finally, Bayle "se défendit sur le point de fait avec une faiblesse qui fut remarquée de tous.--S'il maintient ses dénégations avec fermeté, jamais en revanche il ne désavoue

ni ne blâme l'essentiel des doctrines soutenues. Sans doute il blâme ce qui traite à la glorification du roi de France et de la politique française. Mais il approuve et prend à son compte cette idée fondamentale de l'Avis que l'esprit de satire et de rebellion est toujours condamnable ...;"³⁷

therefore, argued Delvolvé, Bayle must have written l'Avis aux réfugiés.

Not so, responded Robinson in Bayle the Sceptic. "Nothing could be more unlike what he had thus far written. It [l'Avis] was wholly taken up with the problem of sovereignty. Its . . . spirit was political and not moral; Bayle had not yet shown any interest in politics."³⁸ Such are some of the positive statements used by Robinson to introduce his argument against Bayle's authorship of l'Avis. He seems to ignore the parallels and logic developed by Delvolvé, or else he feels that his arguments are not in keeping with the spirit of Bayle's corpus. Moreover, doctrines in harmony with Bayle's thought were likewise espoused by several other members of the moderate party. Bayle himself acknowledged that there were in the reformed party

des bonnes âmes qui sont encore persuadées, malgré les déclamations et les livres de M. Jurieu, qu'il faut aimer ceux qui nous haïssent, prier pour ceux qui nous persécutent, souffrir patiemment pour le nom de Dieu, ne rendre point le mal pour le mal, l'injure pour l'injure, ni écrire des satires³⁹

Moreover, Eric Haase pointed out that many Huguenots arrived in exile still proclaiming fidelity to both France and Louis XIV;⁴⁰ to absolutism. Furthermore, Mme Labrousse confirmed that l'Avis was written "par un partisan convaincu de l'absolutisme--il n'en manquait pas ... parmi les ministres les plus pieux."⁴¹ Or, in other words, the position of the author of l'Avis was not unique among the exiles.

In any case, Robinson lists several additional disparities between the context of l'Avis and Baylian thought. He dwells upon Bayle's lack of knowledge of the English scene to establish his conclusions. He points out that Bayle's "previous treatment of things English had been uniformly vague."⁴² Finding no detailed descriptions of the British political situation anywhere else in the Bayle canon, Robinson naturally assumed that "the elaborate handling of English history and the use of English authorities" were beyond Bayle's immediate command.⁴³ Whereas Delvolve saw a relationship between political aspects of l'Avis and le Commentaire, Robinson felt that the extreme position taken on royal sovereignty by the author of l'Avis nullified any right to tolerance. In addition, there is not a word about tolerance in the entire volume; the whole spirit of l'Avis is opposed to freedom of thought. Hence, in a very straightforward manner, Robinson sums up his argument against Bayle's authorship of the disputed work:

Bayle's erring conscience is entirely absent. The argument throughout is strongly Catholic.⁴⁴ The use of the first three centuries of Christian history is the opposite of that made by Bayle in his Philosophical Commentary. The defense of the killings of the Vaudois is in direct contradiction to the position taken in the reply to Maimbourg. If Bayle could have written France Entirely Catholic⁴⁵ in 1685, and this work in 1690, he was a veritable chameleon.⁴⁶

Thus, Robinson states in no uncertain terms the disparities his research revealed. It now seems appropriate to ask: "Whom shall we believe?" Having examined external evidence, Bastide and Ascoli arrived at different conclusions: Bastide for, Ascoli against attribution to Bayle. Delvolvé saw l'Avis as a logical continuation of Bayle's thought patterns and, therefore, attributed it to him. Robinson, also

approaching the text internally, produced evidence diametrically opposed to that which evolved from Delvolvé's reasoning. Thus it becomes clear that coexistence of ideas has provided inferential, but not conclusive evidence for attribution of l'Avis aux réfugiés, and that use of traditional methods has not produced definitive internal evidence.

Intuitive analysis is much more elusive than the search for ideas or stylistic differences. Nonetheless, some critics⁴⁷ are endowed with what seems to be an innate ability to sense an author's touch in a work. Most of the early attributions of l'Avis aux réfugiés were based upon intuition. To make accurate impressionistic judgments effectively and consistently demands a union of wide learning and "esthetic perception" brought about by a long and concentrated study of an author's works.⁴⁸ For the most part, however, the average critic is "all too often . . . confronted with a passage which seems to offer no point of entry."⁴⁹ If an objective method can be devised for penetrating beneath the surface of a text, the ultimate process of analysis may be considerably shortened. Relating intuition to attribution research, Stephen Wachal states that "intuition can indeed be useful, but . . . in an investigation of this kind [attribution studies] it should be supplemented by the routine examination of as many variables as the limits of practicality permit."⁵⁰ Pinpointing of dominant stylistic traits should provide the necessary complement for such intuitive research.

An analysis of style is defined here as the sum total of the stylistic elements noted below, and the author's choice, consciously or intuitively, of how to use them. Interest in style investigation has grown rapidly since 1930,⁵¹ and penetrating research by such scholars

as Spitzer,⁵² Bally,⁵³ and Bruneau,⁵⁴ has lengthened the ever growing list of possible stylistic determinants. Until 1930, the staple categories of this type of evidence were primarily versification and vocabulary.⁵⁵ Since then imagery, simile, metaphor, rhythm, inflection, grammar, syntax, use of sources, indeed even spelling and punctuation (where it can be established that differences are the author's and not the publisher's choice) have been added. Some of these elements remain in the realm of subjective analysis, simply because no method has yet been devised to count them systematically.⁵⁶ Others, because their structure is such that their presence can be readily identified and counted, have lent themselves to a more objective type of analysis. In cases involving vocabulary, rhythm, inflection, punctuation and even some aspects of syntax, the electronic computer has been programmed to recognize, sort, count, and report occurrences of the specified items.⁵⁷ Thereby the scholar engages a highly efficient but brainless clerk to process "information in such quantities as no man's lifetime or energy could previously have contained."⁵⁸ It is to be noted that the computer scholar's intent is not to have a machine take over the art of literary criticism. Rather, he uses the machine to prepare the way, and, in some cases, find a "point of entry" so that the literary critic, the historian or the investigator of style may get on with his business. Because of its quantitative nature, the automated approach to stylistic analysis has received the names "stylostatistics" or "computational stylistics." Whatever the approach taken to solve an attribution problem, the literary critic and the literary historian must bring to bear all possible external and internal evidence, using the most exact means at their command. To this end, it is conceivable that the process of subjective

judgment could at least partially be formalized or rendered explicit, especially with regard to stylistic identification. The union of human judgment, modern computer technology and statistical procedures might then provide reliable, replicable information about stylistic features of a particular author.

An author's style--conceived of as constant features or combinations of features in his writing habits, or in his choice of words--when analyzed, may reveal facts that a pseudonym might otherwise have kept hidden. In a well known quotation, Baudelaire wrote: "Pour deviner l'âme d'un poète, ou du moins sa principale préoccupation, cherchons dans ses oeuvres quel est le mot ou quels sont les mots qui s'y représentent avec le plus de fréquence. Le mot traduira l'obsession."⁵⁹ Baudelaire uses the term "obsession" to describe that element in the soul of a writer which cannot be permanently disguised by an act of the author's will.

In recent years a considerable body of evidence supporting the view that "an author's individuality is at least partially inherent in the frequency with which lexical and grammatical elements occur in his texts"⁶⁰ has been published. Wachal reviews upwards of 150 documents in which possible objective correlates of style were treated and points out that frequency studies of obviously discernable stylistic elements have been successfully discriminated by objective means.⁶¹ Because of the results reported in past studies, the basic hypothesis of a computational stylistic study--that there is, indeed, a relationship between grammatical and lexical frequencies and the paternity of texts--will be taken as proved in the development of this study.

If Bayle did write l'Avis, he took care to disguise the fact by all means available to a man eager to protect his own life and safety. Nonetheless, even in his own time there were those who assumed it was his. Some were likely swayed by hearsay evidence, some by textual elements, such as recurring word or thought patterns. If one can "fingerprint" a style by discovering those characteristics so inextricable from the writer's personality that he does not even perceive them as pertaining peculiarly to himself and therefore is at no pains to obscure them, it should be possible to determine, through a computational stylistic analysis of l'Avis aux réfugiés, whether or not it was the work of Bayle's pen.

NOTES FOR CHAPTER I

¹Gerald E. Bentley, "Authenticity and Attribution in Jacobean and Caroline Drama," Evidence for Authorship, eds. David Erdman and Ephim Fogel (Ithaca: Cornell University, 1966), p. 180.

²Pierre Bayle, "Nouvelles de la République des Lettres," juillet, 1685, art. iv., in his Oeuvres diverses, ed. Elisabeth Labrousse, 4 Vols. (1727; rpt. Hildesheim: Georg Olms, 1864-1968), I, 327. All quotations from Bayle's Oeuvres diverses are taken from this 4-volume reprint of the 1727 edition. Cited hereafter as OD, I, II, III, IV.

³Avis important aux Réfugiez sur leur prochain retour en France, Donné pour Etrennes à l'un d'eux en 1690. Par Monsieur C.L.A.A. P.D.P. Reprinted in Bayle, OD, II, 579-633. Attributed also to Daniel de Larroque and Paul Péliisson-Fontanier.

⁴Pierre Bayle I, Du Pays de Foix à la cité d'Erasmus (The Hague: Nijhoff, 1963), p. 2]9.

⁵Howard Robinson, Bayle the Sceptic (New York: Columbia University Press, 1931), p. 122. and Labrousse, Pierre Bayle I, 221.

⁶Pierre Jurieu, Lettre [sic] pastorales adressées aux fidèles de France qui gémissent sous la captivité de Babylon [sic] ..., 3^e éd. (Rotterdam: A. Acher, 1686-1689).

⁷According to Desmaizeaux, Vie de Bayle, éd. Beuchot, pre-facing Bayle's Dictionnaire historique et critique, (Amsterdam, 1730), pp. 119 A, 124 B, it was refuted by tronchin du Breuil, Basnage de Beauval and Antoine Coulan. See also Walter Rex, Essays on Pierre Bayle and Religious Controversy, International Archives of the History of Ideas (The Hague: Martinus Nijhoff, 1965), p. 225f.

⁸Ibid., pp. 197-255.

⁹Ibid., p. 225.

¹⁰Mentioned by Georges Ascoli, "Bayle et l'Avis aux réfugiés," Revue d'Histoire littéraire de la France, XX (1913), 521.

¹¹Robinson, p. 529.

¹²Charles Bastide, "Bayle est-il l'Auteur de l'Avis aux réfugiés?," Bulletin de la Société d'Histoire du Protestantisme, LVI (1907), 550.

¹³Ascoli, p. 532.

¹⁴Labrousse, Pierre Bayle I, p. 219.

¹⁵Jean Delvolvé, Religion, critique et philosophie positive chez Pierre Bayle (Paris, 1906), p. 192.

¹⁶Ascoli, pp. 539-544.

¹⁷Labrousse, Pierre Bayle I, 226.

¹⁸Bastide, p. 549.

¹⁹OD, II, 637-685.

²⁰Bastide, p. 549.

²¹Ascoli, p. 522.

²²Delvolvé, p. 191.

²³Bastide, pp. 544-558.

²⁴Ascoli, p. 522.

²⁵Labrousse, Pierre Bayle I, 220.

²⁶Lettre de Basnage à Desmaizeaux, du 19 avril 1707, cited in Delvolvé, p. 192f.

²⁷Labrousse, Pierre Bayle I, 221.

²⁸Delvolvé, p. 178.

²⁹Ibid., p. 183.

³⁰Ibid., pp. 183-184.

³¹Commentaire philosophique sur ces paroles de Jésus-Christ: "Contrains-les d'entrer," OD, II, 355-496.

³²"Toute secte qui s'en prend aux lois des sociétés, et qui rompt les liens de la sûreté publique en excitant des séditions et en prêchant le vol, le meurtre, la haine, le parjure, mérite d'être exterminée par le glaive du magistrat." Ibid., p. 412. Cited in Delvolvé, pp. 184-185.

³³Delvolvé, p. 188.

³⁴Delvolvé suggests that Bayle would have known greater freedom from "la fournaise théologique du Refuge" and would have been able to "jouir en paix du libre commerce des esprits distingués et polis dont Paris était la patrie." Religion, p. 188. He seems to overlook the fact that Bayle knew greater freedom of the press in Holland than he could have possibly known in France. See also Edmond Lacoste, Bayle, novelliste et critique littéraire (Paris: Picart, 1929), p. 63f.

³⁵Labrousse, Pierre Bayle I, p. 189.

³⁶Ibid.

³⁷Ibid., p. 190.

³⁸Robinson, p. 120. Robinson seems to have overlooked Bayle's interest in political affairs mentioned by Desmaizeaux, Vie de Bayle, Dictionnaire, I, xxi; referring to Bayle's desire to keep politics out of les Nouvelles de la république des lettres, or to the fact that politics are not the central issue of any of his other works. Politics--la politique, les politiques--account for only 24 entries in the "Table de matières" concordance to Bayle's complete works.

³⁹Entretiens sur la cabale chimérique, OD, II, 625.

⁴⁰Erich Haase, Einführung in die Literatur des Refuge. Der Beitrag der französischen Protestanten zur Entwicklung analytischer Denkformen am Ende des 17. Jahrhunderts (Berlin: Duncker & Humblot, 1959), pp. 275-276.

⁴¹Labrousse, Pierre Bayle I, 221.

⁴²Robinson, p. 120.

⁴³Ibid.

⁴⁴Notwithstanding the fact that Scriptures quoted were drawn from the Protestant version of the Bible and an extremely well informed awareness of Protestant history was exhibited by the author. It seems only logical that when writing to Protestants an author should use terms--in this case Scriptures--with which Protestants are the most acquainted. Even though each study made of l'Avis which attributes the work to Bayle has used this argument to substantiate its position, I fail to see its import.

⁴⁵Ce que c'est que la France toute catholique sous le règne de Louis-le-grand, OD, II, 336-354.

⁴⁶Robinson, pp. 120-121.

⁴⁷Such as Leo Spitzer, A Method of Interpreting Literature (Northampton, Mass: Smith College, 1949), and Jean Starobinski, L'oeil vivant (Paris: Gallimard, 1961).

⁴⁸Richard Altick, The Art of Literary Research (New York: Norton, 1963), p. 72, likens the literary scholar in an attribution study to an art expert called in to authenticate a museum's new acquisition. After all other tests have proven inconclusive, "he must finally rely upon his knowledge of the way the artist customarily worked. When a specialist who has spent years of his life reading and re-reading his author declares that a disputed piece is genuine, his intuitive expertness must be given respectful attention. Yet no such verdict can ever be regarded as final."

⁴⁹R. A. Sayce, Style in French Prose (Oxford: Oxford Univ. Press, 1953), p. 2.

⁵⁰Stephen Wachal, "Linguistic Evidence, Statistical Inference, and Disputed Authorship," Diss. Wisconsin 1966, p. 315.

⁵¹Helmut A. Hatzfeld, A Critical Bibliography of the New Stylistics Applied to the Romance Literatures, 1900-1952 (Chapel Hill: Univ. of North Carolina Press, 1953), p. iv.

⁵²Leo Spitzer, Stilstudien, 2 vols. (Munich: Huber, 1928); Linguistics and Literary History (Princeton, 1948); A Method of Interpreting Literature (Northampton, Mass., 1949); and "Stylistique et critique littéraire," in Critique, XI (1955), 595-609.

⁵³C. Bally, Traité de stylistique française, 2^e éd., 2 vols. (Heidelberg: Winter, 1919).

⁵⁴C. Bruneau, "La Stylistique," Romance Philology, V (1951), 1-14.

⁵⁵Altick, p. 70.

⁵⁶Such as imagery, simile and metaphor. However, as a scholar reviews a questionable work and associates these (subjective) stylistic elements with those extant in other works by various authors, a counting and sorting process takes place subconsciously. This process does not differ greatly from that employed by the electronic computer, or, as Norman H. Holland so aptly put it: "The computer is the first agency outside the human mind to process symbolic data." "Futures: A Non-Summary of the EDUCOM Symposium on the Computer and Humanistic Studies," Computers and the Humanities, II (Nov. 1967), 59 (Abbreviated hereafter as CHum).

⁵⁷Akin to computer-aided stylistic and literary analysis is the field of computational linguistics which, as its name implies, makes extensive use of the computer's clerical abilities. Louis T. Milic very aptly illustrates the difference between the fields of literature and linguistics vis-à-vis the study of language in his article, "Winged Words: Varieties of Computer Application to Literature," CHum, II (Sept. 1967), 24.

⁵⁸Ibid., p. 25.

⁵⁹Cited by Robert T. Cargo, A Concordance to Baudelaire's 'Les Fleurs du mal' (Chapel Hill: Univ. of North Carolina Press, 1965), p. xi, and in The French Review, XXXIX, No. 5 (April, 1966), 807.

⁶⁰Wachal, "Linguistic Evidence," p. 2.

⁶¹Ibid.

CHAPTER II

THE APPROACH

Several hundred reports of research seeking possible objective correlates of style are scattered through journals and monographs representing such fields as literature, linguistics, biblical studies, journalism, education, statistics, and psychology. Research into that body of literature determined the selection of areas of computational stylistics treated in this project. Works referred to in this chapter, while representative of a larger body of scholarship, are limited to studies concerned with quantifiable indices of style that can be automated, and which are, therefore, directly applicable to the development of procedures used in this attribution study.

In his study of authorship attribution procedures, Wachal reviews 150 previous studies treating objective stylistic correlates. He then groups the problems, procedures, and direction taken by past research into three models: consistency, population, and resemblance. The "consistency model" involves the examination of a work attributed to an author by subjective reasons or on the basis of strong external evidence. After index values for the test works are established, they are compared to see if disputed text values fall in or near the range of those for the known texts.¹ The "population model" involves an examination of samples of a substantial amount of material by different authors, evaluated in a complex probabilistic framework. Finally, the "resemblance model" entails an examination only of works by likely candidates, excluding all other possible authors. As with the consistency model,

attribution is made if index values for one of the hypothesized writers resembles those of the disputed text more closely than the others. This approach is especially effective if the possible-author field has been narrowed to two.²

An attribution study of l'Avis aux réfugiés is suitable to two of these three models, consistency and resemblance. Consequently, the "either-or hypothesis" used in the statistical analysis reported in Chapter IV (pp. 93-94) of this study is based upon the assumptions governing these two models. Past attribution studies of l'Avis aux réfugiés have been made by "strong subjective reasons or on the basis of external evidence." As a result of past research, the field of probable authors has been narrowed to two: Bayle and Larroque, and l'Avis remains in the bibliography of both writers. Since the consistency model is "primarily useful for rejecting works from membership in a canon,"³ and because the major purpose of this study is attribution of l'Avis, which can be done only by eliminating one of the two probable authors, reputed indices from the consistency model provide a valid point of departure for my research. By the same token, since works of unquestionable authenticity--of the same genre, from the same time, and concerning like subjects--are available for the two hypothesized authors, the resemblance model also validates the approach taken in this thesis. In the resemblance model, a set of putative stylistic discriminants drawn from these known works is compared to like indices obtained from the anonymous work. A close resemblance between stylistic elements in the disputed text and those in the known works strongly implies mutual authorship. Conversely, lack of similarity to either or both known works provides a basis for rejection of the hypothesis that either writer is the author of the disputed text.

Concepts established by works in both the consistency and resemblance models suggested five broad categories of putative stylistic discriminants as being most applicable to this study. These five categories consist of: 1) sentence-level measures--number of words, letters, and syllables per sentence, number of letters and syllables per word, and number of letters per syllable; 2) sentence beginnings and endings--classified as to part of speech; 3) function word use--analyzed both individually and in grammatical groups; 4) fixed and variable phrasology--analyzed both individually and in grammatical groups; and 5) vocabulary analysis--examination of word roots and their variations.

Selection of variables for each of the five categories was determined by their automative capabilities--recognizeable and countable incidence; their effectiveness as discriminants as demonstrated in previous studies; their grammatical and syntactical combinations which are likely to identify an author; and their probable use as distinguishable elements which an author could not consciously disguise.

George Udny Yule, a pioneer in the field of computational stylistics, proposed the sentence as an indicator of an author's style in an article published in Biometrika in 1936.⁴ More precisely, he sought to show that the variations of sentence-lengths about their average is a constant characteristic of an author's writing habits. The first few pages of his article treat the problem of punctuation. He quotes Ronald B. McKerrow⁵ who had argued that much, if not most, of the punctuation of sixteenth- and seventeenth-century manuscripts was handled by the compositor, not by the author of a work. After noting that different versions of the texts with which he was working were punctuated differently, Yule concludes that where "punctuation, even as

regards full stops, is largely the work of the compositor, there need be no hesitation in overriding them if necessary: indeed, the use of personal judgment seems unavoidable."⁶ Thus, in preparing his data he revised the punctuation as he saw fit. By doing so he disregarded a principle innate to objective analysis: the objectivity of the evidence provided is only as good as the means used to obtain it. If computational stylistics is to introduce objectivity into literary analysis, then Yule's theory of punctuation revision must be rejected, or at best accepted with reservation, for the following reasons:

(1) Even though it is generally admitted that many authors exercised little care about punctuation in the seventeenth and eighteenth centuries, no definitive study has been made distinguishing those who did from those who did not. A later edition of a given work might well contain corrections suggested or entered by the author.

(2) If a modern critic manipulates punctuation by replacing colons, for example, now with a comma, then with a period, he has increased the size of the "possible-author" field;⁷ assumed that rules (written or common) governing punctuation and sentence composition in earlier periods of writing had already undergone our modern refinement, (for he is, in a sense, modernizing the syntax);⁸ or he has determined that the compositor was not versed in the grammatical usage of his time and punctuated at random.

(3) Attribution based on frequency distribution of sentence lengths where the investigator defines the sentence parameters becomes a somewhat personal venture, dependent upon the knowledge and sensitivity of the scholar.

Yule was not unaware of the high degree of subjectivity these procedures brought to his study. In fact, he assented to the possibility that his study might serve "only as an exploratory piece of work," while hoping it would "still retain interest and value."⁹ It has done both. Yet, as he exposed additional difficulties he had encountered, he continued to combine objective and subjective measures. How should hyphenated words, numbers, the ampersand, and quotations be treated? His decisions on hyphenated words, numbers, and the ampersand were practicable and objective, whereas treatment of quotations posed a greater problem. He considered an author who incorporates brief quotations into the grammatical context of his own sentences to be merely substituting someone else's words. These quotations Yule considered part of the sentence itself. On the other hand, he felt that a complete sentence quoted by an author represents someone else's writing and should be excluded from the analysis. He soon recognized that this seemingly clear distinction became blurred in practice. Rather than make an individual decision with the appearance of each quotation--especially very long ones--he simply left out of his samples "all pages on which this source of trouble was serious."¹⁰ In spite of its shortcomings, Yule's work continues to be cited by scholars and students of computational stylistics.

Word, letter and syllable counts, as they apply to sentence length, vocabulary analysis, and rhythm in both prose and poetic writings have demonstrated discriminating ability.¹¹ When Wachal prepared his data for syllable analysis in English, he determined syllable breaks "simply by reading the text aloud and estimating the number of stresses perceived."¹² Since vocalization and perception of verbal stresses may

vary from reader to reader, a more explicit means of counting and comparing letters and syllables was developed for the present study. Basic rules for syllabification, described in Chapter III (p. 49), were programed into the computer, and accurate replicable results were obtained.

At Columbia University, Louis T. Milic used the IBM 1620 computer as a tool in his attempt to objectify certain traits of Swift's style. In his chapter headed "Connection," he refers to such neutral connectives as "and," "but," and "for," especially as sentence beginnings where they function as transition or reference "fillers." Although accepted and examined by Milic as a possible discriminators of style, he feels this set of variables represents elements of style which the author can consciously control.¹³ This might be true; nevertheless, upon further consideration of sentence beginnings, as a putative index of an author's style, it was felt that, although sentence first words might fall into the category of "conscious ordering" of stylistic elements, it seems logical that, as a writer introduces his thoughts, he is much more aware of how he begins a given sentence than of how he closes it. Moreover, if he consciously observes this element of style in his own writing, it follows that he is likely to be aware of similar or divergent habits among other writers. Having such an awareness, an experienced writer might very well see diversity of sentence beginnings as a means of concealing his identity for any number of reasons. On the other hand, an inexperienced writer, as he develops his own style, might seek to imitate patterns of sentence beginnings he recognizes in celebrated writers. In any case, if a writer, famous or unknown, is more aware of sentence beginnings than of sentence endings, it follows that the latter

are more likely to be unconscious acts and, therefore, "are more likely to reveal something that the writer might deliberately wish to conceal."¹⁴

To an uninstructed layman, one of the most impressive statistically oriented attribution studies was made on the authorship of the disputed Federalist papers by Frederick Mosteller and David Wallace, two professional statisticians.¹⁵ Like many of their predecessors, they recognized human limitations in precision, accuracy, and objectivity.

As a graduate student in 1941, Mosteller became interested in the Federalist authorship problem. In that pre-computer setting, he, Frederick Williams, and their wives, inspired by Yule's research, set out to count sentence-lengths in the known essays of Hamilton and Madison, the two contenders for authorship of the disputed Federalist papers. It wasn't long before they discovered "an important empirical principle: people can't count, at least not very high."¹⁶ Finally, after eliminating their errors and tabulating their calculations, they discovered, to their dismay, that Hamilton's average sentence-length was identical to Madison's. Other expected discriminants proved equally inconclusive and they abandoned the project. It lay dormant until 1962, when Mosteller and David Wallace revived and pursued it with statistical methods and the electronic computer, because "standard methods of historical research [had] not firmly settled this authorship problem."¹⁷ That the Federalist dispute provided them with "a case study that would give [them] an opportunity to compare the more usual methods of discrimination"¹⁸ using complex statistics was of much more importance to Mosteller and Wallace than the literary-historical conclusions they hoped to reach.

The two statisticians concentrated their study on function words: prepositions, conjunctions, and articles. Even though certain parts of speech, e.g., personal pronouns and auxiliary verbs, score highly as function words, Mosteller and Wallace considered them to be potentially dangerous because they "are likely to be related to external details, and inference from them is difficult."¹⁹ From Miller, Newman, and Friedman²⁰ they obtained a list of 363 filler-type words. Even though some of the words from this list were not relevant to the Federalist period, they proved to be very useful, primarily because they were objective with respect to the Federalist problem, and because they relieved the investigators of "a large onus of choice."²¹ To this "unselected" list they added 28 words drawn from their 3000 word samples of 11 Federalist papers by screening out low-frequency words.²² Finally, they constructed an index of 18 Hamilton Federalists and 19 Madison papers external to the Federalist, compiled an additional register of 240 possible discriminators, discarded those they regarded as contextual, and compared the 103 remaining ones to the 98 previously chosen. The "new" terms were then added to the previous lists, raising the total number of possible discriminators to 165. Then came the task of scrutinizing the lists to rid them of terms that might even suggest contextuality. Sifting and culling, they arrived at a final list of 30 words upon which they based their study.²³

A central theme pervades the non-technical sections of Mosteller and Wallace's study and is brought into focus as they summarize authorship and discrimination problems: "The function words of the language are a fertile source of discriminators. . . . Context is a source of risk. We need variables that depend on authors and nothing else.

Some function words come close to this ideal, but most other words do not." They go on to say that "narrow and specialized variables may be of more use than global and meaningful ones."²⁴ Mosteller and Wallace strongly affirm a distinction between the scholar and the critic (the former seeks answers, the latter analyzes them). But not all computer-oriented literary researchers are so definitive in their discrimination. Stephen Parrish,²⁵ Charles Muller,²⁶ and Louis Milic,²⁷ for example, attempt to remove the barrier between "measurement" and "judgment."²⁸

Assuming that the "uniqueness of an author shows consistently in his style, regardless of the subject-matter or the conventions of the medium of the period,"²⁹ and that an author may more readily change words than grammatical structure, Milic concentrated on syntactic, rather than vocabulary patterns. In making this assumption he considered three basic points: "(1) that style reflects personality; (2) that this is an unconscious process; and (3) that in mature writers the process is consistent."³⁰ A careful reading of Swift's works revealed his tendency to use certain word groupings, which Milic discusses in his chapter "Seriation." He calls the making of lists, catalogues, series, or accumulations "the organizing principle of Swift's thought."³¹

Milic chose random samples from four writers (Addison, Johnson, Gibbon, and Macaulay) to serve as controls against which he would test samples of Swift's writing. All words were analyzed and "manually" assigned a **word-class**, consisting of a two-digit number. Each sample text was thus reduced to a series of significant numbers which were then punched onto IBM cards and fed into the computer. Having been programmed to identify the encoded word-classes, to tabulate, correlate, and print their frequency distributions along with a miscellany of comparative

tables, the computer provided the objective information which allowed Milic to create a graphic profile of Swift's style. Finally, he compared this profile to index values obtained from A Letter of Advice to a Young Poet,³² and attributed the work to Swift.

In his attempt to solve the problem of authorship of the Junius Letters, Alvar Ellegård first searched the Oxford and Cambridge matriculation lists, hoping to find a signature which resembled the Junian hand. Unsuccessful in his endeavors and dissatisfied with the historical, biographical, and inconsistent linguistic evidence,³³ he turned to the statistical tests devised by Yule in hopes of finding a solution to the problem through a systematic study of the Letters' language or style. Finding that Yule's tests were not sensitive enough to treat his small samples³⁴ (many of the Junius letters are under 2000 words in length), he devised his own system of stylistic tests. Ellegård made two assumptions in formulating his theory: first, that certain verbal patterns in a particular author's writing habits remain reasonably constant; second, that some aspects of his style are distinct enough to set him apart from his contemporaries. He then defined style in the context of his research as being synonymous with "constant features or combinations of features in an author's way of writing,"³⁵ a definition which makes unnecessary any value judgment of "good" or "bad" style. The characteristic of style he chose to investigate was an author's use of "typical" words and turns of expression.³⁶

Ellegård picked out words used more frequently by the writer in question than by his contemporaries. These he labeled "plus-words." Phrases or expressions were counted only when a plus- or minus-word was part of its composition. For each of these terms he calculated a

"distinctiveness ratio"--a ratio of frequency of occurrence in Junius to that in general usage³⁷ as determined by his control-group samplings.³⁸ Ellegård felt that those words with a distinctiveness ratio greater than 1.5 and less than 0.7 distinguished Junius clearly enough from his contemporaries to compare them with the writings of Francis, the suspected author.

Ellegård's procedures have received both criticism and praise. It has been said that his works contain "errors of exposition and fact," that he used such small samples in his research that "the tests were not sensitive enough to be wholly reliable,"³⁹ that he failed in not using the electronic computer for "sifting and culling," and that he presented forty-five pages of perfectly worked-out tables . . . based on shifty data."⁴⁰ On the other hand, these same critics acclaim the work as a "remarkable achievement,"⁴¹ and commend its author for having "opened the way to further work on the statistics of style."⁴² The reviewer for the Times Literary Supplement best summarized the procedures Ellegård might have followed in making a more convincing study as he wrote: "If only Mr. Ellegård had [sic] more use of his electronic computer and had examined more stylistic features! The introduction of more variables might have reduced the 300⁴³ figure, and convinced us completely that Francis was indeed Junius."⁴⁴

A combination of content and function words characterized the study of Milton's influence on Shelley conducted by Professor Raben of Queen's College.⁴⁵ He sought to show how often, in any sentence, Shelley used Milton's actual words. To Professor Raben's surprise, he found that the number was much greater than he had estimated. An observance of function words alone would certainly have been inappropriate in such a study.

In addition to his research on sentence-level measures, Yule also developed an approach to quantitative vocabulary analysis⁴⁶ in which he compares ratios of words used by one author to those used by others. More specifically, he concentrated on the common noun and its distribution. For all nouns of a given text he computed frequencies of occurrence and constructed a measure of their incidence that he called the "characteristic." This "K" value is expressed as an integer and represents the repetitiveness of the vocabulary of the work being tested. Subjectivity and precise mathematical functions typify both of Yule's studies, which are now acclaimed as precursors of current computer aided literary research.

It was not until 1957, when Paul E. Bennett tested two of Shakespeare's plays for homogeneity of authorship, that Yule's method was again put to use. Bennett felt that fear of the "statistical theory behind Yule's measure" accounted for the literary scholars' hesitation to utilize his approach, but immediately added that "the measure itself is quite simple to use; it can be used in much the same fashion that one uses a calculating machine, without bothering one's head about the mechanics or the theory of the machine."⁴⁷ Bennett then briefly described his method, which differed only in part from that of his guide. Whereas Yule had used a method of sampling, Bennett used the entire vocabulary of the two plays he tested. He hand-counted every common noun and checked his results against a well-established concordance. Finally, he presented in tabular form the results of his research and concluded that "in regard to the aspect of style which we have measured objectively in the two plays, Shakespeare is very similar to himself."⁴⁸ Bennett was very careful to note that Yule's K characteristic measures

only one aspect of style and that "The real desideratum is to develop objective measures of several different significant aspects of style; authorship might then confidently be ascribed when two or three or four of these measures were in substantial agreement."⁴⁹

Additional research in the area of vocabulary analysis was conducted by John N. Pappas. The Institute for Computer Research in the Humanities (ICRH) Newsletter of December, 1966, introduced Pappas's study of a disputed eighteenth-century French text attributed to both Mlle de Lespinasse and Mme Suard.⁵⁰ The procedures described by Pappas constitute a twofold method of stylistic determination and give rise to one of the very few computerized approaches to qualitative stylistics. He first dealt with sentences, paragraphs, and punctuation. Then he measured "the mean and standard deviation of each type of punctuation per sentence and per paragraph."⁵¹ Pappas next turned to the more standard "frequency count" of "the central word list, eliminating the 10% on top of the most frequently used words and those which occur so infrequently as to make their measure invalid."⁵² While preparing his data, he encountered a problem not foreign to all early computer literary scholars: the treatment of accents. In order to distinguish between à and a, où and ou, for example, he found it necessary to use a "key-word in context" concordance program to print, in context, all words of possible double meaning. Key punching special characters into a given text to represent all diacritical markings is at best a very tedious, time-consuming and, in certain instances, unnecessary task. Assuming, for example that we use the asterisk (*), the slash (/), the percent sign (%), the dollar sign (\$) and the plus (+) sign to represent diacritical marks, an encoding of the line "... on sçait combien peu

vous êtes scrupuleux à détronner les Rois, aiant même trouvé les moiens après cela ..." might look like this for proofreading:

ON SC\$AIT COMBIEN PEU VOUS E%TES SCRUPULEUX A* DE/TRO%NER

LES ROIS, Aiant ME%ME TROUVE/ LES MOINS APRE*S CELA ...

The presence of the special characters, rather than aiding the reader, deters him. Only the "*" and, perhaps the "/" in trouvé would be necessary to establish meaning in a lexicographical index of this sentence. In consideration of the fact that most words whose meaning changes with the presence of a given diacritical mark are function words and will therefore have a high frequency rate, it would seem advantageous to encode only these relatively few, but often occurring, words. Had Pappas chosen to prepare his data in this manner, he would have accomplished two things: by eliminating the need to "see certain words in context" he would have reduced his computer costs and data processing time,⁵³ and he would have limited the number of words requiring a key-word in context listing to content words with dual meanings (e.g., être) and verb forms of the first conjugation whose non-accented past participle resembles its present tense forms (e.g., trouvé, trouve). The first 15,000 words of l'Avis aux réfugiés were keypunched with markings similar to those demonstrated above since they were completed before the appearance of Pappas's article. The remaining 185,000 words of text were, however, keypunched in a more simplified, yet effective manner: distinguishing characters accompanied only those words having a potentially high frequency and whose substance is changed by presence or absence of a diacritical mark. A "pre-processing" computer program designed to eliminate the "unnecessary" characters and to expand qu', d', j', n', c', s', and m' by restoring their elided "e's", was written

and used to standardize the input data (s'il and s'ils, however, remained s').⁵⁴ Hence the number of words requiring a key-word in context listing was greatly reduced, although not entirely eliminated.⁵⁵ To facilitate the recognition of words which have evolved from a common root, to count their frequencies, and to list them lexically with their roots, the French ENROOT program was developed⁵⁶ and used in conjunction with a "printed indexing program."⁵⁷ The resulting output data are lists, complete with absolute frequencies, of all occurring forms of a given root. From this output, vocabulary became evident for which a key-word in context listing was required.

A very similar program, called VIA (Verbally Indexed Associations), subsidized by the Office of Naval Research through the System Development Corporation of Los Angeles, was developed over a three-year period by Mrs. Sally Y. Sedelow. In her comments prefacing VIA's description, Mrs. Sedelow defined style as "the patterns formed in the linguistic encoding of information," and her working definition of stylistic analysis as "the perception of these patterns in language."⁵⁸ She then stated the premise upon which her research project was based, namely, "that the choice of information bearing words, as well as all the other patterns which they help form as they are embedded in sentences, paragraphs, speeches, chapters . . . is the necessary province of stylistic analysis."⁵⁹

The procedures employed in VIA evolved from the hypothesis that human perception of important ideas or themes is most often a function of sufficient repetition of words or word patterns to make an impression on the reader. Thus, hoping to pick out important ideas or concepts within a given body of text, Mrs. Sedelow programmed the computer to

print out lexically the vocabulary of a given text with each individual word's accompanying location, according to chapter, paragraph, sentence, and position in sentence. These words were listed by "root-group," with a figure denoting the total number of words included in the group printed next to the last word of the set. Using this machine-generated list, Mrs. Sedelow had to study the vocabulary and identify the terms she wished to examine further as "primary words."⁶⁰ The next step entailed a manual search through synonym dictionaries and thesauri and compilation of an associated-word list. This list, in turn, had to be checked against the original text to see if the new terms occurred in it. Finally, having been fed this "thesaurus," the computer instituted "a rather complicated search for these possible associated-words, searching first for each word associated with the given primary word, next for words linked to such of those associated words as have primary status in their own right."⁶¹ Thus a multi-level investigation occurs, which leads to the isolation of themes within a text.

A second, even less conventional program developed by Mrs. Sedelow is called MAPTEXT⁶² and may be used in connection with VIA (using VIA's output as input) or completely independently of it. Whereas VIA deals primarily with vocabulary and themes developed through word usage, MAPTEXT seeks to reveal patterns of word usage which were hitherto covert. It allows one to visualize a text free from semantic consideration. For example, to study the verb-adverb distribution in a text, each verb and adverb would be assigned a number as the original data deck is prepared for input.⁶³ The resulting output from the computer would furnish a symbolic picture of the text. Numbers representing encoded words would be printed, whereas the uncoded words in the text

would be represented by zeros or dashes. Thus, if a researcher wanted to ascertain the density of occurrence of a particular part of speech, he would only have to read the symbolic output.

Works discussed in this chapter represent but a few of the many reports published on computer-aided literary studies, for, just as the utilization of the computer has known an exponential growth in general, so, also, has its application to the fields of linguistics, stylistics, and literary content-analysis.⁶⁴ Research reported here gave direction to the present project by suggesting examination of such putative indices of an author's style as word and sentence length, syllabification, sentence beginnings and endings, content and function words, fixed and variable word combinations, as well as numerous part of speech and grammatical elements.

Discovering and defining possible objective correlates of style represent a considerable part of the research involved in this project, but two equally important areas remain to be reported: the delineation of procedures, e.g., how these variables were located mechanically, and the results obtained when 843 variables were compared across the test articles. Reporting of these two areas occupies the next two chapters.

NOTES FOR CHAPTER II

¹Wachal, "Linguistic Evidence," p. 5.

²Ibid.

³Ibid.

⁴G. Udny Yule, "On Sentence Length," Biometrika, XXX, 1936, 363-390. Even though Yule was not able to utilize the computer in his study, I have included his article in this "survey" because his method paralleled that of computational stylistics.

⁵"So far as punctuation is concerned, there seems very little evidence that many authors exercised any care about it whatever. After all, even at present, few authors trouble to punctuate their MSS. with any care or consistency. Such punctuation as is found in ordinary MSS. of the sixteenth and seventeenth centuries is indeed most erratic and seldom goes beyond full stops at the end of most of the sentences and some indication of the caesura in verse." An Introduction to Bibliography (Oxford: Clarendon Press, 1927), p. 250.

⁶Yule, "On Sentence Length," p. 365.

⁷Admitting to the subjectivity involved in so editing a text, Yule wrote: ". . . at first I by no means realized the full extent of this difficulty, and when I did often felt myself horribly incompetent to deal with it. I am sure my final decisions could often be contested, and were not infrequently inconsistent with one another." Still he considers his procedures valid "if only as an exploratory piece of work." "On Sentence Length," p. 365.

⁸For example, Le Petit Robert gives the modern definition of "phrase" as "Tout assemblage d'éléments linguistiques capable de représenter pour l'auditeur l'énoncé complet d'une idée conçue par le sujet parlant," whereas its "old" meaning is, "Tour ou construction." Further research revealed that in classical French, "phrase" is defined as an "expression, une façon de parler, locution, tournure." See Oscar Bloch and W. Von Wartburg, Dictionnaire Etymologique de la langue française, 5^e ed. (Paris: Presse univ. de France, 1968), p. 482, and Gaston Cayrou, Le Français classique (Paris: Didier, 1948), p. 659. In the classical period definitions given, "a complete expression of a single thought" is not stipulated. The lack of such regimentation would seemingly allow the compositors greater freedom to edit. However, it must be remembered that the fact that a printer or compositor breathed the same social and political atmosphere as the author for whom he set type, might well have provided him with more structural, as well as ideological insight into a given text than can a reader examining the work three to four hundred years later rightly expect to possess.

⁹Yule, "On Sentence Length," p. 365.

¹⁰Ibid., p. 367.

¹¹Wachal, "Linguistic Evidence," p. 195.

¹²Ibid.

¹³Milic points out that Swift's redundancy "derives from the urge to control meaning." A Quantitative Analysis of the Style of Jonathan Swift (The Hague: Mouton and Co., 1967), p. 121.

¹⁴Milic, "Unconscious Ordering in the Prose of Swift," The Computer and Literary Style, ed. Jacob Lezd (Kent, Ohio: Kent State Univ. Press, 1966), p. 82.

¹⁵Frederick Mosteller and David Wallace, Inference and Disputed Authorship: 'The Federalist' (Reading, Mass: Addison-Wesley, 1964).

¹⁶Ephim G. Fogel, "The Humanist and the Computer: Vision and Actuality," Proceedings of the IBM Literary Data Processing Conference, IBM, Yorktown Heights, New York, 1964, p. 17, quotes from a 1961 article by Mosteller. See also Mosteller and Wallace, Disputed Authorship, p. 7.

¹⁷Ibid., p. 1.

¹⁸Mosteller and Wallace, "Inference in an Authorship Problem: A Comparative Study of Discrimination Methods Applied to the Authorship of The Federalist Papers," paper read at the statistical meetings in Minneapolis, Minnesota, September 9, 1962, p. 1.

¹⁹Mosteller and Wallace, Disputed Authorship, p. 39.

²⁰G. A. Miller, E. B. Newman and E. A. Friedman, "Length-frequency statistics of written English," Information and Control, 1 (1958), 370-389.

²¹Mosteller and Wallace, Disputed Authorship, p. 39. To my knowledge, such a list does not exist for the French language. It was, therefore, necessary to cull, from grammars (e.g., Grevisse), studies on style (e.g., Sayce, Marouzeau), and syntax (e.g., Haase) a file of French function words according to the criteria presented by Mosteller and Wallace. See further explanation of function word selection, infra., pp. 56-57.

²²Ibid., pp. 11-13, 39-42.

²³Ibid., p. 67.

²⁴Mosteller and Wallace, Disputed Authorship, p. 265. These thoughts, from a statistician, parallel those of scholars of the New Criticism using the formalistic approach to literary analysis. See also

the very interesting study, "litterae ex machina," Proceedings of the IBM Literary Data Processing Conference (Yorktown Heights, New York, Sept. 1964), pp. 37-54, in which Alan Markman relates computer techniques to critical theories.

²⁵Stephen M. Parrish, "Computers and the Muse of Literature," Proceedings of the Conference on the Use of Computers in Humanistic Research, Rutgers, Dec. 1964, pp. 14-19, and in Edmund A. Bowles, Computers in Humanistic Research (Englewood Cliffs, New Jersey: Prentice-Hall), pp. 124-134 (revised).

²⁶Charles Muller, Essai de statistique lexicale: "L'illusion comique" de Pierre Corneille (Paris: Klincksieck, 1964), and Etude de statistique lexicale: Le vocabulaire du théâtre de Corneille (Paris: Larousse, 1967).

²⁷Milic, "Winged Words," A Quantitative Analysis, and "Unconscious Ordering."

²⁸One of the more sceptical critics of computational stylistics is Stephen Ullman who calls the statistical method "too crude to catch some of the subtle nuances of style: emotive overtones, evocative resonance, complex and delicate rhythmic effects and the like." He does, however, admit to three important "ancillary" uses: ". . . to establish the authorship of anonymous works," to obtain a "rough indication of the frequency of a particular device, its 'density' in a given work," and in some cases to "reveal a striking anomaly in the 'distribution' of stylistic elements which may thus raise important problems of aesthetic interpretation." Language and Style (Oxford: Basil Blackwell, 1966), pp. 118-121.

²⁹Milic, "Winged Words," p. 77.

³⁰Ibid.

³¹Ibid., p. 83.

³²Milic, A Quantative Analysis, p. 268.

³³The Times Literary Supplement reviewer of Ellegård's two works, Who Was Junius? (Stockholm, 1962), and A Statistical Method for Determining Authorship, The Junius Letters, 1769-1792 (Göteborg, 1962) called the previously accrued linguistic evidence "very unconvincing." January 25, 1963, p. 67.

³⁴Arthur Sherbo and George Zimmer also cautioned against using small samples. In a pilot study they found that one-thousand word samples were inadequate for testing word pattern repetition. This problem has been eliminated in this study by the use of complete representative works in lieu of text samples. See Zimmer, "The Attribution of Authorship: A Computerized Method Evaluated and Compared with Other Methods Past and Future," Diss. Michigan State University, 1968, pp. 29, 32.

³⁵Ellegård, A Statistical Method, p. 9.

³⁶The "mots-clefs" approach was also taken by Pierre Guiraud and reported in his Caractères statistiques du vocabulaire (Paris: Presse univ. de France, 1954), and in his Problèmes et méthodes de la statistique linguistique (Dordrecht: Reidel, 1959), pp. 84-96. M. Guiraud's research, however, pursues a more technical statistical approach than Ellegård's. He examines and criticizes the formulae introduced into literary studies for vocabulary analysis by statisticians and linguists such as Yule, Herdan, and Zipf, and then develops his own formula. One of his principal concerns was the establishment of the "norm" (derived from the statistical "normal curve"), which he established for the end of the nineteenth century based on 1,200,000 words from prose writers of that period.

³⁷For example, the word "uniform" had a relative frequency in Junius' works of 280 per million words. In the control group sample of one million words its relative frequency is 65. Thus the distinctiveness ratio is 280/65 or 4.3, which is "clearly a Junian plus word" which he used about "four times as often as his contemporaries." Ellegård, A Statistical Method, p. 15.

³⁸Professor Ellegård used a control group of "one million words, drawn from about a hundred authors" to establish his plus and minus word list. Ibid., p. 21.

³⁹Times, January 25, 1963, p. 67.

⁴⁰Zimmer, "Attribution of Authorship," p. 28, and his review of Who Was Junius? and A Statistical Method, in Journal of English and Germanic Philology, June, 1963, pp. 688-689.

⁴¹Ibid., p. 688.

⁴²Times, January 25, 1963, p. 67.

⁴³This figure represents the population of potential "Junii." For "sound" attribution in Ellegård's study the number must be smaller than 300.

⁴⁴Times, January 25, 1963, p. 67. The fundamental criticisms of Ellegård's research stem from the inconsistency of his own subjective impressions and manual counts; not from his approach to the problem of attribution.

⁴⁵Joseph Raben, "A Computer Aided Study of Literary Influence: Milton to Shelley," IBM Proceedings, 1964.

⁴⁶The Statistical Study of Literary Vocabulary (Cambridge, 1944).

⁴⁷Paul E. Bennett, "The Statistical Measurement of A Stylistic Trait in Julius Caesar and As You Like It," Shakespeare Quarterly, 8 (1957), 33-34.

⁴⁸Ibid., p. 44.

⁴⁹Bennett, p. 45.

⁵⁰John N. Pappas, "Authentication of an Eighteenth-Century Text," Institute for Computer Research in the Humanities Newsletter, II, Nos. 4 and 5 (Dec. 1966, and Jan. 1967), 3-4, 3-4.

⁵¹Ibid., Dec. 1966, p. 4.

⁵²Ibid.

⁵³Since the time Pappas developed his programs, and, for that matter, since the time I prepared my textual data, typewriter terminals which include upper and lower case letters, punctuation marks (the IBM 026 and 029 keypunches I used required special multiple punches for all punctuation marks except the comma and the period), and diacritical marks have been developed. As Robert Wachal has said, "One would hope that the days of the keypunch are numbered, at least for humanists." "Getting at Style through Statistics," rev. of Statistics and Style, eds. Lubomír Doležel and Richard W. Bailey (New York: American Elsevier, 1969), in CHum, IV (May 1970), 27.

⁵⁴Several other tests were performed by the pre-processing program, most of which were peculiar to the late seventeenth-century texts with which we were dealing. To ensure uniformity in the input data, all 200,000 words were run through this program.

⁵⁵For example, à occurred 1024 times and a 415 times in l'Avis aux réfugiés. Assuming one occurrence of either per line of text, 1439 lines or 24 pages of output would be required for this word alone. Nevertheless, while preparing the input data, for example, I failed to fully realize that encoding past participles according to use--e.g., verbal, adjectival, or nominal--would facilitate comparisons of such stylistic elements as verb-tense, noun to adjective, or adjective to adverb ratios.

⁵⁶See infra., pp. 62-65, for a more detailed description of this program. A grant-in-aid of \$1500 was provided by the Computer Laboratory of Michigan State University for the development of this program. Mary Rafter, a professional programmer at Michigan State University, wrote the preliminary version based on a much less sophisticated, less complex English version developed by John Hafterson of the Learning Systems Institute at M.S.U. See Basic Information and Retrieval System Technical Manual, BIRS 2.0 (Michigan State University, 1968), pp. 1211-1212.

⁵⁷See infra., pp. 61-62. John Morris reported that he found using PIP in conjunction with one of our early versions of ENROOT "very fruitful" in his article, "A Computer-Assisted Study of a Philosophical Text," CHum, III (Jan. 1969), 175-176.

⁵⁸Sally Y. Sedelow, "Stylistic Analysis: Report on the Second Year of Research," System Development Corporation Document, TM-1908/200/000, March 1, 1966, p. 7.

⁵⁹Ibid., p. 8. The technique Mrs. Sedelow used in composing VIA resembles one already in use for some time in the social sciences known as Content Analysis. The General Inquirer system developed at MIT, commencing in 1961, is a well-known example of this technique. Another is the Basic Index and Retrieval System (BIRS) designed to search and retrieve research documents. BIRS served as a point of departure, and its principal programmer, John Hafterson, as guide as we explored the possibilities of adapting parts of the system to the French language and to the authorship attribution of the present study. Even though Mrs. Sedelow's program had some features, e.g., line location of desired vocabulary printed on preliminary output, that were not built into BIRS, we felt that the immediate availability of BIRS, the presence of its writer, and the fact that a program written for another machine would need some revising and appreciably delay the project, made more judicious the use of BIRS.

⁶⁰Defined by Mrs. Sedelow as "words occurring with high frequency relative to the rest of the text." Ibid., p. 15. She concludes her final report under the three-year research grant with a hopeful statement that VIA will one day be completely automated, thereby eliminating the tedious manual establishment of thesauri. "Stylistic Analysis: Report on the Third Year of Research," System Development Corporation Document, TM-1908/300/00, March 1, 1967, p. 91.

⁶¹Sedelow, "Second Year Report," p. 16.

⁶²Ibid.

⁶³It is conceivable that adverbs could be further divided as to time, place, or manner and be encoded accordingly.

⁶⁴Its use in many other areas of humanistic research has also grown.

CHAPTER III

PROCEDURES AND PROGRAM DESCRIPTIONS

Because the data to be used in this attribution study of l'Avis aux réfugiés ultimately consist of frequency figures, making as much use as possible of modern technology and the science of statistics is an eminently sound procedure. Indeed, it is only by making intelligent use of the computer and statistical techniques that two goals of this project--validity and reliability--can be attained. In order to be certain that a different researcher, repeating the same procedures, will obtain identical results, each step of the experiment must be explicitly defined. Wachal points out that "One of the best means of guaranteeing explicitness is to write a set of computer programs which will perform this procedure. To the extent that the computer programs guide the work, the results are replicable, inasmuch as the machine cannot act on any instructions that are not entirely explicit."¹ The literary scholar, then, who chooses to use the computer and statistics as tools for research, has no easy task. He must define literary terms--lexical, syntactical, phonemic, or grammatical--into a mathematical language which can be read and acted upon electronically. Once the selected elements have been identified and their incidence recorded, statistical techniques can be utilized for analysis or inference. Since one of the goals of this project was the development of a procedure for constructing a practicable, sensitive test for authorship, the STAR (1620 authorship report) system, a series of computer programs, was developed.

The STAR system evolved with the co-operation of the Indiana-Purdue Universities at Fort Wayne for use on their IBM 1620 computer.

Although a small and relatively unsophisticated machine, the 1620 is especially suited to natural language analysis for several reasons. First, the structure of the 1620 lends itself to the handling of variable length character strings (i.e., words of varying lengths), characteristic of literary texts. Second, the simplicity of the "assembly language" is advantageous to the process of searching through lists of words. STAR, dealing primarily with vocabulary, syntactical elements and grammatical terms, involves a great deal of this type of processing. Third, STAR must read and print large amount of data (this study put to use a data base of over 1.7 million characters). The instructions to read, print and punch for the 1620 are extremely simple. Whereas only one instruction is required to read a card on the 1620, as many as one hundred-fifty instructions may be necessary on a more sophisticated system such as the IBM 360 or CDC 6500.

Thus the STAR system was written largely in assembly language. SPS (Symbolic Programming System) was chosen as the primary programming language because it provides the fastest possible character manipulation on the 1620. Its major inadequacy lies in its limited ability to be converted to a larger or more advanced system, e.g., it is totally incompatible to a fixed word-length machine. The use of FORTRAN would eliminate this difficulty; however, since FORTRAN is not so efficient as SPS, processing time would be increased at least ten-fold.

The data used by STAR closely resemble the natural language text found in any book, and may be prepared with a minimum of clerical work. Another advantage of the system is its general applicability. It is adaptable to almost any type of material, and is even independent of the natural language of the material used (with one minor exception which is

explained in detail below, page 49, footnote 13). The programs were originally written to study French texts, but any alphabetic language would work as well.²

Basically, the results produced by STAR fall into two categories. First, numerous scores are derived from sentence and word lengths. The second lot consists of a group of scores dependent on the frequency of certain words or phrases. The results are not complicated statistical scores, but simple totals and averages. This output may be subjected to any required degree of statistical analysis.

Section 1 of the program description gives a brief summary of how the input data is prepared. Section 2 provides a description of the five independent programs and sub-routines used in this study and an analysis of how the different scores developed by each program are obtained. Finally, section 3 describes those editing and testing programs which were used to check the accuracy of the other parts of the system. This section also treats programs used to proofread the original data.

Data Description

The format chosen for preparation of the basic text provides one of the basic differences between quantitative authorship studies. In a very broad sense, the many different types of data may be conveniently reduced to two broad groups. The first consists of a pre-processed or encoded data similar to that used by Louis T. Milic in his comprehensive study on the style of Jonathan Swift.³ Before punching any data on cards, Milic "translated" each word of the text to be examined into a two-digit code representing the grammatical part of speech of the

encoded word based upon its function in the sentence. These codes were then punched into the data cards.⁴

The second group is comprised of data punched directly onto cards from the original text. The data prepared for the current study fall into this category. A comparison of this method with that of encoded data readily illustrates that the data processing procedures will be much slower because of the large number of cards to be read. For example, a card of encoded data may contain up to thirty-six words per card; whereas the second type squeezes ten words onto a card.

It is to be noted, however, that data of the second class have one very important advantage over those of the first class: they may be used for any number of different types of studies. The data prepared for this study may be used for almost any type of processing, and, as the reader will note, a wide variety of tests have been run using them. On the other hand, encoding data restricts analysis to the type of processing envisioned when "translated."⁵

In addition to alphabetical rather than digital format of the data, still another important distinction separates the data base used here from the majority of other authorship attribution studies. Most researchers have chosen to take samplings from the authors' works being studied. Alvar Ellegård, for example, draws samplings from approximately 100 authors to establish a basis for his attribution study.⁶ In the present project, complete works, rather than random samples, were chosen as test articles. To be sure, this involved the accumulation and processing of significantly more data, but has produced (and we anticipate it will yet produce) more accurate results (i.e., results less subject to sampling error)⁷ than would have been attained

had sampling of the test articles been employed.

In order to keep a constant check on the accuracy of the data, special identifying information was punched in the first five columns of each card. Columns six through 80 contain the text. Column one contains a code which signifies which author wrote the material.⁸

X--l'Avis aux réfugiés (referred to hereafter as the "unknown")

L and Q--Daniel de Larroque

B--Pierre Bayle

P--Paul Péliisson-Fontanier

D--René Descartes

Columns two through five contain a sequence number which serves to keep the data in the proper order.

The limited character set of the IBM 1620 computer necessitated making some minor modifications of the text. During the keypunching phase, all italics and underlining were ignored. Latin quotations which were not an integral part of the text were omitted. Because many of the accent marks used in French function to change sound rather than basic meaning, because the performing of a content analysis would necessitate the presence of more than a single word, and because omitting the "unnecessary" accents would extensively facilitate the keypunching operation, all accents were omitted except those whose presence changed the basic meaning of a given word.⁹ In addition the use of the apostrophe was ignored except for a few special cases such as l' (which could be either le or la), and l'on. The combination l'auteur, for example, was expanded to l' auteur, allowing the space between the apostrophe and the noun to become a word delimiter. In the case of l'on, it was felt that this particular use of the definite article might be a valid

discriminant, therefore the union remained. Where the apostrophe was clearly the elision of e or i, these letters were restored, e.g., qu' became que; m', t', s', n', were expanded to me, te, se, ne; and s'il was changed to si il, etc. Finally, any parentheses appearing in the text were replaced by commas.

In order to simplify processing, and because the semi-colon, colon, exclamation mark and question mark do not appear on the IBM 1620, punctuation was limited to the period and the comma. Thus the question mark, the colon and the exclamation mark were all punched as periods, while the dash, the semi-colon and all parentheses became commas. These changes involved only minor revisions to the text and greatly simplified both the keypunching and program development.

Finally, to insure accuracy of the text, a few small programs were written to check the exactness of the text with regards to the special characters. These programs are described in section 3. In addition to the aforementioned accuracy checks, a major pre-processing program, written for the CDC 3600, also verified the basic text.¹⁰

Section 2 - Program Descriptions

SENWOL

Sentence and word length as quantitative stylistic indicators have challenged several attribution study scholars.¹¹ Some have claimed success, others have admitted defeat. Proceeding under the assumption that even though sentence-level measures standing alone may not provide sufficient grounds for a conclusive argument, I felt they might at least combine with other "indicators" of internal evidence. Furthermore, discussions with statisticians at Michigan State University and at Indiana

University at Fort Wayne assured me that the methods of past sentence length research and not the theory of its applicability had been questioned. Therefore, feeling quite confident that sentence and word length could be valid discriminants, I proceeded to have SENWOL written. The following items represent the type of data generated by SENWOL:

1. The average number of words in a sentence, where a sentence is defined as a group of words between periods.¹²
2. The average number of letters in a sentence.
3. The average number of letters in a word.
4. The total number of letters, total number of words, and total number of sentences in all the data examined.
5. A provision is also made for the exclusion of words of a given length, e.g., one or two letter words. See SENWOL results, p. 88.

The program obtains these data by analyzing one sentence at a time. That is, the program reads in one complete sentence, counts all the letters and words in that sentence, prints the information for that sentence, and then goes on to the next sentence. When all the data have been read in, the program computes the averages and prints the final totals and averages. Appendix E demonstrates a sample of SENWOL output. Appendix D contains a simplified flowchart of the program.

SYLAN

The syllabication analysis program (SYLAN) and SENWOL are similar in that they produce analagous output. SENWOL counts and stores individual characters: SYLAN works on combinations of alphabetical characters. In order to compute average syllables per sentence and average syllables per word, a sub-routine called BREAK was inserted into

the basic SENWOL program. BREAK takes five basic rules of syllabication and works on the text one rule at a time.¹³ Like SENWOL, it analyzes one sentence at a time.

In the preparation of the data base the article l' was separated from its noun or pronoun (except in the case of l'on) in order to be counted as a separate word. However, l' contains no vowel and thus does not fall within the bounds of the basic rules. It was therefore treated as a special case; as a word having one syllable. The identifying of l' is the first step in a series of word editing performed by SYLAN. The remaining editing procedures serve to eliminate other words which cannot be analyzed by these rules. Specifically, these words are those containing special characters and numbers. For example, A* represents à; LA*, là; OU*, où; DE*S, dès; etc. Numbers which were punched as digits and not in alphabetic representation cannot be analyzed by these rules and were thus returned as special words having no syllables. To be certain that the number of unanalyzable words was insignificant, a tally was made of all such words which showed one such term to be found in every 1500 words. All words which had been edited were then broken down into syllables according to the five rules as referenced above. The first letter of each syllable was specially marked in computer memory with a procedure analogous to underlining that letter. After the word had been completely broken down, the program counted the total number of syllables in the word, added this number to previous totals, and then punched out all the required information as described above pertaining to words, sentences, and the entire article. Again, the output procedure is very similar to that utilized by SENWOL.

An additional check was made to insure that the program properly handled all cases of syllabification. From the initial output from SYLAN, certain exceptions to the five rules were noted and insertions (additional instructions) were incorporated in the program. After the final results had been obtained and checked, the error was less than 1%, which was considered negligible in terms of the large sample size. Appendix G contains a simplified flowchart of the program.

STYLBEND

The ways an author chooses to begin and end a sentence is considered to be a characteristic of his style.¹⁴ Therefore, STYLBEND (Stylistics Beginnings and Endings) was written. This program searches for, prints, and punches the first and last words of each sentence of a text along with its corresponding sequence number which ties it back to the original data base.¹⁵ Because the computer was not programmed to recognize parts of speech, it was necessary to take the data which were thus punched onto machine cards and hand sort them. Fifteen basic parts of speech were chosen as possible beginnings.¹⁶ In order to avoid as much ambiguity and chance for human error as possible, the same parts of speech were used as endings. Once these categories had been established it was necessary to hand sort each of the punched cards into its proper category. In many cases it became necessary to refer back to the basic text in order to see the word in its context and properly classify it.¹⁷ The code/sequence numbers as described above proved invaluable at this point. As soon as the sorting process was completed, a simple counting program was written, which counted the total use of each part of speech. The results depicted each of the selected parts of speech chosen by a

given author to introduce or to terminate the sentences he had written. The accuracy of this program was determined by constant referral to the basic text required by words which could be classed as various parts of speech. In addition a second STYLBEND program was written, the first having been written for the CDC 3600. The second, written for the IBM 1620, produced a slightly different output. Through constant proof-reading of the basic text, minor errors were noted and corrected. The 1620 version of STYLBEND served as a final check to insure accuracy as well as precision.¹⁸ The output from the CDC 3600 computer printed out the total number of occurrences of each word as well as each line on which the word occurred, and separate cards were punched for beginnings and endings. The 1620 output produced two words per card, the first and the last of each sentence plus the location of the sentence beginning. The final tabulations of STYLBEND represent the results obtained from the most current, corrected copy of the text. All discrepancies that had been noted through proofreading were corrected. Appendix H contains a simplified flowchart of the program.

EXSOR

One of the simplest of the quantitative analyses in stylo-statistics is the study of word usage. However, this is not enough. Many times, a study of the use of groups of words or expressions may yield more meaningful results. It is for such a study that this program has been developed.

EXSOR is a program designed to search a sentence-structured text to find given word patterns called expressions. The searching process may be completely automatic, or it may be combined with manual searching

by the user. Before the program itself is described, the input and output specifications will be given in detail.

The input consists of two groups of data: the expressions and the text. They are inserted in that order with an End Of Job¹⁹ card at the end of each group. The text is assumed to be sentence-structured; that is, it must be divided into sentences with a period: the only allowable sentence delimiter. The comma may also be used as punctuation, but all other punctuation marks are treated as separate words. The space is considered a word delimiter along with the period and comma; any number of consecutive spaces are condensed to just one space as the text is read in. The text is punched on cards in columns six through seventy-five; the first five columns contain a sequence number and will be ignored by the computer.

The expressions are of two types. First, an expression may contain only literals. Consider the examples:

PLUT A* DIEU QUE

and

ILS NE NOUS L' AUROIENT JAMAIS FAIT

The first is a fixed expression without variant forms. The second is composed of fixed forms and accompanying grammatical variations which may or may not be selected for testing. EXSOR will find each case that these expressions were used in the text within a single sentence and print out each expression and its frequency. The second type of expression may contain one of the two special variables (WRD and WRDS) which may not be used anywhere in the text. Using these two symbols, we can define a variable expression. As an example, suppose we

wish to find all occurrences of the structure:

ne ... jamais

We are not interested in the words in between or how far apart the two literals are. We would replace the dots with special characters.

WRD causes the program to skip exactly one word of the text, and WRDS causes the continual skipping of words until a match is found for the next literal. In the example above we would input the expression:

NE (WRDS) JAMAIS

A match is first found for the word NE, and then each consecutive word is tested until a match is found for JAMAIS or until the end of the sentence is reached (in no case will an expression overlap the end of a sentence). WRDS is therefore a variable which can stand for any number of words. WRD has a somewhat similar use, but it is restricted to exactly one word. Thus, if we wished to find all cases in which the NE is separated by exactly one word from the JAMAIS, we would input the expression NE WRD JAMAIS. If we wished them separated by exactly two words, we would use the expression NE WRD WRD JAMAIS. Finally, we may combine the two different variables. The expression NE WRD WRD WRDS JAMAIS will find all cases with a separation of at least two words.

Two special illegal constructions are checked for as the expressions are read in:

1. The construction WRDS WRD is arbitrarily chosen to be invalid; it is exactly equal to WRD WRDS and considerable time is saved by not having to deal with both cases. When this condition is encountered, an error

message is typed, the correct expression is substituted, and the program continues.

2. An expression may not begin or end with WRDS or WRD. When this construction is encountered, the user must reload the expression list beginning with the first expression (after correcting or removing the incorrect expression).

The output of the program is straightforward. Each expression is printed with an identification number and a frequency score, both of which are clearly marked. The material may also be punched on cards by the computer if desired.

The program is broken up into two parts. The first of these is very fast and simple. It merely reads and stores all the expressions, using special markers to replace WRD and WRDS, and checking for the two errors mentioned above. Part two does the actual searching. First, it reads in a complete sentence. It then makes one pass through the sentence for each expression, until it has searched for all possible expressions. It then moves on to the next sentence; when it encounters an End of Job card which marks the end of the text, it prints out the results of the search.

By breaking the text up into parts, operating on only one sentence at a time, the search is simplified.

Using the sense switches²⁰ the user may select any or all of the following options.

If sense switch 1 is on, the complete text will be printed.

If sense switch 2 is on, the program will restart itself and automatically accept another new set of data.

If sense switch 3 is on, the memory position of each expression will be typed on the typewriter. This allows the user to see if the expression list is almost full, almost empty, or somewhere in between. Using sense switches 3 and 4, the user may combine manual searching with automatic searching. Certain expressions, due to their variability of use, may need to be sorted by hand. These should be marked with a period on the input card, and the computer will ignore their tabulation as it searches the text. It will, however, print the sequence numbers where the expressions appear in the text, so that manual sorting will be much easier. Sense switch 3 is used to make the computer ignore the expressions as it searches, and sense switch 4 is used if the sequence numbers are to be printed.

Due to the small size of available memory, certain limitations were imposed on the program regarding the amount of material it could handle. Exceeding any of the following limits will not necessarily cause the program to stop, but may give faulty results:

1. The maximum number of expressions may not exceed 800.
2. The total number of words in all expressions may not exceed 800.
3. All sentences must be less than 150 words.
4. All sentences must take less than 15 cards.
5. Average sentence length over the whole text must be greater than one card.

FREQFUN

Another basic type of quantitative stylistic data is the frequency of use of selected words. FREQFUN (Frequency of Function Words) is a program which produces the frequency of a given list of words.

The information thus obtained includes a listing of the words selected for study followed by their absolute and relative frequencies, and a total count of the vocabulary of the text. These figures may then be subjected to any number of statistical analyses.

Two basic factors influenced the development of the program: (1) speed and efficiency of the sorting and counting process, and (2) maximum capacity of the selected word list. For this study a total of 361 words comprised the selected word list. In an attempt to obtain objectivity in our choice of function words, use was made of FAP (File Analysis Program), a component of BIRS. FAP performs word-frequency analysis across (as opposed to merely within) paragraphs of a given file (literary work). It selects as descriptors of a paragraph those terms which make that paragraph a distinct element of the work. The program has three options or methods: method one maximizes on "content" terms, method two, on content and rare terms, and method three, on function words. The scores obtained from this program are a function of the average proportional use of a term and its variability of use in a given document, abstract of a document, or paragraph not exceeding 50 cards in length.²¹

L'Avis aux réfugiés, Cabale Chimérique, and Le Prosélyte abusé (roughly 153,000 words) were fed through FAP, method 3 with a "V" value cutoff established at 2.0. This procedure extracted from the 153,000 taken vocabulary a total of 184 "function" words. For this preliminary analysis, a synonym list was used which equated such terms as le, la, l', les (the synonym list was not, however, used with FREQFUN). The fact that the part of speech categories of these 184 terms, based upon the final FAP-3 tabulations, did not differ greatly from those related by

Mosteller and Wallace was encouraging. In addition to filler words classed as prepositions, conjunctions, pronouns, adjectives, adverbs, and auxiliary verbs, our FAP-3 output indicated that certain nouns (cas, cas-là, cause, considération(s), côté, dommage, face, force(s), moien, moral, personnes, propos, tas), negations (guère(s), non, ni, personne, nul(le), aucun(e), pas), and twenty-one present tense and present or past participial forms other than auxiliaries, were so dispersed throughout the text to qualify as function words. Whereas Mosteller and Wallace had the Miller-Newman-Friedman list of function words to draw upon, and because no such list exists for French function words the FAP afforded the best available method to supplement the statisticians' list. However, because of Mosteller and Wallace's success, after carefully omitting words whose rates could depend dangerously on context, many of these nouns and verbs were removed from the list before the final data were run. Only in the case of pronouns did we depart from their procedures. It was felt that because each of the works to be tested dealt basically with the same topics, pronouns might be valid function rather than content words. Their analysis has therefore been included in the results described in Chapter IV, pp. 135-162. Appendix K lists the function words tested in this project. The program provides locations for approximately two-thousand words averaging five characters in length.²² In order to provide a fast, efficient program, all selected words were read in and stored on the disk--the 1620's external memory. Then, they were ordered according to length so that words from the data base would be compared only to selected words of an equal length. The next step involved alphabetizing, within numeric groups, the now ordered selected word list. Finally, trailing blanks²³ are removed as the words

are stored in memory (this step increases the maximum number of words possible on the selected list by compacting the number of locations required.)

After the selected words have been prepared as described above, the text is read in and the program begins the searching process. A typical card of text is processed as follows: The program searches for the first complete word on the card and counts the number of letters it contains. This word is then compared to all words on the selected word list of an equal length to the one on the text card. After the word is found, or after all words of equal size have been checked, the program proceeds to the next word on the cards, to the next, and so on until each word has been tested against the selected word list.

After all cards of the text have been examined, the program prints out the final frequencies of all words in the format shown in Appendix L. Appendix J presents a simplified flowchart of the program.

Section 3 - Miscellaneous Programs

CANAL

This is the basic program used to convert the format of the data produced by EXSOR and FREQFUN to an arrangement acceptable by the analysis of variance program. The word scores obtained from FREQFUN and the expression scores obtained from EXSOR are grouped individually according to author and article.²⁴ That is to say, all word scores are recorded sequentially in the following format where the denominators P, X, B1, B2 and L, refer to the final test articles as described on page 75 and listed in Appendix A.²⁵

FREQFUN	P	X	B1	Bs	L1
A*	79	1365	1574	344	889
AFIN	1	26	29	4	26
AINSI	1	59	39	13	49
AILLEURS	1	18	15	4	11
ALLER	2	23	70	13	14
ALORS	1	16	9	3	0
APRES	6	56	45	24	52
ASSEZ	5	28	43	6	23
AUCUN	2	53	54	5	37
AUJOURD'HUI	2	17	10	11	17
EXSOR					
A* CAUSE DE	0	12	11	4	7
A* CAUSE QUE	0	8	13	4	0
AFIN DE	0	18	18	4	13
AFIN QUE	1	8	10	0	13
A* PLUS FORTE RAISON	-	3	1	1	1
AU PIS ALLER	0	2	4	2	0
A* WRDS EGARD	0	12	3	3	6
DE PART ET DE AUTRE	1	0	3	1	4

The analysis of variance program requires that all scores for a given word or expression be recorded and analyzed before subsequent word or expression scores may be considered. In order to rearrange the data received from EXSOR and FREQFUN so that the analysis of variance may be performed, all data are read onto the disk in their raw form. The frequencies for each word expression are then retrieved and punched onto cards in the revised format.

Editing

Because of the changes necessitated by the limited character set of the IBM 1620 (as noted above, p. 46), numerous programs were written to check the text for keypunch or copy errors. These programs print out the sentence sequence numbers which mark the location of sentences which have some questionable characteristic. Examples of questionable

characteristics are sentences of fewer than three words, sentences containing numbers, and sentences containing misspelled characters (percent sign, dollar sign, parentheses, etc.). The sentences thus pinpointed were then checked by hand against the original document.

Output Conversion

Numerous smaller programs were written to rearrange the data from two or more programs into a format easier to analyse. Because these programs do not affect any numerical values, but only change their positions on the machine cards, they will not be discussed in detail even though their use may be briefly mentioned elsewhere.

Each of the preceding sections describing the STAR system with the possible exception of Section I (Data description) has extracted selected portions from the data base. It is now time to examine a series of programs designed to work on the entire text as the entity. The programs to be discussed at this point are either modified versions of existing programs, obtained from the Learning Systems Institute at Michigan State University or CDC 3600 versions of STYLBEND and EXSOR. To give a detailed description of these two would be superfluous since their functions are identical to those of the 1620 version discussed above, pp. 51-55.

FAP

The File Analysis Program is a component of the BIRS (Basic Index and Retrieval System) which is designed to perform vocabulary analysis. However, it has one disadvantage for studies of this type:

its basic unit, or sample, is the abstract (paragraph). Previous studies²⁶ have shown that the average paragraph contains too few words to be an adequate sample. Therefore because of its size it is much too small to obtain reliable and consistent results. For this reason, FAP was not used as a principal source of vocabulary scores, but rather as a component program: as a framework for developing the PIP-ENROOT programs described below. Using FAP in this way saved some programming effort and will allow future users of the system the chance to obtain and use vocabulary "content" scores, as provided through FAP output, for vocabulary analysis studies. The exact structure of the FAP program will not be described in detail here;²⁷ however, it is very similar to that of the PIP program described below.

PIP

The Printed Index Program is also a component of the Basic Index and Retrieval System (BIRS) developed at Michigan State University. The PIP program is extremely complex, and a more detailed description of its various options can be found in the BIRS manual.²⁸ Its basic use in this study has been to prepare the data base for use with ENROOT, a subroutine designed to reduce a word to its root as described below. PIP reads in the text, gives the material to ENROOT to be processed, and does all output operations required after ENROOT is finished. The PIP program contains three basic parts.

First, the program reads various parameters which describe the input and output that will be required. We are using only a small part of the PIP program and must specify exactly the processing which we require. The input parameters include an exact description of the data

cards (see pp. 46-47). The output parameters call for a key-word in context index; this is a listing of each word in the vocabulary and its absolute frequency of occurrence as well as its root.

After these parameters have been read, the program processes the text by paragraph. It finds a word, gives the word to ENROOT to be processed, and writes on a scratch tape the output received from ENROOT for that word. After this has been done for each word in the paragraph, the program proceeds to the next group, and continues this process until the complete text has been read in. When all words have been processed, the scratch tape contains each word and its root. PIP next alphabetizes all the words and tallies the total number of occurrences of each individual word. Finally, a listing is printed which shows each root generated in the text along with the absolute frequency of each word returned to that root.

ENROOT

This is a subroutine whose purpose is to reduce words to a basic root form. As explained above, PIP reads in the data base and relays each word to ENROOT in the form of a one to sixteen character array and ENROOT returns the root of each word in the same form. For the purpose of uniformity in this study, we chose the infinitive as a basic root and all words etymologically derived from a given verb are returned to it. Only in cases where the noun or adjective form preceded the verb into the French language were these forms designated as roots. In instances where the orthography of a given term varied, the modern spelling was generally applied to such roots. The original subroutine, as developed by the Learning Systems Institute, was written to handle English texts and is much less sophisticated.

A grant-in-aid from the Computer Laboratory at M.S.U. provided funds for hiring a professional programmer, Miss Mary Rafter, to modify the English ENROOT. We spent several months researching the most efficient method of reducing the 55,551 word vocabulary of l'Avis aux réfugiés (minus its preface) to a final root vocabulary of 2689 words. The same process was then followed for the "known authors." The problems encountered were many. The orthography, both of l'Avis aux réfugiés and of the other texts to be analyzed, varied greatly. At first, it was thought that these differences might stem from regional habits, but further analysis revealed that works published in the same area contained the same variations of spelling. The English ENROOT provided for a synonym list designed to deal with such problems. It then became necessary to find a means of increasing the capacity of the synonym list. The English version allowed locations for 500 synonyms; the capacity of the French version soon rose to 1500²⁹ as we sought greater accuracy. Before long we found ourselves swamped with list upon list of exceptions but still short of the degree of accuracy we were seeking.³⁰ The adding of more lists and the eventual expansion of the synonym list to 2000 pairs of words slowed down the program considerably while increasing its accuracy to an acceptable 95-98%. At this point, the cost and time required for each additional degree of accuracy made continued automation impractical. The entire 166,000 words were run through ENROOT and after ten hours of manual re-ordering and keypunching, the desired 99% accuracy (allowing one percent for human error) was reached.

Most of the instructions in this subroutine involved the searching of various word-lists. For this reason a special subprogram called

EQSRCH (search for terms of equal nature) was written. The subprogram was written in COMPASS because it provides a more efficient searching mechanism than FORTRAN.

A compilation of synonyms containing roots (usually infinitive forms) and corresponding irregular spellings of the same word is essential to ENROOT. Suffixes and prefixes were established based upon lists in Robert and Grevisse³¹ and checked against a reverse alphabetical listing of the entire text vocabulary. Finally, lists were established which isolate or prevent further change or reduction by the basic program. A complete glossary of these special lists will be found in Appendix Y.

The analysis of the vocabulary proceeded one word at a time through four basic steps as follow:

1. MISCELLANEOUS PREPARATION. This section consists of performing preliminary checks on each word to process words that would be erroneously classified by passing through the remaining procedures. The program searches the special lists described in the glossary and returns the word as a root if it is found on any list. It also has special routines to remove (in some cases) a terminal e or terminal s. After this has been completed, ENROOT searches the synonym list for the word. If the word is found there, ENROOT obtains the proper root and stops processing that word.

2. SUFFIX TREATMENT. In this section the program looks for and deals with the suffixes contained in the suffix list. ENROOT compares the end of the word with each entry in the suffix lists, looking first for eight letter suffixes, then seven letter suffixes, then six, and so forth until a suffix is found or until the end of the suffix list

is reached. If a suffix is found, ENROOT removes the suffix and searches the synonym list for the word without its suffix. Whenever a word is found on the synonym list, its proper root is found and processing of the word is completed.

3. PREFIX TREATMENT. Prefix treatment is much the same as suffix treatment: ENROOT compares the beginning of the word with each entry in the prefix list, searching for the longest prefixes first. If a prefix is found, it is removed and the synonym list is searched again; if the word is found there, ENROOT obtains the proper root and stops further processing.

4. INFINITIVE FORMATION. Because the desired root form is generally the infinitive, the program concludes with modifications to the last few letters of the word. ENROOT again searches the special word lists, and either adds letters or drops letters from the end of the word in order to construct the proper verb form. When this is completed, ENROOT returns all roots to PIP for final output at a later time.

Appendix Y is a glossary of word lists used with ENROOT.

ANOVAR

(Analysis of Variance) All of the programs described thus far have dealt with obtaining absolute and relative frequencies of vocabulary, expressions, or sentence and word length. When comparing the frequencies obtained from each of the test articles some differences were obvious, others were slight. In order to objectively analyze these differences a straightforward, uncomplicated statistical approach was sought. After several hours of consultation with statisticians at Michigan State University and Indiana University at Fort Wayne, the

analysis of variance was decided upon because it allows wide divergence from the underlying assumption of normality, and because the extremely large data base used in this study would add even more robustness to the test. The analysis of variance would provide a test of significance of the obtained frequencies based upon three assumptions: normality of the error (within author variation), equality of variance of the errors, and statistical independence of the errors.³² These assumptions are "made in deriving statistical methods and are usually . . . apt to be violated in applications and are introduced only to ease the mathematics of the derivation . . . Statistical methods have been called 'robust' if the inferences are not seriously invalidated by the violation of such assumptions."³³

Henry Scheffé concludes that "Nonnormality has little effect on inferences about means . . ." and that "Inequality of variances in the cells of a layout has little effect on inferences about means if the cell numbers are equal, . . ."³⁴ However, because of unequal cell size in the present study, a test of the homogeneity of variance was performed as illustrated by Robert Steel.³⁵ The test results fell well within the prescribed limits. We therefore proceeded to perform the analysis of variance on the frequencies obtained from the programs described in this chapter using ANOVAR.

A variable is generated measuring the distance of each of the two "known" articles from the "unknown". These two variables are analyzed by ANOVAR. Significance in this analysis indicates that one of the "knowns" is statistically closer to the "unknown" than is the other.

We shall use sentence variables as an example of the form of the data analyzed by ANOVAR. Letting X stand for the average sentence length

of the unknown article, and letting B and L denote like values for the known authors, the variables, $|X-B|$,³⁶ and $|X-L|$ measure the distances of B and L from X with respect to sentence length. A test of the "difference between these two distance variables" is performed by a one-way analysis of variance. Significance of the analysis of variance test indicates that with regard to average sentence length one of the known authors more closely parallels this aspect of style in the "unknown" text.

ANOVAR computes the pooled within variance³⁷ which thus measures the within author variation. This within author variation is compared to the "between author variation" by means of the "F" ratio shown below.

$$F = \frac{\text{between variation}}{\text{within variation}}$$

This study employs two types of statistical tests: the F-test (associated with the analysis of variance), and the t-test. Since in this study, the analysis of variance was applied to data containing only two treatment groups, (X-B and X-L) the F-test and t-test were equivalent. As these two tests can be used interchangeably, programming convenience dictated which of the two tests would be employed in any given instance.

Once the F value is determined it is used to decide which values are most and least significant. Professor Steel³⁸ includes a statistical table in his text which indicates exactly how large the "F" value must be in order to conclude that the test is significant at a given level. In the present study, an F value of 4.17 or significance at the .05 level was deemed adequate.³⁹ ANOVAR uses a subroutine to look up the F value in a table and to determine the exact level of significance. Where .XXXX is the value which was obtained from a table, the fragment "RESULTS SIGNIFICANT AT THE .XXXX LEVEL," represents a typical output statement from ANOVAR.

NOTES FOR CHAPTER III

¹Wachal, "Linguistic Evidence," p. 10.

²See the annotated listing of programs used in this study, Appendix C.

³Louis T. Milic, Quantitative Approach, pp. 151-152.

⁴Encoded data has three special advantages. First, because complete words are not punched on cards, the data is very compact. (Milic punched all his data on approximately 2000 cards. This study has used over 25,000 cards). Second, the data can be punched in fixed format; that is, each card will have data in exactly the same place. Milic, for example, began a new code in every third column. Data in fixed format is extremely simple to process. Third, because the data is compact and easy to process, processing time is reduced.

⁵Milic's data, for example, could not be used to study word lengths.

⁶Alvar Ellegård, A Statistical Method, p. 21. See also Milic, Quantitative Approach, p. 71 and George Zimmer, "Attribution of Authorship," p. 23 for discussion of sample sizes.

⁷The larger sample size resulting from the analysis of entire texts rather than sample extracts from the texts increased the sensitivity of the statistical tests. The average length of the test articles in this study was 8505 words.

⁸See Appendix A for complete listing of works tested.

⁹Proofreading of the original text revealed that a distinction had to be made between à and a, là and la, où and ou, dû and du, dès and des, etc. Where words were noted to be etymologically the same, no attempt was made to distinguish between them. Only after the entire 196,000 words of text employed in this project had been keypunched and the data prepared for final analysis did I realize the extent to which the same words (e.g. entre nous, j'entre, il est entré, l'entrée, . . .) appeared as different parts of speech. This was particularly evident with "-er" verb forms. Failure to make this distinction at an early enough stage of the study limited somewhat the comparative analyses of verb terms originally envisioned. Had, for example, the past participles been coded so as to discern their usage, such comparative studies as imperfect vs. passé composé and/or literary tenses would have been a simple matter without the researcher having to depend on a key-word in context concordance. To be sure, words may often change meaning when used idiomatically or in locutions. This problem is handled through an expression finding program called EXSOR. See Section 3 of this chapter.

¹⁰See discussion of pre-processing program, Chapter II (pp. 32-33).

¹¹The most prominent studies reported are G. Udney Yule, "On Sentence Length as a Statistical Characteristic of Style in Prose," Biometrika, XXX (1939), 363-390; C. B. Williams, "A Note on the Statistical Analysis of Sentence-Length as a Criterion of Literary Style," Statistics and Style, ed. Doležel and Bailey (New York: Elsevier, 1969), pp. 69-75; Hemming Spand-Hanssen, "Sentence Length and Statistical Linguistics," Structures and Quanta (New York: Humanities Press, 1963), pp. 58-73; and Milic Quantitative Approach to the Style of Jonathan Swift, pp. 59-61. See List of References for additional entries.

¹²As previously noted (p. 47), periods and commas are the only punctuation marks included in the data base. The periods indicate complete stops or the introduction of enumerated clauses, except in ten instances where the number of words between such complete stops caused an overflow of memory in the computer. In these cases the text was edited by the insertion of a period at what appeared to be the most logical breaking point. The addition of these periods did not affect the final results as they were inserted only for the expression sorting program, EXSOR (see pp. 51-55).

¹³Maurice Grevisse, Le Bon usage, 7^e éd. (Paris, 1959), p. 57, lists three basic rules for dividing words into syllables as do most introductory phonetics manuals. Professor Clelland E. Jones conveniently sub-divides Grevisse's three rules in his Manual of French Pronunciation (Salt Lake City: Dessert News Press, 1961), p. 7. For BREAK'S purposes, we proceeded from the basic assumption that all syllables in French begin with a consonant and end with a vowel. To this basic rule, then, were added those provided by Professor Jones.

¹⁴Bernard O'Donnell, "Stephen Crane's The O'Ruddy: A Problem in Authorship Discrimination," The Computer and Literary Style, ed. Jacob Leed (Kent, Ohio: Kent State University Press, 1966), p. 111.

¹⁵For example, the word ce occurred 117 times as the first word in a sentence in l'Avis aux réfugiés. Of these 117 occurrences, 5 were adjectives, 21 were the indefinite relative pronouns, ce que and ce qui, and 91 were the indefinite, impersonal pronoun. See Appendix L for sample output.

¹⁶See Chapter IV (p. 95).

¹⁷In cases of possible dual classification, the function of the word in the sentence, how it was used, took precedence over its dictionary classification.

¹⁸The expression "precision without accuracy" is George Zimmer's. "Attribution of Authorship," p. 13.

¹⁹This card is a control card for the IBM 1620 computer only. For a description of this and other features of the IBM 1620, see the IBM Monitor I Reference Manual, p. 10.

²⁰Daniel N. Leeson and Donald L. Dimitry, Basic Programming Concepts and the IBM 1620 Computer (New York: Holt, Rinehart and Winston, Inc., 1962), p. 290.

²¹Mathematically this function is represented as

$$\text{Method 1: } V_j = s_j^2 / m_j$$

$$\text{Method 2: } V_j = \sqrt{s_j^2 / m_j}$$

$$\text{Method 3: } V_j = m_j / \sqrt{s_j^2}$$

where m_j denotes the average proportional use of term $j(t_j)$ in the abstracts (paragraphs or other designated units) of a given file and s_j^2 the variance of proportional use. V_j , then, represents the "value" of t_j as an indexing term in that file. See BIRS, Chapter 13.

²²The figure five represents the average word length of the 290,000 words of data base.

²³The maximum number of positions a word may occupy in the computer is sixteen. For words of less than sixteen alphabetic characters, the locations left unoccupied are called trailing blanks.

²⁴Article as used here refers to the individual works tested. See Appendix A.

²⁵It is to be noted that space limitations do not permit reproducing here all the word/expression scores thus recorded. The scores shown above are sample drawn from the original output.

²⁶Milic, Quantitative Approach, pp. 71-72; Ellegård, A Statistical Method, p. 9.

²⁷For a more detailed explanation of the nature, function and possible uses of the FAP program see BIRS, Chapter 13.

²⁸Ibid.

²⁹The final version of ENROOT has locations for 2000 synonyms and has been somewhat streamlined, but still deals with several lists of exceptions. A much more simplified version, based upon our past successes and failures is being written; too late, however, for use with this study. Appendix X is the final synonym list used in this project.

³⁰After September of 1969, the programmer with whom I had been working was no longer available and it became necessary to transfer all programs, data, results, etc. to Indiana-Purdue University at Ft. Wayne. The computer laboratory facilities were made available for my use, and the lead programmer/operator, Mr. Paul Gabriel, assumed the task of

adapting ENROOT (and the other programs which had previously been written for the CDC 3600) to the CDC 6500 computer at Lafayette, Indiana through an IBM 360 terminal.

³¹Paul Robert, ed., Petit Robert (Paris, 1967), pp. 1954-1956 and Grevisse, Le Bon Usage, pp. 78-96.

³²Henry Scheffé, The Analysis of Variance (New York: Wiley and Sons, 1959), p. 331.

³³Ibid., p. 360.

³⁴Ibid., p. 345.

³⁵Robert G. D. Steel and James H. Torrie, Principles and Procedures of Statistics (New York: McGraw-Hill, 1960), p. 82.

³⁶Vertical lines indicate "absolute value."

³⁷The variance of a sample/population is a statistical measure of the variation or range of the observations within the given sample/population. It is defined by the formula $v = (x_i - \bar{X})^2$. In general, a large variability will give a large value for the variance, v , and, conversely, a small value of v reflects a small variability in the data.

³⁸Steel, Principles and Procedures, (New York: McGraw-Hill, 1960), p. 440.

³⁹Significance at the .05 level means that 5% of all possible samples may lead to the erroneous rejection of a true hypothesis. Or, in the words of Professor Jerome Li, "it . . . is the probability that a Type I error rejection of a hypothesis that is true may be made on the basis of a single sample." Jerome C. R. Li, Introduction to Statistical Inference (Ann Arbor, Mich: Edwards Bros., 1959), p. 49.

CHAPTER IV

TEST RESULTS AND ANALYSIS

Past computational stylistic studies of authorship attribution have for the most part concentrated on one aspect of style. Yule first analyzed sentence lengths and later developed a statistical indicator for noun distribution in a given text. Ellegård expanded Yule's vocabulary study, establishing "plus" and "minus" words, while making little distinction between content and function terms. Statisticians Mosteller and Wallace concentrated on function words. O'Donnell studied basic syntactical patterns which might be taught in an introductory English writing course. Finally, Milic concentrated his study on structural grammar.

One purpose of this project has been to combine, for the first time, a series of variables and procedures, some of which had heretofore been individually tested, and to add to them other variants which may prove to be valid stylistic discriminants. To call this study a grammatical, vocabulary, or syntactical approach would be to misname it, for it encompasses all of these stylistic elements. The great amount of data to be analyzed made nearly complete automation the most feasible approach. As the project progressed, experience indicated additional areas in which increased automation would even further facilitate the final analysis.¹

Thus, an essential part of this study is the establishment of automated procedures and combinations of variables where no previous pattern existed. In addition, program development, selection of variables, analysis of data, and articulation of interpretive patterns

augment the contribution of this study. Moreover, this investigation is unique because it utilizes complete texts of French prose to distinguish between stylistic features of French authors.² Finally, since an author's style is a combination of an almost unlimited number of characteristics, programs designed for use in this study were developed to examine as many combinations of variables as practicable. Eight hundred and forty-three test items were delineated in the project. Possible style discriminants were tested independently and finally compared collectively. Their total represents a profile, while perhaps only partial, at least more complete than could be produced by single-characteristic studies of an author's style. This profile should serve to distinguish one author from another.

The 843 test items were grouped into five major areas as discussed in Chapter II (p. 21). They are represented here with the computer program which produced their results. Each program has been given a mnemonic title which indicates its function and facilitates its discussion. The test items will be discussed in the following order:

1. Sentence-level measures (SENWOL)
2. Sentence beginnings and endings (STYLBEND)
3. Function word frequencies and usage (FREQFUN)
4. Expression frequencies and usage (EXSOR)
5. Vocabulary analysis (ENROOT)

Results from each category will be presented in tabular form showing statistical differences and implied similarities between Bayle's, Larroque's and the unknown author's styles. Also included in each appropriate table are t-test values and their confidence levels. As previously explained in Chapter III (p. 67), the one-way analysis of

variance which compares two items is statistically equivalent to the t-test. Whereas the significant F-value for the analysis of variance suggests similarity, a significant t-value indicates differences rather than equality in treatment means.³ Results tabulated in this chapter, with the exception of those from SENWOL and STYLBEND, which include both statistical tests, are presented with their level of significance based upon the t-value for the sake of uniformity and clarity. After each of the 843 test items was subjected to the t-test, tables of significance were consulted to determine which of the obtained values were likely to occur due to chance or sampling error fewer than five times out of one hundred. Most of the tests which showed significance indicated a possible error of less than one, rather than five, out of a hundred. For a sample size of 30 or greater, a t-value of 2.58 is necessary to establish an .01 level of confidence, while a value of 1.96 is generally sufficient to indicate an .05 level.⁴ SENWOL and STYLBEND results will be presented with both analysis of variance and t-test scores. Variables whose analysis fell below the .05 confidence level are not reported in this chapter.

For the final analysis, 843 variables were tested across five articles. Descartes' Méditations, Péliſson's "Discours sur les oeuvres de Monsieur Sarasin," and articles by Bayle and Larroque from Les Nouvelles de la république des lettres served as controls and were not included in the final analysis.⁵ Because of Ascoli's convincing argument seen earlier in Chapter I (p. 6) refuting attribution of l'Avis to Péliſson, and because preliminary tests revealed that his style resembles that of the author of l'Avis much less than that of either Bayle or Larroque, Péliſson's article was omitted from the final testing.

Five test articles were finally chosen:

1. P - The preface to l'Avis aux réfugiés was retained to test the hypothesis that its writer was not the same as the author of l'Avis itself.
2. X - l'Avis aux réfugiés was tested apart from its preface.
3. B1 - Cabale Chimérique provided a text known to be by Bayle, in which he denies having written l'Avis and defends himself against the attacks of Jurieu. The subject matter, genre, and length are almost identical to l'Avis.
4. B2 - Réponse d'un nouveau converti was attributed to Bayle because it exposes ideas of tolerance with which he was in sympathy. He never avowed its authorship, but then he never openly denied it either. The Réponse is supposedly a prelude to l'Avis, having appeared in 1689. Retained as a final test article because if it is in reality a miniature Avis, there should be strong parallels between l'Avis and its forerunner.
5. L1 - Le prosélyte abusé was chosen from among Larroque's works made available for this study by the Bibliothèque Nationale, because it most closely resembles l'Avis and Bayle's test work in content, genre, and length.

The data tabulated and interpreted from test articles in the study were directed into the five previously listed categories, for each of which a computer program had been devised. The purpose was to arrange the data in a format which facilitates further analysis of the relationships between articles.

The results from each of the five categories will be presented in tabular form followed by tentative conclusions, where warranted, and by a summary of statistical values which suggest them.

Sentence Level Measures

Tables 4:1 through 4:9 present distributions of frequencies of words, syllables and alphabetical characters per sentence. Distributions of the following information are summarized:

1. average number of words per sentence,
2. average number of syllables per sentence,
3. average number of letters per sentence,
4. average number of letters per word,
5. average number of syllables per word, and
6. average number of letters per syllable.

Additional distributions are presented which group into genres the works to be tested. Finally, summary tables are presented of all data regarding sentence level measures.

Table 4:1 shows the division of each article into units of one hundred sentences. Even though articles X, B1, and L1, contain 1811, 1839, and 1176 sentences respectively, the table only includes the first 1000 sentences of each. Because values beyond the 1000 sentence listing did not significantly change the established pattern, there was no need to include the final averages. Their presence would merely have extended two of the five columns. The first column shows the division into one-hundred-sentence units. The remaining five columns list the average number of words per sentence per unit. Denominators P, X, B1, B2, and L1 correspond to the final five articles tested as described above.

The major articles in this study were tested in units of one hundred sentences for internal homogeneity. Table 4:1 represents results from these internal tests of homogeneity with reference to

TABLE 4:1. AVERAGE NUMBER OF WORDS PER SENTENCE

Sentence Numbers	P	X	B1	B2	L1
1-100	40.667	30.950	31.510	50.330	30.430
101-200	-	32.530	32.100	50.650	30.860
201-300	-	32.440	30.060	45.770	29.510
301-400	-	30.740	30.070	-	31.880
401-500	-	28.440	31.040	-	23.470
501-600	-	33.410	36.690	-	33.620
601-700	-	32.000	31.770	-	33.120
701-800	-	31.210	28.740	-	36.400
801-900	-	31.880	28.110	-	31.990
901-1000	-	31.510	31.250	-	26.990
Final Average	40.667	30.549	31.338	49.059	30.783

average number of words per sentence. Line one of the sentence number column shows the average number of words per sentence in the first one hundred sentences, line two shows the second one hundred sentences, and each following line deals with 100 incidents of occurrence. The preface, P, contains only eighty-four sentences; therefore its values are found only in the first line of the columns. Because the preface and B2 are short, their consistency values are not as readily apparent as those in X, B, and L1. Consistency within a given author's work is evidenced by small deviation between one-hundred-sentence units. Had these works been written by more than one author (cf., especially X) there would not likely be the consistency which is present, unless the collaborators had

intentionally designed their work to match each other's style.

Table 4:1 shows consistency from unit to unit of B1 except at the 501-600 sentence level where the author brings together three of the major topics of his work in a most unclear, complex series of dependent clauses and phrases.⁶ In the B2 column, variations are widely divergent from each of the other representative works. A follow-up study might pursue the possibility as suggested by Robinson⁷ that Bayle did not write the Réponse d'un nouveau converti. On the other hand, the relative shortness of the article may not have permitted a pattern to develop. Finally, in the L1 column, three groups of sentences disturb the internal homogeneity pattern for no apparent reason. The final average figures reflect the mean values for the complete articles and are typical of the visual closeness of authors X, B1, and L1 in the sentence level tests.

Table 4:2 presents an overview of absolute frequencies of number of sentences, words, syllables, and letters in the test articles.

The author column presents last names of each author whose work was tested in this group: preface, unknown, Péliſson, Bayle, and Larroque. The word "preface," used in tables of this study, refers to the work and to its author. Numbers one through nine refer to works by Bayle listed by title in Appendix A. Numbers one through twenty-three refer to works by Larroque similarly listed. Numbers listed under sentences, words, syllables, and letters represent absolute frequencies of occurrence in each category for each article. These figures serve as raw data for the analyses to follow.

Only data from SENWOL include tests of Péliſson's work. These sentence level results revealed a greater difference between Péliſson

TABLE 4:2. OVERVIEW OF ABSOLUTE FREQUENCIES

Author		Sentences	Words	Syllables	Letters
Preface		84	3418	5583	14835
Unknown		1811	55551	91610	248328
Péllisson		322	12628	20659	55365
Bayle	1	1839	57787	92948	245521
Bayle	2	287	14147	23309	62662
Bayle	3	46	2077	3375	8981
Bayle	4	105	4608	7607	20096
Bayle	5	54	2484	4183	11236
Bayle	6	36	1341	2253	5970
Bayle	7	131	5345	8686	22802
Bayle	8	27	1117	1858	4910
Bayle	9	21	813	1352	3541
B Totals		2546	89495	145571	385719
Larroque	1	1176	36190	59003	157130
Larroque	2	154	6822	11333	30121
Larroque	3	55	2144	3462	9267
Larroque	4	70	3463	5669	15293
Larroque	5	49	1977	3234	8678
Larroque	6	26	972	1560	4133
Larroque	7	140	5073	8543	22252
Larroque	8	75	3027	5073	13348
Larroque	9	37	1735	2998	7836
Larroque	10	48	2028	3359	8916
Larroque	11	43	1746	2965	7665
Larroque	12	35	1341	2251	5911
Larroque	13	63	2189	3416	8917
Larroque	14	131	4865	7838	21199
Larroque	15	27	850	1357	3564
Larroque	16	155	4883	7802	21083
Larroque	17	162	4874	8222	21823
Larroque	18	112	3757	6530	17043
Larroque	19	140	5191	8703	22739
Larroque	20	171	6255	10338	27196
Larroque	21	173	5624	9280	24667
Larroque	22	84	2711	4208	11146
Larroque	23	35	1170	1829	4849
L Totals		3161	108898	178973	474776

and the author of l'Avis than between either Bayle or Larroque and the unknown author. Since these results confirm the rather conclusive argument presented by Ascoli reported on page 6 of this study, Péliisson's work was dropped from further testing.

Table 4:3 records the average number of words per sentence and comparative statistical values for thirty-five test articles.

The article column remains the same in Tables 4:3 through 4:6 and refers to the works listed by title and author in Appendix A.

The second column lists the average number of words per sentence per article. Of some significance is the marked divergence between the mean number of words per sentence of B1 when compared to averages of B2 through B9. A similar difference is readily notable when comparing L1 to L2 through L23. Presence of these distinct differences, apparent also in Tables 4:4 and 4:5 prompted the regrouping of test articles according to genre as shown in Table 4:7.

Figures defining "All of B" and "All of L" summarize the mean sentence length of articles by Bayle and Larroque respectively. This summary mean emphasizes the already observed divergence between articles B1 and articles B2 through B9, as well as L1 and his remaining test works.

Statistical calculations at the bottom of the table are ANOVAR (analysis of variance) program results, designed to determine the likelihood that either B or L wrote the disputed article. All statistical tests in this study are based on the hypothesis that either Bayle or Larroque wrote l'Avis aux réfugiés. Using the "All of B" and "All of L" mean values, the computer calculated an F-value of 4.31 representing a level of significance or confidence level of .04 in favor of the

TABLE 4:3. AVERAGE NUMBER OF WORDS PER SENTENCE

Article		Average Words/Sentence
Preface		40.667
Unknown		30.549
Péllisson		39.217
Bayle	1	31.338
Bayle	2	49.059
Bayle	3	45.152
Bayle	4	43.886
Bayle	5	46.000
Bayle	6	37.250
Bayle	7	40.802
Bayle	8	41.370
Bayle	9	38.714
All of B		35.151
Larroque	1	30.783
Larroque	2	44.299
Larroque	3	38.982
Larroque	4	49.471
Larroque	5	40.347
Larroque	6	37.385
Larroque	7	36.236
Larroque	8	40.360
Larroque	9	46.892
Larroque	10	42.250
Larroque	11	40.605
Larroque	12	38.314
Larroque	13	34.746
Larroque	14	37.137
Larroque	15	31.481
Larroque	16	31.503
Larroque	17	30.086
Larroque	18	33.545
Larroque	19	37.079
Larroque	20	36.579
Larroque	21	32.509
Larroque	22	32.274
Larroque	23	33.429
All of L		34.450
Mean of /X-B/ is		10.958
Mean of /X-L/ is		6.720
Calculated F-value is		4.312
Level of significance is		.044
Statistically significant in favor of /X-L/		

conclusion that Larroque was the author of l'Avis. This tentative conclusion was later invalidated by a regrouping of the test articles as shown in Table 4:7.

Table 4:4 shows the average number of syllables per sentence of each work analyzed.

As noted in previous tables, the article column refers to numbered works named in Appendix A. The second column lists the average number of syllables per sentence per article. Again it is apparent that the average number of syllables per sentence of article B1 diverges sharply from the means of B2 through B9. This pattern of divergence is also repeated in the L1 vs. L2 through L23 figures. The mean values defining "All of B" and "All of L" summarize the average number of syllables per sentence in the works by Bayle and Larroque used in this study. This summary mean emphasizes the already mentioned differences between articles B1 and articles B2 through B9, as well as L1 and his remaining articles.

At the bottom of the table are found statistical calculations computed by the analysis of variance program. The mean of X-B was statistically compared to the mean of X-L. The comparison revealed a statistical difference with a calculated F-value of 3.95 and an .05 level of significance or confidence level in favor of the conclusion that Larroque was the author of l'Avis aux réfugiés. This tentative conclusion was later invalidated by a regrouping of test articles as shown in Table 4:7.

TABLE 4:4. AVERAGE NUMBER OF SYLLABLES PER SENTENCE

Article		Average Syllables/Sentence
Preface		66.464
Unknown		50.585
Péllisson		64.158
Bayle	1	50.543
Bayle	2	81.216
Bayle	3	73.370
Bayle	4	72.448
Bayle	5	77.463
Bayle	6	62.583
Bayle	7	66.305
Bayle	8	68.815
Bayle	9	64.381
All of B		57.176
Larroque	1	50.173
Larroque	2	73.591
Larroque	3	62.945
Larroque	4	80.986
Larroque	5	66.000
Larroque	6	60.000
Larroque	7	61.021
Larroque	8	67.640
Larroque	9	81.027
Larroque	10	69.979
Larroque	11	68.953
Larroque	12	64.314
Larroque	13	54.222
Larroque	14	59.832
Larroque	15	50.259
Larroque	16	50.335
Larroque	17	50.753
Larroque	18	58.304
Larroque	19	62.164
Larroque	20	60.456
Larroque	21	53.642
Larroque	22	50.095
Larroque	23	52.257
All of L		56.619
Mean of /X-B/ is		17.993
Mean of /X-L/ is		10.801
Calculated F-value is		3.955
Level of significance is		.053
Statistically significant in favor of /X-L/		

Table 4:5 lists the average number of letters per sentence of each work analyzed.

As noted in previous tables, the article column refers to selected works delineated in Appendix A. In Table 4:5 the second column records the average number of letters per sentence in each test article. Once again, a distinct difference is observed when the mean values of B1 and B2 through B9 as well as L1 and L2 through L23 are compared. Figures concidental to "All of B" and "All of L" in this case summarize the average number of letters per sentence over all Bayle and Larroque articles tested. Once more this summary mean emphasizes the divergence, observed also in Tables 4:3 and 4:4, between articles B1 and articles B2 through B9. A similar distinction is reaffirmed when averages of L1 and L2 through L23 are compared with their summary mean.

Statistical calculations at the bottom of the table are analysis of variance program results which determine the likelihood that either Bayle or Larroque wrote the disputed article, based upon the hypothesis that one or the other was, in reality its author. Using "All of B" and "All of L" mean values, the computer calculated an F-value of 4.349 which represents a level of significance or confidence level of .04 in favor of the conclusion that Larroque was the author of l'Avis aux réfugiés. Again this tentative conclusion was invalidated by a regrouping of test articles as shown in Table 4:7.

TABLE 4:5. AVERAGE NUMBER OF LETTERS PER SENTENCE

Article		Average Letters/Sentence
Preface		176.607
Unknown		137.122
Péllisson		171.941
Bayle	1	133.508
Bayle	2	218.334
Bayle	3	195.239
Bayle	4	191.390
Bayle	5	208.074
Bayle	6	165.833
Bayle	7	174.061
Bayle	8	181.852
Bayle	9	168.619
All of B		151.500
Larroque	1	133.614
Larroque	2	195.591
Larroque	3	168.491
Larroque	4	218.471
Larroque	5	177.102
Larroque	6	158.962
Larroque	7	158.943
Larroque	8	177.973
Larroque	9	211.784
Larroque	10	185.750
Larroque	11	178.256
Larroque	12	168.886
Larroque	13	131.540
Larroque	14	161.824
Larroque	15	132.000
Larroque	16	136.019
Larroque	17	134.710
Larroque	18	152.170
Larroque	19	162.421
Larroque	20	159.041
Larroque	21	142.584
Larroque	22	132.690
Larroque	23	138.543
All of L		150.198
Mean of /X-B/ is		45.560
Mean of /X-L/ is		26.378
Calculated F-value is		4.349
Level of significance is		.043
Statistically significant in favor of /X-L/		

Table 4:6 presents the average number of letters per word, syllables per word, and letters per syllable of each text analyzed in this study.

Names and numbers in the article column refer to works listed by title and author in Appendix A. Figures in the second and third columns represent the average word length, determined first by the number of alphabetical characters per word and then by the number of syllables per word. In the fourth column are listed the average number of letters per syllable. Whereas a distinct divergence of mean values appeared in Tables 4:3, 4:4, and 4:5, where letters and syllables were compared at the sentence level, no such difference is evident when these items are observed at the word level. The numbers defining "All of B" and "All of L" summarize the mean values of articles B1 through B9 and L1 through L23 in the letters/word, syllables/word, and letters/syllable columns.

A visual analysis of the averages reveals only slight differences between values in each column. In Bayle's works, the average number of letters per word, for example, ranges only from 4.2 to 4.5. Larroque's averages, on the other hand, extend from 4.0 to 4.5. In the average syllables/word column, even a greater degree of consistency appears in the representations of Bayle's corpus. This is not so, however, in the case of Larroque, whose averages range from 1.5 to 1.7. Finally, in the letters per syllable column, general consistency appears in all represented works, the overall range varying only from 2.5 to 2.7.

When the computed mean values did not make readily apparent a significant divergence, the value of using a statistical analysis to sharpen the distinctiveness ratio became more evident. For each summary value of Table 4:6 statistical calculations were performed, yielding the

TABLE 4:6. AVERAGE NUMBER OF LETTERS PER WORD, SYLLABLES PER WORD, AND LETTERS PER SYLLABLE

Article	Average Letters/Word	Average Syllables/Word	Average Letters/Syllable
Preface	4.343	1.634	2.657
Unknown	4.489	1.656	2.711
Péllisson	4.384	1.636	2.680
Bayle 1	4.260	1.613	2.641
Bayle 2	4.450	1.655	2.688
Bayle 3	4.324	1.625	2.661
Bayle 4	4.361	1.651	2.642
Bayle 5	4.523	1.684	2.686
Bayle 6	4.452	1.680	2.650
Bayle 7	4.266	1.625	2.625
Bayle 8	4.396	1.663	2.643
Bayle 9	4.355	1.663	2.619
All of B	4.310	1.627	2.650
Larroque 1	4.340	1.630	2.663
Larroque 2	4.415	1.661	2.658
Larroque 3	4.322	1.615	2.677
Larroque 4	4.416	1.637	2.698
Larroque 5	4.389	1.636	2.683
Larroque 6	4.252	1.605	2.649
Larroque 7	4.386	1.684	2.605
Larroque 8	4.410	1.676	2.631
Larroque 9	4.516	1.728	2.614
Larroque 10	4.396	1.656	2.654
Larroque 11	4.390	1.698	2.585
Larroque 12	4.408	1.679	2.626
Larroque 13	4.074	1.561	2.610
Larroque 14	4.357	1.611	2.705
Larroque 15	4.193	1.596	2.626
Larroque 16	4.318	1.598	2.702
Larroque 17	4.477	1.687	2.654
Larroque 18	4.536	1.738	2.610
Larroque 19	4.380	1.677	2.613
Larroque 20	4.348	1.653	2.631
Larroque 21	4.386	1.650	2.658
Larroque 22	4.111	1.552	2.649
Larroque 23	4.144	1.563	2.651
All of L	4.360	1.643	2.653
Mean of /X-B/ is	.119	.019	.060
Mean of /X-L/ is	.148	.041	.064
Calculated F-value is	.511	3.875	.160
Level of significance is	.513	.055	.694

results shown at the bottom of the table. None of the three tests produced positive results pointing to the authorship of l'Avis aux réfugiés. In order to be considered significant the calculated F-value would have to be 4.17 or greater. Only the values for average syllables per word approach this figure. Again tentative conclusions drawn from these tests were later invalidated by a regrouping of test articles as shown in Table 4:7.

Because word level tests as represented in Table 4:6 revealed no significant differences, one and two letter words, primarily functional in nature, were omitted from the test data so that a higher degree of significance or distinction between the test authors' prose rhythm might be obtained. However, even though a greater divergence did begin to reveal itself, differences were so slight that it was impossible to draw any conclusions from them. Tabulated values of the average number of letters per word, the average number of syllables per word and the average number of letters per syllable with one and two letter words removed are found in Appendix F.

Table 4:7 divides the test articles into two groups: expository criticism and literary criticism, and shows statistically significant differences obtained after the articles were regrouped with regard to sentence lengths.

Tables 4:3 through 4:5 showed a distinct numerical difference between the mean values of the three variables tested. In addition, slight variability was noted in the test results reported in Table 4:6. The consistency of this differentiation suggested the presence of another stylistic feature as cause of the distinction. In Tables 4:3 through 4:5 the change of mean seemed to coincide with a change in

TABLE 4:7. REGROUPING OF TEST ARTICLES

Groups		Articles
B-1	B1 and B2	(<u>Cabale Chimérique</u> and <u>Réponse</u>)
B-2	B3 through B9	(<u>Nouvelles de la république des lettres</u>)
L-1	L1	(<u>Le Prosélyte abusé</u>)
L-2	L2 through L23	(<u>Nouvelles de la république des lettres</u>)
Significant differences for the variable: Number of words per sentence		
Articles	t-value	Conf. Level (CL)
B-1 & L-1	2.520	.025
B-1 & B-2	4.657	.001
L-1 & L-2	4.675	.001

authorial intent, i.e., when the author stopped editorializing and began criticizing literary works. To test this hypothesis further, works which represent editorializing or expository criticism were grouped separately from those drawn from literary analysis.

Thus, the group B-1 comprises articles B1 and B2, and the group B-2 consists of B3 through B9. Likewise L-1 is L1, and L2 through L23 become L-2.

T-tests were performed on grouped articles denominated B-1 and L-1, B-1 and B-2, and L-1 and L-2. A t-value of 2.520, with a confidence level of .025 was obtained from the test of B-1 and L-1. This confidence level indicates that there is a statistical difference between the two authors which could occur fewer than two and one-half times in each one hundred samples, yielding a 97.5% chance that the noted difference was not due to sampling error. Values of 4.657 and 4.675, both yielding a confidence level of .001 resulted from t-tests

performed on groups B-1 vs. B-2 and L-1 vs. L-2 respectively. The .001 significance level suggests better than a 99% chance--999 in one thousand--that the noted difference was not due to sampling error or chance.

T-tests for significant differences thus performed on SENWOL data yielded substantiation for the hypothesis that grouping according to authorial intent was justified. Results from the regrouping tests tend to support Sherbo's hypothesis that "there is one style for the criticism of poetry, a second for the considerations on corn laws, a third for introducing a new periodical to the public, a fourth for reviewing books . . ."8 In addition, whereas the analysis of variance had failed to distinguish between authors B and L9, the new groupings made computational differences in styles of the two writers more readily apparent.

B-1 and L-1 are examples of expository criticism, while B-2 and L-2 represent literary criticism. A higher level of confidence was obtained from between genres tests than from those within genres. This fact served as the basis for further use of genre groupings in seeking a solution for the disputed authorship of l'Avis aux réfugiés.

Table 4:8 shows the average number of words, syllables, and letters per sentence; the average number of letters and syllables per word; and the average number of letters per syllable after articles dealing with literary criticism had been eliminated.

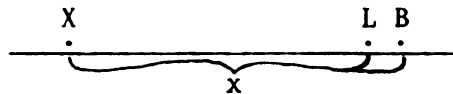
The first column provides a shortened form of the article title followed by the letter abbreviation which has been assigned for the purpose of presenting these data. The following six columns summarize the average number of words per sentence, syllables per sentence, letters per sentence, letters per word, syllables per word, and letters per

TABLE 4:8. AVERAGE NUMBER OF WORDS, SYLLABLES, AND ALPHABETIC CHARACTERS PER SENTENCE; AVERAGE NUMBER OF LETTERS PER WORD, SYLLABLES PER WORD, AND LETTERS PER SYLLABLE

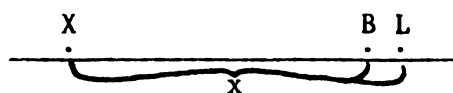
Article		Average Wrd/Sen	Average Syl/Sen	Average Let/Sen	Average Let/Wrd	Average Syl/Wrd	Average Let/Syl
Preface	P	40.667	66.464	176.607	4.343	1.634	2.657
<u>L'Avis</u>	X	30.549	50.585	137.122	4.489	1.656	2.711
<u>Cabale</u>	B1	31.338	50.543	133.508	4.260	1.613	2.641
<u>Réponse</u>	B2	49.059	81.216	218.334	4.450	1.655	2.688
<u>Prosélyte</u>	L	30.783	50.173	133.614	4.340	1.630	2.663

syllable for each article. These data accumulate the information presented in Tables 4:3 through 4:6 for the articles listed.

Before the test articles were regrouped (see Table 4:7), a linear representation of sentence level measures, e.g., words per sentence, looked like



where X, L, and B represent the styles of the unknown author, Larroque, and Bayle with respect to words, letters, and syllables per sentence. The distance x between B (or L) and X was significantly measurable with Larroque's values resembling those of l'Avis more closely than Bayle's. However, L and B were so similar that the analysis of variance could not distinguish between them. Similarly, a linear representation of word level measures, e.g., letters per word, appeared as



where the values for Bayle appear closer to those of l'Avis but still not statistically distinguishable from Larroque's. After the test articles were regrouped, separation between X, B, and L became much less obvious. Some values for Bayle and Larroque, e.g., words, syllables, and letters per sentence, now approached those of the unknown author so closely that statistical tests could not distinguish them. Still, as shown in Table 4:9 it was possible to statistically differentiate between all of them in four of six tests. In addition, three of six SENWOL tests showed significant divergence between the unknown text and its preface.

Summary and Conclusions of SENWOL Results

Table 4:9 is a summary of the SENWOL results based on the regrouped article totals.

The column denominated "Test Variable" lists six stylistic discriminators used to analyze test articles on the basis of sentence level and individual word measures. Column $P \neq X$ (read P not equal to X) contains the t-values obtained from comparing the test variables in l'Avis with those of its preface. $P \neq X$ symbolizes the presence of a statistically significant difference between articles with regard to corresponding test variables. Confidence level (CL) columns indicate the degree of confidence with which it may be judged that differences are not due to sampling error or to chance. Figures in the $X \neq B1$ or $B2$ column corresponding to words per sentence, syllables per sentence, and letters per sentence represent a statistically significant difference between X and B2; whereas the t-value for variables letters per word, syllables per word, and letters per syllable were derived from the X to B1 comparisons.

TABLE 4:9. SENWOL SUMMARY

Test Variable	P \neq X	CL	X \neq L1	CL
Words Per Sentence	2.931	.005	-	-
Syllables Per Sentence	2.801	.01	-	-
Letters Per Sentence	2.557	.025	-	-
Letters Per Word	-	-	4.892	.001
Syllables Per Word	-	-	2.704	.01
Letters Per Syllable	-	-	4.227	.001

Test Variable	X \neq B1 or B2	CL	B1 \neq L1	CL
Words Per Sentence	8.957	.001	-	-
Syllables Per Sentence	8.883	.001	-	-
Letters Per Sentence	8.653	.001	-	-
Letters Per Word	9.748	.001	3.702	.001
Syllables Per Word	6.764	.001	3.272	.005
Letters Per Syllable	7.021	.001	1.962	.05

In order to use t-test results to suggest a probable author, it was first necessary to establish a hypothesis against which tests of difference could be made. Since both Bayle's and Larroque's complete works include l'Avis aux réfugiés, and because literary critics and historians have seriously considered no other writer as its author, the initial hypothesis that either Bayle or Larroque wrote l'Avis seemed both justified and logical. To name a probable author under this hypothesis required that two conditions be satisfied. Because the possible-author field had been reduced to two candidates, the first

condition required that, for any given variable, there be a statistically significant difference between Bayle and Larroque.¹⁰ The second condition required the presence of a significant difference between one of the supposed writers and the unknown author, but not between the unknown author and the remaining candidate. When these conditions were met, in view of the original hypothesis, the writer whose use of the stylistic element in question did not differ significantly from that of the unknown author was advanced as the probable author of l'Avis aux réfugiés.

There remains, however, another possible inference that may be drawn from t-test results. If, for example, significant values reveal a decided difference between B and X, as well as between L and X, whether or not B differs from L is immaterial. Whenever both Bayle and Larroque differ significantly from the unknown author, the probable author is listed as neither (N). On the basis of these criteria, t-tests reported in Table 4:9 were performed after regrouping test articles. Final statistical results from SENWOL data showed three tests, i.e., letters per word, syllables per word, and letters per syllable, which revealed statistically significant results meeting the conditions just stated ($X \neq B1$ and $X \neq L1$). In all, the SENWOL summary table shows nine values significant at the .001 level, two at the .005 level, two at the .01 level, one at the .025 and one at the .05 level, for a total of fifteen significant values of a possible twenty-four. Ten of these values, all having at least an .01 confidence level, suggest that neither Bayle nor Larroque authored the unknown text.

In summary, then the SENWOL data provided information which led to two conclusions: First, there is, in fact, a quantitative difference

between expository and literary criticisms in sentence level measures, and second, the number of observed and calculated differences in this section indicate considerable evidence that the author of l'Avis aux réfugiés and Bayle employ different writing practices in their use of SENWOL variables.

Sentence Beginnings and Endings

Tables 4:10 through 4:28 present results obtained from the STYLBEND program which located and isolated first and last words of each sentence in thirty-four original test articles. Supplemental hand sorting placed raw STYLBEND data into fifteen grammatical categories as listed below.

adjectives	gerunds	nouns
adverbs	infinitives	numbers
articles	interrogatives	prepositions
conjunctions	Latin	pronouns
exclamations	negations	verbs (finite)

A simple counting and averaging program then summed the absolute frequencies for each part of speech and calculated the percentage of occurrence of each part of speech as a sentence beginning or ending for each test article. In order to construct a coherent pattern for presentation of results and to facilitate the hand sorting procedure, part of speech categories for both initial and terminal words were kept the same. As might well be expected, some zero values were obtained because of this decision (e.g., articles, interrogatives). Nevertheless, since zero values did not affect the final statistics for this set of variables, they were ignored, and the categories remained.

Tables presenting STYLBEND results fall into three groups.

Tables 4:10 through 4:24 record frequencies, percentage of use, and statistical values relating to the use of each of fifteen parts of speech as sentence beginnings and/or endings for all thirty-four test works.

Tables 4:25 and 4:26 summarize data drawn only from articles related to expository criticism. These five articles (see Table 4:7) had revealed themselves as being of a genre more like l'Avis aux réfugiés than those of literary criticism.

Finally, Tables 4:27 and 4:28 summarize significant values obtained from the STYLBEND program, their t-test results, their levels of confidence and a probable author hypothesis.

Table 4:10 shows absolute frequencies, percentage of use, and statistical values relating to the use of adjectives as sentence beginnings and endings in thirty-four test articles.

Article columns identify position of the adjective as either initial or terminal. Alphabetical characters (P, X, B, or L) and digits following them in article columns identify works tested as listed in Appendix A. In frequency columns appear the number of times an adjective is used in its corresponding article as either a sentence beginning or ending. In order to correct for distortions of statistical results due to differences in article lengths, percentage of use values were calculated and are presented in percent columns.

Results of the test for internal homogeneity of variance and ANOVAR (analysis of variance) results for all values in Table 4:10 follow the first and last word values of L23. The hypothesis of equal variance supposes that use of the prescribed part of speech as a sentence first word is so consistent that discrimination between individual

TABLE 4:10. ADJECTIVES AS SENTENCE BEGINNINGS AND ENDINGS

Article	Frequency	Percent	Article	Frequency	Percent
First P	4	4.762	Last P	10	11.905
First X	82	4.528	Last X	279	15.406
First B 1	100	5.435	Last B 1	254	13.804
First B 2	9	3.136	Last B 2	45	15.679
First B 3	2	4.348	Last B 3	8	17.391
First B 4	12	11.321	Last B 4	15	14.151
First B 5	4	7.273	Last B 5	8	14.545
First B 6	5	13.889	Last B 6	9	25.000
First B 7	16	12.214	Last B 7	17	12.977
First B 8	4	14.815	Last B 8	6	22.222
First B 9	4	19.048	Last B 9	2	9.524
First L 1	60	5.102	Last L 1	183	15.561
First L 2	5	3.247	Last L 2	25	16.234
First L 3	4	7.143	Last L 3	7	12.500
First L 4	8	11.429	Last L 4	12	17.143
First L 5	7	14.286	Last L 5	9	18.367
First L 6	4	15.385	Last L 6	4	15.385
First L 7	5	3.597	Last L 7	26	18.705
First L 8	7	9.211	Last L 8	5	6.579
First L 9	2	5.405	Last L 9	10	27.027
First L 10	2	4.167	Last L 10	11	22.917
First L 11	0	0.000	Last L 11	4	9.302
First L 12	3	8.571	Last L 12	6	17.143
First L 13	2	3.175	Last L 13	15	14.815
First L 14	1	.763	Last L 14	25	19.084
First L 15	0	0.000	Last L 15	4	14.815
First L 16	8	5.161	Last L 16	29	18.710
First L 17	8	4.938	Last L 17	39	24.074
First L 18	4	3.571	Last L 18	22	19.643
First L 19	8	5.714	Last L 19	0	0.000
First L 20	12	6.977	Last L 20	33	19.186
First L 21	10	5.780	Last L 21	16	9.249
First L 22	1	1.190	Last L 22	16	19.048
First L 23	1	2.857	Last L 23	3	8.571

Hypothesis of Equal Variance Fails with t-value of 2.777

Mean of /X-B/ is	5.985	3.410
Mean of /X-L/ is	2.997	4.887
Calculated F-value is	4.539	1.023
Level of significance is	.039	.320

Statistically significant in favor of /X-L/

works by the same author is unlikely.¹¹ For adjectives as sentence beginnings the hypothesis of homogeneity of variance was rejected as witnessed by the statement, "Hypothesis of equal variance fails with t-value of 2.777." Stylistically speaking, then, Bayle and Larroque are not consistent in their use of adjectives as sentence beginnings; but which, if either, most resembles the unknown author?

Below the test for internal variance results are found the ANOVAR results where the average of all B percent values is subtracted from the X percent value to obtain the mean absolute¹² difference between the articles thus tested. The "slash" marks on either side of /X-B/ and /X-L/ define these differences as absolute values. Resulting figures were used to calculate the F statistic. The computed F-value of 4.539 for adjectives used as sentence beginnings represents a confidence level of .039 in favor of the conclusion that Larroque was the author of l'Avis aux réfugiés. Just as several other tentative conclusions were later negated by a regrouping of the test articles shown in Table 4:7, so it was with this one (see Table 4:27). Statistical computations for adjectives used as sentence endings yielded no significant results.

Table 4:11 shows absolute frequencies, percentage of use, and statistical values relating to use of adverbs as sentence beginnings and endings in thirty-four test articles.

Article columns identify the position of the adverb as either the first or last word in sentences tested in the preface to l'Avis aux réfugiés, l'Avis itself, or in works of Bayle (B) and Larroque (L). Numbers immediately next to these author denominators further identify works tested as listed in Appendix A. The actual number of times an adverb is used in the adjacent article as either a beginning or ending

TABLE 4:11. ADVERBS AS SENTENCE BEGINNINGS AND ENDINGS

Article	Frequency	Percent	Article	Frequency	Percent
First P	1	1.190	Last P	4	4.762
First X	44	2.430	Last X	66	3.644
First B 1	26	1.413	Last B 1	75	4.076
First B 2	9	3.136	Last B 2	12	4.181
First B 3	3	6.522	Last B 3	4	8.696
First B 4	4	3.774	Last B 4	2	1.887
First B 5	2	3.636	Last B 5	1	1.818
First B 6	0	0.000	Last B 6	0	0.000
First B 7	3	2.290	Last B 7	4	3.053
First B 8	1	3.704	Last B 8	0	0.000
First B 9	0	0.000	Last B 9	1	4.762
First L 1	26	2.211	Last L 1	41	3.486
First L 2	1	.649	Last L 2	8	5.195
First L 3	1	1.786	Last L 3	2	3.571
First L 4	0	0.000	Last L 4	6	8.571
First L 5	0	0.000	Last L 5	1	2.041
First L 6	0	0.000	Last L 6	1	3.846
First L 7	2	1.439	Last L 7	4	2.878
First L 8	2	2.632	Last L 8	3	3.947
First L 9	4	10.811	Last L 9	5	13.514
First L 10	1	2.083	Last L 10	3	6.250
First L 11	0	0.000	Last L 11	1	2.326
First L 12	0	0.000	Last L 12	1	2.857
First L 13	0	0.000	Last L 13	1	1.587
First L 14	2	1.527	Last L 14	5	3.817
First L 15	1	3.704	Last L 15	2	7.407
First L 16	1	.645	Last L 16	3	1.935
First L 17	7	4.321	Last L 17	4	2.469
First L 18	1	.893	Last L 18	3	2.679
First L 19	2	1.429	Last L 19	38	27.143
First L 20	3	1.744	Last L 20	7	4.070
First L 21	5	2.890	Last L 21	8	4.624
First L 22	0	0.000	Last L 22	0	0.000
First L 23	0	0.000	Last L 23	1	2.857
Mean of /X-B/ is		1.626			2.066
Mean of /X-L/ is		1.805			2.754
Calculated F-value is		.087			.159
Level of significance is		.766			.694

is tabulated in the frequency column. In order to correct for distortions of statistical results due to differences in article lengths, percentage of use values were calculated and are presented in percent columns.

At the bottom of the table are located statistical calculations from ANOVAR used to determine the likelihood that either Bayle or Larroque wrote the disputed article, based upon the hypothesis that one or the other did, in reality, write it. The absence of a "Hypothesis of equal variance fails" statement in this table and in Tables 4:14 through 4:24 indicates that homogeneity of internal variance does exist. Using mean absolute values of /X-B/ and /X-L/, an F-value was calculated for adverbs used both as beginnings and endings. Lack of significant results indicates that all authors tested use adverbs as beginnings and endings, with such similarity that statistical tests used could reveal no substantial differences. The .76 level of significance means that there is a seven out of ten possibility that any difference observed is due to sampling error or to chance and not due to actual differences in writing habits. Both the .76 and .69 confidence levels fall far below the .05 considered acceptable in this study.

Table 4:12 shows absolute frequencies, percentage of use, and statistical values relating to use of articles (definite and indefinite) as sentence beginnings and endings in thirty-four test articles.

Columns designated "Article" identify the position of definite or indefinite articles used in either initial or terminal positions. Denominators P, X, B1 to B9 and L1 to L23 identify works tested as listed in Appendix A. In frequency columns appear the number of times articles were used in their corresponding works as either beginnings

TABLE 4:12. ARTICLES AS SENTENCE BEGINNINGS AND ENDINGS

Article	Frequency	Percent	Article	Frequency	Percent
First P	12	14.286	Last P	0	0.000
First X	126	6.957	Last X	0	0.000
First B 1	150	8.152	Last B 1	0	0.000
First B 2	16	5.575	Last B 2	0	0.000
First B 3	2	4.348	Last B 3	0	0.000
First B 4	16	15.094	Last B 4	0	0.000
First B 5	14	25.455	Last B 5	0	0.000
First B 6	7	19.444	Last B 6	0	0.000
First B 7	20	15.267	Last B 7	0	0.000
First B 8	9	33.333	Last B 8	0	0.000
First B 9	0	0.000	Last B 9	0	0.000
First L 1	102	8.673	Last L 1	0	0.000
First L 2	19	12.338	Last L 2	0	0.000
First L 3	11	19.643	Last L 3	0	0.000
First L 4	10	14.286	Last L 4	0	0.000
First L 5	8	16.327	Last L 5	0	0.000
First L 6	3	11.538	Last L 6	0	0.000
First L 7	12	8.633	Last L 7	0	0.000
First L 8	17	22.368	Last L 8	0	0.000
First L 9	3	8.108	Last L 9	0	0.000
First L 10	7	14.583	Last L 10	0	0.000
First L 11	3	6.977	Last L 11	0	0.000
First L 12	9	25.714	Last L 12	0	0.000
First L 13	2	3.175	Last L 13	0	0.000
First L 14	4	3.053	Last L 14	0	0.000
First L 15	1	3.704	Last L 15	0	0.000
First L 16	9	5.806	Last L 16	0	0.000
First L 17	11	6.790	Last L 17	0	0.000
First L 18	12	10.714	Last L 18	0	0.000
First L 19	18	12.857	Last L 19	0	0.000
First L 20	18	10.465	Last L 20	0	0.000
First L 21	29	16.763	Last L 21	0	0.000
First L 22	8	9.524	Last L 22	0	0.000
First L 23	5	14.286	Last L 23	0	0.000

Hypothesis of Equal Variance Fails with t-value of 3.705

Mean of /X-B/ is	9.550
Mean of /X-L/ is	5.687
Calculated F-value is	2.658
Level of Significance is	.109

or endings. As might well be expected from basic rules of French syntax, the frequency column for articles used as sentence endings lists only zeros. However, such was not the case for articles used as beginnings. In order to correct for distortions of statistical results due to differences in lengths of the works analyzed, percentage of use values were calculated and are tabulated in the percent column.

Statistical results for this variable begin with rejection of the hypothesis of equal internal variance in articles by Larroque. Had there been internal consistency of use of definite and indefinite articles at the acceptable .05 confidence level, the t-value would have had to be less than 1.96.

Analysis of variance results showed divergency in use of definite and indefinite articles in initial and terminal positions in the writings of Larroque. However, when the mean absolute values of /X-B/ and /X-L/ for this variable were compared, the contrast was not great enough to register a statistically significant difference. The calculated F-value of 2.658 attests to this conclusion.

Table 4:13 records absolute frequencies, percentage of use, and statistical values relating to use of conjunctions as sentence beginnings and endings in thirty-four test articles.

The position of the test variable, conjunctions, as either initial or terminal, is defined in article columns. Designators P, X, B1 to B9 and L1 to L23 identify works tested as listed in Appendix A. In frequency columns appear the number of times conjunctions were used in their corresponding works as either beginnings or endings. Figures in the percent column represent percentage of use of conjunctions as first and last words of sentences in works tested.

TABLE 4:13. CONJUNCTIONS AS SENTENCE BEGINNINGS AND ENDINGS

Article	Frequency	Percent	Article	Frequency	Percent
First P	17	20.238	Last P	2	2.381
First X	613	33.849	Last X	11	.607
First B 1	542	29.457	Last B 1	23	1.250
First B 2	83	28.920	Last B 2	2	.697
First B 3	2	4.348	Last B 3	1	2.174
First B 4	9	8.491	Last B 4	1	.943
First B 5	5	9.091	Last B 5	0	0.000
First B 6	3	8.333	Last B 6	0	0.000
First B 7	7	5.344	Last B 7	0	0.000
First B 8	3	11.111	Last B 8	0	0.000
First B 9	2	9.524	Last B 9	1	4.762
First L 1	383	32.568	Last L 1	5	.425
First L 2	57	37.013	Last L 2	1	.649
First L 3	10	17.857	Last L 3	0	0.000
First L 4	23	32.857	Last L 4	0	0.000
First L 5	12	24.490	Last L 5	0	0.000
First L 6	7	26.923	Last L 6	0	0.000
First L 7	30	21.583	Last L 7	1	.719
First L 8	21	27.632	Last L 8	2	2.632
First L 9	9	24.324	Last L 9	0	0.000
First L 10	11	22.917	Last L 10	0	0.000
First L 11	4	9.302	Last L 11	0	0.000
First L 12	9	25.714	Last L 12	0	0.000
First L 13	17	26.984	Last L 13	0	0.000
First L 14	49	37.405	Last L 14	3	2.290
First L 15	12	44.444	Last L 15	0	0.000
First L 16	46	29.677	Last L 16	1	.645
First L 17	51	31.481	Last L 17	0	0.000
First L 18	34	30.357	Last L 18	0	0.000
First L 19	47	33.571	Last L 19	0	0.000
First L 20	48	27.907	Last L 20	2	1.163
First L 21	45	26.012	Last L 21	0	0.000
First L 22	39	46.429	Last L 22	0	0.000
First L 23	12	34.286	Last L 23	0	0.000

Hypothesis of Equal Variance Fails with t-value of 3.080

Mean of /X-B/ is	21.113	1.024
Mean of /X-L/ is	7.280	.624
Calculated F-value is	25.690	1.882
Level of significance	.000	.177

Statistically significant in favor of /X-L/

Statistical results for the conjunction-use variable begin with the statement, "Hypothesis of equal variance fails with t-value of 3.080." As with two previous variables, adjectives and articles, whose statistics report similar results (see Tables 4:10 and 4:12), Larroque's inconsistent use of conjunctions in the initial sentence position caused rejection of the hypothesis of homogeneity of variance.

Even though the internal test for homogeneity revealed that Larroque used conjunctions inconsistently as sentence first words, when the mean absolute value /X-L1/ of all his works tested was compared to the corresponding mean for B, /X-B1/, a strong significant difference appeared. As shown in Table 4:13, the calculated F-value of 25.690 confirms the visually observable difference between mean absolute differences. The 25.690 F-value represents a confidence level of .000085 in favor of the conclusion that Larroque wrote l'Avis aux réfugiés. However, just as several other tentative conclusions were later negated or revised because of a regrouping of test articles, so was this one (see Table 4:29).

Use of conjunctions as sentence endings is not common in French syntax, although both Grevisse and Le Bidois suggest that pourtant, ainsi, donc, en effet, and cependant are permitted as sentence closers.¹³ None of the writers tested used these conjunctive forms so frequently as to distinguish his writings from the others.

Table 4:14 lists absolute frequencies, percentage of use, and statistical values relating to use of exclamations as sentence beginnings and endings in thirty-four test articles.

Columns denominated "Article" identify position of the test variable, exclamations, as either initial or terminal. Designators P, X,

TABLE 4:14. EXCLAMATIONS AS SENTENCE BEGINNINGS AND ENDINGS

Article	Frequency	Percent	Article	Frequency	Percent
First P	0	0.000	Last P	0	0.000
First X	5	.276	Last X	0	0.000
First B 1	13	.707	Last B 1	8	.435
First B 2	2	.697	Last B 2	0	0.000
First B 3	0	0.000	Last B 3	0	0.000
First B 4	0	0.000	Last B 4	0	0.000
First B 5	0	0.000	Last B 5	0	0.000
First B 6	0	0.000	Last B 6	0	0.000
First B 7	0	0.000	Last B 7	0	0.000
First B 8	0	0.000	Last B 8	0	0.000
First B 9	0	0.000	Last B 9	0	0.000
First L 1	16	1.361	Last L 1	4	.340
First L 2	0	0.000	Last L 2	0	0.000
First L 3	0	0.000	Last L 3	0	0.000
First L 4	0	0.000	Last L 4	0	0.000
First L 5	0	0.000	Last L 5	0	0.000
First L 6	0	0.000	Last L 6	0	0.000
First L 7	1	.719	Last L 7	0	0.000
First L 8	0	0.000	Last L 8	0	0.000
First L 9	0	0.000	Last L 9	0	0.000
First L 10	0	0.000	Last L 10	0	0.000
First L 11	0	0.000	Last L 11	0	0.000
First L 12	0	0.000	Last L 12	0	0.000
First L 13	1	1.587	Last L 13	0	0.000
First L 14	0	0.000	Last L 14	0	0.000
First L 15	0	0.000	Last L 15	0	0.000
First L 16	0	0.000	Last L 16	0	0.000
First L 17	0	0.000	Last L 17	0	0.000
First L 18	0	0.000	Last L 18	0	0.000
First L 19	0	0.000	Last L 19	0	0.000
First L 20	0	0.000	Last L 20	0	0.000
First L 21	0	0.000	Last L 21	0	0.000
First L 22	0	0.000	Last L 22	0	0.000
First L 23	0	0.000	Last L 23	0	0.000
Mean of /X-B/ is		.309			.048
Mean of /X-L/ is		.363			.014
Calculated F-value is		.353			.782
Level of significance is		.563			.612

B1 to B9 and L1 to L23 identify works tested as named in Appendix A. In frequency columns appear the number of times that exclamations were used in their corresponding works as either beginnings or endings. Because the exclamations group contains both exclamations and interjections, frequency of use varies between first and last words. The STYLBEND program uses periods as sentence delimiters. Therefore, since exclamation marks were keypunched as periods, frequency of occurrence of true interjectory forms, e.g., ah, ha, oh, or quoi is reflected in both beginnings and endings; whereas exclamatory forms, e.g., "Voilà que," "que je suis content," or "combien je souffre," not followed by a period are reflected only in first word totals. In order to correct for distortions of statistical results due to differences in article lengths, percentage of use values were calculated and are tabulated in percent columns.

At the bottom of the table are located statistical results which, for the exclamations variable, revealed no statistically significant differences between treatment means tested. The .5 and .6 levels of significance indicate a 50% chance that any observed difference was due either to sampling error or to chance.

Table 4:15 shows absolute frequencies, percentage of use, and statistical values relating to use of gerunds as sentence beginnings and endings in thirty-four test articles.

Again, article columns identify position of the test variable as either initial or terminal. Alphabetical characters (P, X, B1, B2 or L1) and digits next to them identify works tested as listed in Appendix A. The number of times gerunds are used as first or last words in a sentence by a given author, in a given work, is noted in the frequency columns. Correction for distortions of statistical results caused by

TABLE 4:15. GERUNDS AS SENTENCE BEGINNINGS AND ENDINGS

Article	Frequency	Percent	Article	Frequency	Percent
First P	1	1.190	Last P	0	0.000
First X	2	.110	Last X	2	.110
First B 1	3	.163	Last B 1	5	.272
First B 2	2	.697	Last B 2	0	0.000
First B 3	0	0.000	Last B 3	0	0.000
First B 4	0	0.000	Last B 4	0	0.000
First B 5	0	0.000	Last B 5	0	0.000
First B 6	0	0.000	Last B 6	0	0.000
First B 7	0	0.000	Last B 7	0	0.000
First B 8	0	0.000	Last B 8	0	0.000
First B 9	0	0.000	Last B 9	0	0.000
First L 1	3	.255	Last L 1	0	0.000
First L 2	0	0.000	Last L 2	0	0.000
First L 3	0	0.000	Last L 3	0	0.000
First L 4	0	0.000	Last L 4	1	1.429
First L 5	0	0.000	Last L 5	0	0.000
First L 6	0	0.000	Last L 6	0	0.000
First L 7	0	0.000	Last L 7	1	.719
First L 8	1	1.316	Last L 8	0	0.000
First L 9	0	0.000	Last L 9	0	0.000
First L 10	0	0.000	Last L 10	0	0.000
First L 11	0	0.000	Last L 11	0	0.000
First L 12	0	0.000	Last L 12	0	0.000
First L 13	0	0.000	Last L 13	0	0.000
First L 14	0	0.000	Last L 14	0	0.000
First L 15	0	0.000	Last L 15	0	0.000
First L 16	0	0.000	Last L 16	0	0.000
First L 17	1	.617	Last L 17	0	0.000
First L 18	0	0.000	Last L 18	1	.893
First L 19	0	0.000	Last L 19	0	0.000
First L 20	0	0.000	Last L 20	0	0.000
First L 21	0	0.000	Last L 21	0	0.000
First L 22	0	0.000	Last L 22	0	0.000
First L 23	0	0.000	Last L 23	0	0.000
Mean of /X-B/ is		.156			.116
Mean of /X-L/ is		.176			.213
Calculated F-value is		.052			.968
Level of significance is		.815			.665

differences in article lengths was made by calculating percentage of use figures which are listed in percent columns.

Analysis of variance performed on the mean of transformed variables /X-B/ and /X-L/ yielded no statistically significant differences.

Table 4:16 shows absolute frequencies, percentage of use, and statistical values relating to use of infinitives as sentence beginnings and endings in thirty-four test articles.

As in previous tables presenting STYLBEND results, article columns identify position of test variable as either the first or last word in a sentence. Alphabetical characters (P, X, B1, B2, and L1) and figures next to them coincide with author/work tested list in Appendix A. In frequency columns appear the number of times that infinitives were used in listed works as either sentence beginnings or endings. In order to compensate for different article lengths, which otherwise would cause distortions of statistical results, percentage of use values were calculated and are tabulated in percent columns.

Statistical results for infinitives variable are found at the bottom of the table. As evidenced by calculated F-values of .407 and .199 for infinitives as beginnings and endings respectively, no statistically significant difference exists for this variable.

TABLE 4:16. INFINITIVES AS SENTENCE BEGINNINGS AND ENDINGS

Article	Frequency	Percent	Article	Frequency	Percent
First P	0	0.000	Last P	3	3.571
First X	3	.166	Last X	92	5.080
First B 1	2	.109	Last B 1	105	5.707
First B 2	0	0.000	Last B 2	11	3.833
First B 3	0	0.000	Last B 3	2	4.348
First B 4	0	0.000	Last B 4	3	2.830
First B 5	0	0.000	Last B 5	3	5.455
First B 6	0	0.000	Last B 6	0	0.000
First B 7	0	0.000	Last B 7	3	2.290
First B 8	0	0.000	Last B 8	0	0.000
First B 9	0	0.000	Last B 9	0	0.000
First L 1	2	.170	Last L 1	68	5.782
First L 2	0	0.000	Last L 2	3	1.948
First L 3	0	0.000	Last L 3	2	3.571
First L 4	0	0.000	Last L 4	1	1.429
First L 5	0	0.000	Last L 5	2	4.082
First L 6	0	0.000	Last L 6	1	3.846
First L 7	0	0.000	Last L 7	4	2.878
First L 8	0	0.000	Last L 8	4	5.263
First L 9	0	0.000	Last L 9	4	10.811
First L 10	0	0.000	Last L 10	3	6.250
First L 11	0	0.000	Last L 11	3	6.977
First L 12	0	0.000	Last L 12	0	0.000
First L 13	0	0.000	Last L 13	0	0.000
First L 14	0	0.000	Last L 14	7	5.344
First L 15	1	3.704	Last L 15	3	11.111
First L 16	0	0.000	Last L 16	10	6.452
First L 17	0	0.000	Last L 17	9	5.556
First L 18	0	0.000	Last L 18	1	.893
First L 19	0	0.000	Last L 19	8	5.714
First L 20	0	0.000	Last L 20	10	5.814
First L 21	0	0.000	Last L 21	13	7.514
First L 22	0	0.000	Last L 22	5	5.952
First L 23	0	0.000	Last L 23	1	2.857
Mean of /X-B/ is		.153			2.584
Mean of /X-L/ is		.305			2.252
Calculated F-value is		.407			.199
Level of significance is		.534			.662

Table 4:17 presents absolute frequencies, percentage of use, and statistical test results relating to use of interrogatives as sentence beginnings and endings in thirty-four test articles.

In the article columns of Table 4:17, the position of the test variable as either initial or terminal, and the author/number code of test articles, as identified in Appendix A, are listed. The frequency column contains the number of times in each work tested interrogatives (interrogative adjectives, adverbs, or pronouns) occurred as sentence beginnings and/or endings. In order to compensate for different article lengths, which otherwise would cause distortions of statistical results, percentage of use values or relative frequencies were calculated and are shown in percent columns. None of the test articles produced even one example of an interrogative in the terminal position. STYLBEND was designed to identify and isolate words immediately preceding and following periods. The words thus isolated and punched on IBM cards, were hand sorted and placed into the appropriate part of speech category. Therefore, should an author have concluded a statement with ... comment?, ... combien?, ... pourquoi?, or ... n'est-ce pas?, or should any interrogative term have been used as interjectory interrogatives, they would have been tabulated as sentence endings. One word interjectory interrogatives were to have been counted as both a first and a last word, but none occurred in the 200,000 words of text analyzed.

Analysis of variance statistics which utilized the mean absolute difference between transformed variables (X-B and X-L) generated an F-value of 3.890. The calculated F-value of 3.890 yields a confidence level of .054 in favor of the hypothesis that Bayle wrote l'Avis aux réfugiés. Even though this level of significance suggests a better than

TABLE 4:17. INTERROGATIVES AS SENTENCE BEGINNINGS AND ENDINGS

Article	Frequency	Percent	Article	Frequency	Percent
First P	1	1.190	Last P	0	0.000
First X	49	2.706	Last X	0	0.000
First B 1	55	2.989	Last B 1	0	0.000
First B 2	7	2.439	Last B 2	0	0.000
First B 3	2	4.348	Last B 3	0	0.000
First B 4	0	0.000	Last B 4	0	0.000
First B 5	0	0.000	Last B 5	0	0.000
First B 6	0	0.000	Last B 6	0	0.000
First B 7	0	0.000	Last B 7	0	0.000
First B 8	0	0.000	Last B 8	0	0.000
First B 9	0	0.000	Last B 9	0	0.000
First L 1	20	1.701	Last L 1	0	0.000
First L 2	0	0.000	Last L 2	0	0.000
First L 3	0	0.000	Last L 3	0	0.000
First L 4	0	0.000	Last L 4	0	0.000
First L 5	0	0.000	Last L 5	0	0.000
First L 6	0	0.000	Last L 6	0	0.000
First L 7	2	1.439	Last L 7	0	0.000
First L 8	0	0.000	Last L 8	0	0.000
First L 9	0	0.000	Last L 9	0	0.000
First L 10	0	0.000	Last L 10	0	0.000
First L 11	0	0.000	Last L 11	0	0.000
First L 12	0	0.000	Last L 12	0	0.000
First L 13	0	0.000	Last L 13	0	0.000
First L 14	0	0.000	Last L 14	0	0.000
First L 15	0	0.000	Last L 15	0	0.000
First L 16	0	0.000	Last L 16	0	0.000
First L 17	0	0.000	Last L 17	0	0.000
First L 18	0	0.000	Last L 18	0	0.000
First L 19	0	0.000	Last L 19	0	0.000
First L 20	0	0.000	Last L 20	0	0.000
First L 21	0	0.000	Last L 21	0	0.000
First L 22	0	0.000	Last L 22	0	0.000
First L 23	0	0.000	Last L 23	0	0.000

Mean of /X-B/ is 2.047

Mean of /X-L/ is 2.569

Calculated F-value is 3.890

Level of significance is .054

Statistically significant in favor of /X-B/

94% chance that the observed difference between the treatment means of /X-B/ and /X-L/ was not due to a sampling error or to chance, this tentative conclusion in favor of Bayle was later negated because of the regrouping of the test articles, as shown in Table 4:7.

Table 4:18 records absolute frequencies, percentage of use, and statistical test results relating to use of Latin words as sentence beginnings and endings in thirty-four test works.

Article columns identify the position of the test variable as either initial or terminal, and, by means of denominators P, X, B, and L and numbers adjacent to them, specify the works tested as listed in Appendix A. In frequency columns appear the number of times Latin terms occurred in the corresponding articles as first or last words. Relative frequency or percentage of use figures are listed in the percent column. These figures were used in statistical computations in order to avoid distortions in the analysis of variance results due to differences in article lengths.

At the bottom of the table are located statistical results from ANOVAR used to determine the likelihood that either Bayle or Larroque wrote the disputed article. For the Latin variable, computed F-values of .156 and 3.710 do not reflect a statistically significant difference between Bayle, Larroque and the unknown author at the confidence level required in this study.

Although these tests did not reveal a statistically significant difference between authors X, B, and L's frequency of use of Latin words as noted above, data gathered by the STYLBEND program and reported in Table 4:18 present additional information that might warrant further study. An examination of the frequency columns reveals a similarity in

TABLE 4:18. LATIN WORDS AS SENTENCE BEGINNINGS AND ENDINGS

Article	Frequency	Percent	Article	Frequency	Percent
First P	0	0.000	Last P	0	0.000
First X	7	.387	Last X	13	.718
First B 1	4	.217	Last B 1	12	.652
First B 2	0	0.000	Last B 2	2	.697
First B 3	0	0.000	Last B 3	0	0.000
First B 4	3	2.830	Last B 4	6	5.660
First B 5	0	0.000	Last B 5	0	0.000
First B 6	0	0.000	Last B 6	1	2.778
First B 7	1	.763	Last B 7	4	3.053
First B 8	0	0.000	Last B 8	0	0.000
First B 9	0	0.000	Last B 9	0	0.000
First L 1	4	.340	Last L 1	6	.510
First L 2	0	0.000	Last L 2	0	0.000
First L 3	0	0.000	Last L 3	0	0.000
First L 4	0	0.000	Last L 4	0	0.000
First L 5	0	0.000	Last L 5	0	0.000
First L 6	0	0.000	Last L 6	0	0.000
First L 7	0	0.000	Last L 7	0	0.000
First L 8	0	0.000	Last L 8	0	0.000
First L 9	1	2.703	Last L 9	1	2.703
First L 10	0	0.000	Last L 10	0	0.000
First L 11	0	0.000	Last L 11	0	0.000
First L 12	0	0.000	Last L 12	0	0.000
First L 13	0	0.000	Last L 13	0	0.000
First L 14	1	.763	Last L 14	2	1.527
First L 15	0	0.000	Last L 15	0	0.000
First L 16	3	1.935	Last L 16	2	1.290
First L 17	0	0.000	Last L 17	0	0.000
First L 18	0	0.000	Last L 18	0	0.000
First L 19	0	0.000	Last L 19	0	0.000
First L 20	0	0.000	Last L 20	0	0.000
First L 21	0	0.000	Last L 21	1	.578
First L 22	0	0.000	Last L 22	0	0.000
First L 23	0	0.000	Last L 23	0	0.000
Mean of /X-B/ is		.589			1.366
Mean of /X-L/ is		.505			.723
Calculated F-value is		.156			3.710
Level of significance is		.697			.060

the relative and absolute frequency of use of the test variable by the author of l'Avis aux réfugiés and by Bayle. Author X introduced seven sentences with Latin terms, while terminating thirteen in the same fashion. Likewise, in article B1, Bayle began four and ended twelve sentences with Latin. Using STYLBEND grouped-data listings, Latin quotations were readily located in context. A comparison of their use revealed that of seven Latin terms used by author X as beginnings, six are short quotations which occupy the entire sentence, while one is part of a phrase used to introduce an argument. All four of the Latin beginnings in article B1 introduce short, full-sentence quotations. Seven additional latinate constructions in X are parts of final quotations or phrases used to conclude an argument. Likewise, Bayle's introduction of eight additional quotations to substantiate his arguments parallels the pattern established in l'Avix aux réfugiés.

Table 4:19 presents absolute frequencies, percentage of use, and statistical values relating to use of negative forms as sentence beginnings and endings in thirty-four test articles.

As in previous tables presenting STYLBEND results, article columns identify position of the test variable as either the first or last word in a sentence. Denominators P, X, B, and L and digits next to them coincide with the author/work list in Appendix A. In frequency columns appear the number of times that negative forms were used in listed works as either sentence beginnings or endings. In order to compensate for different article lengths which would otherwise cause distortions of statistical results, percentage of use values were calculated and are tabulated in percent columns.

TABLE 4:19. NEGATIONS AS SENTENCE BEGINNINGS AND ENDINGS

Article	Frequency	Percent	Article	Frequency	Percent
First P	0	0.000	Last P	0	0.000
First X	36	1.988	Last X	22	1.215
First B 1	39	2.120	Last B 1	38	2.065
First B 2	11	3.833	Last B 2	0	0.000
First B 3	2	4.348	Last B 3	2	4.348
First B 4	0	0.000	Last B 4	1	.943
First B 5	2	3.636	Last B 5	1	1.818
First B 6	1	2.778	Last B 6	0	0.000
First B 7	1	.763	Last B 7	0	0.000
First B 8	0	0.000	Last B 8	0	0.000
First B 9	0	0.000	Last B 9	1	4.762
First L 1	22	1.871	Last L 1	23	1.956
First L 2	0	0.000	Last L 2	0	0.000
First L 3	1	1.786	Last L 3	1	1.786
First L 4	0	0.000	Last L 4	0	0.000
First L 5	0	0.000	Last L 5	0	0.000
First L 6	0	0.000	Last L 6	0	0.000
First L 7	0	0.000	Last L 7	2	1.439
First L 8	1	1.316	Last L 8	2	2.632
First L 9	0	0.000	Last L 9	0	0.000
First L 10	0	0.000	Last L 10	1	2.083
First L 11	0	0.000	Last L 11	0	0.000
First L 12	0	0.000	Last L 12	1	2.857
First L 13	0	0.000	Last L 13	1	1.587
First L 14	0	0.000	Last L 14	0	0.000
First L 15	1	3.704	Last L 15	3	11.111
First L 16	1	.645	Last L 16	2	1.290
First L 17	2	1.235	Last L 17	0	0.000
First L 18	0	0.000	Last L 18	1	.893
First L 19	0	0.000	Last L 19	2	1.429
First L 20	0	0.000	Last L 20	3	1.744
First L 21	0	0.000	Last L 21	1	.578
First L 22	0	0.000	Last L 22	0	0.000
First L 23	0	0.000	Last L 23	0	0.000
Mean of /X-B/ is		1.551			1.473
Mean of /X-L/ is		1.678			1.289
Calculated F-value is		.254			.071
Level of significance is		.623			.786

Statistical results for the negations variable are found at the bottom of the table. Using mean absolute differences of /X-B/ and /X-L/ the computer calculated an F-value of .254 and .071 for first and last words respectively. For the negations variable, observed differences, represented by the tabulated F-values, were not statistically significant enough to formulate a hypothesis of attribution to either Bayle or Larroque.

Table 4:20 lists absolute frequencies, percentage of use, and statistical values relating to use of nouns as sentence beginnings and endings in thirty-four test articles.

The first column of Table 4:20 identifies the test variable's position as either initial or terminal and denotes the authors and works tested as listed in Appendix A. The number of times that nouns were used in test articles as sentence beginnings and endings appears in frequency columns. In order to correct for distortions of statistical results due to differences in article lengths, percentage of use or relative frequency values were calculated and are presented in percent columns.

After the first and last word values for article L23, statistical results for all figures in Table 4:20 are summarized. Again, the analysis of variance program calculated the F-statistic. Computed F-values of .112 and 2.85 are not statistically significant and therefore do not suggest attribution to either Bayle or Larroque for variables represented in Table 4:20.

TABLE 4:20. NOUNS AS SENTENCE BEGINNINGS AND ENDINGS

Article	Frequency	Percent	Article	Frequency	Percent
First P	0	0.000	Last P	46	54.762
First X	20	1.104	Last X	1093	60.353
First B 1	10	.543	Last B 1	1050	57.065
First B 2	4	1.394	Last B 2	186	64.808
First B 3	0	0.000	Last B 3	24	52.174
First B 4	5	4.717	Last B 4	66	62.264
First B 5	3	5.455	Last B 5	34	61.818
First B 6	1	2.778	Last B 6	20	55.556
First B 7	2	1.527	Last B 7	83	63.359
First B 8	0	0.000	Last B 8	17	62.963
First B 9	0	0.000	Last B 9	13	61.905
First L 1	28	2.381	Last L 1	648	55.102
First L 2	3	1.948	Last L 2	98	63.636
First L 3	1	1.786	Last L 3	42	75.000
First L 4	0	0.000	Last L 4	44	62.857
First L 5	0	0.000	Last L 5	34	69.388
First L 6	0	0.000	Last L 6	17	65.385
First L 7	9	6.475	Last L 7	80	57.554
First L 8	5	6.579	Last L 8	49	64.474
First L 9	0	0.000	Last L 9	13	35.135
First L 10	1	2.083	Last L 10	24	50.000
First L 11	0	0.000	Last L 11	29	67.442
First L 12	0	0.000	Last L 12	21	60.000
First L 13	2	3.175	Last L 13	29	46.032
First L 14	0	0.000	Last L 14	74	56.489
First L 15	0	0.000	Last L 15	11	40.741
First L 16	0	0.000	Last L 16	96	61.935
First L 17	8	4.938	Last L 17	92	56.790
First L 18	0	0.000	Last L 18	73	65.179
First L 19	2	1.429	Last L 19	77	55.000
First L 20	2	1.163	Last L 20	105	61.047
First L 21	15	8.671	Last L 21	123	71.098
First L 22	0	0.000	Last L 22	52	61.905
First L 23	0	0.000	Last L 23	24	68.571
Mean of /X-B/ is		1.580			3.473
Mean of /X-L/ is		1.814			7.131
Calculated F-value is		.112			2.857
Level of significance is		.738			.097

Table 4:21 shows absolute frequencies, percentage of use, and statistical values relating to use of numbers as sentence beginnings and endings in thirty-four test articles.

The article column of Table 4:21 identifies the test variable's position as the first or last element of a sentence. Denominators P, X, B, and L and numbers adjacent to them coincide with the author/work list in Appendix A. In frequency columns appear the number of times that numbers occurred in listed works as sentence beginnings or endings. Percentage of use values were calculated to correct for distortions of statistical results due to differences in article lengths and are tabulated in percent columns.

At the bottom of Table 4:21 are found statistical results for all values in the table. Using mean absolute differences of /X-B/ and /X-L/, the computer calculated F-statistics of 1.539 and 1.263 for numbers as sentence first and last words respectively. Falling well below the F or 4.17 required for statistical significance, 1.539 and 1.263 figures do not suggest with any degree of certainty that either Bayle or Larroque wrote the disputed work.

TABLE 4:21. NUMBERS AS SENTENCE BEGINNINGS AND ENDINGS

Article	Frequency	Percent	Article	Frequency	Percent
First P	1	1.190	Last P	2	2.381
First X	37	2.043	Last X	20	1.104
First B 1	20	1.087	Last B 1	20	1.087
First B 2	1	.348	Last B 2	2	.697
First B 3	4	8.696	Last B 3	0	0.000
First B 4	0	0.000	Last B 4	1	.943
First B 5	0	0.000	Last B 5	1	1.818
First B 6	0	0.000	Last B 6	2	5.556
First B 7	8	6.107	Last B 7	7	5.344
First B 8	0	0.000	Last B 8	1	3.704
First B 9	0	0.000	Last B 9	1	4.762
First L 1	17	1.446	Last L 1	16	1.361
First L 2	0	0.000	Last L 2	0	0.000
First L 3	0	0.000	Last L 3	1	1.786
First L 4	0	0.000	Last L 4	0	0.000
First L 5	0	0.000	Last L 5	0	0.000
First L 6	0	0.000	Last L 6	0	0.000
First L 7	25	17.986	Last L 7	1	.719
First L 8	0	0.000	Last L 8	2	2.632
First L 9	0	0.000	Last L 9	0	0.000
First L 10	6	12.500	Last L 10	0	0.000
First L 11	18	41.860	Last L 11	1	2.326
First L 12	0	0.000	Last L 12	3	8.571
First L 13	15	23.810	Last L 13	2	3.175
First L 14	0	0.000	Last L 14	0	0.000
First L 15	4	14.815	Last L 15	0	0.000
First L 16	0	0.000	Last L 16	1	.645
First L 17	8	4.938	Last L 17	0	0.000
First L 18	8	7.143	Last L 18	0	0.000
First L 19	0	0.000	Last L 19	1	.714
First L 20	15	8.721	Last L 20	1	.581
First L 21	0	0.000	Last L 21	1	.578
First L 22	1	1.190	Last L 22	0	0.000
First L 23	3	8.571	Last L 23	0	0.000
Mean of /X-B/ is		2.620			1.927
Mean of /X-L/ is		6.431			1.250
Calculated F-value is		1.539			1.263
Level of significance is		.222			.269

Table 4:22 gives absolute frequencies, percentage of use, and statistical values relating to use of prepositions as sentence beginnings and endings in thirty-four test articles.

As in previous tables presenting STYLBEND results, article columns identify the position of the variable as either initial or terminal. Alphabetical characters, P, X, B, or L, and numbers next to them identify works tested as listed in Appendix A. In frequency columns appear the number of times a preposition is used in its corresponding article as either a sentence beginning or ending. Percentage of use values were calculated in order to correct for distortions of statistical results due to differences in article lengths and are tabulated in percent columns.

Because rules of French syntax regarding position of the preposition are even more rigid than those governing conjunctions, it would seem natural to expect zero values in all positions of the last word section of Table 4:22. Nevertheless, four such occurrences are recorded in l'Avis aux réfugiés. STYLBEND grouped-data listing shows that the words in question are located in lines 1077, 1512, 4548 and 3450 of the basic text as shown below:

- X1077 distinguo, Les rois sont-ils dépendans de Dieu
 seul. (?)¹⁴ C'est selon. (:) Si ... Si ...
- X1512 encore mieux peu de tems après. Ils regardent
 assez long-tems ...
- X4548 universitez et des parlemens, fut cassé bien-tôt
 après. En sorte que ...
- X3450 ... ou de n'avoir pas eu la force de crier contre.
 On ne peut ...

Of four cases quoted, the first, although a preposition, is used as a

TABLE 4:22. PREPOSITIONS AS SENTENCE BEGINNINGS AND ENDINGS

Article	Frequency	Percent	Article	Frequency	Percent
First P	2	2.381	Last P	0	0.000
First X	135	7.454	Last X	4	.221
First B 1	136	7.391	Last B 1	0	0.000
First B 2	35	12.195	Last B 2	0	0.000
First B 3	5	10.870	Last B 3	0	0.000
First B 4	7	6.604	Last B 4	0	0.000
First B 5	3	5.455	Last B 5	0	0.000
First B 6	0	0.000	Last B 6	0	0.000
First B 7	8	6.107	Last B 7	0	0.000
First B 8	0	0.000	Last B 8	0	0.000
First B 9	4	19.048	Last B 9	0	0.000
First L 1	93	7.908	Last L 1	0	0.000
First L 2	17	11.039	Last L 2	0	0.000
First L 3	2	3.571	Last L 3	0	0.000
First L 4	7	10.000	Last L 4	0	0.000
First L 5	6	12.245	Last L 5	0	0.000
First L 6	4	15.385	Last L 6	0	0.000
First L 7	9	6.475	Last L 7	0	0.000
First L 8	6	7.895	Last L 8	0	0.000
First L 9	0	0.000	Last L 9	0	0.000
First L 10	7	14.583	Last L 10	0	0.000
First L 11	9	20.930	Last L 11	0	0.000
First L 12	3	8.571	Last L 12	0	0.000
First L 13	7	11.111	Last L 13	0	0.000
First L 14	7	5.344	Last L 14	0	0.000
First L 15	1	3.704	Last L 15	0	0.000
First L 16	16	10.323	Last L 16	0	0.000
First L 17	7	4.321	Last L 17	0	0.000
First L 18	14	12.500	Last L 18	0	0.000
First L 19	13	9.286	Last L 19	0	0.000
First L 20	18	10.465	Last L 20	0	0.000
First L 21	13	7.514	Last L 21	0	0.000
First L 22	15	17.857	Last L 22	0	0.000
First L 23	3	8.571	Last L 23	0	0.000
Mean of /X-B/ is			4.324		
Mean of /X-L/ is			3.946		
Calculated F-value is			.075		
Level of significance is			.781		

coordinating conjunction, having a series of "if" statements following it. If the original punctuation had been retained, the particular occurrence would not have been counted, but because all colons were key-punched as periods, and because periods were recognized by the computer as sentence delimiters, this somewhat irregular usage was isolated and counted.

The second and third "prepositions" are in reality used as adverbs and may be classed as such. However, their usage when coupled with a second adverb of time (peu de temps, bientôt) forces the reader to supply an object, even if this object does consist of an idea or a series of events.

The final term classed as a preposition appears to be a true breach of accepted French syntax, for once again the reader is forced to supply an object. Although each of these usages, excepting perhaps the fourth, might be classified as a different part of speech, the fact remains that common ground rules were established before the data were run, and X was the only one of the authors tested to use these forms in the ways described.

At the bottom of Table 4:22 are found the statistical results computed by the analysis of variance program. Once again the results fail to show statistical evidence suggesting either Bayle or Larroque as the probable author of the disputed text. No values are printed for the last-word column because the great number of zero occurrences made further calculation impossible.

Table 4:23 shows absolute frequencies, percentage of use, and statistical values relating to use of pronouns as sentence beginnings and endings in thirty-four test articles.

TABLE 4:23. PRONOUNS AS SENTENCE BEGINNINGS AND ENDINGS

Article	Frequency	Percent	Article	Frequency	Percent
First P	44	52.381	Last P	2	2.381
First X	531	29.321	Last X	61	3.368
First B 1	681	37.011	Last B 1	88	4.783
First B 2	82	28.571	Last B 2	9	3.136
First B 3	23	50.000	Last B 3	4	8.696
First B 4	42	39.623	Last B 4	2	1.887
First B 5	21	38.182	Last B 5	1	1.818
First B 6	18	50.000	Last B 6	2	5.556
First B 7	62	47.328	Last B 7	2	1.527
First B 8	10	37.037	Last B 8	0	0.000
First B 9	11	52.381	Last B 9	1	4.762
First L 1	355	30.187	Last L 1	82	6.973
First L 2	51	33.117	Last L 2	7	4.545
First L 3	26	46.429	Last L 3	0	0.000
First L 4	22	31.429	Last L 4	1	1.429
First L 5	16	32.653	Last L 5	2	4.082
First L 6	8	30.769	Last L 6	1	3.846
First L 7	39	28.058	Last L 7	9	5.475
First L 8	14	18.421	Last L 8	3	3.947
First L 9	18	48.649	Last L 9	0	0.000
First L 10	13	27.083	Last L 10	3	6.250
First L 11	9	20.930	Last L 11	0	0.000
First L 12	10	28.571	Last L 12	2	5.714
First L 13	17	26.984	Last L 13	4	6.349
First L 14	67	51.145	Last L 14	7	5.344
First L 15	6	22.222	Last L 15	1	3.704
First L 16	70	45.161	Last L 16	7	4.516
First L 17	58	35.802	Last L 17	7	4.321
First L 18	39	34.821	Last L 18	3	2.679
First L 19	50	35.714	Last L 19	4	2.857
First L 20	54	31.395	Last L 20	7	4.070
First L 21	56	32.370	Last L 21	4	2.312
First L 22	20	23.810	Last L 22	2	2.381
First L 23	11	31.429	Last L 23	3	8.571
Mean of /X-B/ is		13.082			2.088
Mean of /X-L/ is		6.510			1.890
Calculated F-value is		6.510			.133
Level of significance is		.017			.718
Statistically significant in favor of /X-L/					

Article columns of Table 4:23 identify the test variable's position as either a sentence beginning or ending. Denominators P, X, B, or L and numbers adjacent to them coincide with the author/work list in Appendix A. Frequencies with which pronouns occurred in listed works as sentence initiators or terminators appear in frequency columns. In order to correct for distortions of statistical results due to differences in article lengths, percentage of use or relative frequency values were calculated and are presented in percent columns.

Following tabulated values for L23 are found ANOVAR results. The mean absolute differences /X-B/ and /X-L/ were used to calculate an F-value of 6.255 for the sentence beginning statistics and a .133 figure for sentence endings. The 6.255 value is statistically significant at the .01 level in favor of the conclusion that Larroque wrote the disputed text. On the other hand, the .133 value for pronouns used as sentence endings is not significant enough to suggest a probable author. Unlike other significant ANOVAR results, the conclusion suggested in Table 4:23 was confirmed, rather than rejected after the regrouping of articles noted in Table 4:7.

Table 4:24 lists absolute frequencies, percentage of use, and statistical values relating to use of conjugated verbs as sentence beginnings and endings in thirty-four test articles.

Table 4:24 is the final table which lists values for all thirty-four test articles. As in Tables 4:10 through 4:23, article columns identify the test variable's position as either initial or terminal. Letters of the alphabet, P, X, B, or L, and numbers next to them coincide with the author/work list in Appendix A. The number of times that conjugated verbs occurred as sentence beginnings or endings in the

TABLE 4:24. CONJUGATED VERBS AS SENTENCE BEGINNINGS AND ENDINGS

Article	Frequency	Percent	Article	Frequency	Percent
First P	1	1.190	Last P	15	17.857
First X	121	6.681	Last X	148	8.172
First B 1	59	3.207	Last B 1	162	8.804
First B 2	26	9.059	Last B 2	18	6.272
First B 3	1	2.174	Last B 3	1	2.174
First B 4	8	7.547	Last B 4	9	8.491
First B 5	1	1.818	Last B 5	6	10.909
First B 6	1	2.778	Last B 6	2	5.556
First B 7	3	2.290	Last B 7	11	8.397
First B 8	0	0.000	Last B 8	3	11.111
First B 9	0	0.000	Last B 9	1	4.762
First L 1	45	3.827	Last L 1	100	8.503
First L 2	1	.649	Last L 2	12	7.792
First L 3	0	0.000	Last L 3	1	1.786
First L 4	0	0.000	Last L 4	5	7.143
First L 5	0	0.000	Last L 5	1	2.041
First L 6	0	0.000	Last L 6	2	7.692
First L 7	5	3.597	Last L 7	11	7.914
First L 8	2	2.632	Last L 8	6	7.895
First L 9	0	0.000	Last L 9	4	10.811
First L 10	0	0.000	Last L 10	3	6.250
First L 11	0	0.000	Last L 11	5	11.628
First L 12	1	2.857	Last L 12	1	2.857
First L 13	0	0.000	Last L 13	11	17.460
First L 14	0	0.000	Last L 14	8	6.107
First L 15	0	0.000	Last L 15	3	11.111
First L 16	1	.645	Last L 16	4	2.581
First L 17	1	.617	Last L 17	11	6.790
First L 18	0	0.000	Last L 18	8	7.143
First L 19	0	0.000	Last L 19	10	7.143
First L 20	2	1.163	Last L 20	4	2.326
First L 21	0	0.000	Last L 21	6	3.468
First L 22	0	0.000	Last L 22	9	10.714
First L 23	0	0.000	Last L 23	3	8.571
Mean of /X-B/ is		4.194			2.308
Mean of /X-L/ is		5.986			2.844
Calculated F-value is		10.043			.329
Level of significance is		.003			.576
Statistically significant in favor of /X-B/					

listed works appear in frequency columns. The final columns, denominated "Percent," list percentage of use or relative frequencies which were calculated and used in order to correct for distortions of statistical results due to differences in article lengths.

At the bottom of the table are presented statistical results as computed by the ANOVAR program. The calculated F-value of 10.043 represents a confidence level of .003 in favor of the conclusion that Bayle wrote l'Avis aux réfugiés. Like other tentative conclusions, the one drawn in Table 4:24 was negated by a regrouping of test articles as shown in Table 4:7.

Table 4:25 summarizes percentage of use of each of the fifteen parts of speech as sentence beginnings in the final five test articles.

The part of speech column lists fifteen part of speech categories used as sentence initiators. The denominators P, X, B1, B2 and L1 symbolize the preface, l'Avis aux réfugiés, Bayle 1, Bayle 2, and Larroque 1 as listed in Appendix A. Figures in columns P, X, B1, B2 and L1 correspond to the percentage of sentences in a given work which begin with one of the fifteen parts of speech listed. For example, the author of the preface began 4.762 percent of his sentences with an adjective, none with infinitives, and 52.381 percent with pronouns.

The fact that all column totals do not equal one hundred percent is due to a computer rounding error and is of no significance.

TABLE 4:25. PERCENTAGE OF USE SUMMARY FOR SENTENCE BEGINNINGS

Part of Speech	P	X	B1	B2	L1
adjectives	4.762	4.528	5.435	3.136	5.102
adverbs	1.190	2.430	1.413	3.136	2.211
articles	14.286	6.957	8.152	5.575	8.673
conjunctions	20.238	33.849	29.457	28.920	32.568
exclamations	0.000	.276	.707	.697	1.361
gerunds	1.190	.110	.163	.697	.255
infinitives	0.000	.166	.109	0.000	.170
interrogatives	1.190	2.706	2.989	2.439	1.701
Latin	0.000	.387	.217	0.000	.340
negations	0.000	1.988	2.120	3.833	1.871
nouns	0.000	1.104	.543	1.394	2.381
numbers	1.190	2.043	1.087	.348	1.446
prepositions	2.381	7.454	7.391	12.195	7.908
pronouns	52.381	29.321	37.011	28.571	30.187
verbs	1.190	6.681	3.201	9.059	3.827
Totals	99.998	100.000	99.995	99.960	100.001

Table 4:26 summarizes percentage of use of each of the fifteen parts of speech as sentence endings in the final five test articles.

The part of speech column lists the fifteen part of speech categories used as sentence terminators. The denominators P, X, B1, B2 and L1 symbolize the preface, l'Avis aux réfugiés, Bayle 1, Bayle 2, and Larroque 1 as listed in Appendix A. In columns P, X, B1, B2, and L1 are listed the percentage of sentences in a given work which end with

TABLE 4:26. PERCENTAGE OF USE SUMMARY FOR SENTENCE ENDINGS

Part of Speech	P	X	B1	B2	L
adjectives	11.905	15.406	13.804	15.679	15.561
adverbs	4.762	3.644	4.076	4.181	3.486
articles	0.000	0.000	0.000	0.000	0.000
conjunctions	2.381	.607	1.250	.697	.425
exclamations	0.000	0.000	.435	0.000	.340
gerunds	0.000	.110	.272	0.000	0.000
infinitives	3.571	5.080	5.707	3.833	5.782
interrogatives	0.000	0.000	0.000	0.000	0.000
Latin	0.000	.718	.652	.697	.510
negations	0.000	1.215	2.065	0.000	1.956
nouns	54.762	60.353	57.065	64.808	55.102
numbers	2.381	1.104	1.087	.697	1.361
prepositions	0.000	.221	0.000	0.000	0.000
pronouns	2.381	3.368	4.783	3.136	6.973
verbs	17.857	8.172	8.804	6.272	8.503
Totals	100.000	99.998	100.000	100.000	99.999

adjectives, adverbs, conjunctions, and other parts of speech as listed. The author of l'Avis aux réfugiés, for example, terminated 15.406 percent of his sentences with adjectives, .607 percent with conjunctions, none with exclamations, 60.353 percent with nouns, 3.368 percent with pronouns, and 8.172 percent with finite verbs. Bayle, on the other hand, recorded percentages of 13.804, 1.250, .435, 57.065, 4.783, and 8.804 for the same variables. The significance of these differences is

discussed with the statistical t-test results reported after Table 4:28.

The fact that all column totals do not equal one-hundred is again due to a computer rounding error and is insignificant.

Statistical Summary of Sentence Beginning Variables

The information in Tables 4:25 and 4:26 served as input data for the final analysis of STYLBEND variables. Analysis of variance performed on transformed variables /X-B/ and /X-L/ indicated which of the two authors (B or L) was closer to X in his use of a particular stylistic element. Since the format of the data after the articles were regrouped did not permit further use of the analysis of variance program, and since the t-test had proven effective in observing differences between expository and literary criticisms, it was adopted as the statistical measure for the remainder of the study.

Tables 4:10 through 4:24 provided the data shown in Tables 4:25 and 4:26. The values in these two tables were essential to performing t-tests on the STYLBEND variables across the final five test articles. A summary of t-values and confidence levels relating to use of STYLBEND variables as sentence beginnings comprises Table 4:27.

The first column of Table 4:27 lists the fifteen part of speech categories used in STYLBEND analyses. Column headings X:P, X:B1, X:B2, X:L1, and B1:L1 symbolize author/article combinations for which the hypothesis of no difference was tested. Figures in each of these columns are the calculated t-values. A t-value of greater than 1.96 means that a significant difference does exist between the two test authors' quantitative use of the variable in question. On the other hand, a non-significant t-value--smaller than 1.96--prescribes that the

TABLE 4:27. SUMMARY OF SENTENCE BEGINNINGS

Part of Speech	X:P	CL	X:B1	CL	X:B2	CL	X:L1	CL	B1:L1	CL	Prob. Auth
adjectives	.1007	-	1.2589	-	1.0756	-	.7205	-	.3976	-	-
adverbs	.7292	-	2.2395	.025	.7084	-	.3860	-	1.6417	-	-
articles	2.5268	.025	1.3653	-	.8663	-	1.7256	-	.5046	-	-
conjunctions	2.5886	.01	2.8533	.005	1.6477	-	.7254	-	1.8074	-	-
exclamations	.4822	-	1.8566	-	1.1485	-	3.4657	.001	1.7952	-	-
gerunds	2.4340	.025	.4298	-	2.1159	.05	.9449	-	.5534	-	-
infinitives	.3733	-	.4653	-	.6900	-	.0288	-	.4517	-	-
interrogatives	.8470	-	.5147	-	.2604	-	1.7864	-	2.2162	.05	-
Latin	.5709	-	.9323	-	1.0550	-	.2045	-	.6392	-	-
negations	1.3047	-	.2805	-	1.9622	.05	.2266	-	.4734	-	-
nouns	.9683	-	1.8770	-	.4283	-	2.7109	.01	4.4126	.001	B
numbers	.5449	-	2.3300	.025	2.0001	.05	1.1975	-	.8726	-	-
prepositions	1.7553	-	.0728	-	2.7344	.01	.4563	-	.5227	-	-
pronouns	4.4941	.001	4.9334	.001	.2594	-	.5064	-	3.8489	.001	L
verbs	2.0045	.05	4.8490	.001	1.4662	-	3.3274	.001	.9102	-	N

hypothesis of identical style should not be rejected. The presence of a non-significant t-value between X and B1, for example, does not affirm that their styles are, in fact, identical, but merely that the observed difference in quantitative use of the variable in question was not great enough to distinguish between them at the level of confidence prescribed in this study. For each significant t-value a corresponding confidence level (CL) has been listed. A dash (-) in the CL columns indicates a non-significant t-value.

Conclusions on Sentence Beginnings

The last column of Table 4:27 suggests a probable author for those variables which have significantly different values. Conditions which must be satisfied in order to name a probable author were delineated on pages 93-94 of this chapter. A dash (-) in this column means that no statistically significant differences were observed, or that prescribed conditions for attribution were not satisfied. One glance at the probable author column reveals that, considered individually, the results of the sentence-beginning-variables tests were anything but conclusive: one variable, nouns, suggested Bayle as the probable author; one, pronouns, suggested Larroque; and a third, finite verbs, indicated that neither Bayle nor Larroque wrote l'Avis aux réfugiés. At first these results seemed disturbing, especially in view of the fact that the ANOVAR results had earlier shown five variables with significantly distinguishing values (see Tables 4:10, 4:13, 4:17, 4:23, and 4:24). Analysis of variance performed before the test articles were regrouped according to genres, suggested that Larroque's quantitative use of adjectives, conjunctives, and pronouns in the

initial sentence position resembled that of l'Avis. Interrogative and finite verb use, on the other hand, pointed to Bayle as its author. Only one of these five variables, pronouns, remained unchanged by the regrouping of test articles.

Statistical Summary of Sentence Ending Variables

Table 4:28 displays t-values, confidence levels and probable author designations for each of the fifteen part of speech categories used as sentence endings. The columnar structure of the table is the same as Table 4:27, where the first column lists the part of speech categories, and the columns headed X:P, X:B1, X:B2, X:L1, and B1:L1, symbolize the author/article combinations for which the hypothesis of no difference was tested. Again, figures in each of these columns are calculated t-values for corresponding variables. Next to each significant t-value (greater than 1.96) is listed the confidence level (CL), which represents the degree of probability that the observed difference was due to sampling error or chance. Where no significant difference exists, a dash (-) is noted. As in Table 4:27, the final column lists a probable author for those variables demonstrating required significantly different values. Conditions established on pages 93 and 94 of this chapter must be satisfied in order to name Bayle (B), Larroque (L), or neither Bayle nor Larroque (N) as the probable author. A dash (-) in this column means that there were no statistically significant differences observed or that prescribed conditions for attribution were not satisfied. Even though author X ended two percent more of his sentences with adjectives than did Bayle, and almost four percent more than the author of the preface, these differences were not great enough to

TABLE 4:28. SUMMARY OF SENTENCE ENDINGS

Part of Speech	X:P	CL	X:B1	CL	X:B2	CL	X:L1	CL	BI:L1	CL	Prob. Auth
adjectives	.8725	-	1.3703	-	.1192	-	.1147	-	1.3369	-	-
adverbs	.5309	-	.6768	-	.4466	-	.2270	-	.8213	-	-
articles	0.0000	-	0.0000	-	0.0000	-	0.0000	-	0.0000	-	-
conjunctions	1.9252	-	2.0212	.05	.1794	-	.6666	-	2.3037	.025	L
exclamations	0.0000	-	2.8091	.005	0.0000	-	2.4836	.025	.4027	-	N
gerunds	.3047	-	1.1140	-	.5633	-	1.1400	-	1.7891	-	-
infinitives	.6194	-	.8377	-	.9086	-	.8328	-	.0873	-	-
interrogatives	0.0000	-	0.0000	-	0.0000	-	0.0000	-	0.0000	-	-
Latin	.7792	-	.2405	-	.0392	-	.6974	-	.4937	-	-
negations	1.0161	-	2.0208	.05	1.8771	-	1.6243	-	.2082	-	-
nouns	1.0231	-	2.0175	.05	1.4374	-	2.8438	.005	1.0601	-	N
numbers	1.0678	-	.0505	-	.6297	-	.6269	-	.6748	-	-
prepositions	.4312	-	2.0171	.05	.7969	-	1.6127	-	0.0000	-	-
pronouns	.4934	-	2.1596	.05	.2037	-	4.5081	.001	2.5437	.025	N
verbs	3.0948	.005	.6850	-	1.1082	-	.3204	-	.2862	-	-

generate a high enough t-value to yield the acceptable .05 confidence level.

Conclusions on Sentence Endings

Four variables were, however, used with a quantitative difference great enough to produce attributable values: conjunctions, exclamations, nouns, and pronouns. Because Bayle's use of conjunctive forms differed significantly from both Larroque's and the unknown author's, Larroque was named as probable author for this variable. The remaining three significant values, exclamations, nouns, and pronouns, suggest that neither Bayle nor Larroque wrote l'Avis aux réfugiés. In addition to those variables which satisfied the probable author conditions, two additional variables, negations and prepositions, provided t-values significant at the .05 level. These results also imply a difference between Bayle's writing habits and those of the unknown author.

An overall summary of sentence beginnings and endings variables shows one characteristic in favor of Bayle as author of l'Avis, two in favor of Larroque, and four in favor of neither Bayle nor Larroque. In addition, a total of five sentence beginning (adverbs, conjunctions, numbers, pronouns and finite verbs) and six sentence ending variables (conjunctions, exclamations, negations, nouns, prepositions, and pronouns), or eleven of the thirty variables, showed a statistically significant difference between Bayle and author X.

Function Word Frequencies and Usage

Characteristic of all authorship attribution studies is the examination of vocabulary. During the past decade of computer-aided literary and linguistic studies, scholars have distinguished between words which are generally semantic bound--content words--and those which have no significant meaning--function words. Mosteller and Wallace provided a "general purpose list" of function words for the English language.¹⁵ Since no such list was available for French at the onset of this project, it was necessary to construct one. The most logical approach seemed to merely transform Mosteller and Wallace's list into a French equivalent. However, since several words, e.g., "any," "down," "only," "shall," "should," "than," and "would," are not as orthographically definitive in French as in English, the statisticians' list provided only a foundation (although, admittedly a very helpful one) upon which the final list of 199 French function words was constructed.¹⁶ Absolute frequencies for each of the final test words were generated and are reported in Appendix K. Tables 4:29 through 4:41 present the results obtained by comparing the 199 words across five final test articles. For each word, ten comparisons were made although only the results which most directly led to naming a probable author under the conditions prescribed on pages 93 and 94 are tabulated in this chapter.¹⁷

The tables presenting FREQFUN results have been directed into two categories: individual words and grouped words. The first category comprises Tables 4:29 through 4:36. Tables 4:37 through 4:41 contain grouped-word results.

Table 4:29 presents a list of twenty-two words whose quantitative use in the test articles suggests Larroque as probable author of l'Avis aux réfugiés.

In order to qualify for this table each of the twenty-two words had to pass the hypothesis-of-no-difference test. That is to say, no statistically significant difference was observed between their quantitative use in l'Avis and Le Prosélyte. In addition to this screening test, which implies, but does not prove similarity, each word had to satisfy conditions prescribed earlier (pp. 93-94) for advancing Larroque as probable author. Words meeting these requirements are listed alphabetically in the column denominated "Word." Figures in B1 \neq L1 and X \neq B1 columns are calculated t-values. Those in CL columns represent corresponding confidence levels or degrees of probability that the observed difference was due merely to sampling error and not to stylistic differences. Again it was rewarding to note that, whereas the generally accepted confidence level is .05, ninety percent of the values reported in Table 4:29 yielded significance levels of better than .03.

TABLE 4:29. FUNCTION WORDS, X = L1

Word	B1 \neq L1	CL	X \neq B1	CL
à	2.496	.025	2.823	.01
ainsi	3.312	.001	2.217	.05
aller - all	4.116	.001	4.686	.001
avoir	3.690	.001	5.198	.001
déjà	2.592	.01	2.077	.05
depuis	3.667	.001	3.892	.001
devoir - cond. only	1.972	.05	2.585	.01
donner	3.062	.005	4.190	.001
en	3.801	.001	2.935	.005
enfant	3.737	.001	2.151	.05
etc.	2.930	.005	2.676	.01
faire	3.080	.005	3.719	.001
falloir	3.108	.005	2.434	.025
force	2.238	.05	3.643	.001
le	3.460	.001	4.193	.001
ni	2.484	.025	2.104	.05
non	3.093	.005	2.556	.025
point - neg.	3.824	.001	2.927	.005
pouvoir	3.013	.005	3.635	.001
prendre	2.681	.01	2.877	.005
quand	3.382	.001	2.900	.005
venir	2.480	.025	2.642	.01

Table 4:30 lists thirty-two words whose quantitative use in test articles suggests Bayle as probable author of l'Avis aux réfugiés.

As in Table 4:29, each of the words listed in the "Word" column had to satisfy three conditions in order to qualify for the $X = B1$ list. Symbolically, these conditions are $X = B1$, $B1 \neq L1$, and $X \neq L1$. $X = B1$ means that, when the word's quantitative use in l'Avis (X) was compared to its use in Cabale (B1), no significant difference was observed. The second condition required that there be a significant difference between the words used in Cabale (B1) and in Le Prosélyte (L1), and, finally, if its use in l'Avis differed significantly from that of Le Prosélyte, the word, having satisfied all three conditions, was included in the $X = B1$ list.

Columns headed $B1 \neq L1$ and $X \neq L1$ contain calculated t-values representative of the second and third conditions noted above. Next to each significant t-value are figures which represent corresponding confidence levels (CL).

Words marked with an asterisk (*) revealed zero values (no occurrences) in one or more test articles. The large sample size statistically justified scoring these words in spite of the presence of zero values. Also, if a given function word is used by one author twenty times in a 35,000 word text and not at all by another in a work of comparable size and content, it seems only logical that such a word be accepted as a valid discriminator. Alors, for example, occurred sixteen times in l'Avis and nine times in Cabale but not at all in Larroque's Le Prosélyte. Both statistically and intuitively, this usage suggests similarity between Bayle and the unknown author and a significant difference between Larroque and the other two writers.

TABLE 4:30. FUNCTION WORDS, X = B1

Word	B1 ≠ L1	CL	X ≠ L1	CL
alors*	2.374	.025	3.229	.005
auparavant	2.047	.05	2.353	.025
beaucoup	2.592	.01	2.975	.005
c'est-à-dire*	2.961	.005	3.425	.001
cause - noun	3.213	.005	3.114	.005
celui	4.246	.001	5.035	.001
cependant	3.371	.001	3.253	.005
certain	4.517	.001	4.863	.001
chacun	4.243	.001	2.660	.01
comme	5.700	.001	5.223	.001
comment	4.822	.001	4.915	.001
dans	4.008	.001	3.558	.001
devoir - all	4.460	.001	5.571	.001
elles	5.280	.001	3.889	.001
être	2.080	.05	3.589	.001
femme	2.203	.05	3.124	.005
main	2.275	.025	2.194	.05
mais	3.606	.001	2.807	.005
même	3.651	.001	2.163	.05
moien	2.939	.005	3.119	.005
ne	2.515	.025	2.435	.025
notre	3.826	.001	3.696	.001
on	8.547	.001	9.440	.001
parce	7.168	.001	7.072	.001
petit	4.475	.001	3.504	.001
pire*	2.374	.025	2.796	.01
pourtant	5.114	.001	4.769	.001
se	5.344	.001	4.921	.001
son	4.068	.001	2.712	.01
tel	5.013	.001	3.730	.001
très	3.024	.005	3.161	.005
trouver	4.015	.001	3.214	.005

Likewise, the connective c'est à dire and the adjective pire were not used by Larroque in his 35,000 word text.

Table 4:31 presents seventy-four function words whose quantitative use by Bayle differed significantly from that of the author of l'Avis aux réfugiés.

The symbol $X \neq B1$ in the table title describes results obtained and tabulated when the words listed in Table 4:30 were compared quantitatively in l'Avis and Cabale. In the t-value column are listed results of the t-test for difference followed by the significance, or confidence, level. Only those words for which a significant difference was observed are listed. A complete list of 199 function words tested comprises Appendix K. Of seventy-four words listed in Table 4:31, twenty-two were presented earlier (Table 4:29) as they satisfied the conditions which, for these particular variables, advanced Larroque as probable author of l'Avis. To those twenty-two words are added fifty-two others which serve to distinguish X from B. The total list is made up of adjectives, adverbs, articles, connectives, negations, nouns, prepositions, pronouns, and verbs: nine different part of speech categories. An analysis of individual terms which constitute each of these categories reveals eight adjectives, ten adverbs, two articles, eleven connectives, five negations, three nouns, eleven prepositions, nine pronouns, and fifteen verbs. The mean number of terms for each category is approximately eight. Only article, noun, and verb groups deviate notably from an otherwise consistent pattern. The fact that a similar divergence exists in figures derived from comparison of B1 and L1 (Table 4:32) but not in those from X:B1, L1 comparisons suggests that even though adjectives, nouns, and verbs are generally content bound, they

TABLE 4:31. FUNCTION WORDS, X \neq B1

Word	t-Value	CL	Word	t-Value	CL
à	2.823	.005	il	19.704	.001
ainsi	2.217	.05	ils	13.989	.001
aller	4.686	.001	je	24.285	.001
autant	3.035	.005	le	4.193	.001
avoir	5.198	.001	leur	13.841	.001
bien	2.255	.025	lors	4.484	.001
ce	2.084	.05	lorsque	7.902	.001
cela	2.169	.05	l'on	2.364	.025
chaque	3.770	.001	moins	1.963	.05
chez	3.043	.005	mon	11.141	.001
contre	5.479	.001	monsieur	19.691	.001
croire	4.135	.001	néanmoins	2.029	.05
de	4.877	.001	ni	2.104	.05
déjà	2.077	.05	non	2.556	.025
demander	3.002	.005	nous	3.236	.005
depuis	3.892	.001	ou	4.594	.001
dessus	2.225	.05	où	2.468	.025
devoir - cond.	2.585	.01	par	2.216	.05
dire	4.402	.001	parmi	2.369	.025
donner	4.190	.001	pas	2.525	.025
elle	5.735	.001	plus	4.614	.001
en	2.935	.005	point - neg.	2.927	.005
enfant	2.151	.05	pour	2.113	.05
et	2.693	.01	pourquoi	3.328	.001
etc.	2.676	.01	pourvoir	3.635	.001
faire	3.719	.001	prendre	2.877	.005
falloir	2.434	.025	puis	5.496	.001
force	3.643	.001	puisque	7.497	.001
grand	3.337	.001	quand	2.900	.005
homme	6.816	.001	que	4.795	.001
hors	2.177	.05	quelque	1.991	.05
ici	2.825	.005	répondre	2.498	.025
savoir	6.297	.001	un	4.454	.001
si	4.766	.001	venir	2.642	.01
sous	r.119	.001	voir	4.120	.001
tout	2.824	.005	votre	18.496	.001
tout-à-fait	2.139	.05	vous	28.093	.001

may be valid quantitative, as well as qualitative discriminants. Generally speaking, content words, when tested across two or three works yield statistically significant differences. Such results may have some qualitative value, but little or no quantitative value since the observed difference is obviously due to subject matter and not to stylistic patterns. On the other hand, if two pieces of writing treat the same topic, but one of them is written, for example, in a verbal and the other in a nominal style, preponderance of verbal forms becomes a discriminating factor. The presence of a quantitative difference of word-classes suggests a fruitful area of further qualitative research which is beyond the scope of the present quantitative study.

Table 4:32 lists ninety-five function words whose quantitative use in Cabale Chimérique (B1) differed significantly from their use in Le Prosélyte abusé (L1).

In order to propose either Bayle or Larroque as probable author of l'Avis aux réfugiés, the either/or hypothesis requires that a significant difference first exist between Bayle and Larroque for the variable in question. Table 4:32 contains all of the 199 function words which yielded a statistically significant difference of use with a confidence level of .05 or better when B1 and L1 were compared. The ninety-five terms again form nine part of speech categories which reveal the same general consistency and divergence noted in Table 4:31. Forty-eight of the ninety-five terms (51%) are of the type which Mosteller and Wallace considered "good discriminators," e.g., their use is totally independent of content.

For twenty-seven of these forty-eight terms the t-test for difference between B1 and L1 revealed variances significant at the .001

TABLE 4:32. FUNCTION WORDS, B1 \neq L1

Word	B1 \neq L1	CL	Word	B1 \neq L1	CL
à	2.496	.025	de	5.336	.001
ainsi	3.312	.001	déjà	2.592	.01
aller - all	4.116	.001	depuis	3.667	.001
alors	2.374	.025	devoir - all	4.460	.001
après	3.057	.005	devoir - cond.	1.972	.05
aujourd'hui	2.612	.01	dire - all	6.306	.001
auparavant	2.047	.05	donner - all	3.062	.005
aussi	2.151	.05	elle	3.972	.001
autrefois	2.772	.01	elles	5.280	.001
avant	2.002	.05	en	3.801	.001
avoir	3.690	.001	enfant	3.737	.001
beaucoup	2.592	.01	entre - prep.	2.424	.025
bientôt	2.389	.025	etc.	2.930	.005
c'est-à-dire	2.961	.005	être - all	2.080	.05
cause - noun	3.213	.005	faire - all	3.080	.005
ce	3.584	.001	falloir - all	3.108	.005
cela	3.721	.001	femme	2.203	.05
celui	4.246	.001	force	2.238	.05
cependant	3.371	.001	homme	2.380	.025
certain	4.517	.001	ici	4.839	.001
chacun	4.243	.001	il	11.143	.001
comme	5.700	.001	ils	6.519	.001
comment	4.822	.001	je	8.944	.001
croire - all	5.389	.001	le, la, l', les	3.460	.001
dans	4.008	.001	leur	3.957	.001
lors	2.092	.05	pouvoir - all	3.013	.005
main	2.275	.025	prendre - all	2.681	.01
mais	3.606	.001	propos	2.846	.005
malgre	3.514	.001	puis	4.770	.001
même	3.641	.001	quand	3.382	.001
mettre - all	2.931	.005	que	2.300	.025
moien	2.939	.005	quelque	3.723	.001
mon	5.610	.001	regarder - all	2.266	.025
monsieur	3.759	.001	répondre - all	2.392	.025
ne	2.515	.025	se	5.344	.001
ni	2.484	.025	son	4.068	.001
non	3.093	.005	sous	2.121	.05
notre	3.826	.001	sur	2.351	.025
nous	10.598	.001	tant	2.095	.05
on	8.547	.001	tel	5.013	.001

TABLE 4:32. (Continued)

Word	B1 ≠ L1	CL	Word	B1 ≠ L1	CL
parce	7.168	.001	tellement	2.742	.01
personne	2.166	.05	très	3.024	.005
petit	4.475	.001	trouver - all	4.015	.001
pire	2.374	.025	venir - all	2.480	.025
plus	3.932	.001	voir - all	2.362	.025
point - neg.	3.824	.001	votre	3.100	.005
pour	4.294	.001	vous	8.572	.001
pourtant	5.114	.001			

(1/1000) level. Overall results of the FREQFUN tests comparing Bayle and Larroque were especially rewarding in view of the fact that the entire either/or hypothesis is based on ability to distinguish between the two suggested authors. The sentence-level (SENWOL and STYLBEND) measures suggested more similarity than divergence in stylistic patterns. FREQFUN comparisons, on the other hand, revealed a greater number of differences between the two writers, and thereby provided more evidence for attribution.

Table 4:33 presents a list of forty-two function words whose frequency of use in l'Avis aux réfugiés differed significantly from their use in Cabale (B1) and Le Prosélyte (L1).

As in other tables of FREQFUN results, function words which showed significant differences are listed alphabetically in the word column. In Table 4:33 t-values which attest to observed differences are listed under the symbols which characterize comparisons made. X:B1 symbolizes comparison of values from l'Avis to those from Cabale; X:L1 symbolizes values from l'Avis compared to those from Le Prosélyte. Next to each t-value is listed the level of confidence (CL) at which the

TABLE 4:33. FUNCTION WORDS, X \neq B1 OR L1

Word	X:B1	CL	X:L1	CL	Least Significant
autant	3.035	.005	3.193	.005	equal
bien	2.255	.025	2.415	.025	equal
ce	2.084	.05	5.422	.001	equal
cela	2.169	.05	5.659	.001	X:B1
chaque	3.770	.001	4.195	.001	equal
chez	3.043	.005	2.258	.025	X:L1
contre	5.479	.001	4.782	.001	equal
croire - all	4.135	.001	9.069	.001	equal
ce	4.877	.001	9.506	.001	equal
demander - all	3.002	.005	3.564	.001	equal
dire - all	4.402	.001	10.253	.001	equal
elle	5.735	.001	3.633	.001	equal
homme	6.816	.001	3.912	.001	equal
ici	2.825	.005	2.760	.01	X:L1
il	19.704	.001	6.373	.001	equal
ils	13.989	.001	6.286	.001	equal
je	24.285	.001	13.742	.001	equal
leur	13.841	.001	8.351	.001	equal
lors	4.484	.001	2.106	.05	X:L1
lorsque	7.902	.001	6.254	.001	equal
l'on	2.364	.025	2.253	.025	equal
moins	1.963	.05	2.306	.025	equal
mon	11.141	.001	5.099	.001	equal
monsieur	19.691	.001	16.557	.001	equal
néanmoins	2.029	.05	2.808	.005	equal
nous	3.236	.005	7.543	.001	X:B1
ou	4.594	.001	4.422	.001	equal
parmi	2.369	.025	3.303	.001	X:B1
plus	4.614	.001	7.657	.001	equal
pour	2.113	.05	2.468	.025	equal
pourquoi	3.328	.001	3.737	.001	equal
puis	5.496	.001	9.582	.001	equal
puisque	7.497	.001	5.933	.001	equal
que	4.795	.001	6.554	.001	equal
quelque	1.991	.05	2.052	.05	equal
répondre - all	2.498	.025	4.652	.001	X:B1
savoir - all	6.297	.001	4.265	.001	equal
si	4.766	.001	3.625	.001	equal
tout	2.824	.005	3.145	.005	equal
voir - all	4.120	.001	6.107	.001	equal
votre	18.496	.001	13.724	.001	equal
vous	28.093	.001	18.816	.001	equal

difference was observed. In the column denoted "Least Significant" are listed the author/article combinations (X:B1 or X:L1) whose observed differences, even though within the acceptable .05 range, yielded the lesser degree of confidence. For example, the t-value calculated from the comparison of ce in X and L1 yielded a confidence level of .001 (one chance in one thousand that the two test values came from the same population).¹⁸ For the X and B1 comparison, on the other hand, the confidence level is .05. This five in one hundred figure suggests a higher probability of error and is, therefore, less significant. For this reason the X:B1 .05 value is listed in the "Least Significant" column. The word "equal" in this column means that the observed differences can be accepted with the same degree of confidence for each of the comparisons made. A summary of the notations in this column reveals four words, cela, nous, parmi, and répondre, whose observed difference in the X:L1 comparison can be accepted with greater confidence than in the X:B1 comparison. Similarly, there are three words, chez, ici, and lors, whose comparison of quantitative use in X and B1 revealed a greater degree of confidence. Confidence levels for the remaining thirty-five words were equal.

Table 4:34 lists twenty-six words whose quantitative use in the preface to l'Avis aux réfugiés differed significantly from their use in the text itself.

Even though results tabulated above have no direct bearing on the question, "Is l'Avis aux réfugiés the product of Bayle's hand?" they serve to support the theory mentioned earlier, Chapter I (p. 4), that the preface was not the work of the same pen which wrote l'Avis. Once again significant terms have been listed alphabetically in the "Word"

TABLE 4:34. FUNCTION WORDS, X \neq P

Word	P:X	CL	Word	P:X	CL
assez	2.301	.025	monsieur	3.273	.005
autre	2.813	.005	ne	2.009	.05
avoir - all	5.236	.001	ni	2.220	.05
car	2.263	.025	notre	8.396	.001
dès	1.974	.05	qui	2.264	.025
en	2.100	.05	quoi	2.396	.025
encore	2.164	.05	répondre - all	3.934	.001
et	2.123	.05	temps	3.111	.005
excepté	5.814	.001	trop	4.321	.001
faire - all	2.470	.025	voir - all	2.144	.05
ils	2.201	.05	votre	4.621	.001
je	12.704	.001	vouloir - all	1.967	.05
mon	7.939	.001	vous	6.898	.001

column, and statistical t-values with corresponding confidence levels (CL) complete the table. The same basic part of speech pattern among significant terms exists in this table as in Tables 4:31 through 4:33. That there is no heavy concentration of one part of speech category is an indication of the general stability of the function word list.²¹

Verbal Summary of Individual Function Words

Table 4:35 is a summary of Tables 4:29 through 4:34 and presents four lists of function words whose frequency of use in l'Avis, Le Prosélyte, and Cabale is similar (X = L and X = B) or significantly different (X not equal B1 or L, and X not equal P).

The column headings of Table 4:35 symbolize authors and works whose quantitative function word usage was compared. Denominators X, L, B and P correspond to the unknown author, Larroque, Bayle, and author of the preface respectively. In order to qualify for the X = L column,

TABLE 4:35. SUMMARY OF SIGNIFICANT FUNCTION WORDS

X = L	X = B	X Not Equal B1 or L1	X Not Equal P
à	alors	autant	assez
ainsi	beaucoup	bien	autre
aller - all	cause - noun	ce	avoir - all
avoir - all	celui	cela	car
déjà	cependant	chaque	dès
depuis	c'est-à-dire	chez	en
devoir - cond.	certain	contre	encore
donner - all	chacun	croire - all	et
en	comme	de	excepté
enfant	comment	demander - all	faire - all
etc.	dans	dire - all	ils
faire - all	devoir - all	elle	je
falloir - all	elles	homme	mon
force	être - all	ici	monsieur
le	femme	il	ne
ni	main	ils	ni
non	mais	je	notre
point	même	leur	qui
pouvoir - all	moien	lors	quoi
prendre - all	ne	lorsque	répondre - all
quand	notre	l'on	temps
venir - all	on	moins	trop
	parce	mon	voir - all
	petit	monsieur	votre
	pire	néanmoins	vouloir - all
	pourtant	nous	vous
	se	ou	
	son	parmi	
	tel	plus	
	très	pour	
	trouver - all	pourquoi	
		puis	
		puisque	
		que	
		quelque	
		répondre - all	
		savoir - all	
		si	
		tout	
		voir - all	
		votre	
		vous	
Total - 22	Total - 32	Total - 42	Total - 26

twenty-two words listed had to satisfy the conditions prescribed for establishing Larroque as probable author. For example, quantitative use of the preposition A* (à, au, aux, auquel, auxquels) revealed a statistically significant t-value of 2.496²⁰ between Bayle and Larroque. A similar difference (2.823) was observed between Bayle and the unknown author. By virtue of the either/or hypothesis, the lot for all words revealing significant differences between B1 and L1 as well as B1 and X fell to Larroque. Likewise, all words which revealed significant differences between B1 and L1 and L1 and X are listed as Baylian terms and tabulated under heading X = B. For the first time in the project a number of test items pointed to greater similarities in X and B than in X and L: individual function word tests suggested thirty-two terms were used quantitatively the same by Bayle and the unknown author as opposed to twenty-two used commonly by Larroque and the author of l'Avis. On the other hand, forty-two of 199 function words showed a significant difference between the unknown author and both hypothesized writers. These forty-two words are found under the column heading "X Not Equal B1 or L1." The final column of Table 4:34 lists all those words from l'Avis and its preface which were observed to have a significant difference of quantitative use. They number twenty-six.

An analysis of the four lists in Table 4:34 reveals a number of adjectives, nouns, pronouns, and verbs which, according to Mosteller and Wallace, might be content-inspired. These words were included in the final list because the FAP program²¹ had shown that their distribution resembles that of function, rather than content words. Nevertheless, because Mosteller and Wallace emphasized that "personal pronouns and auxiliary verbs, especially with respect to mood and tense" should not

be included since they "are likely to be related to external details,"²² all words whose use might depend on context were omitted from the $X = L1$, $X = B1$, and $X \neq B1$ or $L1$ lists in a further-refinement experiment. Although there were fewer words overall suggesting that Bayle did not author l'Avis aux réfugiés, there were, likewise, proportionately fewer words used commonly in l'Avis and Cabale. The overall percentage of change favored someone else as being the unknown author. The $X = L1$ list was reduced by 60% (22 to 9), the $X = B1$ list was reduced by 54% (32 to 15); whereas the $X \neq B1$ or $L1$ list was reduced by only 38% (42 to 26). Thus, the case against Bayle's authorship of l'Avis was only strengthened by following the admonitions of Mosteller and Wallace. Since this experiment did not significantly alter the overall individual function word results, the refined lists, complete with their calculated t-values and confidence levels, are included in Appendix N, O and P rather than in the text proper.

Statistical Summary of Individual Function Words

Table 4:36 presents a summary of probable author and author/article combinations for which the hypothesis-of-no-difference was tested. This summary also includes the number of variables tested in each combination whose confidence levels ranged from .001 to .05.

When the quantitative use of 199 function words was compared across the final five test articles, statistically significant differences were observed with significance levels that afford a much greater degree of confidence than had been hoped for. Tables 4:29 through 4:34 show individual words whose use differed significantly within a given author/article combination. Table 4:36 presents the total number of

TABLE 4:36. FREQFUN INDIVIDUAL WORDS--STATISTICAL SUMMARY

	X = L1*	X = B1*	X:B1
	4 @ .001 CL	14 @ .001 CL	37 @ .001 CL
	6 @ .005 CL	7 @ .005 CL	11 @ .005 CL
	3 @ .01 CL	3 @ .01 CL	4 @ .01 CL
	3 @ .025 CL	2 @ .025 CL	8 @ .025 CL
	6 @ .05 CL	5 @ .05 CL	14 @ .05 CL
Totals	22	31	74
	B1:L1	X ≠ B1 or L1*	X:P
	49 @ .001 CL	25 @ .001 CL	9 @ .001 CL
	14 @ .005 CL	4 @ .005 CL	3 @ .005 CL
	6 @ .01 CL	1 @ .01 CL	0 @ .01 CL
	15 @ .025 CL	5 @ .025 CL	5 @ .025 CL
	11 @ .05 CL	7 @ .05 CL	9 @ .05 CL
Totals	95	42	26

Probable Author Summary: B = 31, L = 22, N = 42.

*Confidence level reported is lesser of the two.

words from each author/article comparison significant at the .001, .005, .01, .025, or .05 confidence levels. Groups denominated X = L1 and X = B1 summarize statistical results which suggest Larroque or Bayle as probable author of l'Avis aux réfugiés. In these two instances, as in the X ≠ B1 or L1 comparisons, the degree of probability figure (CL) which allows the greatest chance for error is reported as explained on page 146. Even so, a visual comparison of the X = L1 and X = B1 significance levels suggests Bayle, rather than Larroque, as the probable author with a substantially greater degree of confidence. However, X = L1

and X = B1 results, based as they are upon the either Bayle or Larroque hypothesis, merely show which of the two is closer to the unknown author in his use of these variables. They do not show how close either is. For that reason summary results from comparisons X:B1, B1:L1, and X \neq B1 or L1 are presented. T-tests of difference revealed more than twice as many function words used differently in X and B1 than were used similarly; thirty-seven opposed to fourteen at the .001 level, for example. The probable author summary at the bottom of Table 4:36 suggests the conclusion that must be drawn from the individual-function-word analysis: although Bayle's quantitative use of the function words tested resembles that of the unknown author more than Larroque's, the observed differences for these variables adduce that Bayle did not write l'Avis aux réfugiés.

FREQFUN-Grouped Data

After 199 function words were compared individually across the final five test articles, they were assigned to one or more of twenty-nine part of speech, usage, or origin categories as shown in Appendix Q. Whenever a word appeared in more than one category, e.g., coordinating conjunctions of Latin origin and coordinating conjunctions of liaison, its values served each group independently, but were not counted twice when the categories were combined. Groups into which the function words were directed are those delineated by Grevisse in his general discussions on individual part of speech classifications.²³

Grouping of function words by category permits testing of low frequency terms which might otherwise have passed unnoticed. Such words may not individually be discriminators because they do not occur often

enough to register statistically or intuitively; whereas they may, in reality, be part of a larger pattern whose whole can be measured. A writer may be conscious of his choice of individual words, replacing one by another of the same or different meaning, while maintaining the grammatical framework of the sentence. Milic argues that a "writer's mind projects a syntactical mold for a sentence at the same time as it handles the complex interaction between word and idea. That is, while the sentence is mentally in suspension, a certain outline is planned for it, in which the individual places for words are designed in grammatical rather than semantic terms."²⁴ Milic uses this argument to justify his structural grammar approach to stylistic studies. His reasoning is plausible and suggests an additional postulate: to this grammatical framework may be affixed any of a large number of different words of the same part of speech category, e.g., adverbs of manner; the limiting factor being the author's vocabulary. If, as Milic suggests, this framework has become habitual and proceeds "below the level of conscious thought,"²⁵ and if this mold into which the mature author fits his vocabulary is stable, words filling these slots will form a categorical rather than semantic pattern. Conceivably an author could have such a large active vocabulary that it would not be necessary for him to repeat the same word often enough for its quantitative use to be discriminative. However, when figures for the term in question are combined with those of its synonyms, formal, rather than content, discrimination may become evident. Thus such terms as exprès, pis, auparavant, autrement, autour, and dedans contributed their individually insignificant quantitative values to the grammatical category to which each belongs. That these categorical groupings proved to be good discriminators is

demonstrated by the results presented in Tables 4:37 through 4:41:

FREQFUN grouped data results.

Table 4:37 presents a list of the twenty-nine FREQFUN part of speech categories, statistical results for author/article combinations which were compared, and a probable author designation for each significant group.

Twenty-nine part of speech categories suggested by Grevisse are tabulated in the column headed "Word Group." Abbreviations C C and S C stand for coordinating and subordinating conjunctions. In addition to providing a means for recording the greatest possible number of tokens of an author's function-word vocabulary, grouping of these terms by type provides a means by which function words, which could possibly be content bound, may be evaluated both independently and collectively. When like-genres, treating like-topics, are compared, each word--type or token--chosen by the author serves to distinguish him from, or associate him with, his companion writers.²⁶ Thus, the 184 function words derived from the FAP, method-three, were combined with the FREQFUN individual-words list, and all were directed into their appropriate categories. Coordinating and subordinating conjunction groups contain only one-word connectives; results for compounds are included in EXSOR results (see pp. 162-193). Appendix Q provides a complete list, by category of function words tested.

In Table 4:37, the denominators $X \neq B1$, $X \neq L1$, and $B1 \neq L1$ symbolize statistically significant differences observed when articles denoted in columns two, three, and four were compared. Figures under each of these column headings are calculated t-values which affirm the observed differences. A dash (-) in the $X \neq B1$, $X \neq L1$, or $B1 \neq L1$

TABLE 4:37. FREQFUN GROUPED DATA RESULTS

Word Group	X \neq B1	X \neq L1	B1 \neq L1	Least Sign CL Comb	Prob Auth
C C - Latin origin	2.254	5.966	3.997	.025 X:B1	N
C C - alternation	4.594	4.422	-	.001 Equal	N
C C - cause	2.137	3.044	-	.05 X:B1	N
C C - consequence	3.000	-	2.456	.025 B1:L1	L
C C - explication	-	2.118	-	.05 X:B1,L1	-
C C - liaison	-	-	3.295	.001 X:B1,L1	-
C C - opposition	3.229	-	3.422	.005 X:B1	L
C C - transition	-	-	-	-	-
C C - all	-	5.844	4.924	.001 Equal	B
S C - all	6.286	-	5.989	.001 Equal	L
Conj - all	-	5.777	5.985	.001 Equal	B
Prep - Latin origin	2.852	10.114	7.697	.005 X:B1	N
Prep - adverbial	-	-	-	-	-
Prep - all	3.076	8.785	6.154	.005 X:B1	N
Adv - Latin origin	-	3.575	3.779	.001 Equal	B
Adv - affirmation	4.968	5.309	-	.001 Equal	N
Adv - composed	3.701	2.150	-	.05 X:L1	N
Adv - doubt	-	2.677	-	.01 B1:L1,X	-
Adv - manner	-	5.682	6.240	.001 Equal	B
Adv - negation	-	2.720	2.103	.05 B1:L1	B
Adv - place	-	3.747	3.319	.001 Equal	B
Adv - quantity	-	-	-	-	-
Adv - time	2.477	4.677	2.476	.025 Equal	N
Adv - all	10.535	37.447	45.719	.001 Equal	N
Adj - poss indef demon	2.970	-	-	.005 X:L1	-
Art - def and indef	3.425	5.052	2.077	.05 B1:L1	N
Nouns	10.950	8.210	-	.001 Equal	N
Pronouns	2.996	4.387	-	.005 X:B1	N
Verbs - all	9.875	9.085	-	.001 Equal	N
Totals	B = 6	L = 3	N = 13	Insignificant = 7	

Confidence Level Shown is for the Least Significant t-Value Tabulated.

columns means that the quantitative difference between treatment means of the variable in question was not distinct enough to generate the minimum 1.96 t-value required for an .05 level of confidence. Figures in the column headed "Least Sign CL Comb" represent the significance levels (CL) which acknowledge divergent usage of the variable in question, with the least amount of confidence. Next to each least-significant value is the author/article combination which produced it. An analysis of this column reveals eight variables (coordinating conjunctions of Latin origin, cause, explication, liaison, and opposition; prepositions of Latin origin; all prepositions; and pronouns) which suggest that Bayle's writing, more so than Larroque's resembles that of l'Avis with regard to coordinating conjunction, preposition, and pronoun use. In the probable author column, the letters B, L, and N represent Bayle, Larroque, and neither Bayle nor Larroque. A dash (-) in this column means that there were no statistically significant differences observed or that conditions for attribution described on pages 93-94 were not met. The summary results for this table are quite revealing. Only six variables point to Bayle as author of l'Avis aux réfugiés; three suggest Larroque. On the other hand, thirteen of the twenty-nine part of speech categories imply that neither Bayle nor Larroque authored the disputed text. In all, seventeen of twenty-nine FREQFUN grouped data variables revealed statistically significant differences between Bayle and the unknown author. Eight of these variables (coordinating conjunctions of alternation, subordinating conjunctions, adverbs of affirmation, adverbs composed of two or more parts of speech, all adverbs, articles, nouns and verbs) disclosed differences significant at the .001. Stated otherwise, for eight of these twenty-nine variables,

there is one chance in one thousand that the observed difference is due to chance. Six (coordinating conjunctions of consequence and opposition; prepositions of Latin origin and all prepositions; and possessive, indefinite and demonstrative adjectives) were significant at the .005 (5/1000); two (coordinating conjunctions of Latin origin and adverbs of time) were significant at the .025 (25/1000); and only one (coordinating conjunctions of cause) was significant at the acceptable .05 (5/100) level. An analysis of the $X \neq L1$ column revealed that twenty-one FREQFUN grouped-data variables differentiated between Larroque and the writer of l'Avis.

In order to discover whether or not the significant grouped-data variables constituted a general part-of-speech pattern, each of the groups for which a quantitative difference was observed was listed separately, according to probable author. Tables 4:38 and 4:39 present the groups whose divergent use in the $X:B1$, $X:L1$, and $B1:L1$ comparisons suggest Bayle and Larroque respectively as authors of l'Avis aux réfugiés.

TABLE 4:38. FUNCTION-WORD GROUPS WHICH SUGGEST BAYLE WROTE L'AVIS AUX REFUGIES

Function Word Group	$X \neq L1$	CL	$B1 \neq L1$	CL
Coordinating conj. - all	5.844	.001	4.924	.001
Conjunctions - all	5.777	.001	5.985	.001
Adverbs - Latin origin	3.575	.001	3.779	.001
Adverbs - manner	5.682	.001	6.240	.001
Adverbs - negation	2.720	.01	2.103	.05
Adverbs - place	3.757	.001	3.319	.001

As in previous FREQFUN results tables the test variable is listed in the left-hand column. The other four columns contain statistical results derived from the frequency of use of test variables in articles symbolized by the column headings $X \neq L1$ and $B1 \neq L1$. The columns denominated CL tabulate the degree of probability, or confidence level at which the observed difference can be accepted as having occurred by stylistic design rather than by chance. An analysis of the tabulated variables reveals that the six individual groups do fall into a pattern: the quantitative use of single word conjunctions and adverbs differed significantly in l'Avis and Le Prosélyte and in Cabale and Le Prosélyte, but not in l'Avis and Cabale. Therefore, by virtue of the either Bayle or Larroque hypothesis, quantitative use of these syntactical elements suggests Bayle as the probable author of l'Avis aux réfugiés.

TABLE 4:39. FUNCTION-WORD GROUPS WHICH SUGGEST LARROQUE WROTE L'AVIS AUX REFUGIES

Function Word Group	$X \neq B1$	CL	$B1 \neq L1$	CL
Coordinating conj. - consequence	3.000	.005	2.456	.025
Coordinating conj. - opposition	3.229	.005	3.422	.001
Subordinating conj. - all F	6.286	.001	5.989	.001

The three groups which constitute Table 4:39 do little to refute the results of Table 4:38. Of twenty-nine function-word groups, only three suggest similarity between Larroque and the unknown author with regard to the variables listed. Moreover, they represent only one basic

syntactical element: single-word connectives. It seems logical that when stylistic features of a mature author, who does not attempt to disguise his writing, are compared from work to work, more than just one analogous characteristic should reveal itself. Larroque had no impelling motive for concealing his identity by purposely changing his style.²⁷

It follows then, that if Larroque wrote l'Avis aux réfugiés, his conscious choice of stylistic patterns, as well as his unconscious habits, in Le Prosélyte abusé should resemble those of l'Avis aux réfugiés.

That the statistical results of Table 4:39 reveal only three groups, or one syntactical element common to L and X, can only suggest that either Larroque's stylistic habits had not matured to a point of stability, or else he did not author l'Avis aux réfugiés. T-test for difference results listed in Tables 4:37 and 4:40 suggest the latter conclusion.

Twenty-one, or 72% of the twenty-nine function word groups presented in Table 4:37, disclosed significant differences between X and L1 as tabulated in the $X \neq L1$ column. Thirteen of the twenty-one significantly divergent groups just mentioned were similarly distinct when X and B1 were compared. These thirteen groups are listed in Table 4:40.

Column denominators in Table 4:40 are the same as those used in other tables which report FREQFUN results. Again it was rewarding to note that the majority of the significant values yielded confidence levels in the .001 - .005 range.

TABLE 4:40. FUNCTION-WORD GROUPS WHICH SUGGEST NEITHER BAYLE NOR LARROQUE WROTE L'AVIS AUX REFUGIES

Function Word Group	X \neq B1	CL	X \neq L1	CL
Coord. conj. - Latin origin	2.254	.025	5.966	.001
Coord. conj. - alternation	4.594	.001	4.422	.001
Coord. conj. - cause	2.137	.05	3.044	.005
Prepositions - Latin origin	2.852	.005	10.114	.001
Prepositions - all	3.076	.005	8.785	.001
Adverbs - affirmation	4.968	.001	5.309	.001
Adverbs - composed	3.701	.001	2.150	.05
Adverbs - time	2.477	.025	4.677	.001
Adverbs - all	10.535	.001	37.447	.001
Articles - def. & indef.	3.425	.001	5.052	.001
Nouns	10.950	.001	8.210	.001
Pronouns	2.996	.005	4.387	.001
Verbs	9.875	.001	9.085	.001

Statistical Summary of Grouped Function Words

Table 4:41 is a statistical summary of the confidence level values for the FREQFUN grouped data probable author designations. The X = L1 column is a statistical summary of Table 4:39 and contains the confidence level figures for the three groups listed in that table. Likewise, column X = B1 presents similar data for its six groups drawn from Table 4:38. Finally, the grammatical/part of speech categories whose significance values are posted in the X \neq B1 or L1 column are

TABLE 4:41. FREQFUN GROUPED WORDS PROBABLE AUTHOR SUMMARY

	X = L1	X = B1	X \neq B1 or L1
	1 @ .001	5 @ .001	5 @ .001
	1 @ .005	0 @ .005	3 @ .005
	1 @ .025	0 @ .025	2 @ .025
	0 @ .01	0 @ .01	0 @ .01
	0 @ .05	1 @ .05	3 @ .05
	—	—	—
Totals	3	6	13
Summary:	L = 3	B = 6	N = 13

found in Table 4:40. All confidence level figures reported in Table 4:41 are extracted from the "Least Sign CL Comb" column of Table 4:37 and are listed in descending order from the most to the least significant. In satisfying the conditions prescribed for probable attribution the B1:L1 (B1 compared to L1) and X:L1 differences yielded the same, .001, level of confidence. It may therefore be concluded with the same degree of certainty that, for the particular variables in question, Bayle differs from Larroque, and Larroque, in turn, differs from the unknown author. Figures in column X \neq B1 or L1 also reveal strong significant differences, although not quite so one-sided as in the X = B1 listings. Six of thirteen, or 46% of the listed confidence levels, provide an equal degree of confidence that X differs from both B1 and L1. Even though significant differences were observed for all tabulated variables, corresponding confidence levels for six of the remaining seven t-values suggest that it may be concluded with a greater

degree of confidence that Larroque's quantitative use of the corresponding six variables differs from the unknown author's than do Bayle's.

Function Word Conclusions

Both individual-word and grouped-word tests performed on FREQFUN data provided strong internal evidence, choice and use of function words --whether viewed as the conscious application of an active vocabulary or as the unconscious ordering of word patterns to a syntactical framework fashioned by experience--that reveals divergence, rather than similarity, between Bayle, Larroque, and the author of l'Avis aux réfugiés.

Expression Frequencies and Usage

In the concluding remarks to their chapter "Words and Their Distribution," Mosteller and Wallace stated, "Phrases are attractive, and were it possible to choose a moderately large list, they could be handled as words."²⁸ Sayce described several uses of prepositional, conjunctive, and interjective phrases, figures of speech, clichés, and proverbs while referring to them as individual characteristics of style.²⁹ Muller expressed the utility, indeed, the desirability of isolating phrases in stylistic studies.³⁰ Finally, Milic recognized the stylistic value of word combinations and verged upon a study of phrases as he examined three-word patterns in the works of Jonathan Swift.³¹ Studies involving identification, location, and tabulation of locutions and other quantifiable word combinations have, for the most part, been limited in the past to making indexes. Muller pointed out that most index compilers, including scholars at the well known research computing

center in Besançon, have included in their lists "un certain nombre de locutions qu'ils n'analysent pas," and that "la liste de ces locutions est très variable d'un index à l'autre."³² The ordering and quantifying of a general list of locutions and other quantifiable expressions is one of the unique contributions of the present study.

In selecting items to appear on the "expression list" for this study, account was taken of problems of selection and measurement involved in making such a list as suggested by Mosteller and Wallace: "The measurement problem is more difficult to carry out, and the available data would be much less adequate. Choosing a pool of phrases is not easy. Possibly, the index could be used . . ."³³ Problems they expose may be resolved into three basic questions: first, "Is available data adequate for such a study?" The words "available data" apply not only to sample size, which must be extremely large so that word patterns will appear with a measurable consistency and predictability, but also to the existence of established lists of expressions upon which to build.³⁴ The second question asks, "Is the researcher prepared to assemble, cull, and sort possible discriminators from non-standardized indexes, grammars, and style manuals?" Because of the almost infinite number of word combinations, resolution of this problem alone seems overwhelming. The third question deals with quantification: "Is the researcher prepared to count and tabulate frequencies of even a moderately large list?" Here the stylistic analyst is faced with making a choice: either he must note each new expression while remembering and scoring those already listed, or he must automate his study. The first choice demands that the researcher's mind be primed as he examines his text line upon line, sentence upon sentence.³⁵ The second choice, while

relieving the researcher of the constant pressure of lineal concentration, requires development of a complex, sophisticated computer program which will search a natural language text, and locate, isolate, count, and print out, in key-word-in-context form, a selected list of expressions.

Each of the problems presented above was resolved in the present study. Since the data base consisted of complete texts ranging from 15,000 to 58,000 words in length, the question of small sample size was eliminated.³⁶ Finding appropriate lists of expressions, on the other hand, was not an easy task. Many phrases from the Besançon studies as mentioned by Muller, served as judgmental devices for selection of items which appear in the general expression list for this study. Additional locutions and word combinations of a grammatical nature were drawn from Grevisse³⁷ and Haase.³⁸ Finally, by far the largest group came from a careful reading of the works to be tested.³⁹ In all, 661 expressions⁴⁰ were noted and keypunched. EXSOR⁴¹ was written to locate, isolate, and count the frequency of occurrence of 661 word combinations. From that point on, as suggested by Mosteller and Wallace, expressions were "handled as words." After all duplicates and extremely low frequency (one or two total occurrences) combinations were eliminated, the final list totaled some 554 expressions as listed in Appendix T. Tables 4:42 through 4:54 present results obtained by comparing frequencies of the 554 expressions across the final test articles. As with function words, ten comparisons were made for each expression. Nevertheless, only results which led most directly to naming a probable author under the conditions prescribed on pages 93 and 94 are tabulated in this section.⁴²

Tables presenting EXSOR results have been directed into two categories: individual expressions and grouped expressions. Results for the first group constitute Tables 4:42 through 4:49. Tables 4:50 through 4:54 contain grouped expression results.

Table 4:42 lists twelve expressions whose quantitative use in the test articles suggests Larroque as probable author of l'Avis aux réfugiés.

TABLE 4:42. EXPRESSIONS, X = L1

Expression	X ≠ B1	CL	B1 ≠ L1	CL
à propos	1.960	.05	3.191	.005
avant que	2.502	.025	3.080	.005
avant que de	2.065	.05	2.148	.05
depuis que	2.274	.025	2.138	.05
en effet	2.780	.01	3.106	.005
il étoit	2.680	.01	2.135	.05
le public	4.905	.001	4.356	.001
ne WRD point	3.468	.001	3.322	.001
ne WRD WRDS point	2.479	.025	2.976	.005
non plus	4.815	.001	2.846	.005
ou que	3.069	.005	2.047	.05
tant que	2.212	.05	2.047	.05

The symbol, X = L1, used in conjunction with the word "Expressions" in the table title, implies that use of expressions in l'Avis aux réfugiés and in Larroque's Le Prosélyte abusé is similar. In addition,

the equal sign designates a probable author with regard to variables listed in the table headed $X = L1$. In order to qualify for inclusion in this table, each of twelve expressions had to pass a hypothesis-of-no-difference test. That is to say, when their quantitative use in l'Avis (X) and in Le Prosélyte (L) were compared, no significant difference was observed. In addition to this screening test, two more conditions had to be met: a significant difference had to exist between the expressions' use in X and in Cabale (B1). Secondly, because of the either/or hypothesis, Bayle's quantitative use of the expressions in question had to differ from Larroque's. Expressions which met these requirements are listed in the column headed "Expression". Figures in the $X \neq B1$ and $B1 \neq L1$ columns are calculated t-values signifying a quantitative difference of use of corresponding variables in articles compared. In CL columns are confidence levels or degrees of probability that the observed difference was due to sampling error rather than to authorial design. Sixteen, or 67%, of twenty-four calculated values yielded significance levels which sustained the observed differences in quantitative use of expressions tested, with a greater degree of confidence than the acceptable .05 level affords.

Whereas Table 4:42 listed twelve expressions which imply similarity between Larroque and the unknown author, Table 4:43 presents twenty-two word combinations whose quantitative use suggests that Bayle might have written l'Avis aux réfugiés.

As in Table 4:42, each word combination listed in the expression column had to satisfy three conditions in order to qualify for the $X = B1$ list. Symbolically, these conditions are $X = B1$, $B1 \neq L1$, and $X \neq L1$. $X = B1$ means that the quantitative use of a variable--an

TABLE 4:43. EXPRESSIONS, X = B1

Expression	X ≠ L1	CL	B1 ≠ L1	CL
à ce que	4.393	.001	3.492	.001
bien que	2.978	.005	3.444	.001
car WRDS si	3.275	.005	2.295	.025
c'est	2.652	.01	3.478	.001
c'est que	3.935	.001	2.931	.005
ce que	5.531	.001	4.122	.001
ce qui	3.549	.001	3.116	.005
comme si	2.449	.025	2.241	.025
de même	2.197	.05	2.276	.025
de même que	2.781	.01	2.421	.025
l'autre	4.381	.001	3.047	.005
ne WRDS plus que	2.102	.05	2.422	.025
ne WRD WRDS que	3.385	.001	3.098	.005
non pas que	2.200	.05	2.256	.025
par conséquent	3.331	.001	3.918	.001
parce que	7.072	.001	7.288	.001
plus WRD WRDS que	3.764	.001	2.732	.01
que de	1.961	.05	2.724	.01
quoi que	4.278	.001	3.268	.005
quoi que WRDS que	3.509	.001	2.795	.01
sur tout	4.326	.001	3.123	.005
tant WRD WRDS que	2.291	.025	2.066	.05

expression, in this case--in l'Avis (X), when compared to its use in Cabale (B1), revealed no significant difference. The second condition, $B1 \neq L1$, required that there be a significant difference between the use of the variable in Cabale (B1) and Le Prosélyte (L1). Finally, if the expression's use in l'Avis (X) differed significantly from its use in Le Prosélyte (L1), having satisfied all three conditions, it was included in the $X = B1$ list.

Columns headed $B1 \neq L1$ and $X \neq L1$ contain calculated t-values which affirm that the second and third conditions noted above were met. Next to each significant t-value are figures which represent the level of significance at which the difference was observed. Once again a high percentage of the significant values recorded degrees of probability at better than the .05 level: eighteen at the .001 level, eight at .005, five at .01, eight at .025, and only five at .05. Stated otherwise, 89% of the calculated values for the twenty-two word combinations listed in Table 4:43 revealed statistical differences between the variables' use in the works tested which could occur fewer than two and one-half times in each one hundred samples.

Table 4:44 lists sixty-eight expressions whose quantitative use in l'Avis aux réfugiés differed significantly from their use in Cabale Chimérique.

The symbol $X \neq B1$ in the table title describes results obtained and tabulated when expressions listed in Table 4:44 were compared quantitatively in l'Avis and Cabale. In columns denominated "X B1 - Absolute Freq," the actual number of times each word combination occurred in articles X and B1 is recorded. Results of the t-test for difference are listed in the t-value column. As in all FREQFUN and

TABLE 4:44. EXPRESSIONS, X \neq B1

Expression	X Absolute	B1 Freq	t-Value	CL
à mesure que	0	4	1.961	.05
à moins que	6	0	2.498	.025
à moins que WRDS ne	4	0	2.040	.05
à-propos	2	9	2.046	.05
à propos	8	2	1.960	.05
à WRDS egard	31	15	2.494	.025
au dela	4	0	2.040	.05
au dessus	11	3	2.212	.05
auprès de	0	4	1.961	.05
autant que	15	4	2.610	.01
autant WRDS que	44	18	3.459	.001
avant que	6	19	2.502	.025
avant que de	3	11	2.065	.05
avant que WRDS ne	0	4	1.961	.05
c'est selon	4	0	2.040	.05
ce ne étoit pas	0	4	1.961	.05
ce ne seroit pas	0	4	1.961	.05
ce que c'est que	1	7	2.066	.05
ce sera	7	1	2.178	.05
ce sont	11	2	2.568	.025
de sorte que	23	9	2.587	.01
depuis que	4	14	2.274	.025
en attendant que	0	4	1.961	.05
en ce temps-là	4	0	2.040	.05
en effet	10	1	2.780	.01
en même temps	7	0	2.699	.01
en même tems	0	9	2.942	.005
entant que	4	0	2.040	.05
esprit de satire	4	0	2.040	.05
hors de	9	2	2.177	.05
il est	74	108	2.256	.025
il étoit	7	22	2.680	.01
il faudrait	4	20	3.170	.005
il sera	1	7	2.066	.05
jamais WRDS ne	19	37	2.259	.025
jusqu'à	25	11	2.453	.025
jusqu'au	8	1	2.393	.025
l'an	19	4	3.223	.005
l'on	103	75	2.364	.025
le droit d'examiner	4	0	2.040	.05

TABLE 4:44. (Continued)

Expression	X Absolute	B1 Freq	t-Value	CL
le public	5	38	4.905	.001
lors que	0	35	5.801	.001
moins	85	64	1.963	.05
ne WRD pas	298	376	2.500	.025
ne WRD WRDS pas	393	505	3.159	.005
ne WRD WRD WRDS pas	147	198	2.383	.025
ne WRD point	117	73	3.468	.001
ne WRD WRDS point	169	132	2.479	.025
ne WRD WRD WRDS plus	142	95	3.361	.001
non pas	21	10	2.086	.05
non plus	30	3	4.815	.001
où l'on	10	3	2.013	.05
ou que	16	3	3.069	.005
ou WRD WRDS ou	58	36	2.462	.025
par la voie	6	0	2.498	.025
par rapport	6	0	2.498	.025
pas même	1	7	2.066	.05
plus WRDS plus	65	40	2.644	.01
pour ce qui est	7	1	2.178	.05
prince légitime	4	0	2.040	.05
puis que	0	44	6.505	.001
que ce soit	12	4	2.079	.05
quelque un	2	16	3.217	.005
repos public	5	0	2.281	.025
si WRD WRDS si	22	55	3.589	.001
supposé que	0	5	2.192	.05
tant que	11	3	2.212	.05
un tel	5	18	2.617	.01

EXSOR statistical results, the smallest acceptable t-value considered significant is 1.96. Only variables whose comparison of use revealed a value of 1.96 or greater are tabulated. The final column, denoted CL, records confidence levels at which observed differences are admitted. Fifty-seven percent of the calculated t-values in this table yielded confidence levels of better than .025. Thirty, or forty-three percent,

of the sixty-eight word combinations revealed a five-in-one-hundred probability that observed differences are due to sampling error.

If Mosteller and Wallace's criteria for function word selection were applied to the list of expressions in Table 4:44, six word combinations marked with an asterisk (*) would have to be eliminated, since their use might be related to context. Since l'Avis and Cabale treat the same subject, it seemed likely that their "content" expression usage should be similar; therefore these six expressions were included. Nevertheless, four of the six--prince légitime, repos public, le droit d'examen, and esprit de satire--were not used at all by Bayle. On the other hand, le public and il est, the remaining two content-related expressions, were used much more often by Bayle than by the unknown author as illustrated in the columns listing absolute frequencies.⁴³

Table 4:45 records sixty-nine expressions whose quantitative use in Cabale Chimérique differed significantly from their use in Le Prosélyte abusé.

Table 4:42 presented a list of expressions whose comparative use in test articles suggested a stylistic similarity between Larroque and the author of l'Avis aux réfugiés. Table 4:43 revealed a similar, although somewhat longer list which allied Bayle and the unknown author. In order to suggest stylistic similarity and thereby propose either Larroque or Bayle as probable author of the disputed text, it was necessary that a significant difference first exist between Bayle and Larroque for the variable in question.⁴⁴ Table 4:45 contains those expressions which yielded a statistically significant difference of use with an .05 or better degree of probability when Cabale (B1) and Le Prosélyte (L1) were compared. The columnar structure for this table is

TABLE 4:45. EXPRESSIONS, B1 \neq L1

Expression	B1 Absolute	L1 Freq	t-Value	CL
à cause que	13	0	2.854	.005
à ce que	11	23	3.492	.001
à moins que	0	3	2.189	.05
à moins que WRDS ne	0	3	2.189	.05
à peu près	0	3	2.189	.05
à propos	2	10	3.191	.005
à-propos	9	0	2.374	.025
au contraire	4	13	3.217	.005
avant que	19	1	3.080	.005
avant que de	11	1	2.148	.05
bien public	7	0	2.094	.05
bien que	14	26	3.444	.001
car WRDS si	25	29	2.295	.025
cause que	17	1	2.873	.005
c'est	127	123	3.478	.001
c'est que	29	37	2.931	.005
c'est WRDS ce que	10	14	1.996	.05
ce ne est pas	15	2	2.266	.025
ce qui	77	79	3.116	.005
ce qui est	10	14	1.996	.05
ce que	147	148	4.122	.001
ce que c'est que	7	0	2.094	.05
c'est-à-dire	14	0	2.961	.005
comme si	5	10	2.241	.025
dans le tems que	1	5	2.256	.025
d'ailleurs que	0	3	2.189	.05
de même	4	9	2.276	.025
de même que	2	7	2.421	.025
de rien	10	0	2.503	.025
depuis que	14	2	2.138	.05
en cas que	5	9	1.982	.05
en conscience	0	8	3.574	.001
en effet	1	8	3.106	.005
en WRDS lieu	15	21	2.445	.025
encore de	1	6	2.567	.025
étant que	0	3	2.189	.05
fort mal	0	3	2.189	.05
il est WRDS à	37	12	2.017	.05
il étoit	22	5	2.135	.05
il falloit	0	6	3.095	.005

TABLE 4:45. (Continued)

Expression	B1 Absolute	L1 Freq	t-Value	CL
il sera	7	0	2.094	.05
jusqu'à ce que	1	5	2.256	.025
l'autre	37	45	3.047	.005
le public	38	2	4.356	.001
mal à propos	0	5	2.826	.005
ne WRD point	73	78	3.322	.001
ne WRD WRDS point	132	120	2.976	.005
ne WRDS plus que	8	14	2.422	.025
ne WRDS que de	37	12	2.017	.05
ne WRD WRDS que	686	514	3.098	.005
ne WRDS que parce que	1	5	2.256	.025
non pas que	1	5	2.256	.025
non plus	3	10	2.846	.005
où l'on	3	19	4.613	.001
ou que	3	7	2.047	.05
par conséquent	5	18	3.918	.001
par rapport	0	4	2.527	.025
parce que	23	70	7.288	.001
plus WRDS plus	40	9	2.898	.005
plus WRD WRDS que	178	77	2.732	.01
puis que	44	84	6.309	.001
que de	91	33	2.724	.01
quoi que	18	29	3.268	.005
quoi que WRDS que	11	19	2.795	.01
sur tout	14	24	3.123	.005
tant que	3	7	2.047	.05
tant WRD WRDS que	44	15	2.066	.05
un tel	18	0	3.358	.001

the same as for Table 4:41 described above except for author/article denominators B1 and L1. Visual observations of absolute frequency columns may be misleading because figures presented do not compensate for differences in article lengths. For example, when the fact that article B1 contains 57,630 words and L1 36,201 is taken into consideration the already observable difference between absolute frequencies of bien que

becomes more pronounced. In a like manner, when two frequencies, e.g., for indefinite relatives ce qui and ce que, appear nearly equal, compensating for different article lengths may, in fact, reveal a significant difference. Calculated t-values and corresponding confidence levels affirm the presence of a quantitatively significant difference of use for variables listed. Seventy-one percent, or forty-nine, of sixty-nine B1 \neq L1 expressions yielded significance at better than the .05 level.

Table 4:46 lists eight expressions whose quantitative use in l'Avis aux réfugiés differed significantly from similar use in Cabale Chimérique and Le Prosélyte abusé.

As in other tables of EXSOR results, word combinations which revealed significant differences of use are listed alphabetically in the expression column. Table 4:46 t-values which attest to observed differences are listed under symbols which characterize comparisons made.

X:B1 symbolizes a comparison of values from l'Avis to those from Cabale;

X:L1 represents values from l'Avis compared to those from Le Prosélyte.

Next to each t-value is listed the level of confidence (CL) at which the difference was observed. In the column denominated "Least Significant" are listed author/article combinations (X:B1 or X:L1) whose observed differences, even though within the acceptable .05 range, yielded a lesser degree of confidence. For example, the t-value calculated from the comparison of the combination où l'on in X and L1 yielded a confidence level of .005, or five chances in one thousand that the two test values came from the same population. For the X and B1 comparison, on the other hand, the confidence level is .05. This five in one hundred figure suggests a higher probability of error and is, therefore, less significant than the .005 level. For this reason the symbol X:B1 is

TABLE 4:46. EXPRESSIONS, X \neq B1 OR L1

Expression	X:B1	CL	X:L1	CL	Least Significant
autant wrds que	3.459	.001	3.348	.001	Equal
jusqu'à	2.453	.025	2.034	.05	X:L1
ne WRD WRDS pas	3.159	.005	2.169	.05	X:L1
ne WRD WRD WRDS pas	2.383	.025	2.622	.01	X:B1
ne WRD WRD WRDS plus	3.361	.001	3.122	.005	X:L1
où l'on	2.013	.05	2.873	.005	X:B1
pas même	2.066	.05	3.035	.005	X:B1
plus WRDS plus	2.644	.01	4.805	.001	X:B1

listed in the Least Significant column. The word "equal" in this column means that the observed differences can be accepted with the same degree of confidence for each of the comparisons made. A summary of the notations in this column reveals four expressions, ne WRD WRD WRDS pas, où l'on, pas même, and plus WRDS plus, whose observed difference in the X:L1 comparison can be accepted with greater confidence than in the X:B1 comparison. Similarly, there are three word-patterns, jusqu'à, ne WRD WRDS pas, and ne WRD WRD WRDS plus, whose comparison of quantitative use in X and B1 revealed a greater degree of confidence. For only one expression, autant WRDS que, were the confident levels equal.

Table 4:47 records forty-three expressions whose quantitative use in the preface to l'Avis aux réfugiés differed significantly from their use in the text itself.

TABLE 4:47. EXPRESSIONS, X \neq P

Expression	X Absolute	P Freq	t-Value	CL
à bon droit	0	1	4.031	.001
à l'égard du	1	1	2.675	.01
attendu que	0	1	4.031	.001
au dela de	2	1	2.041	.05
avoir suspecte	0	1	4.031	.001
c'est ce qui	1	2	4.512	.001
c'est WRDS ce qui	3	2	3.273	.005
ce ne sont pas	1	1	2.675	.01
ce que c'est que	1	1	2.675	.01
ce qui	68	11	3.094	.005
dans le cas	2	1	2.041	.05
de part et d'autre	0	1	4.031	.001
dès avant que	0	1	4.031	.001
depuis peu	3	2	3.273	.005
devant cela	0	1	4.031	.001
Dieu aidant	0	1	4.031	.001
en WRD WRDS que	334	35	3.042	.005
en WRD WRDS venue	2	1	2.041	.05
entre autres choses	2	1	2.041	.05
étant en lieu de	0	1	4.031	.001
faute de	1	1	2.675	.01
faux pas	1	1	2.675	.01
il est question	2	1	2.041	.05
il étoit	7	2	2.109	.05
il sera	1	2	4.512	.001
jusques dans	2	2	3.784	.001
lors que	0	1	4.031	.001
ne point	5	2	2.579	.01
ne WRDS rien WRDS que	27	5	2.380	.025
pendant que	13	3	2.218	.05
pire WRDS que WRDS ne	2	1	2.041	.05
puis que	0	2	5.701	.001
quant à	1	1	2.675	.01
quelqu'un	2	1	2.041	.05
quoi de	0	1	2.041	.05
secouer le joug	2	1	2.041	.05
si WRD WRDS si	22	4	2.093	.05
si WRD WRDS si WRD WRDS si	5	3	3.838	.001
supposé que	0	1	4.031	.001
sur quoi	5	3	3.838	.001
tel que	2	1	2.041	.05
tel qui	0	1	4.031	.001
un tas de	1	1	2.675	.01

The quantitative use of twenty-six function words, listed in Table 4:34, provided evidence in support of the theory that the preface to l'Avis aux réfugiés and the text proper were not written by the same person. Forty-three word-patterns, listed alphabetically in the expression column of Table 4:47, lend additional substance to this thesis. When comparing the number of times each expression occurred, consideration must be given to the fact that the two test works, X and P, are of different lengths. Figures in the absolute frequency columns record the actual number of times each word-pattern was used. The final two columns contain calculated t-values and corresponding confidence levels for observed differences. Expressions marked with an asterisk (*) occurred so infrequently that, even though their statistics show significance, their value as discriminators is questionable. Nevertheless, because each of them occurred only in the preface, their distinguishing value cannot be totally discredited.

Verbal Summary of Individual Expressions

Table 4:48 is a verbal summary of Tables 4:42 through 4:47 and presents four lists of expressions whose frequency of use in l'Avis, Le Prosélyte, and Cabale suggests similarity ($X = L1$ and $X = B1$) or demonstrated a significant difference ($X \neq B1$ or $L1$ and $X \neq P$).

Column headings of Table 4:48 symbolize authors and works whose quantitative use of 554 expressions was compared. Denominators X, L, B, and P correspond to the unknown author, Larroque, Bayle, and author of the preface respectively. Expressions in $X = L1$ and $X = B1$ columns satisfied conditions for naming a probable author prescribed on pages 93 and 94. Therefore, by virtue of the either Bayle or Larroque

TABLE 4:48. SUMMARY OF SIGNIFICANT EXPRESSIONS

X = L1	X = B1	X ≠ B1 or L1
à propos avant que avant que de depuis que en effet il étoit le public ne WRD point ne WRD WRDS point non plus	à ce que bien que car WRDS si c'est c'est que ce que ce qui comme si de même de même que	autant WRDS que jusqu' à ne WRD WRDS pas ne WRD WRD WRDS pas ne WRD WRD WRDS plus où l'on pas même plus WRDS plus
ou que tant que	l'autre ne WRDS plus que ne WRD WRDS que non pas que par conséquent parce que plus WRD WRDS que que de quoi que quoi que WRDS que sur tout tant WRD WRDS que	
Total = 12	Total = 22	Total = 8
X ≠ P	X ≠ P	X ≠ P
à bon droit à l'égard du attendu que au dela de avoir suspecte c'est ce qui c'est WRDS ce qui ce ne sont pas ce que c'est que ce qui dans le cas de part et d'autre dès avant que depuis peu devant cela	dieu aidant en WRD WRDS que en WRD WRDS venue entre autres choses étant en lieu de faute de faux pas il est question il étoit il sera jusques dans lors que ne point ne WRDS rien WRDS que pendant que	pire WRDS que WRDS ne puis que quant à quelqu' un quoi de secouer le joug si WRD WRDS si si WRD WRDS si WRD WRDS si supposé que sur quoi tel que tel qui un tas de
Total = 43		

hypothesis, expressions which point to one of the hypothesized writers identify him with the unknown author with respect to those particular variables. A summary of significant function words, Table 4:35, lists test items which point to more similarities in X and B than in X and L. Likewise, the summary of significant expressions, Table 4:48, records twenty-two expressions identifying Bayle with the author of l'Avis, opposed to twelve which link Larroque to the disputed text. A comparative analysis of the X = L1 and X = B1 lists reveals that 50%, or six expressions used in common by X and Larroque, are conjunctive in nature; three are negations; one is prepositional; and two--il était and le public--are most likely related to context. On the other hand, eighteen, or 82% of X = B1 word patterns are connectives. The remaining four expressions form no apparent pattern.

In the X \neq B1 or L1 column of Table 4:48 are listed eight word combinations whose quantitative use differed in all three major test articles. Four of these eight expressions are negative forms; four are connectives. The import of the patterns into which listed word groups fall is discussed in the EXSOR grouped-data analysis. The final column of Table 4:48 lists all expressions from l'Avis and its preface which were observed to have a significant difference of quantitative use. They number forty-three. Again, those expressions whose use might depend on content are marked with an asterisk (*).

Statistical Summary of Individual Expressions

Table 4:49 presents a statistical summary of probable author and author/article combinations for which the hypothesis-of-no-difference was tested using 554 expression variables. The summary also includes

the number of expressions tested in each combination whose confidence levels ranged from .001 to .05.

TABLE 4:49. EXSOR INDIVIDUAL EXPRESSIONS--STATISTICAL SUMMARY

	X = L1*	X = E1*	X:B1
	2 @ .001 CL	4 @ .001 CL	9 @ .001 CL
	1 @ .005 CL	7 @ .005 CL	6 @ .005 CL
	1 @ .01 CL	3 @ .01 CL	7 @ .01 CL
	2 @ .025 CL	3 @ .025 CL	17 @ .025 CL
	6 @ .05 CL	5 @ .05 CL	30 @ .05 CL
Totals	12	22	69
	B1:L1	X ≠ B1 or L1*	X:P
	13 @ .001 CL	1 @ .001 CL	18 @ .001 CL
	19 @ .005 CL	1 @ .005 CL	4 @ .005 CL
	3 @ .01 CL	1 @ .01 CL	8 @ .01 CL
	16 @ .025 CL	1 @ .025 CL	1 @ .025 CL
	20 @ .05 CL	4 @ .05 CL	12 @ .05 CL
Totals	71	8	43

Probable author summary: B = 22, L = 12, N = 8.

*Confidence level reported is lesser of the two.

When the quantitative use of 554 word combinations was compared across the final test articles, statistically significant differences were observed which yielded significance levels that afforded a greater degree of confidence than the acceptable .05 level. Nevertheless statistical values from EXSOR data test results did not afford as high an average confidence level as FREQFUN results. Tables 4:42 through 4:48 present individual expressions whose quantitative use differed

significantly within a given author/article combination. Table 4:49 brings together the total number of word patterns from each author/article comparison significant at the .001, .005, .01, .025, or .05 confidence levels. Groups denominated $X = L1$ and $X = B1$ summarize statistical results which suggest Larroque or Bayle as probable author of l'Avis aux réfugiés. In these two groups, as in the $X \neq B1$ or $L1$ comparisons, the degree of probability figure (CL), as explained on page 146, and which allows the greater chance for error, is reported. A visual comparison of $X = L1$ and $X = B1$ significance levels again suggests Bayle, rather than Larroque, as probable author, with a substantially greater degree of confidence. Add to this the fact that Bayle had ten more word combinations in common with X than did Larroque, and Bayle's link with l'Avis becomes stronger than Larroque's. Even so, $X = L1$ and $X = B1$ results, based as they are upon the either Bayle or Larroque hypothesis, merely show which of the two is closer to the unknown author in his use of a given variable. They do not show how close either is to the disputed text. For that reason, summary results from the comparisons $X:B1$, $B1:L1$, and $X \neq B1$ or $L1$ are also presented in Table 4:49. As in the function word results, t-tests of difference revealed more than three times as many expressions used differently in X and B1 than were used similarly. The probable author summary at the bottom of the table, based solely upon the $X = L1$, $X = B1$, and $X \neq B1$ or $L1$ figures, reveals that Bayle's quantitative use of expressions, more so than Larroque's, resembles that of the author of l'Avis aux réfugiés. Twelve expressions suggest Larroque, and twenty-two point to Bayle as the probable author. Nevertheless, because three times as many, or sixty-nine, word patterns were used differently by Bayle and X, it would

be difficult to conclude that Cabale Chimérique and l'Avis aux réfugiés are works by the same author.

EXSOR-Grouped Data

Utilization of categorical groupings of function words as stylistic discriminants was discussed on pages 152 and 153. In conjunction with the rationale already presented, two additional factors justify the procedure with respect to word-pattern usage. The first derives from the question: if an author's style is designated nominal or verbal, does it necessarily follow that adjective or adverbial locutions and phrases will also be dominant characteristics of his stylistic habits? Or will noun and verb word-patterns preponderate? In order to answer this question objectively, a quantitative comparison must be made, and in order to make such a comparison, a frequency count of each type of phrase or locution is required. The second factor involves orthographical and meaning changes in structurally similar expressions. For example, à l'égard de, à l'égard du, and à l'égard des differ in meaning from à son égard and à cet égard, but not from each other. Furthermore, the possessive adjective son and the demonstrative cet further define the essence of the à ... égard combination. Individual identification, enumeration, and classification are prerequisite, therefore, to the assignment of an expression to a part of speech or usage category. In addition to the content variability noted above, there also exists a difference in part of speech classification: à l'égard de, du, des are prepositional locutions in which the noun, égard, combines with the prepositions à and de, functions as a connective, and in doing so loses most of its nominal significance. On the other hand, the

expressions, à son égard and à cet égard import to the noun its fullest meaning. Nevertheless, Le Petit Robert classifies the à ... égard combinations as adverbial locutions⁴⁵ because of their function and position in the French grammatical/syntactical pattern. Grouping of individual expressions into part of speech or usage categories, then, affords the researcher opportunity to conduct both a quantitative and qualitative analysis of a writer's word-pattern habits.

In order to count occurrences of each expression described above, the literals à l'égard de, à l'égard du, and à l'égard des as well as the literal-variable combination à WRDS égard were included among the 554 word combinations tested. Because the variable, WRDS, permits the computer to skip any number of words within sentence parameters until it comes to the literal égard, all occurrences of the à l'égard group were registered in the à WRDS égard totals, as well as individually. By summing totals for the first group and subtracting them from the literal-variable expression total, an exact count was obtained for each group. This final figure was added to similar calculations in the appropriate part of speech or usage group.

Even though a concentrated effort was made to eliminate all duplicates, the sheer number of word combinations tested suggests the possibility that some figures might still be duplicated. For example, the preposition jusqu'à and the conjunction jusqu'à ce que were both tested separately. Nevertheless, since the computer recognizes characters and not parts of speech, each occurrence of jusqu'à was included in the totals for jusqu'à ce que. Hence, final figures for jusqu'à were adjusted before its values were combined in the appropriate preposition category. However, in the case of limiting negation word-patterns,

ne WRD WRDS que and ne WRDS rien WRDS que, because the variable, WRD, means at least one word--and WRDS means any number of words--must be present between literals ne ... que, it is conceivable that the frequency count of one of these two word-patterns might be inaccurate. The cause of the error, should it exist, would most likely be the presence of a pleonastic ne. Because the number of times that such an error is likely to appear is very small, and because the proportion of error remains constant for all test articles, printing of a key-word-in-context list for further refinement of this and similar variables did not seem warranted.

Inclusion of such parallel expressions as ne WRD pas, and ne WRD WRDS pas provided a means for examining how many times the adverbial negation was used with only a finite verb separating the literals, ne ... pas, and how often object pronouns or the adverb, même, were used in conjunction with the negated verb. The expression ne WRD WRDS pas recorded 393 occurrences in l'Avis aux réfugiés; whereas ne WRD pas occurred 298 times in the same article. Accordingly, it may be concluded that fifteen negated finite verbs in l'Avis aux réfugiés were accompanied by grammatical variables, e.g., object pronouns or même.⁴⁶ The figure 393, representing the absolute frequency of the literal ne ... pas, was combined with totals of all other adverbial negations for final statistical computations. The examples presented above are representative of procedures followed in developing the expression list used in this study. In addition, they illustrate some problems encountered as well as some advantages of including both individual and grouped word-patterns in a stylistic study.

Inasmuch as expressions obtained from a close reading of test articles were heterogeneous, it was impossible to limit their classification to the two general categories--"fixed expressions," such as en effet, and "variable expressions," such as à (l') (son) (cet) égard (de) --originally anticipated. As the list grew, the need for a more refined, standardized classification procedure became evident. Grevisse again served as a source and supplied six--negations, adverbial, conjunctive, interjective, prepositional, and verbal locutions--of nine broad categories.⁴⁷ The remaining three groups--adjective, noun, and pronoun phrases--developed as word-patterns, which did not precisely fit into one of the standard categories, were noted from the test articles and grouped according to usage or structure. In turn, conjunctive, prepositional, verbal, and negations groups were further sub-divided as follows:

CONJUNCTIVE PHRASES

- a. Coordinating conjunctions
- b. Subordinating conjunctions of cause, comparison, concession, condition, consequence, purpose, time

PREPOSITIONAL PHRASES

- a. Prepositional phrases which terminate with a noun
- b. Prepositional phrases which terminate with a verb
- c. Prepositional phrases which terminate with either a noun or a verb
- d. Miscellaneous prepositional phrases⁴⁸

VERB PHRASES

- a. Verbs followed by a conjunction
- b. Verbs with impersonal pronoun subject
- c. Miscellaneous verb constructions⁴⁹

NEGATIONS

- a. Adjectival negations
- b. Adverbial negations
- c. Connective and limiting negations
- d. Pronominal negations

After quantitative use of each sub-division listed above was compared across the final test articles, individual frequencies for nine broad categories were totaled and fed again into the computer. As it had done with each sub-division, the machine performed statistical t-tests on nine major categories.

Table 4:50 presents a general overview of statistical results derived from quantitative comparisons of EXSOR grouped data. Four additional Tables, 4:51 through 4:54 complete reporting of EXSOR results.

Twenty-nine entries in the column headed "Expression Group" represent groups of word-patterns whose quantitative use was compared across five final test articles. Abbreviations "C C" and "S C" stand for coordinating conjunctions and subordinating conjunctions. Likewise, "E" and "F" represent results obtained from EXSOR and FREQFUN data respectively, and "E + F" means EXSOR and FREQFUN data were combined. Once again, denominators $X \neq B1$, $X \neq L1$, and $B1 \neq L1$ in the second, third and fourth columns symbolize statistically significant differences observed when word-pattern groups were compared. Even though ten comparisons were made for each variable, only those which most directly bear on the authorship of l'Avis aux réfugiés are reported in this table.⁵⁰ Figures in columns two, three and four are calculated t-values which affirm the observed difference. A dash (-) in the $X \neq B1$, $X \neq L1$, or $B1 \neq L1$ columns means that no significant difference was observed. Stated affirmatively, the dash implies, but does not certify, similarity in quantitative use for its corresponding variable. Eleven of twenty-nine test groups revealed significant differences between Bayle and the unknown author. Six of these groups--adverbial phrases, comparatives and superlatives, subordinating conjunctions of cause and

TABLE 4:50. EXSOR GROUPED DATA RESULTS

Expression Group	X ≠ B1	X ≠ L1	B1 ≠ L1	Least CL	Sign Comb	Prob Auth
Adj phrases	-	2.206	-	.025	L1:B1,X	-
Adv phrases	5.474	-	3.065	.005	B1:L1	L
Inter phrases	-	-	-	-	-	-
Noun phrases	2.808	5.508	3.394	.005	X:B1	N
Pronoun phrases	2.296	3.787	-	.025	X:B1	N
Comp and superlatives	4.030	6.011	2.621	.01	B1:L1	N
C C - EXSOR	2.104	-	-	.05	L1:B1,X	-
C C - EXSOR & FREQFUN	-	4.548	4.593	.001	Equal	B
S C - cause	4.625	11.190	7.206	.001	Equal	N
S C - comparison	-	-	-	-	-	-
S C - concession	1.972	3.849	2.072	.05	X:B1	N
S C - condition	-	-	-	-	-	-
S C - consequence	-	-	3.124	.005	X:B1,L1	-
S C - purpose	-	2.311	-	.025	L1:B1,X	-
S C - time	4.961	1.980	2.478	.05	X:L1	N
S C - all EXSOR	5.045	10.418	5.871	.001	Equal	N
S C - EXSOR & FREQFUN	-	8.154	8.078	.001	Equal	B
Misc prepositionals	-	2.677	2.503	.025	B1:L1	B
Prep plus noun	4.070	3.460	-	.001	Equal	N
Prep plus verb	-	-	-	-	-	-
Prep plus noun or verb	-	-	-	-	-	-
Verb plus conj	-	2.960	2.448	.025	B1:L1	B
Verb plus imper pron	-	-	-	-	-	-
Verbs - misc const	3.164	2.725	-	.01	X:L1	N
Adjectival neg	-	-	-	-	-	-
Adverbial neg	-	2.496	2.877	.025	X:L1	B
Connective neg	-	-	-	-	-	-
Pronominal neg	-	-	-	-	-	-
All negations	-	2.244	2.433	.025	X:L1	B
Totals	B = 6	L = 1	N = 9	Insignificant = 13		

Confidence level shown is for the least significant t-value tabulated.

time, all EXSOR subordinating conjunctions combined, and prepositions followed by nouns--yielded differences at the .001 confidence level. Only two--EXSOR coordinating conjunctions and subordinating conjunctions

of concession--recorded the lower, but still acceptable .05 level. Confidence levels for the remaining three significant variables fall in the .01 - .005 range.

Figures in the $X \neq L1$ column affirm the presence of a significant difference between Larroque and the unknown author with regard to their quantitative use of seventeen grouped data variables. Thus, 38% of the variables listed in Table 4:50 distinguished Bayle, and 58% separated Larroque from the unknown author. These results again suggest that even though the writings of both authors vary significantly from l'Avis aux réfugiés, Larroque's writing habits differ more than Bayle's. Information provided in the final two columns of Table 4:50 further corroborates this observation. Numbers in the least significant confidence level combination column represent the significance levels (CL) which acknowledge divergent usage of the corresponding variable with the least amount of confidence. Next to each confidence level value is the author/article combination which produced it. In the probable author column, the letters B, L, and N again represent Bayle, Larroque, and neither Bayle nor Larroque. A dash (-) in this column means that there were no statistically significant differences observed, or that conditions for attribution prescribed on pages 93 and 94 were not satisfied. The probable author summary totals show six variables which point to Bayle as author of l'Avis aux réfugiés: only one suggests Larroque. On the other hand, nine of the grouped-data variables imply that neither of the hypothesized writers wrote the controversial pamphlet.

In order to discover whether or not the significant EXSOR grouped-data variables formed a general part of speech or usage pattern, each variable for which a quantitative difference was observed was

listed separately, according to probable author. Tables 4:51 and 4:52 present groups whose divergent use in X:B1, X:L1, and B1:L1 comparisons suggest Bayle and Larroque respectively as authors of l'Avis aux réfugiés.

TABLE 4:51. EXPRESSION GROUPS WHICH SUGGEST BAYLE AS AUTHOR OF L'AVIS AUX REFUGIES

Expression Group	X \neq L1	CL	B1 \neq L1	CL
Coordinating conj. - E + F	4.548	.001	4.593	.001
Subordinating conj. - E + F	8.154	.001	8.078	.001
Misc. prepositionals	2.677	.01	2.503	.025
Verb + conjunction	2.960	.005	2.448	.025
Adverbial negations	2.496	.025	2.877	.005
All negations	2.244	.025	2.433	.025

As in previous tables presenting EXSOR results, the test variable is listed in the left-hand column. The other four columns contain statistical results derived from frequency of use of test variables in articles symbolized by column headings X \neq L1 and B1 \neq L1. Columns denominated CL tabulate the degree of probability or confidence level at which observed differences can be accepted. A cursory review of six test variables listed in the expression-group column reveals three of nine broad categories of expression classification: conjunctives, prepositionals, and negations. Nevertheless, as was mentioned earlier, four prepositional locutions which constitute the miscellaneous-prepositionals group are conjunctive in nature. When this fact is taken into consideration, the X = B1 grouped-data list shows greater

consistency. The number of stylistic variables Bayle and X have in common is reduced. Only one expression group suggested similarity between Le Prosélyte abusé (L1) and l'Avis aux réfugiés: adverbial phrases. Table 4:52 presents the statistical results which confirm a difference of quantitative use of this variable when article combinations X:B1 and B1:L1 were compared.

TABLE 4:52. EXPRESSION GROUPS WHICH SUGGEST LARROQUE AS AUTHOR OF L'AVIS AUX REFUGIES

Expression Group	X \neq B	CL	B1 \neq L1	CL
Adverbial phrases	5.474	.001	3.065	.005

Statistical results recorded in Table 4:52 corroborate an already observed difference. FREQFUN grouped-data results revealed an overall difference in adverb use between X and B1, which yielded an .001 confidence level. The 5.474 t-value in the X \neq B1 column of Table 4:52 also yields a .001 level of significance. The fact that the adverbial phrase group includes, by far, the greatest number of test expressions (143) adds to the significance of this observed stylistic difference between Cabale Chimérique (B1) and l'Avis aux réfugiés (X).

Table 4:53 is a summary of all expression groups which registered significant differences of quantitative use when X was compared to B1 and to L1.

Column denominators in Table 4:53 are the same as those used in other tables which report FREQFUN and EXSOR results. Nine EXSOR grouped-data variables, listed in the expression group column, disclosed significant differences in X to B1 and X to L1 comparisons. Columns headed

TABLE 4:53. EXPRESSION GROUPS WHICH SUGGEST THAT NEITHER BAYLE NOR LARROQUE WROTE L'AVIS AUX REFUGIES

Expression Group	X \neq B1	CL	X \neq L1	CL
Noun phrases	2.808	.005	5.508	.001
Pronoun phrases	2.296	.025	3.787	.001
Comparatives & superlatives	4.030	.001	6.011	.001
Subord. conj. - cause	4.625	.001	11.190	.001
Subord. conj. - concession	1.972	.05	3.839	.001
Subord. conj. - time	4.961	.001	1.980	.05
Subord. conj. - all EXSOR	5.045	.001	10.418	.001
Prepositions + noun	4.070	.001	3.460	.001
Verbs - miscellaneous	3.164	.005	2.725	.01

X \neq B1 and X \neq L1 contain calculated t-values which attest to observed differences of quantitative use. Next to each t-value is its corresponding significance level. Even though only nine, or thirty-one percent, of grouped-data variables disclosed differences, these nine variables represent six of nine broad categories delineated on page 185. Moreover, fourteen significant values, yielding confidence levels of .005 - .001 (one to five chances in one thousand for error), provide stable internal evidence that both Bayle and Larroque differ from the unknown author.

Statistical Summary of Grouped Expressions

Table 4:54 is a statistical summary of confidence level values for EXSOR grouped-data probable author designations. The X = L1 column

lists the least significant confidence level recorded in Table 4:52. Since only one variable, adverbial phrases, satisfied prescribed conditions for suggesting Larroque as probable author of l'Avis aux réfugiés, only one of five significance level values is represented. Column X = B1 records confidence level figures for six expression groups which met conditions required to advance Bayle as probable author. Finally, nine values, representative grouped word-patterns which indicate that neither Bayle nor Larroque authored the disputed text, comprise the X \neq B1 or L1 column. All confidence levels reported in Table 4:54 correspond to figures in the Least Sign CL Comb column of Table 4:50 and are listed in descending order from most to least significant. The probable author summary, represented by the symbols L = 1, B = 6, and N = 9, again indicates that neither Bayle nor Larroque wrote l'Avis aux réfugiés. Moreover, an analysis of the confidence levels recorded in Table 4:54 substantiates this conclusion.

TABLE 4:54. EXSOR GROUPED EXPRESSIONS PROBABLE AUTHOR SUMMARY

	X = L1	X = B1	X \neq B1 or L1
	0 @ .001	2 @ .001	4 @ .001
	1 @ .005	0 @ .005	1 @ .005
	0 @ .01	0 @ .01	1 @ .01
	0 @ .025	4 @ .025	1 @ .025
	0 @ .05	0 @ .05	2 @ .05
Totals	1	6	9
Summary:	L = 1	B = 6	N = 9

Conclusions for EXSOR Data

Thus, like FREQFUN individual and grouped-data findings, results of EXSOR grouped-data tests provide additional internal evidence--choice and quantitative use of word-patterns--against the claim that Bayle wrote l'Avis aux réfugiés. The second series of tests reported in this section, EXSOR individual expressions, listed twenty-two word combinations which advanced Bayle as author of l'Avis; twelve suggested Larroque and eight neither Bayle nor Larroque. Nevertheless, because three times as many individual expressions distinguished X and B1 as indicated similarity, test results from this section (individual expressions) tend to substantiate, rather than negate, the conclusion suggested above.

Vocabulary Analysis

Each section of this chapter has dealt, in its own way, with vocabularies of test articles. Part one treated lengths of individual vocabulary terms and combined quantitative relationships between letters, syllables, and words as sentence components. Attention at that point was focused on elements which evoke rhythm, rather than meaning. Part two concentrated on individual word use as applied to grammatical patterns of sentence first and last words. Section three again treated only part of the total vocabulary as frequencies of a selected group of function words were compared. Finally, part four reported comparative measures of specified word-patterns. It is now time to examine vocabulary relationships on a type-token basis.⁵¹ Methods and results reported in this section differ from those in each of the preceding ones

in two ways: first, all previous tests were performed on averages or frequencies weighted by article length. As a result, their statistical value as stylistic discriminators could be verified by a t-test. Tests performed on data described in this section, however, (with the exception of the last two comparisons), are not based on averages or relative frequencies, and, in some cases, are completely independent of article length. For these reasons, the t-test, valid only when testing differences in treatment means, was utilized only in the last two comparisons reported in this section. Instead of the t-measure of difference, individual statistical values were generated for the entire type-token vocabularies of each test article as suggested by Miller,⁵² Guiraud,⁵³ Herdan,⁵⁴ and Yule.⁵⁵ The fact that figures representing total (in lieu of partial) type and token vocabularies are utilized in between-author comparisons constitutes the second way in which this section differs from earlier ones.

In order to convert raw data (natural language texts) into a more viable form for the type-token analyses reported below, the ENROOT program, described on pages 62-65, was utilized. ENROOT generated a list of word roots from each test article. In addition, the program registered and caused to be printed each word in the text, complete with its frequency of occurrence--in conjunction with its corresponding root. Appendix Z provides a sample of ENROOT output. Roots and derivations were then totaled for each author. These data, besides alluding to a measure of vocabulary richness for each author, served also as input data for two established statistical measures of comparative word usage: logarithmic type-token ratios proposed by Herdan and Guiraud, and Yule's "characteristic K." In order to compare mathematically type-token

statistical values across five test articles, sophisticated statistical procedures which are beyond the scope and purpose of this project would be required. Therefore, even though statistical measures are utilized to put the final results into a format which permits comparison, the final judgment as to similarity or difference between compared works must be a subjective one.

Table 4:55 presents a summary of ENROOT totals which served as input data for the comparisons mentioned above.

TABLE 4:55. SUMMARY OF ENROOT COMPARATIVE DATA

Variable	P	X	B1	B2	L1
Total number of words	3,418	55,551	57,787	14,147	36,143
Total function words	2,884	44,400	48,572	11,765	30,143
Total content words	534	11,151	9,215	2,382	6,047
Total number of roots	695	2,689	2,426	1,376	1,800
Total number of variations	1,054	6,672	5,791	2,613	4,046

In the "Variable" column are listed five vocabulary information sources provided by ENROOT. Numeric values of each listed variable are found in columns denominated P, X, B1, B2, and L1. Each alphanumeric column heading refers to the final test articles as delineated on page 75. Values for the first three variables--total number of words, function words, and content words--were derived from SENWOL and FREQFUN results as well as from ENROOT. The fact that these three different sources provided concordant results served as an additional check on accuracy of the programs. Values for function words represent absolute

frequencies of occurrence for function words listed in Appendix K whereas figures showing content word frequencies are differences obtained from subtracting total function words from total word occurrences. No distinction was made between content and rare--only one or two total occurrences--terms. The line "Total number of roots" records the number of different roots used by each author, generated by ENROOT. The final line in Table 4:55 shows the total number of observed variations for all roots. In essence, this figure represents the active vocabulary of an author with respect to the work from which the values are drawn. To be sure, the number of different roots used is an indicator of vocabulary size for a writer. Nevertheless, an author who has command of several derivatives from the same root possesses a much richer vocabulary than one who adheres to or deviates but slightly from the lexical norm. An analysis of the figures in Table 4:55 reveals that 2689 roots in l'Avis represent 4.8%, 2426 roots in Cabale represent 4.1%, and 1800 roots in Le Prosélyte represent 4.9% of the total word occurrences for the respective texts. On the other hand, a similar comparison reveals that root variations comprise 12%, 10%, and 10.9% of the token vocabulary for the same three works. Even though the observed difference is not overwhelming, the fact remains that a difference does exist. The two-percent differential in variations between Bayle and the author of l'Avis represents a difference of 881 derivatives. This number becomes even more significant when one considers that Bayle's article, Cabale Chimérique, is more than two thousand words longer than l'Avis aux réfugiés.

Writing about vocabulary ratios, Herdan stated that "as the sample increases in size, more and more words will be brought into play,

so that for infinitely great samples the ratio of words will approach to the true ratio $W1/W2$,"⁵⁶ where $W1$ and $W2$ represent the words used by two different, mature writers. Using Herdan's work as a reference point, Pierre Guiraud developed a quantitative means for gauging the richness of an author's vocabulary in his well-known work, Les Caractères statistiques du vocabulaire.⁵⁷ Accordingly, his formula expresses "la richesse du vocabulaire, R " as a function of the vocabulary of a text (V) and the total number of words, or the length of the text (N). It is expressed mathematically:

$$R = \frac{V}{\sqrt{N}}$$

For texts of fewer than 15,000, or greater than 40,000, words he modifies the formula to read:

$$R = \frac{V}{\sqrt{N}} \pm \log \frac{L}{N}$$

The second formula was utilized in the present study for articles P, X, B1, and B2. Because article L1 fell within the 15,000-40,000 word limits, logarithmic correction factors were not necessary.⁵⁸ In all computations involving $\log L$, the figure 24,000, suggested by Guiraud as the lexical range of the average author, was used. Values of R were computed for each test article. First, values representing the number of roots, and second, the number of root variations were substituted for V . Results of these computations are found in Table 4:56.

According to Guiraud the figure 20.50 represents the vocabulary richness for the average author.⁵⁹ Values tabulated above seem to

TABLE 4:56. VOCABULARY RICHNESS OF FIVE TEST ARTICLES

	P	X	B1	B2	L1
Roots	13.82	12.94	11.82	11.73	9.47
Variations	20.52	29.84	25.70	22.19	21.27

substantiate the reasoning that, although the number of roots called into use by an author is an indicator of his control of the language, a more accurate index is his ability to use derivative forms. A comparison of the X, B1, and L1 values for the variations-index reveals that each of the writers tested exhibited a rich vocabulary, according to the norm proposed by Guiraud. Nevertheless, both root and variations computations substantiate already mentioned differences between X, B, and L.

Since the number of roots and the size of an author's active vocabulary, represented by the number of root derivatives he uses, are both measures of a writer's control of a language, it follows that they should be valid stylistic discriminants. When works to be tested are of different lengths, however, vocabulary type-token figures must, in some way be weighted to make them relative to article lengths. For past tests this process posed little difficulty because the frequency of a word, for example, was directly proportional to article length. On the other hand, the rate at which a writer introduces new roots does not remain constant, but decreases as length of the work increases. The vocabulary/article-length relationship, or type/token ratio has been observed and discussed by each of the linguists mentioned in this section. Of especial interest to researchers who apply the TTR to stylistic discrimination studies is the following statement by Herdan:

By the term 'type/token ratio' we understand the ratio of the vocabulary in a given text, or sample from it, to the total number of words comprised in the text or the sample. That quantity changes, in general, with the size of the piece of literary work, the vocabulary increasing with the text length--but by no means proportional to it--in such a way that the quantity decreases, on the whole, with increasing sample size. It cannot, therefore, in this form serve as a characteristic of style, which must be independent of the text length. The logarithmic type/token ratio, i.e. the log type/log token, on the other hand, remains sensibly constant for samples of different size from a given literary text and is, therefore, suitable to serve as a style characteristic.⁶⁰

Thus, since the logarithmic type/token ratio "remains sensibly constant," this principle was applied to ENROOT data drawn from works tested in this study. First, the log-ratio constant was computed for each test article.⁶¹ This value served as a point of reference, and the number of roots which should occur for a standard article size of 45,000 words was computed. The value of 45,000 words was chosen because it is midway between the lengths of B1 and L1. Moreover, a 45,000 word sample is large enough to eliminate errors due to small sample sizes. All calculations were performed on the CDC 6500 computer in order to obtain maximum accuracy.⁶² Final values, calculated for both roots and variations, are recorded in Table 4:57.

Column denominators for Table 4:57 are the same as for all tables in this section. The first two variables listed in the left-hand column denote the vocabulary type whose projected values were calculated. Corresponding predicted totals are listed under respective author/article column headings. The variable, "Roots peculiar to each author" does not represent a calculated logarithmic projection. Recorded values for this variable represent the absolute number of roots unique to P, X, B1, B2, and L1 as shown. There were, for example, 738

TABLE 4:57. LOGARITHMIC TYPE/TOKEN PROJECTED VALUES AND UNIQUE ROOTS

Variable	P	X	B1	B2	L1
Roots per 45,000 words	5524*	2309	2031	3301**	2103
Variations per 45,000 words	9559*	5630	4753	6774**	4807
Roots peculiar to each author	30	738	588	129	357

*Results unreliable because of small sample size.

**Reliability is questionable because of small sample size.

roots peculiar to l'Avis aux réfugiés, 588 unique to Cabale Chimérique, and 357 found solely in Le Prosélyte abusé. It seems logical to assume that such a large number of unique roots in X and B1--works of similar lengths and which ostensibly treat the same subject--would imply that they were not produced by the same hand. Had Bayle written both articles, root and variation values would most likely have been more nearly coextensive. Thus, figures in Table 4:57 readily affirm the pattern established in previous sections as well as in Tables 4:55 and 4:56: Bayle and Larroque differ significantly from X.

As reported in Chapter II (p. 30) of this study, G. Udny Yule developed a method for precise and objective measurement of a significant trait of literary style: word-frequency distribution. In an article published in the Journal of Applied Mathematics and Physics in 1955, Gustav Herdan defended Yule's statistical approach to vocabulary analysis:

Yule's 'Characteristic' K of the word-frequency distribution of a linguistic text is derived under the assumption that the occurrence of a word was governed by a law of chance . . .

However, the constant K can be derived without such an assumption, which has not only the advantage of obviating adverse criticism . . . , but of showing K to be an easily interpretable, useful and interesting characteristic of a linguistic text.⁶³

While affirming that the statistical theory is "admittedly difficult," Paul E. Bennett stressed that "the measure itself is quite simple to use."⁶⁴

Yule's "K" measures repetitiveness of the vocabulary of a given literary work, expresses the result as a simple numerical value--such as 116.39, 113.28, or 120.10--and is completely independent of article length.⁶⁵ Even though Yule and Bennett applied the characteristic "K" to frequency distribution of nouns, it may just as well be applied to verbs, adjectives, adverbs, function words, content words or even word roots. For that reason, and since a "particular style may be characterized by a constant relation between uniformity and diversity in the number of iterations of the items of vocabulary,"⁶⁶ Yule's formula was applied to four vocabulary variables whose frequencies were calculated by ENROOT.⁶⁷ Table 4:58 records the results obtained when ENROOT data were subjected to Yule's word-frequency distribution test.

TABLE 4:58. "K" CHARACTERISTICS OF ENROOT DATA

Variable	P	X	B1	B2	L1	Prob. Auth.
K char. on total words	116.39	113.28	120.10	116.14	112.35	L
K char. on roots	213.32	112.38	140.91	211.67	203.18	N
K char. on content words	33.37	38.97	104.56	77.37	85.28	N
K char. on function words	417.67	432.91	405.44	411.09	385.46	N

Column one, labeled, "Variable," lists four vocabulary categories--called "items" by Herdan--to which Yule's test was applied. Column denominators P, X, B1, B2, and L1 again refer to test articles described on page 75. A comparative analysis of values tabulated above revealed a striking similarity between X and L1 when the measure was calculated for the total token vocabulary. The characteristic for Bayle, (B1), on the other hand, diverged decidedly from each of the others and from X and L1 particularly. Accordingly, L1, symbolizing Larroque, was listed in the probable author column for this variable. Such was not the case, however, for the remaining three variables. In each case calculated values, based upon actual word-frequency distributions and not upon projections from text samples, disclosed a marked difference between X and both B1 and L1. For this reason, N, meaning neither Bayle nor Larroque, was recorded in the probable author column for each of the last three variables in Table 4:58. Evidence set forth in this series of tests is in harmony with results of other tests reported in this chapter. Furthermore, this test provides an objective measurement of one significant aspect of literary style: word-frequency distribution; and these results, as tabulated above, can serve only to suggest that Bayle did not write l'Avis aux réfugiés.

As noted above, (p. 194), two quantitative comparisons whose data source was the ENROOT program utilized the t-test: verbals vs. non-verbals and content words vs. function words. Table 4:59 presents results of these comparisons.

Column labels for Table 4:59 are the same as for other tables in this chapter which report t-test results: $X \neq L1$, $X \neq B1$, and $B1 \neq L1$ symbolize the presence of a statistically significant difference between

TABLE 4:59. RESULTS OF VERBAL TO NON-VERBAL AND CONTENT WORD TO FUNCTION WORD COMPARISONS

Variable	X \neq L1	CL	X \neq B1	CL	B1 \neq L1	CL	Prob. Auth.
Verbals to Non-Verbals	2.824	.005	4.851	.001	-	-	N
Content Words to Function Words	12.761	.001	18.090	.001	3.084	.001	N

authors/works represented, and CL stands for confidence level. Numbers in columns two through six are calculated t-values and corresponding levels of significance. The dash (-) in the B1 \neq L1 and CL columns for verbals to non-verbals means that Bayle and Larroque did not differ significantly in their use of this variable in articles tested. Once again a marked difference is observed between quantitative stylistic habits manifest by Bayle in Cabale Chimérique and by the unknown author in l'Avis aux réfugiés.

Summary of Test Results

Table 4:60 is a summary table which brings together results of five categories of test variables analyzed and reported in this chapter.

In the first column of Table 4:60 are listed test categories for which five corresponding computer programs were written. These programs generated data results described in this chapter and summarized in columns two through eight of Table 4:60. Figures in the column denoted "Number of Variables" represent the number of individual comparisons made within corresponding categories. The third column lists the number

TABLE 4:60. SUMMARY OF TEST RESULTS

Category	Number of Variables	Number of Discrim	X=B	X=L	X≠B	X≠L	X≠B or L
Sentence Level Measures	6	3	0	0	3	3	3
Sentence Beginnings and Endings	30	13	1	2	11	6	4
Function Words							
a) Individual	184	120	31	22	74	88	42
b) Grouped	29	25	6	3	17	21	13
Expressions							
a) Individual	554	129	22	12	68	69	8
b) Grouped	29	19	6	1	11	17	9
Vocabulary Analysis	11	10	0	1	9	9	8
Totals	843	319	66	41	193	203	87

of variables in each group whose difference of quantitative use yielded a confidence level of .05 or better. In the X = B and X = L columns are summarized the total variables in each category which suggested Bayle or Larroque as probable author of l'Avis aux réfugiés. The column labeled X ≠ B or L brings together the variables which demonstrated a statistically significant difference at at least the .05 confidence level when X was compared to both B and L. Likewise, columns denominated X ≠ B and X ≠ L list the variables in each category which disclosed marked differences between Bayle or Larroque and the unknown author. Finally, individual summary figures are totaled at the bottom of the table.

Not unexpectedly, the pool of 843 stylistic variables (potential discriminants) was reduced to 319 actual discriminants as their quantitative use was compared across the test articles. The fact that all 843

variables, or even half of them, were not discriminators is likely attributable to the fact that two or three prosateurs writing in the same language on similar topics use several stylistic elements commonly. Moreover, the more their family, religious and educational backgrounds resemble each other, the more likely are similarities of linguistic style to increase.⁶⁸ Approximately 64% of the stylistic elements examined showed no appreciable discriminating ability in works tested.⁶⁹ Variables whose generated t-values yielded confidence levels of better than .05 were considered stylistic discriminators; their quantitative use in the test articles was distinct enough to be a distinguishing feature, as opposed to those variables which were either common to all works tested or else were used too infrequently to register statistically.

Of discriminants which distinguished between X and B at the .05 confidence level, 73% indicated that Bayle did not write l'Avis aux réfugiés. At the .01 level, the percentage of discriminators drops to 68, but the chance for error is also reduced from five-in-one-hundred to one-in-one-hundred. Were only the discriminants at the .001 level (one chance in one thousand for error) to be considered, percentage of discriminants opposing Bayle's authorship would again be 72. Table 4:61 summarizes the percentage of significant variables which reject Bayle's and Larroque's authorship of l'Avis at the .05, .01, .005, and .001 confidence levels. Percentages in Table 4:61 are calculated from the total number of discriminants which distinguish the writings of Bayle or Larroque from l'Avis: $X = B$, $X \neq B$; $X = L$, $X \neq L$.

One might now be prompted to ask, "Just how meaningful are these results?" In a similar situation, Milic affirmed that "by consistent

TABLE 4:61. PERCENTAGE OF DISCRIMINANTS WHICH REJECT BAYLE AND LARROQUE AS PROBABLE AUTHORS OF L'AVIS AUX REFUGIES

	Significance Level			
	.05	.01	.005	.001
Bayle	73%	68%	70%	72%
Larroque	82%	86%	87%	91%

adherence to sound and standard procedures of objective demonstration, we can reach a conclusion in the form of a probability statement about which we can feel such certainty as is consistent with the soundness of the original assumptions and the reliability of the data. In other words, the conclusions based on objective procedure have all the reliability of statistical demonstration, which is almost always more than that of mere speculation and conjecture."⁷⁰ In the present study, selection of stylistic elements to be tested across l'Avis and works by Bayle and Larroque was based upon accepted procedures employed by successful past attribution studies in English literature.⁷¹ Moreover, subjection to statistical analysis of data gathered through mechanical counting methods followed established statistical techniques. These procedures provided an objective demonstration of distribution of stylistic elements in works tested. The soundness of the original assumptions for using the statistical tests employed was demonstrated in Chapter III (pp. 65-67). Finally, the assumption that either Bayle or Larroque wrote l'Avis aux réfugiés was based upon inferential internal and external evidence from past studies of the problem.

Notwithstanding the observed statistical differences between works tested, one sceptical of quantitative findings may argue that the "tone" or "spirit" of l'Avis is nevertheless Bayle's. But there is no convincing method of demonstrating a difference of feeling about a work. The subjectivity demanded in making such a judgment emphasizes once more the uncertainty of internal evidence which requires personal evaluation. Statistical results produced in this thesis are free from such subjective evaluation and are therefore more reliable: another researcher, using the same method, testing the same variables would arrive at the same results.

What, then, is the conclusion to be drawn from these statistical data? L'Avis aux réfugiés, though sharing some stylistic elements with Bayle's works, is nonetheless statistically distinguishable from them. The preface to l'Avis stands apart quantitatively from both the main text and Bayle's works. Finally, Larroque's works reveal even fewer similarities to l'Avis and its preface (with a greater degree of statistical confidence, as shown in Table 4:61) than do Bayle's. These test results represent strong, objective, internal evidence against Bayle's or Larroque's authorship of l'Avis aux réfugiés.

NOTES FOR CHAPTER IV

¹For example, with the advent of computers with diacritical marking capabilities, hand sorting and checking against key-word in context concordances could be minimized. Also, instead of hand sorting the statistically significant variables, cutting and pasting for final typing, a minor addition to the program could cause the machine to perform the task of rearranging the data for tabular presentation.

²Several authorship attribution studies have been made in English, but none have as yet appeared in French, and most certainly not with the degree of automation incorporated in this study.

³The analysis of variance tests the hypothesis-of-no-difference between author-means for selected variables.

⁴George W. Snedecor, Statistical Methods (Ames, Iowa: Iowa State University Press, 1962), p. 46.

⁵Except in the case of SENWOL and STYLBEND whose results are discussed on pp.

⁶Within this group of 100 sentences, Bayle refutes accusations Jurieu had leveled against him, challenges Jurieu to bring forth competent witnesses, defends the integrity of those who had sent him the information on "le projet de paix," while assuring the reader that this was his first contact with the project, and finally expresses positivism toward those who would be peace-makers as opposed to those who constantly seed trouble, criticize, and meddle in political affairs. OD, II, 649-651.

⁷According to Delvolvé, "... Bayle fit paraître un écrit où leur [les réformés] esprit d'intolérance, dont il avait commencé à sentir les atteintes, est dénoncé et sévèrement condamné: c'est la Réponse d'un nouveau converti à la lettre d'un réfugié.... sous le masque d'un nouveau converti ..." Religion, Critique, p. 176. Robinson, however, refutes this hypothesis. Bayle the Sceptic, p. 119.

⁸Arthur Sherbo, "The Uses and Abuses of Evidence," Evidence For Authorship, David Erdman and Ephim Fogel, eds. (Ithaca, New York: Cornell Univ. Press, 1966), p. 8.

⁹The significant values presented in Tables 4:3 through 4:5 show distinction between author X when compared to B and L1, not between B and L proper.

¹⁰In this argument, the proportion of use of the stylistic characteristic in question is represented by the names Bayle, Larroque, and the unknown author.

¹¹Three of the thirty variables subjected to this test failed: adjectives, articles, and conjunctions, when used as sentence beginnings. All others indicated homogeneity of authorship within the works known to be by Bayle and Larroque.

¹²The word "absolute" is used here in its statistical sense, meaning plus and minus signs are ignored. The same meaning applies to the discussions of statistical results for Tables 4:10 through 4:24.

¹³Georges Le Bidois and Robert Le Bidois, Syntaxe du français moderne, II 2e éd., (Paris: Editions Picard, 1967), p. 234. See also Grevisse, Le Bon Usage, p. 936.

¹⁴Punctuation in parentheses represents the actual punctuation of the text before it was keypunched. See pp. 21-22 of this study for discussion of keypunch punctuation.

¹⁵Mosteller and Wallace, Disputed Authorship, p. 38.

¹⁶See Chapter III, pp. 56-57 of this study for procedures used in deriving the list of French function words.

¹⁷Six combinations which were not included in Tables 4:29 through 4:41 compare quantitative use of function words between P and L1, P and B1, P and B2, B1 and B2, B2 and L1, and X and B2.

¹⁸Population in this case would be an author's literary corpus.

¹⁹Mosteller and Wallace, Disputed Authorship, p. 249.

²⁰The 2.496 t-value yields a confidence level of .025.

²¹See Chapter III, pp. 60-61.

²²Mosteller and Wallace, Disputed Authorship, p. 39.

²³Grevisse, Le Bon Usage, pp. 168-967.

²⁴Milic, A Quantitative Approach, p. 78.

²⁵Ibid.

²⁶Even authors having the same educational, social, and literary backgrounds are likely to differ in their use of pronouns if one discusses Protestant theology, and the other defends the position of a heretic with reference to Protestant theology.

²⁷According to Ascoli, Larroque had no apparent need to conceal his identity. In fact, he is purported to have often said, "J'ai dit [ou] j'ai prouvé cela dans mon Avis aux Réfugiés." Ascoli argues, however, that Larroque avowed authorship of the controversial tract in order to gain favor in the French religious/political circle. "Bayle et l'Avis aux réfugiés." pp. 531-533.

²⁸Mosteller and Wallace, Disputed Authorship, p. 45.

²⁹Sayce, pp. 48-52, 69-88.

³⁰Muller, Essai de Statistique Lexicale, pp. 125-129. Muller located such locutions as à peine, de grâce, and grâce à, in order to accurately classify them as parts of speech. Muller objected to the current application of locutions in indexes produced at le Centre de Besançon, not to their use as discriminators.

³¹Milic, A Quantitative Approach, pp. 204-225.

³²Muller, Essai de Statistique Lexicale, p. 126f.

³³Mosteller and Wallace, Disputed Authorship, p. 45.

³⁴Mosteller and Wallace, for example, added to a pool of basic function words established by Miller, Newman and Friedman. Ibid., p. 39.

³⁵Even with utmost care and concentration one's eyes and powers of observation fall short of perfection. For example, as I read the works to be tested, I carefully noted the quantifiable word combinations. The list was then key-punched on IBM cards and run through an alpha-numeric sorter. This process alphabetized the entire list and made identification of duplicates a simple matter. Invariably I found double, even triple listings.

³⁶Even with the large sample sizes utilized in this study, several expressions did not record frequencies high enough to be considered valid discriminators.

³⁷Le Bon Usage.

³⁸A Haase, Syntaxe française du XVII^e siècle, 5^e éd. (Munich: Max Hueber, 1965).

³⁹See note 36.

⁴⁰The word "expression", used in connection with this study, includes locutions, idioms, and other quantifiable word combinations tested.

⁴¹See Chapter III, pp. 51-55 for a description of this program and Appendix S for its flowchart.

⁴²The six author-article combinations which were not included in Tables 4:42 through 4:54 compare quantitative use of test expressions in P and L1, P and B, P and B2, B1 and B2, B2 and L1, and X and B2.

⁴³Direct comparison of absolute frequencies is justified in this case because l'Avis aux réfugiés and Cabale Chimérique are very nearly the same length.

⁴⁴This stipulation is required by the "either Bayle or Larroque hypothesis" described on pp.

⁴⁵Le Petit Robert, p. 545.

⁴⁶Had the expression ne WRDS meme pas been included, the part of speech delineation could have been even more refined.

⁴⁷Le Bon Usage, pp. 509-511, 759-1055.

⁴⁸Four phrases which form this category--en effort, entre autres, hors de doute, outre cela--all function in the text as connectives. Only two of them, entre autres and outre cela occurred frequently enough to register statistically when tested alone. See Table 4:50 for the frequency of use comparative results and Appendix V for the absolute frequencies recorded from each test article.

⁴⁹This group consists primarily of partial idiomatic expressions. Because the verb proper varied as to tense and person, it was impractical to program all possible verb forms; therefore, only elements which remain constant were counted.

⁵⁰In addition to exclusions recorded in note 42 above, X:P comparison results have been omitted from Table 4:50 because the preface is so short that accuracy of its EXSOR results is questionable.

⁵¹George A. Miller defined the type-token ratio as the ratio of the number of different words (types) to the total number of words (tokens) in a passage. Language and Communication (New York: McGraw-Hill, 1951), p. 122.

⁵²Ibid., pp. 88-94, 120-124.

⁵³Guiraud, pp. 19-71.

⁵⁴Gustav Herdan, Language as Choice and Chance (Groningen, Holland: Noordhoff, 1956), pp. 12-22; and Type-Token Mathematics (The Hague: Mouton & Co., 1960), pp. 21-42.

⁵⁵Yule, The Statistical Study, pp. 35-82.

⁵⁶Herdan, Language, p. 17.

⁵⁷Guirand, pp. 52-53, 69.

⁵⁸Because the total occurrences, rather than "mots forts" were used, the value of N, not 2N, was used in this study. See Guiraud, p. 68f.

⁵⁹Ibid., pp. 53, 69.

⁶⁰Herdan, Type-Token Mathematics, p. 26.

⁶¹Mathematically this constant is represented by the equation

$$K = \frac{\log \text{ roots}}{\log \frac{\text{total}}{\text{occurrences}}}$$

⁶²Calculations with logarithms can lead to rounding errors unless extreme caution is used. To avoid such errors, a special programming technique called "double precision" was used by Paul Gabriel as he wrote the program to obtain the values for this analysis. The double precision technique provides for computations with a thirty-six decimal-place accuracy. Even using this technique it is impossible to produce reliable results from a small sample size. For this reason, the values for the preface and B2 are accepted with a lesser degree of reliability than for the larger samples.

⁶³Gustav Herdan, "A New Derivation and Interpretation of Yule's 'Characteristic' K," Journal of Applied Mathematics and Physics, VI (1955), 332.

⁶⁴Bennett, "Statistical Measurement," p. 33.

⁶⁵Ibid., p. 35. See also Herdan, "New Derivation," p. 333.

⁶⁶Ibid., p. 334. See also Bennett, "Statistical Measurement," p. 34.

⁶⁷Yule's formula is expressed

$$K = 10,000 \left(\frac{S2 - S1}{(S1)^2} \right)$$

where S1 = $\sum fx$ or the number, $\sum fx$, of nouns, for example, which occurred x times in a given work; and S2 = $\sum fx^2$ or the frequency of term x times the square of the number of times term x occurred. The number, 10,000 is introduced simply to avoid the inconvenience of handling very small decimal figures. See Yule, Statistical Study, pp. 47, 57-68.

⁶⁸Both Bayle and Larroque were sons of intelligent, if not scholarly, Protestant ministers. Each received his formative education under Protestant tutelage, and each reviewed for the Nouvelles de la république des lettres, although Bayle's experience in this latter capacity greatly exceeded Larroque's. See Labrousse, Pierre Bayle I, pp. 18-22, 122, 219 f59.

⁶⁹Of the 165 function words Mosteller and Wallace started with in their study of the Federalist Papers, only 30 were discriminants. Thus, 82% of their test variables were rejected for having no appreciable discriminating ability. Disputed Authorship, p. 67.

⁷⁰Milic, Quantitative Approach, p. 244.

⁷¹O'Donnell, "Stephen Crane's The O'Ruddy, a Problem in Authorship Discrimination," The Computer and Literary Style; Mosteller and Wallace, Disputed Authorship; Ellegård, A Statistical Method; Milic, Quantitative Approach, for example.

CHAPTER V

SUMMARY AND CONCLUSIONS

The principal aim of this investigation has been to discover whether or not Pierre Bayle wrote l'Avis aux réfugiés. To this end, a series of computer programs modeled on those used successfully in author identification studies in English literature were developed. New elements were indexing of French function words and word-patterns and analyzing of root-and-variation vocabularies found in works by Bayle and Daniel de Larroque--the most frequently mentioned candidates for attribution of l'Avis--and in the disputed text. For the first time, stylistic methods were applied to an attribution study in French literature.

An examination of the roles of the historian and the literary critic shows that both must be concerned with accurate attribution. The problem of identifying the author of a text applies particularly to the works of Pierre Bayle. Of his eight-volume corpus, only the Dictionnaire historique et critique and his correspondence were signed. Bayle was, however, but one of many exiled Huguenots who expressed their feelings and philosophies in anonymous writings. But, because of his "attachement absolu à la vérité, et son respect de toutes les raisons qui prétendent y conduire,"¹ he was often criticized by the orthodox Calvinist faction, especially as fashioned by Jurieu. Such was the case when Jurieu denounced Bayle as "auteur de l'Avis aux réfugiés, et complice d'une redoutable Cabale ourdie contre le parti réforme tout entier."² Between the time l'Avis first appeared and the time of Jurieu's accusation,

nearly a year elapsed. Even though the attacks leveled against Huguenot policies and politics in l'Avis stirred discussions among both conservative and zealot factions of the exiles, it was only after Jurieu published his Examen d'un libelle intitule: l'Avis aux réfugiés³ that attention was focused upon Bayle as its possible author.

There ensued "une guerre de pamphlets" which eventually brought a reprimand to both Bayle and Jurieu from the Dutch Consistory. Bayle consistently denied authorship of l'Avis, even though his "passion de la vérité" would not permit him to disavow agreement with many of its tenets. It has been argued that because his philosophy parallels that espoused in l'Avis, and because some critics assert that he had a hand in its publication,⁴ "il est tout à fait indubitable qu'on doit en tenir Bayle pour entièrement responsable."⁵ Not all scholars involved in attribution studies of l'Avis aux réfugiés have, however, arrived at the same conclusion, nor have they been so positive in their attribution. Bastide and Ascoli found conflicting external evidence regarding attribution of l'Avis aux réfugiés. Besides the fact that these two writers disagreed upon their selection of authors, the evidence they offered concerning attribution, much of which came from the same sources, was hearsay and inferential.

In addition to the efforts of Bastide and Ascoli, Delvolvé, Robinson and Mme Labrousse each devoted chapters of their studies on Bayle to l'Avis aux réfugiés and its attribution. Delvolvé and Robinson, each approaching the text internally, arrived at different conclusions. Mme Labrousse attributed l'Avis to Larroque but held Bayle responsible because of the assumption that he "y introduisit nombre d'additions et de changements--ne fût-ce que parce que Larroque était un

assez piètre écrivain."⁶ In any case, her reasoning and conclusions differ substantially from those offered by Delvolvé and Robinson. She finds l'Avis aux réfugiés harmonious with Bayle's published opinions and sufficiently unlike the opinions of any other contemporary. What is more, she feels it is the kind of work that Bayle could indeed have written. From this combination of circumstances, she places the work in the Bayle canon. Circumstantial evidence is supportive, not definitive. Delvolvé took a positive approach as he sought ideological similarities in le Commentaire and l'Avis. His argument seems to be the most convincing of the three mentioned above, because he leads his reader, step by step, through plausible reasoning.⁷ When considered in the light of the arguments presented in Chapter I (p. 10) of this thesis, however, the evidence he adduced becomes inferential, or at best, circumstantial. Robinson approached the question differently. Rather than attempting to direct his readers' thoughts through ideological comparisons, he merely pointed out recurring Baylian themes which are either nonexistent or refuted in l'Avis aux réfugiés. Even so, many of Robinson's conclusions, even though more firmly grounded--but less enticing--than Delvolvé's, lack objectivity

The existence of conflicting interpretations which lead to divergent conclusions points out that intuitive analyses, even those which concentrate on ideas drawn from the text, are often elusive. The ability to sense an author's touch in a work might guide a scholar in locating ideas, themes, or structures congruent to his writing habits, but does not constitute sufficient evidence for definitive attribution. The fact that three scholars, each well versed in the philosophies and works of Pierre Bayle, disagree in their attributions of l'Avis aux réfugiés

justified a new approach to the problem. None of the past studies on l'Avis aux réfugiés treated style; only Robinson alluded to it as he wrote: "The work seems to lack . . . the manner of Bayle. There is no sprightliness of attack, no abundant use of striking illustrations and penetrating psychological reflections."⁸

In the present study, with the aid of computer technology, an objective, quantitative analysis of stylistic traits in l'Avis aux réfugiés has been made and compared to similar usage in works known to be by Bayle and Daniel de Larroque. Ostensibly, there is no limit to the number of stylistic determinants which may be used by gifted writers. Moreover, some aspects of style remain in the realm of subjective analysis, because no method to count them systematically has yet been devised. Others, however, when structure allows their incidence to be readily identified and counted, have lent themselves to objective analysis. Because of its quantitative nature, this approach to analysis of a literary text is called computational stylistics. The electronic computer has provided the literary scholar with a means of processing quantities of information such as a researcher might not previously have assessed during a whole professional career. Hence, as scholars gain more understanding of computer capabilities and become more versed and imaginative in its utilization, they will be able to supplement intuitive discoveries with measurement data, often arriving at the same conclusions from diverse approaches--especially in attribution studies.

In the present study, a highly efficient assistant--the computer--was engaged to search over 250,000 words of natural language text. Eight hundred forty-three stylistic variables were located, isolated, counted, and printed out, so that their quantitative use might be

compared from one work to another. Test variables were grouped into five major categories--sentence level measures, sentence beginnings and endings, function words, expressions, and vocabulary analysis--for each of which a computer program was devised. Figures representing quantitative use of each variable were statistically compared across l'Avis aux réfugiés, its preface, and works by Bayle and Larroque. Chapters two and three contain a description of methods and specific procedures followed.

Results obtained from variables comparisons were reported in the fourth chapter. Table 4:60 summarized findings for the 843 test variables within the boundaries of the five categories cited above. No attempt was made to evaluate the relative importance of any given category or individual variable. In order to weight the variables, psycholinguists, grammarians, ideologists, etymologists, or stylists, exercising judgment based upon knowledge, experience, or intuition, would find it hard to agree on which variable or group of variables should be weighted most heavily. Hence, non-weighted totals were reported, and no subjective judgments of weighting were made, since one by-product of the study was to be the development of an objective method of stylistic analysis.

Results for each variable, considered singly, even though obtained through objective procedures, might be of minor importance. Nevertheless, if a sufficient number of variables form a consistent pattern, the analogy of the easily broken separate twigs and the much greater resistance of a number of these twigs bound together would seem to apply.⁹ Half of the sentence-level variables--three of six--indicated that Bayle did not write l'Avis; none suggested that he wrote it. Quantitative use of eleven of thirty sentence beginning and ending variables in l'Avis

differed significantly from similar use of these variables in Bayle's works. Only one variable, nouns used as sentence first words, implied similar usage by Bayle and the unknown author. Almost half of the individual function words tested and sixty percent of the function-word groups reject the hypothesis that Bayle is the author, while a mere one out of six individual words and one of five groups support the hypothesis. Although observed differences for word-pattern categories are not so pronounced as function word heterogeneity, a significant variance of usage does, nonetheless, exist. Sixty-eight of 554, or roughly one of every eight expressions tested, were used differently by Bayle and the unknown author. On the other hand, only four of every one hundred expressions were used similarly by the two writers. When quantitative usage of word-patterns by group was compared in the test articles, eleven of twenty-nine, or one group in three, disclosed differences of use by Bayle and the author of l'Avis aux réfugiés. Again, only six, or roughly one group in five, implied like usage. Similar variations were observed when Larroque's works were compared to the disputed text.

The final category whose quantitative characteristics were observed and compared was labeled "Vocabulary Analysis." Nine of eleven tests in this group uncovered significant differences between Bayle's and the unknown author's vocabulary usage. None suggested similarities. Bayle's vocabulary was shown to be rich, according to Guiraud's standard; it is richer than Larroque's, but less rich than the unknown author's. The foregoing valuation was substantiated by Guiraud's vocabulary-richness measure, Herdan's logarithmic type/token analysis, and Yule's "characteristic K." The final two tests involved type-token vocabularies of the test works and revealed that Bayle used a significantly greater

percentage of verbals and function words than did either Larroque or the unknown author. In all, 36% of 843 variables showed discriminating ability in works tested. Of those discriminants which distinguished between Bayle's works and l'Avis aux réfugiés, 73% revealed differences of quantitative use significant at the .05 and 72% at the .001 confidence levels.

The statistical results thus summarized would seem to refute Delvolvé's assertion that "tous les témoignages internes sont favorables à l'hypothèse qui fait Bayle auteur de l'Avis aux réfugiés."¹⁰ Internal evidence includes stylistic analysis; findings from the computational stylistic analysis performed on l'Avis aux réfugiés and Bayle's work justify rejection of the hypotheses that Bayle wrote the controversial pamphlet. Furthermore, conclusions prompted by this study illustrate the need in attribution studies to examine stylistic elements as objectively as possible. However, this new method contributes to, but does not obviate intuitive and ideological analyses.

Past scholarship has argued that many ideas and doctrines exposed in l'Avis aux réfugiés are congruent to Bayle's thinking. Existence of such ideological parallelisms has led some scholars to accept Bayle as its author. Likewise, because similar stylistic traits do exist in Bayle's works and in l'Avis, a researcher concentrating on similarities in ideology might well be stylistically deceived by superficial parallels in writing habits. Stylistic parallelisms, like ideological similarities are interesting and attractive, but sometimes deceiving, unless pursued in depth.

It was mentioned earlier that some scholars¹¹ feel that Bayle's failure to utterly denounce l'Avis aux réfugiés constituted an admission

of guilt. Furthermore, they add, this act of omission weakened and caused to go unheeded his constant denials of authorship of the disputed work. But, if he had denounced the work, if he had ignored truths he felt should be recognized by the Huguenot zealot element,¹² what then would his friends and critics have said of his "passion de la vérité," or of his integrity? In his attacks on Bayle, Jurieu used the claim that Bayle wrote l'Avis as a catalyst toward a more serious accusation: that of participation in a political cabal. Because consequences of involvement in a plan to overthrow William of Orange were more serious than those of writing a controversial pamphlet, it seems only logical that Bayle should concentrate his arguments on the more important accusation. Walter Rex points out that Bayle "very successfully proved it [the cabal] to be a figment of Jurieu's imagination."¹³ Regarding the claim that he authored l'Avis aux réfugiés, Bayle, himself, wrote: "je me contente de dire publiquement ce que j'ai dit en particulier toutes les fois que l'occasion s'en est présentée, c'est que je ne suis pas l'Auteur de ce libelle."¹⁴ Thus Bayle issued a strong denial of authorship of l'Avis, but was restrained from totally denouncing the pamphlet by his respect for the truth and his need to concentrate on Jurieu's accusation.

Each bit of evidence plays its part in causing the acceptance or rejection of a hypotheses. Computational stylistic data produced by this study take their place alongside evidence produced by Bastide and Robinson to reject the hypothesis that Pierre Bayle wrote l'Avis aux réfugiés. Moreover, data obtained from the comparison of Larroque's quantitative use of the variables tested disclosed that his writing habits diverge even more sharply from those of the unknown author than

do Bayle's. This thesis adds support to the claim that Bayle did not write l'Avis aux réfugiés, but it also indicates for the first time that Daniel de Larroque can be even more readily dismissed as a possible author.

The most significant by-product of this project was the development of an objective approach to stylistic analysis of French literature by means of thirty-seven computer programs as noted in Appendix C. These programs provided the means for searching a literary text to recognize, locate, isolate, record, and tabulate quantifiable stylistic traits. If the computer programs used in this thesis are conceived of in the model of a bull's-eye, with the specific task of attribution of l'Avis aux réfugiés in the center, there remain two outer rings. The application of these techniques to French attribution studies would be the interior ring, and the use of attribution studies in general the outer ring. The methods used will vary in their utility with differing studies, and will in no case be foolproof. Nevertheless, as a tool for analysis, they are fast becoming an important means of producing useful, even essential, evidence.

NOTES FOR CHAPTER V

¹Delvolvé, p. 9.

²Ibid., p. 196.

³Pierre Jurieu, Examen d'un libelle contre, la religion, contre l'Etat et contre la Révolution d'Angleterre; intitulé: 'Avis important aux réfugiés sur leur prochain retour en France,' ouvrage précédé de l'Avis important au public (La Haye, 1691).

⁴This statement made by Mme Labrousse is based upon a conclusion drawn from a second-hand source as noted above, p. 4, which involved Louis, "correcteur chez Moetjens." See also Ascoli, p. 544.

⁵Labrousse, Pierre Bayle I, p. 221. With all due respect to Mme Labrousse, whose works on Bayle are truly monuments of erudition and intelligence, it would appear that in taking such a narrow stand she deprives Bayle of a principle which he continually strove to offer all men: their agency, or in this case the opportunity to espouse principles or ideas which they deem to be true while rejecting those they view as false. By holding Bayle entirely accountable for l'Avis aux réfugiés she is in sum saying he is responsible even for those elements, such as French politics, Louis' wars, rebellion, and intolerance, which he spoke out against in le Commentaire, La France toute catholique, and Cabale chimérique.

⁶Ibid., p. 221. In a footnote which treats testimony concerning attribution of l'Avis, Mme Labrousse suggests that even her convictions may not be quite so strong as she suggests in her text. She records several testimonies from Bayle's closest friends and discusses Larroque's supposed avowal of authorship. To the latter she responds, "De toutes façons, Larroque eut souvent un comportement si tortueux qu'on ne saurait accorder un crédit décisif à ses dires." Pierre Bayle I, p. 220f. In addition, the following statements from La Chimère de la cabale de Rotterdam seem to exonerate Larroque: "Quoi qu'il en soit, it est faux que Mr. de Larroque soit le Confident de Mr. Bayle pour le livre de l'Avis aux Réfugiés ... Il est faux qu'il ait eu part au manège de la seconde édition. Il est faux qu'il soit l'Agent de la prétendue Cabale." OD, II, 735.

⁷As I began the present study, I was captured by Delvolvé's reasoning and was predisposed to accept that Bayle wrote l'Avis. Nevertheless, through continued research I have come to see that, although very well conceived, logically presented and extremely interesting, Delvolvé's reasons for attributing the work to Bayle are based on circumstantial evidence.

⁸Robinson, p. 121.

⁹This analogy was suggested by Sherbo, "Uses and Abuses," pp. 10-11.

¹⁰Delvolvé, p. 189.

¹¹Delvolvé, pp. 189-191; and Labrousse, Pierre Bayle I, pp. 221-223.

¹²For example, Bayle saw the continual need to follow the admonition of Paul, "Examine all things and retain that which is true," but even more important, to examine dispassionately and look inward: "Au fond quel mal peut-on craindre de l'Avis aux réfugiés? Car ou ce qu'il nous reproche est vrai, ou il est faux. S'il est faux, deux mots de négative suffisent pour en arrêter tous les effets. S'il est vrai, ce n'est point du livre que nous peut venir du mal, mais de notre propre doctrine, et si celle-ci ne peut pas nous faire du mal le livre ne le peut point non plus." Entretiens sur la cabale chimérique, OD, II, 652.

¹³Walter Rex, Essays on Pierre Bayle and Religious Controversy (The Hague: Nijhoff, 1965) p. 227f.

¹⁴Cabale Chimérique, OD, II, 637.

LIST OF REFERENCES

LIST OF REFERENCES

- Altick, Richard. The Art of Literary Research. New York: Norton, 1963.
- Anon. Avis Important aux Réfugiez sur Leur Prochain Retour en France, Donné pour Etrennes à l'un d'eux en 1690. Paris: Veuve de Gabriel Martin, 1692. (Attributed to Pierre Bayle, Daniel de Larroque and Paul Péliisson-Fontanier.)
- Anon. "The Statistics of Style," review of Alvar Ellegård Who Was Junius? and A Statistical Method for Determining Authorship, The Junius Letters, 1769-1772, Times Literary Supplement, 25 Jan. 1963, p. 67.
- Ascoli, Georges. "Bayle et l'Avis aux réfugiés," Revue d'Histoire littéraire de la France, XX (1913), 517-545.
- Bally, C. Traité de stylistique française. 2^e éd., 2 vols. Heidelberg: Winter, 1919.
- Bastide, Charles. "Bayle est-il l'Auteur de l'Avis aux réfugiés?" Bulletin de la Société d'Histoire du Protestantisme, LVI (1907), 544-558.
- Bayle, Pierre. Dictionnaire historique et critique 4 vols. in-fol. 4^e éd. avec la vie de l'auteur par Mr. Desmaizeaux. Amsterdam: P. Brunel, 1730.
- _____. Oeuvres diverses de Mr. Pierre Bayle, Professeur en Philosophie, et en Histoire, à Rotterdam: Contenant tout ce que cet Auteur a publié sur des matières de Théologie, de Philosophie, de Critique, d'Histoire, et de Littérature; excepté son Dictionnaire Historique et Critique. 4 vols. La Haye, 1727-1731. Reprinted with introduction by Elisabeth Labrousse at Hildesheim: Georg Olms, 1964-1968.
- Bennett, Paul E. "The Statistical Measurement of a Stylistic Trait in Julius Caesar and As You Like It," Shakespeare Quarterly, VIII, No. 1, (Winter, 1957), 33-50.
- Bentley, Gerald E. "Authenticity and Attribution in Jacobean and Caroline Drama," Evidence for Authorship. Eds. David Erdman and Ephim Fogel. Ithaca: Cornell University, 1966, pp. 179-187.
- Bowles, Edmund A. (ed). Computers in Humanistic Research. Englewood Cliffs, New Jersey: Prentice-Hall, 1967.
- Bruneau, C. "La Stylistique," Romance Philology V (1951), 1-14.

- Cargo, Robert T. A Concordance to Baudelaire's 'Les Fleurs du mal'. Chapel Hill: University of North Carolina Press, 1965.
- Cayrou, Gaston. Le Français classique. Paris: Didier, 1948.
- Delvolvé, Jean. Religion, critique et philosophie positive chez Pierre Bayle. Paris: Félix Alcan, 1906.
- Desmaizeaux. "Vie De Bayle," preface to Pierre Bayle's Dictionnaire historique et critique, ed. Beuchot. Amsterdam, 1730.
- Ellegård, Alvar. A Statistical Method for Determining Authorship: The Junius Letters, 1769-1772. Gøthenburg, 1962.
- Fogel, Ephim G. "The Humanist and the Computer: Vision and Actuality," Proceedings of the IBM Literary Data Processing Conference. Yorktown Heights, New York, Sept. 1964, pp. 11-24.
- Freeman, Bryant C. Review of Cargo, Robert T. A Concordance to Baudelaire's "Les Fleurs du mal." Chapel Hill: University of North Carolina Press, 1965 and Quémada B. et K. Menemencioglu. Baudelaire, "Les Fleurs du mal": Concordances, Index et Relevés statistiques. Paris: Larousse, 1965, The French Review, XXXIX, No. 5 (April, 1966), 807-808.
- Grevisse, Maurice. Le Bon usage. 7^e éd. Paris, 1959.
- Guiraud, Pierre. Les Caractères statistiques du vocabulaire. Paris: Presses Universitaires de France, 1954.
- Haase, A. Syntaxe française du XVII^e siècle. Translated by M. Obert. 5^e edition. Paris: Delagrave, 1965.
- Haase, Erich. Einführung in die Literatur des Refuge. Der Beitrag der französischen Protestanten zur Entwicklung analytischer Denkformen am Ende des 17. Jahrhunderts. Berlin: Dunker and Humblot, 1959.
- Hatzfeld, Helmut A. A Critical Bibliography of the New Stylistics, 1900-1952. University of North Carolina Studies in Comparative Literature, No. 5. Chapel Hill, 1953.
- Herdan, Gustav. "A New Derivation and Interpretation of Yule's 'Characteristic' K," Journal of Applied Mathematics and Physics, VI, (1955), 332-334.
- _____. Language As Chance and Choice. Groningen, 1956.
- _____. Type-Token Mathematics, A Textbook of Mathematical Linguistics. The Hague: Mouton and Company, 1960.
- IBM 1620 Monitor I System Reference Manual. San Jose, California: IBM, Product Publications Dept., 1963.

- Jones, Clelland E. Manual of French Pronunciation. Salt Lake City: Deseret News Press, 1961.
- Jurieu, Pierre. Examen d'un libelle contre la religion, contre l'Etat et contre la Révolution d'Angleterre; intitulé: 'Avis important aux réfugiés sur leur prochain retour en France,' ouvrage précédé de l'Avis important au public. La Haye, 1691.
- _____. Lettre [sic] pastorales adressées aux fidèles de France qui gemissent sous la captivité de Babylon [sic] . . . , 3^e ed. Rotterdam: A. Acher, 1686-1689.
- Labrousse, Elisabeth. Pierre Bayle, Tome 1: Du Pays de Foix à la cité d'Erasmus. Archives internationales d'Histoire des idées 1. La Haye: Martinus Nijhoff, 1963.
- _____. Pierre Bayle, Tome II: Hétérodoxie et Rigourisme. Archives internationales d'Histoire des idées 6. La Haye: Martinus Nijhoff, 1964.
- Lacoste, Edmond. Bayle Nouvelliste et critique littéraire. Paris: Picart, 1929.
- Le Bidois, Georges and Le Bidois, Robert. Syntaxe du français moderne. 2 vols. 2nd ed. Paris: Editions Picard, 1967.
- Leed, Jacob (ed.). The Computer and Literary Style. Kent, Ohio: Kent State University Press, 1966.
- Leeson, Daniel N. and Dimitry, Donald L. Basic Programming Concepts and the IBM 1620 Computer. New York: Holt, Rinehart, and Wilson, Inc., 1962.
- Li, Jerome C. R. Introduction to Statistical Inference. Ann Arbor, Michigan: Edwards Bros., 1959.
- Markman, Alan. "Litterae Ex Machina," Proceedings of the IBM Literary Data Processing Conference, Yorktown Heights, New York, Sept. 1964, pp. 37-54.
- McKerrow, Ronald B. An Introduction to Bibliography. Oxford: Clarendon Press, 1927.
- Milic, Louis T. A Quantitative Approach to the Style of Jonathan Swift. The Hague: Mouton and Co., 1967.
- _____. "Unconscious Ordering in the Prose of Swift," The Computer and Literary Style. Ed. Jacob Leed. Kent, Ohio: Kent State University Press, 1966, pp. 79-106.
- _____. "Winged Words: Varieties of Computer Applications in Literature," Computers and the Humanities, II, (Sept. 1967), 24-31.
- Miller, George A. Language and Communication. New York: McGraw-Hill, 1951.

- Miller, George A., Newman, E. B., and Friedman, E. A. "Length-frequency statistics of written English," Information and Control, I (1958), 370-389.
- Mosteller, Frederick, and Wallace, David L. Inference and Disputed Authorship: The Federalist. Reading, Massachusetts: Addison-Wesley, Inc., 1964.
- _____. "Inference in an Authorship Problem: A Comparative Study of Discrimination Methods Applied to the Authorship of The Federalist Papers." Minneapolis, Minnesota (Sept. 9, 1962).
- Morris, John. "A Computer-Assisted Study of a Philosophical Text," Computers and the Humanities, III, (Jan. 1969), 175-176.
- Muller, Charles. Essai de statistique lexicale: "L'illusion comique" de Pierre Corneille. Paris: Klincksieck, 1964.
- _____. Etude de statistique lexicale: Le vocabulaire du théâtre de Corneille. Paris: Larousse, 1967.
- O'Donnell, Bernard. "Stephen Crane's The O'Ruddy, a Problem in Authorship Discrimination," The Computer and Literary Style. Ed. Jacob Leed. Kent, Ohio: Kent State University Press, 1966, pp. 107-115.
- Pappas, John N. "Authentication of an Eighteenth Century Text," Institute for Computer Research in the Humanities Newsletter. II, Nos. 4 and 5 (Dec. 1966 and Jan. 1967), 3-4, 3-4.
- Parrish, Stephan M. "Computers and the Muse of Literature," Proceedings of the Conference on the Use of Computers in Humanistic Research. Rutgers, Dec. 1964.
- Raben, Joseph. "A Computer Aided Study of Literary Influence: Milton to Shelley," IBM Proceedings of the Literary Data Processing Conference, Yorktown Heights, New York, Sept. 1964.
- Rex, Walter. Essays on Pierre Bayle and Religious Controversy. The Hague: Martinus Nijhoff, 1965.
- Robert, Paul. Dictionnaire Etymologique de la langue française. 5^e éd. Paris: Presse universitaire de France, 1968.
- _____. (éd.). Petit Robert, Dictionnaire alphabétique et analogique de la langue française. Paris, 1967.
- Robinson, Howard. Bayle the Sceptic. New York: Columbia University Press, 1931.
- Sayce, R. A. Style in French Prose. Oxford: Oxford University Press, 1953.

- Scheffé, Henry. The Analysis of Variance. New York: Wiley and Sons, 1959.
- Sedelow, Sally Y. "Stylistic Analysis: Report on the Second Year of Research," System Development Corporation Document. TM-1908/200/00, 1 March, 1966.
- Sherbo, Arthur. "The Uses and Abuses of Evidence," Evidence for Authorship. Eds. David Erdman and Ephim Fogel. Ithaca: Cornell University, 1966, pp. 6-24.
- Snedecor, George W. Statistical Methods. Ames, Iowa: Iowa State University Press, 1962.
- Spitzer, Leo. Stilstudien. 2 vols. Munich: Huber, 1928.
- _____. Linguistics and Literary History. Princeton, 1948.
- _____. A Method of Interpreting Literature. Northhampton, Mass: Smith College, 1949.
- _____. "Stylistique et critique littéraire," Critique, XI (1955), 595-609.
- Starobinski, Jean. L'Oeil vivant. Paris: Gallimard, 1961.
- Steel, Robert G. D. and Torrie, James H. Principles and Procedures of Statistics. New York: McGraw-Hill, 1960.
- Ullmann, Stephen. Language and Style. Oxford: Alden Press, 1966.
- Wachal, Stephen. "Linguistic Evidence, Statistical Inference and Disputed Authorship." Diss. Wisconsin, 1966.
- _____. "Getting at Style through Statistics," review of Statistics and Style. Eds. Lubomír Doležal and Richard W. Bailey. New York: American Elsevier. Computers and the Humanities, IV (May 1970) 362-372.
- Williams, C. B. "A Note on the Statistical Analysis of Sentence Length as a Criterion of Literary Style," Statistics and Style. Eds. Doležal and Bailey. New York: American Elsevier, 1969, pp. 69-75.
- Yule, G. Udny. "On Sentence Length as a Statistical Characteristic of Style in Prose," Biometrika, XXX (1936), 363-390.
- _____. The Statistical Study of Vocabulary. Cambridge: Cambridge University Press, 1944.
- Zimmer, George. "The Attribution of Authorship: A Computerized Method Evaluated and Compared with Other Methods Past and Future." Diss. Michigan State University, 1968.

- _____. Review of Ellegård, Who Was Junius? and A Statistical Method,
Journal of English and Germanic Philology, June 1963, pp. 688-689.

APPENDICES

APPENDIX A

APPENDIX A

LIST OF ARTICLES TESTED

** ** INDEX TO ARTICLES FOR RICHARD V. WALL ** ** 22 JUN 70
CONTENTS OF (WALLPHD,1,1,999),MT(385)

TITLE	TAPE	RECORDS	LINE	NUMBERS	DATE	TITLE OF ARTICLE
UNKNOWN	1	313	X0001	X4699	JAN 1690	LA AVIS AUX REFUGIEZ . . .
RAYLF 1	314	669	R0001	R4443	1691	CARALE CHIMERIQUE
RAYLF 2	670	768	R5000	R6129	DEC 1688	REPONSE D'UN NOUVEAU CONVERTI
RAYLF 3	770	776	R6200	R6358	JUL 1686 1	REFLEXIONS SUR LES DIFFERENS DE LA
RAYLF 4	777	796	R6400	R6756	JUL 1686 2	TRAITE SUR LE SERPENT QUI TENTA EVE
RAYLF 5	797	803	R7200	R7395	JUL 1686 4	TRAITE DE LA BENEDICTION NIPTIALE
RAYLF 6	804	807	R7700	R7802	JUL 1686 7	HISTOIRE DES GUERRES DE LA MAISON
RAYLF 7	808	832	R7900	R8302	JUL 1686 8	SUPPLEMENT A* DIVERS ENDROITS DE CES
RAYLF 8	833	834	R8400	R8485	JUL 1686 9	LA MORALE DU MONDE
RAYLF 9	835	837	R8600	R8661	JUL 1686 10	TRAITE DES EVEQUES, ET DE LEURS
LARROQ 1	838	1052	L0001	L2842	1684	LE PROSELYTE ABUSE
LARROQ 7	1053	1077	L5500	L5897	JUL 1687 1	ECONOMIE DIVINE
LARROQ 8	1078	1091	L6000	L6234	JUL 1687 7	HISTOIRES DES REVOLUTIONS
LARROQ 9	1092	1096	L6300	L6435	JUL 1687 8	EXTRAIT D'UNE LETTRE DE M. LARROQUE
LARROQ A	1097	1106	L6500	L6659	JUL 1687 9	LA PAIX DES BONNES AMES
LARROQ B	1107	1123	L6801	L6939	JUL 1687 11	APOLOGIE POUR L'EGLISE ANGLICANE
LARROQ C	1124	1129	L7001	L7105	JUL 1687 13	NOUVELLES ACCUSATIONS CONTRE M.
LARROQ D	1030	1135	L7200	L7359	SEP 1687 1	TRAIT DE L'AIMAN
LARROQ E	1136	1153	L7400	L7779	SEP 1687 2	REFLEXIONS NOUVELLES SUR LES CAUSES
LARROQ F	1154	1159	L7900	L7965	SEP 1687 3	REPONSE DE M. LARROQUE A* LA
LARROQ G	1160	1177	L8000	L8375	SEP 1687 4	DISCOURS SUR UNE MEDAILLE DE FURIA
LARROQ H	1178	1194	L8500	L8884	SEP 1687 5	HISTOIRE DE LA MORT DES PERSECUTEURS
LARROQ I	1195	1210	L9000	L9304	SEP 1687 6	DEUX TRAITTEZ D'OUSSERIUS
LARROQ J	1211	1227	Q1000	Q1403	JUL 1688 1	HISTOIRE GENERAL DES CONCILES
LARROQ K	1228	1241	Q1500	Q1980	JUL 1688 3	C'EST A* DIRE, REFUTATION DE CE QUE
LARROQ L	1242	1255	Q2000	Q2432	JUL 1688 4	HISTOIRES DE PHILIPPE DE VALOIS
LARROQ M	1256	1261	Q2500	Q2696	JUL 1688 5	DEMONSTRATION DE LA VERITE ET DE LA
LARROQ N	1262	1264	Q2800	Q2885	JUL 1688 6	HISTOIRE DU DIVORCE DE HENRY VI
LARROQ 2	1265	1278	L3000	L3527	JAN 1689 1	DE LA VERITABLE RELIGION
LARROQ 3	1279	1285	L4000	L4163	JAN 1689 3	HISTOIRE DE LA MONARCHIE FRANCAISE
LARROQ 4	1286	1297	L4500	L4769	JAN 1689 4	ESSAIS DE PHYSIQUE
LARROQ 5	1298	1306	L5000	L5151	JAN 1689 5	LES OEUVRES POSTHUMES DE M. CLAUDE
LARROQ 6	1307	1313	L5300	L5375	JAN 1689 6	HISTOIRE D'UNE DAME CHRETIENNE
PELISO 1	NOT TAPED		P0001	P1005	1694	LETTRE DE MONSIEUR PELISSON
DESCAR 1	NOT TAPED		D0001	D2180	1647	MEDITATIONS METAPHYSIQUES

APPENDIX B

APPENDIX B

INPUT DATA - TEXT SAMPLE

*\$ABSTRACT

PG 583 X

X 284 JF NE VOUS LE DIS PAS, MONSIEUR, POUR VOUS INSULTER. A* DIEU NE PLAISE.
 X 285VOUS SCAVEZ MES SENTIMENTS. VOUS NE IGNOREZ PAS QUE NE AIANT AUCUNE PART AUX
 X 286AFFAIRES PUBLIQUES, JE AI VU AVEC UNE EXTREME REGRET CETTE SUITE DE
 X 287EVENEMENTS, ET CETTE FATALE NECESSITE, PAR LAQUELLE LA FRANCE SE EST PRIVEE
 X 288DE TANT DE HONNETES GENS, ET DE PERSONNES DE MERITE, QUI ONT ETE CHERCHER
 X 289UN ASYLE DANS LES PAYS ETRANGERS. DE SORTE QUE SI JE VOIS AVEC PLAISIR QUE
 X 290L'ANNEE 1689, NE A POINT REPONDU A* VOS PREDICTIONS, CE NE EST NULLEMENT
 X 291A* CAUSE DU PREJUDICE QUE VOUS EN RECEVEZ, MAIS A* CAUSE QUE ON DOIT ETRE
 X 292BIEN-AISE, EN FAVEUR DE LA RAISON ET DU BON SENS, QUE LA SUPERSTITION DES
 X 293NOMBRES, ET LA CREDULITE POPULAIRE, SOIT DEMENTIE PAR DES EXPERIENCES
 X 294PALPABLES QUI PUISSENT AUTANT LA AFFOBLIR, QUE ELLE SE SEROIT FORTIFIEE
 X 295PAR LES EVENEMENTS A* QUOI VOUS VOUS ETIEZ ATTENDUS. ET POUR VOUS MONTRER
 X 296QUE CE EST LA* LE VERITABLE SUJET DE MA JOIE, VOICE DE*S LE PREMIER JOUR DE
 X 297L'AN 1690, UNE LETTRE OU* JE VOUS FELICITE DE TOUT MON COEUR DES
 X 298FAVORABLES DISPOSITIONS, QUE ON DIT ETRE DANS L'ESPRIT DU ROI POUR LE
 X 299RETABLISSEMENT DE VOTRE PARTI. JE NE VOUS ASSURE PAS QUE TOUT LE MONDE SE
 X 300EN REJOISSA. IL SE TROUVERA TOUJOURS DES IGNORANS ET DE FAUX SAVANS, QUI
 X 301CONDAMNERONT LA TOLERANCE DE VOTRE SECTE DANS LE ROYAUME DU ROI
 X 302TRES-CHRETIEN, ET DU FILS AINE DE L'EGLISE. MAIS JE VOUS REPOND QUE EN
 X 303GENERAL TOUT CE QUE IL Y A DE PLUS RAISONNABLE DANS LES TROIS ORDRES DU
 X 304ROYAUME, APPROUVERONT QUE ON VOUS LAISSE UNE HONNETE LIBERTE, PUISQUE IL
 X 305NE A PAS SEMBLE BON AU SAINT ESPRIT DE SECONDER LES INTENTIONS QUE ON A
 X 306EUES DE VOUS REVENIR A* L'EGLISE CATHOLIQUE. VOUS NE SCAURIEZ CROIRE LE
 X 307PLAISIR QUE JE RESENS PAR AVANCE, EN ME IMAGINANT QUE VOUS NE SEREZ PAS
 X 308DES DERNIERS A* REVENIR. JE NE PARLE PRESQUE DE AUTRE CHOSE AVEC MES AMIS,
 X 309ET JE NE VOIS GUERRE DE GENS QUI NE AYENT PERDU, PAR LA SUPPRESSION DE L'
 X 310EDIT DE NANTES, QUELQUE PERSONNE QUE ILS AIMOIENT, ET QUE ILS ESTIMOIENT
 X 311AVEC BEAUCOUP DE JOIE DES NOUVELLES FAVORABLES QUE ON DEBITE SUR VOTRE
 X 312SUJET. AINSI, MONSIEUR, PREPAREZ-VOUS, TOUT TANT QUE VOUS ETES, A* RECEVOIR
 X 313A* VOTRE RETOUR EN FRANCE, MILLE CARESSES ET MILLE EMBRASSEMENTS DE CEUX
 X 314MEMES QUI SONT ATTACHEZ AVEC UN ZELE INVIOLENT A* LA COMMUNION DE L'
 X 315EGLISE CATHOLIQUE.

*\$ABSTRACT

PG 583 X

X 316 MAIS PERMETTEZ-MOI DE VOUS AVERTIR DE UNE CHOSE, VOUS, MONSIEUR, ET TOUS
 X 317VOS CONFREES REFUGIEZ EN DIVERS PAYS ETRANGERS. CE EST DE FAIRE UNE ESPECE
 X 318DE QUARANTAINE AVANT QUE DE METTRE LE PIED EN FRANCE, AFIN DE VOUS PURIFIER
 X 319DU MAUVAIS AIR QUE VOUS AVEZ HUME DANS LES LIEUX DE VOTRE EXIL, ET QUI VOUS
 X 320A INFECTEZ DE DEUX MALADIES TRES DANGEREUSES, ET TOUT-A*FAIT ODIEUSES. L'
 X 321UNE EST L'ESPRIT DE SATYRE. L'AUTRE UN CERTAIN ESPRIT REPUBLICAIN QUI
 X 322NE VA PAS A* MOINS QUE A* INTRODUIRE L'ANARCHIE DANS LE MONDE, LE PLUS
 X 323GRAND FLEAU DE LA SOCIETE CIVILE. VOILA DEUX POINTS SUR LESQUELS JE PRENS
 X 324LA LIBERTE DE VOUS PARLER EN AMI. COMMENCONS PAR VOTRE ESPRIT DE SATYRE.

*\$ABSTRACT

PG 583 X

X 325 LA FACILITE QUE VOUS AVEZ TROUVEE DANS LES PAYS ETRANGERS DE FAIRE
 X 326IMPRIMER IMPUNEMENT TOUT CE QUE IL VOUS A PLU, A PRODUIT PARMI VOUS UNE SI
 X 327GRANDE QUANTITE DE AUTEURS, QUE IL NE Y A PAS DE APPARENCE QUE AUCUNE SECTE
 X 328VOUS DISPUTE JAMAIS LE PREMIER RANG DE FECONDITE EN CE GENRE LA*. CES
 X 329AUTEURS SONT FORT DIFFERENTS LES UNS DES AUTRES EN CAPACITE, MAIS ILS SE

APPENDIX C

APPENDIX C

ANNOTATED PROGRAM LIST

General Use

- COUNTS (Count Synonyms) Written in SPS for the IBM 1620
- This program reads in the synonym list and counts the total number of synonyms (see Chapter III for information about the synonym list in the ENROOT description).
- EDIT 1 (Edit Program 1) Written in FORTRAN for the CDC 6500
- Because punctuation is vital to the sentence level tests, this program was used to check for any irregular punctuation such as " . . . " or " , . " so that these could be visually checked.
- EDIT 3 (Edit Program 3) Written in SPS for the IBM 1620
- This program lists the text with each sentence separated by a blank line so that unusually long or short sentences can be checked for keypunching errors in punctuation.
- EDIT 4 (Edit Program 4) Written in SPS for the IBM 1620
- This program prints the sequence numbers of all sentences of exactly 1 or 2 words so that they may be checked for errors.
- STD (Standard Deviation) Written in FORTRAN for the CDC 6500
- This program computes the mean and standard deviation of a set of numbers.
- SENWOL (Sentence and Word Lengths) Written in SPS for the IBM 1620
- This program generated all the basic data for the sentence level tests, as described in Chapter III.
- SEQT (Sequence Text) Written in SPS for the IBM 1620
- This program was used to insert sequence numbers and "ABSTRACT" cards into the newly keypunched text (used on the Pelisson and Descartes material).

- FREQFUN (Frequency of Function Words) Written in SPS for the IBM 1620
- This program generates the basic data for the function word analysis described in Chapter III.
- SEQTC (Sequence Check Text) Written in SPS for the IBM 1620
- This program sequence checks the line numbers on the text.
- STYLBEND (Stylistic Beginnings and Endings) Written in SPS for the IBM 1620
- This program finds and punches the first and last word of each sentence. See Chapter III for a complete description.
- SENDIS (Sentence Distribution) Written in FORTRAN for the CDC 6500
- This program tests internal variance by grouping sentence level data into 100-sentence groups within each article.
- SENDAT (Sentence Level Data Analysis) Written in FORTRAN for the CDC 6500
- This program prints out a distribution chart based on the number of words per sentence for articles X, B1, and L1.
- FUNW (Function Word Analysis) Written in FORTRAN for the CDC 6500
- This program performs an analysis of function words similar to that of Mosteller and Wallace: elimination by repeated t-tests to obtain valid discriminators.
- EXSOR (Expression Sorting Program) Written in SPS for the IBM 1620
- This is the SPS equivalent of the CDC 3600 STYLEFRAZ program which computes absolute frequencies of expressions. See Chapter III for a complete description.

General Use - French

- SYLAN (Sylabification Analysis) Written in SPS for the IBM 1620
- This program generates the sylabification data. See Chapter III for a complete description.

- DATA 1 (Data Conversion Program 1) Written in FORTRAN for the CDC 6500
- This program compacts the data from FREQFUN for any given word: it puts the frequencies from all 33 articles on two cards. It performs a similar procedure on the EXSOR data.
- DATA 2 (Data Conversion Program 2) Written in FORTRAN for the CDC 6500
- This program puts total values from FREQFUN and EXSOR groups into a common format so that they may be analyzed by one program.
- SBS (STYLBEND Summary) Written in BAL for the IBM 360/20
- This program reads in the raw data from STYLBEND and computes totals for each part of speech. These are printed and punched for further analysis.
- FUNNY (Function Word Analysis) Written in FORTRAN for the CDC 6500
- This program does the t-tests and other final statistical analyses on the function word data.
- TOTFUN (Compute Total Function Words) Written in FORTRAN for the CDC 6500
- This program reads in the function word data and computes the total number of function words in each article.
- MISC (Miscellaneous Tests) Written in FORTRAN for the CDC 6500
- This program performs the t-tests and other final statistical analyses on all the miscellaneous tests derived from the ENROOT data.
- SANAL (SEWOL Analysis) Written in FORTRAN for the CDC 6500
- This program prints out all reports and does all t-tests and other final statistical analysis for the SEWOL data.
- STAN (STYLBEND Analysis) Written in FORTRAN for the CDC 6500
- This program prints out all reports and does all t-tests and other final statistical analysis for the STYLBEND data.

PIP (Printed Index Program) Written in FORTRAN for the CDC
3600 and 6500

This program produces basic ENROOT data and is described in detail in Chapter III.

FAP (File Analysis Program) Written in FORTRAN for the CDC
3600 and 6500

This program was used in conjunction with the PIP program. See Chapter III for a complete description.

Specific to this Study

CANAL (Convert Data for Analysis) Written in SPS for the IBM 1620

The purpose of this program was to reorganize data for easier processing. See Chapter III for a complete description.

STAND (Standardization) Written in FORTRAN for the CDC 6500

This program was used in association with FAP; it converted FAP data from standardized form to non-standardized form.

ANOVAR (Analysis of Variance) Written in FORTRAN for the CDC 6500

This program, discussed at length in Chapter III, performs the analysis of variance on various sets of data.

EDIT 2 (Edit Program 2) Written in SPS for the IBM 1620

The text was keypunched from manuscripts which contained a period after most numbers. This program finds all occurrences of a number followed by a period so that these periods will not affect the sentence level tests.

HTOR (Convert H to R) Written in BAL for the IBM 360/20

This program was used to modify the ENROOT source deck when converting from the CDC 3600 to the CDC 6500; it changed certain word lists from the 3600 form to the corresponding 6500 form.

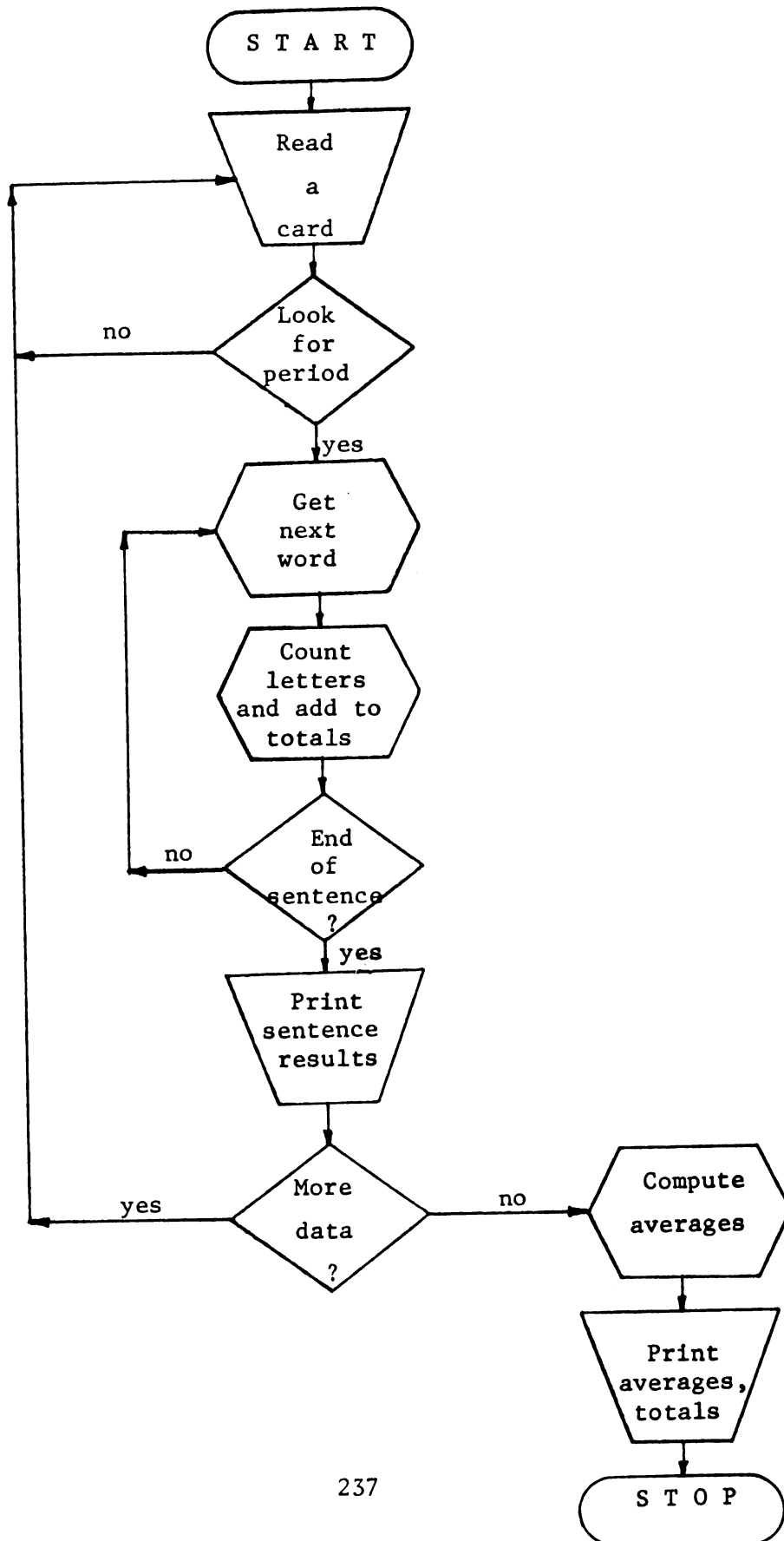
CAT (Form Categories) Written in FORTRAN for the CDC 6500

This program was used to group the words from FREQFUN into the categories described in Chapter IV. It performed a similar function on the expressions from EXSOR.

APPENDIX D

APPENDIX D

SENWOL FLOWCHART



APPENDIX E

APPENDIX E
SAMPLE SENWOL OUTPUT

SEQ. NO. XXXXX	TOT. WRDS XXX	TOT LET XXX	LET. FOR WORDS XXXXXXXXX ...
X3916	009	036	331372656
X3917	008	028	33234319
X3918	030	120	324152565238214632637534225379
X3920	011	042	12338346246
X3921	034	145	22256432972364323724B256124624135D
X3923	018	096	3521552C9732C64349
X3925	004	022	1056
X3925	034	161	232574435926362182936326B324483927
X3928	016	088	0372942824G57125
X3929	008	043	10438764
X3930	012	060	37236A827137
X3931	004	018	8244
X3931	014	053	29447423332424
X3933	055	212	2264442732114222843824472657356222634459151634535238214
X3937	027	129	244637574683039A24243333445
X3940	021	098	366255458242439382160
X3941	011	043	24317248417
X3942	024	110	24364252256A624833762548
X3944	044	195	2422335414370861626228230225A1465252242734C9
X3947	024	108	582322364412543033992206
X3950	050	227	222355352162A483446325244263C23A437271828372370366
X3953	023	106	82828342252370861726343
X3956	023	106	2380236262628252622326B
X3957	029	137	3234435613A2513684227483BA226
X3960	036	149	253544269234422632101366426367348236
X3963	034	134	1666255453245462818622613213454417
X3965	048	207	24435423724253251362322D2342953834CB24231861524A
X3969	015	078	1334763A3582660
X3970	029	152	133A262372768445460439342820C
X3973	008	032	23223307
X3973	020	079	42243633252468432736
X3975	029	126	23564144324308741632085452424

APPENDIX F

APPENDIX F

TABLE 4:6A. AVERAGE NUMBER OF LETTERS AND SYLLABLES PER WORD AND
LETTERS PER SYLLABLE, LESS ONE AND TWO LETTER WORDS

Article	Average Let/Wrd	Average Syl/Wrd	Average Let/Syl
Preface	0.000*	0.000*	0.000*
Unknown	5.588	1.503	3.718
Bayle 01	5.491	1.521	3.611
Bayle 02	5.542	1.510	3.669
Bayle 03	5.429	1.445	3.758
Bayle 04	5.611	1.586	3.538
Bayle 05	5.677	1.580	3.594
Bayle 06	5.685	1.651	3.443
Bayle 07	5.421	1.513	3.582
Bayle 08	5.425	1.507	3.600
Bayle 09	5.728	1.636	3.502
All of B	5.509	1.525	3.614
Larroque 01	5.475	1.485	3.686
Larroque 02	5.551	1.518	3.657
Larroque 03	5.471	1.495	3.659
Larroque 04	5.504	1.464	3.760
Larroque 05	5.566	1.518	3.666
Larroque 06	5.325	1.474	3.613
Larroque 07	5.737	1.648	3.482
Larroque 08	5.537	1.598	3.466
Larroque 09	5.698	1.688	3.376
Larroque 10	5.688	1.567	3.630
Larroque 11	5.760	1.692	3.405
Larroque 12	5.595	1.625	3.443
Larroque 13	5.252	1.435	3.659
Larroque 14	5.466	1.429	3.826
Larroque 15	5.530	1.572	3.518
Larroque 16	5.400	1.428	3.782
Larroque 17	5.644	1.606	3.514
Larroque 18	5.777	1.696	3.406
Larroque 19	5.644	1.618	3.488
Larroque 20	5.463	1.517	3.602
Larroque 21	5.562	1.589	3.500
Larroque 22	5.162	1.332	3.874
Larroque 23	5.725	1.519	3.768
All of L	5.524	1.525	3.624

TABLE 4:6A. (CONTINUED)

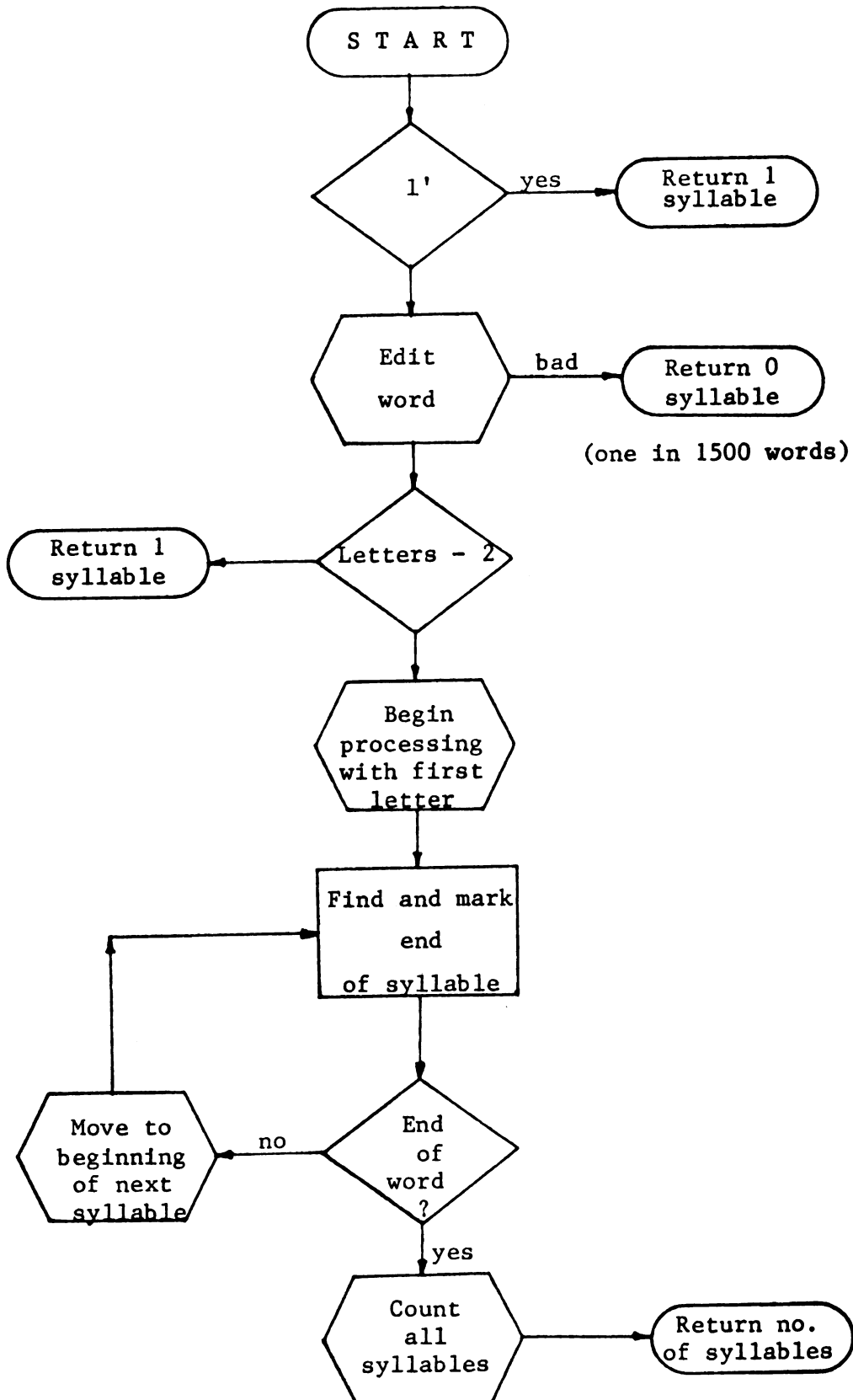
Article	Average Let/Wrd	Average Syl/Wrd	Average Let/Syl
Mean of / X-B /	.109	.059	.138
Mean of / X-L /	.127	.082	.155
Calculated F-value	.273	.965	.207
Level of Significance	.611	.644	.656

*The Preface was not tested separately for these variables, therefore "0" values are present.

APPENDIX G

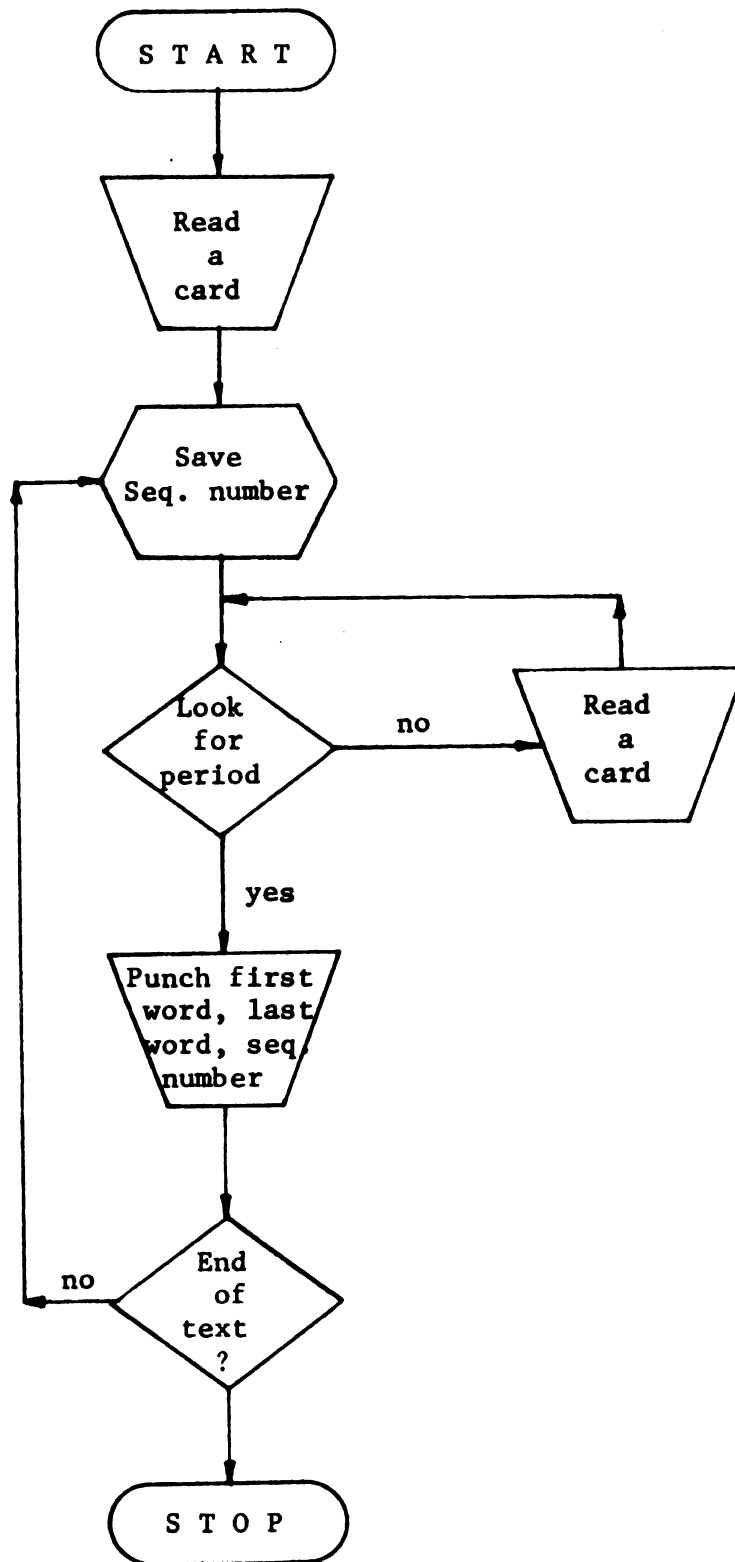
APPENDIX G

SYLAN-SUBROUTINE BREAK FLOWCHART



APPENDIX H

APPENDIX H
STYLBEND FLOWCHART



APPENDIX I

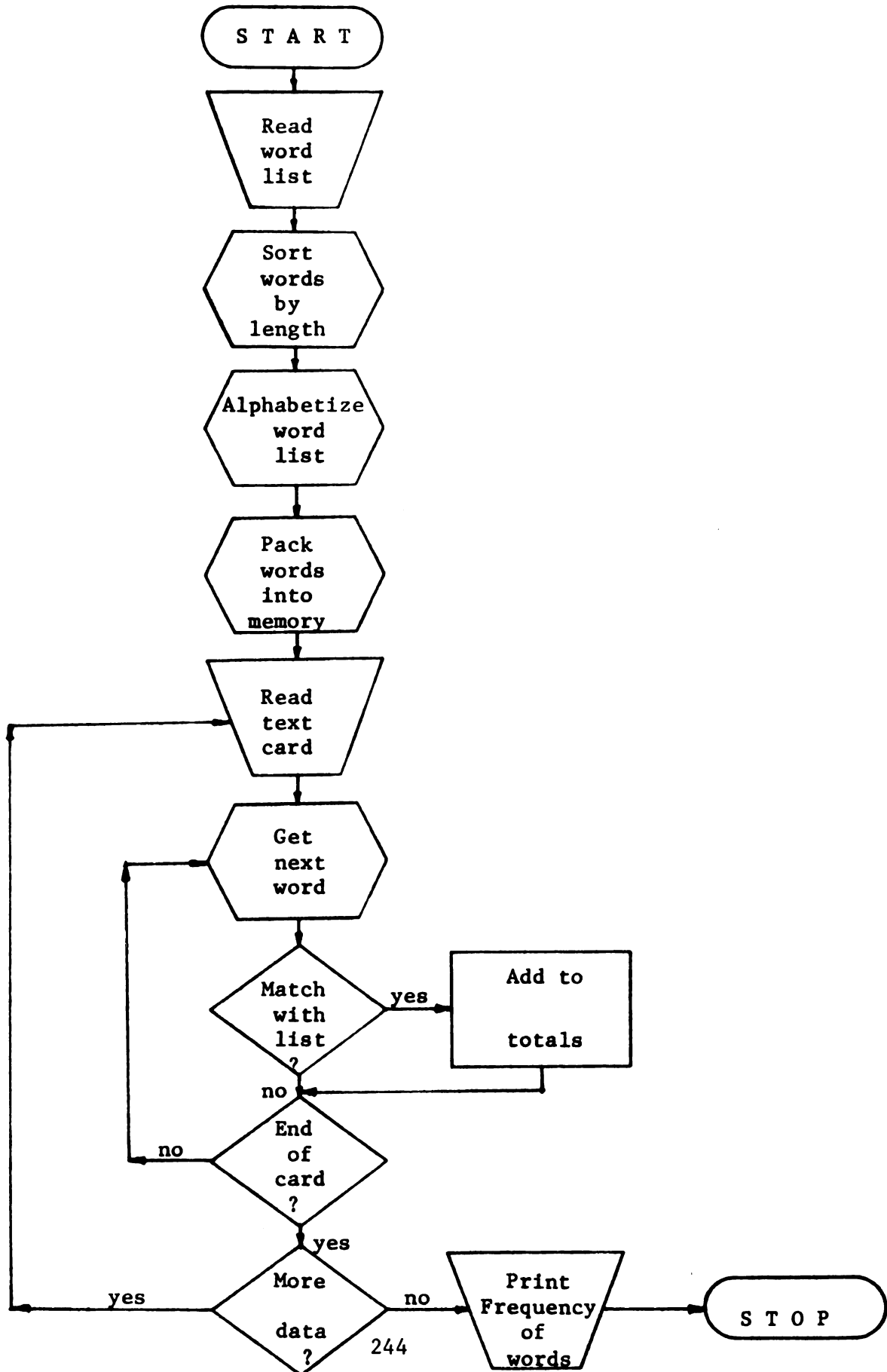
SAMPLE STYLBEND OUTPUT

243

APPENDIX J

APPENDIX J

FREQFUN FLOWCHART



APPENDIX K

APPENDIX K

FUNCTION WORD ABSOLUTE FREQUENCIES

WORD	P	X	B1	B2	L1
-----	----	----	----	----	----
A*	79	1365	1574	344	889
AFIN	1	26	29	4	26
AINSI	1	59	39	13	49
AILLEURS	1	18	15	4	11
ALLER	2	23	70	13	14
ALORS	1	16	9	3	0
APRES	6	56	45	24	52
ASSEZ	5	28	43	6	23
AUCUN	2	53	54	5	37
AUJOURD'HUI	2	17	10	11	17
AUPARAVANT	0	2	3	2	7
AUSSI	2	84	72	24	65
AUSSI-BIEN	0	5	3	1	0
AUSSI-TOT	0	4	2	0	1
AUTANT	2	47	23	5	11
AUTRE	2	189	193	50	145
AUTREFOIS	0	10	4	4	11
AVANT	1	23	39	1	13
AVEC	11	158	138	42	94
AVOIR	128	1294	1629	341	876
BAS	0	7	3	1	1
BEAU	0	16	17	0	11
BEAUCOUP	5	37	34	4	8
BIEN	6	139	186	36	122
BIEN-AISE	0	2	8	2	1
BIENTOT	1	8	16	2	2
BON	6	87	91	14	45
C'EST-A*-DIRE	0	18	14	1	0
CAR	1	114	129	36	94
CAS	1	33	32	4	13
CAUSE	5	46	49	12	11
CE	76	993	1130	292	832
CECI	0	9	15	2	14
CELA	4	80	114	42	116
CELUI	2	102	120	14	128
CEUX	7	145	125	22	87
CEPENDANT	1	9	9	3	20
CERTAIN	1	31	36	17	57
CHACUN	1	16	8	2	24
CHAQUE	0	27	6	2	0
CHEZ	0	3	17	5	8
CHOSE	5	80	105	23	86
COMBIEN	1	11	9	6	12
COMME	7	153	151	77	176
COMMENT	0	20	22	6	45
CONTRE	9	191	103	48	63
COTE	2	19	22	1	8
CROIRE	6	63	123	29	147
DANS	17	362	364	99	310
DAVANTAGE	0	4	5	0	6
DE	281	4764	4497	1157	2477

WORD	P	X	B1	B2	L1
-----	-----	-----	-----	-----	-----
DEDANS	0	1	3	0	0
DEJA	3	18	34	3	8
DEMANDER	2	17	41	9	31
DEPUIS	3	28	68	7	16
DESSUS	0	16	6	1	6
DEVANT	2	9	14	1	15
DEVOIR	5	97	120	52	131
DEVOIR COND	0	8	23	2	6
DE*S	3	15	20	7	10
DIRE	6	192	299	73	310
DONC	2	71	69	10	35
DONNER	5	92	164	24	66
DONT	7	73	102	30	51
FLLE	2	108	41	34	114
ELLES	0	26	17	8	43
EN	64	794	950	190	482
ENCORE	0	76	74	23	64
ENFANT	0	14	5	3	17
ENFIN	0	35	22	13	17
ENSUITE	0	8	2	0	1
ENTENDRE	2	15	23	5	13
ENTRE	3	52	37	10	40
ENVERS	1	11	5	2	1
ET	100	1308	1224	323	796
ETC	2	10	27	6	4
ETRE	84	1412	1563	387	1062
EXCEPTE	4	3	4	0	0
FAIRE	45	500	648	125	330
FALLOIR	4	97	139	13	53
FEMME	0	4	8	2	13
FEU	1	17	9	9	2
FOIS	0	29	26	13	20
FORCE	1	38	13	15	18
FORT	2	75	98	19	60
GRAND	10	132	87	27	64
HOMME	2	44	140	20	61
HORS	1	9	2	2	3
ICI	3	37	68	6	9
IL	65	901	2007	234	797
ILS	16	451	127	64	168
JAMAIS	1	74	103	19	46
JE	86	342	1374	126	553
JOUR	0	26	40	3	28
JUSQUE	3	46	38	7	13
LAQUELLE	1	24	19	7	18
LA*	5	36	41	12	21
LA*-DESSUS	0	5	10	6	12
LE	238	4291	4087	1062	2778
LEQUEL	3	23	38	5	22
LEUR	17	434	120	71	124
LOIN	1	13	17	5	5
LORS	1	8	40	12	13
LORSQUE	0	60	0	0	0
L@ON	4	103	75	27	45
MAIN	1	31	33	12	9
MAINTENANT	0	6	2	0	1
MAIS	10	196	187	64	171

WORD	P	X	B1	B2	L1
MAL	1	50	40	11	20
MALGRE	0	9	3	2	13
MEME	9	145	124	38	123
METTRE	2	61	86	12	29
MIEUX	1	37	41	4	28
MOIEN	1	22	21	4	2
MOINS	5	85	64	11	35
MON	13	22	187	7	49
MONSIEUR	2	3	411	62	185
NE	55	1175	1219	263	853
NE-EST-CE-PAS	0	14	6	5	0
NEANMOINS	0	22	11	4	3
NI	0	80	113	14	46
NON	2	64	40	16	48
NONOBTANT	0	6	2	1	2
NOTRE	32	113	116	17	119
NOUS	18	268	207	41	322
NUL	1	18	20	3	12
ON	32	412	461	109	497
OR	0	23	23	4	25
OU	16	243	159	50	93
OUTRE	0	12	16	4	8
OU*	6	90	131	27	64
PAR	16	358	314	99	223
PARCE	1	23	24	7	70
PARLER	2	82	85	19	41
PARMI	2	30	15	7	4
PAS	20	425	521	93	311
PASSER	4	41	51	14	26
PENDANT	3	19	16	4	10
PERSONNE	4	75	84	11	34
PETIT	1	38	53	9	6
PEU	5	54	61	10	41
PEUTETRE	1	11	17	2	19
PIRE	1	12	9	2	0
PLUS	19	413	304	97	126
PLUSIEURS	4	34	23	8	9
PLUTOT	1	14	20	2	17
POINT	12	192	145	41	142
POUR	25	428	511	96	228
POURQUOI	1	15	41	3	30
POURTANT	0	7	6	0	27
POURVU	0	15	9	0	3
POUVOIR	23	359	280	81	229
PRENDRE	6	79	49	21	52
PRES	1	5	6	2	6
PRESTQUE	2	12	12	1	13
PROPOS	1	9	3	3	10
PUIS	3	16	68	22	90
PUISQUE	0	54	0	3	0
QUAND	3	83	52	15	61
QUE	158	2415	2859	659	1913
QUEL	5	34	36	10	14
QUELQUE	12	116	154	31	54
QUELQUEFOIS	0	10	6	2	6
QUELQUES-UNS	2	12	7	3	2
QUI	58	694	793	187	517

WORD	P	X	B1	B2	L1
-----	---	---	---	---	---
QUOI	8	54	56	19	37
REGARDER	3	21	15	8	20
REPONDRE	8	31	56	48	55
RESTE	3	19	15	11	9
RIEN	5	64	70	23	45
SANS	5	178	150	26	80
SAVOIR	9	89	202	29	106
SELON	4	55	39	11	31
SEULEMENT	5	39	34	9	20
SE	28	510	546	120	225
SI	20	299	443	91	264
SON	26	402	465	81	208
SOUS	2	69	30	7	32
SOUVENT	1	16	13	7	4
SUR	18	224	263	76	128
TANT	7	71	77	33	31
TEL	3	49	71	18	9
TELLEMENT	1	5	12	2	0
TEMPS	9	50	64	22	43
TOUJOURS	3	43	42	10	40
TOUT	23	491	424	80	251
TOUT-A*-FAIT	1	17	7	0	4
TRES	5	49	49	21	12
TROP	8	27	22	7	21
TROUVER	9	102	121	20	36
UN	40	802	1027	224	645
VENIR	7	53	87	13	33
VOICI	0	7	13	8	10
VOILA	1	31	24	9	20
VOIR	13	114	192	44	155
VOTRE	0	345	6	99	15
VOULOIR	2	120	155	25	81
VOUS	1	796	15	311	73
Y	21	258	280	66	182

APPENDIX L

APPENDIX L

TABLE 4:34A. FREQFUN STATISTICAL COMPARISON, X:L1

88 WORDS

Word	t-Value	CL	Word	t-Value	CL
alors	3.229	.005	leur	8.351	.001
auparavant	2.353	.025	lors	2.106	.05
autant	3.193	.005	lorsque	6.254	.001
beaucoup	2.975	.005	l'on	2.253	.025
bien	2.415	.005	main	2.194	.05
c'est-à-dire	3.425	.001	mais	2.807	.005
cause	3.114	.005	même	2.163	.05
ce	5.422	.001	moien	3.119	.005
ceci	2.102	.05	moins	2.306	.025
cela	5.659	.001	mon	5.099	.001
celui	5.035	.001	monsieur	16.557	.001
cependant	3.253	.005	ne	2.435	.025
certain	4.863	.001	n'est-ce pas	3.020	.005
chacun	2.660	.01	néanmoins	2.029	.05
chaque	4.195	.001	notre	3.696	.001
chez	2.258	.025	nous	7.543	.001
chose	3.261	.005	on	9.440	.001
comme	5.223	.001	ou	4.422	.001
comment	4.195	.001	parce	7.072	.001
contre	4.782	.001	parmi	3.303	.001
croire	9.069	.001	petit	3.504	.001
dans	3.558	.001	peut-être	2.677	.01
de	9.506	.001	pire	2.796	.01
demander	3.564	.001	plus	7.657	.001
devant	2.311	.05	plusieurs	2.485	.025
devoir	5.571	.001	pour	2.468	.025
dire	10.253	.001	pourquoi	3.737	.001
elle	3.633	.001	pourtant	4.769	.001
elles	3.889	.001	pourvu	1.978	.05
envers	2.206	.05	puis	9.582	.001
être	3.589	.001	puisque	5.933	.001
femme	3.124	.005	que	6.554	.001
feu	2.580	.01	quelque	2.052	.05
homme	3.912	.001	qui	2.325	.025
ici	2.760	.01	répondre	4.652	.001
il	6.374	.001	sans	2.778	.01
ils	6.286	.001	savoir	4.265	.001
je	13.742	.001	se	4.921	.001
jusque	2.738	.01	si	3.625	.001
là-dessus	2.627	.01	son	2.712	.01

TABLE 4:34A. (CONTINUED)

Word	t-Value	CL	Word	t-Value	CL
tel	3.730	.001	un	4.022	.001
tout	3.145	.005	voir	6.107	.001
très	3.161	.005	votre	13.724	.001
trouver	3.214	.005	vous	18.816	.001

APPENDIX M

APPENDIX M

TABLE 4:34B. FREQFUN STATISTICAL COMPARISON P:B1

31 WORDS

Word	P≠B1	CL	Word	P≠B1	CL
autre	2.777	.01	ne	1.991	.05
avoir - all	3.150	.005	ni	2.588	.01
beaucoup	1.969	.05	notre	8.507	.001
car	2.394	.025	ou	2.053	.05
dire - all	2.758	.01	plusieurs	2.089	.05
encore	2.093	.05	quelques-uns	2.174	.05
et	3.153	.005	quoi	2.411	.025
excepté	5.471	.001	regarder - all	2.048	.05
grand	2.028	.05	répondre - all	2.411	.025
homme	2.170	.05	reste - verb	2.048	.05
hors	2.093	.05	seulement	1.969	.05
il	4.936	.001	temps	2.511	.025
ils	2.922	.005	trop	5.030	.001
jamais	2.055	.05	un	2.634	.01
leur	3.482	.001	vouloir - all	2.355	.025
monsieur	4.529	.001			

APPENDIX N

APPENDIX N

TABLE 4:41A. REFINED FUNCTION WORD LIST RESULTS, X = L1

9 WORDS

Word	X:B1	CL	B1:L1	CL	Least CL	Sign Comb*
à	2.823	.005	2.496	.025	.025	B1:L1
ainsi	2.217	.05	3.312	.001	.05	X:B1
déjà	2.077	.05	2.592	.01	.05	X:B1
depuis	3.892	.001	3.667	.001	.001	EQUAL
etc.	2.676	.01	2.930	.005	.01	X:B1
falloir	2.434	.025	3.108	.005	.025	X:B1
ni	2.104	.05	2.484	.025	.05	X:B1
non	2.556	.025	3.093	.005	.025	X:B1
quand	2.900	.005	3.382	.001	.005	X:B1

*See Chapter IV, (p. 188), for explanation of this column.

APPENDIX O

APPENDIX O

TABLE 4:41B. REFINED FUNCTION WORD LIST RESULTS, X=B1

15 WORDS

Word	X:L1	CL	B1:L1	CL	Least Sign CL Comb*
alors	3.229	.005	2.374	.025	.025 B1:L1
auparavant	2.353	.025	2.047	.05	.05 B1:L1
beaucoup	2.975	.005	2.592	.01	.01 B1:L1
celui	5.035	.001	4.246	.001	.001 EQUAL
cependant	3.253	.005	3.371	.001	.005 X:L1
c'est-à-dire	3.425	.001	2.961	.005	.005 B1:L1
comme	5.223	.001	5.700	.001	.001 EQUAL
comment	4.915	.001	4.822	.001	.001 EQUAL
dans	3.558	.001	4.008	.001	.001 EQUAL
mais	2.807	.005	3.606	.001	.005 X:L1
ne	2.435	.025	2.515	.025	.025 EQUAL
parce	7.072	.001	7.168	.001	.001 EQUAL
pire	2.796	.01	2.374	.025	.025 B1:L1
pourtant	4.769	.001	5.114	.001	.001 EQUAL
très	3.161	.005	3.024	.005	.005 EQUAL

*See Chapter IV, (p. 188), for explanation of this column.

APPENDIX P

APPENDIX P

TABLE 4:41C. REFINED FUNCTION WORD LIST RESULTS, X ≠ B1 OR L1

26 WORDS

Word	X:B1	CL	X:L1	CL	Least CL	Sign Comb*
autant	3.035	.005	3.193	.005	.005	EQUAL
bien	2.255	.025	2.415	.025	.025	EQUAL
ce (pron.)	2.084	.05	5.422	.001	.05	X:B1
cela	2.169	.05	5.659	.001	.05	X:B1
chaque	3.770	.001	4.195	.001	.001	EQUAL
chez	3.043	.005	2.258	.025	.025	X:L1
contre	5.479	.001	4.782	.001	.001	EQUAL
de	4.877	.001	9.506	.001	.001	EQUAL
ici	2.825	.005	2.760	.01	.01	X:L1
lors	4.484	.001	2.106	.05	.05	X:L1
lorsque	7.902	.001	6.254	.001	.001	EQUAL
l'on	2.364	.025	2.253	.025	.025	EQUAL
moins	1.963	.05	2.306	.025	.05	X:B1
néanmoins	2.029	.05	2.808	.005	.05	X:B1
ou	4.594	.001	4.422	.001	.001	EQUAL
parmi	2.369	.025	3.303	.001	.025	X:B1
plus	4.614	.001	7.657	.001	.001	EQUAL
pour	2.113	.05	2.468	.025	.05	X:B1
pourquoi	3.328	.001	3.737	.001	.001	EQUAL
puis	5.496	.001	9.582	.001	.001	EQUAL
puisque	7.497	.001	5.933	.001	.001	EQUAL
que	4.795	.001	6.554	.001	.001	EQUAL
quelque	1.991	.05	2.052	.05	.05	EQUAL
si	4.766	.001	3.625	.001	.001	EQUAL
tout	2.824	.005	3.145	.005	.005	EQUAL
voir	4.120	.001	6.107	.001	.001	EQUAL

*See Chapter IV, (p. 188) for explanation of this column.

APPENDIX Q

APPENDIX Q

FUNCTION WORDS IN GROUPS

COORD CONJ L-O

ET
OU
NI
QUAND
SI
QUE
MAIS
CAR
COMME
DONC
OR

COORD CONJ LIA

ET
ETC
NI
PUIS
ENSUITE
ALORS
AUSSI
COMME
AVEC

COORD CONJ CAUSE

CAR
EFFECTIVEMENT
BIEN

COORD CONJ CON

DONC
AUSSI
PARTANT
ALORS
AINSI
ENFIN
CONSEQUEMMENT

COORD CONJ TRAN

OR

COORD CONJ OP

MAIS
ET
CEPENDANT
TOUTEFOIS
NEANMOINS
POURTANT
QUOIQUE
SIMON
ENCORE
SEULEMENT

COORD CONJ ALT

OU

COORD CONJ EXPL

CEST-A*-DIRE
SOIT

SUBORD CONJ

PUISQUE
LORSQUE
AFIN
POURVU
PARCE

PREP LATIN ORIG

A*
AVANT
AVEC
CONTRE
DE
DERRIERE
DE*S
EN
ENTRE
ENVERS
JUSQUE
OUTRE
PAR
POUR
SANS
SOUS
SUR
VERS

PREP GENERAL

A*
APRES
AVANT
AVEC
CHEZ
CONCERNANT
CONTRE
DANS
DE
DEPUIS
DERRIERE
DE*S
DEVANT
DURANT
EN
ENTRE
ENVERS
EXCEPTE
HORMIS
HORS
JUSQUE
MALGRE
MOYENNANT
OUTRE
PAR
PARMI
PENDANT
POUR
PRES
PROCHE
SANS
SELON
SOUS
SUIVANT
SUR
TOUCHANT
VERS
VOICI
VOILA
DEDANS
DELA
DESSOUS
DESSUS
ENVIRON
ES
EX
IN
JOIGNANT
NONOBSANT

PREP AS ADVERBS

AVEC
APRES
AVANT
CONTRE
DEPUIS
DERRIERE
DEVANT
ENTRE
HORS
OUTRE
PARMI
PROCHE
SANS
SELON
A*
DANS
SUR
SOUS
DEDANS
DESSUS
DESSOUS

ADV LATIN ORIGIN

AILLEURS
APRES
ASSEZ
AVANT
AVEC
BIEN
DEMAIN
EN
ENSEMBLE
LA*
LOIN
MAL
MIEUX
MOINS
OU*
PIS
PLUS
QUAND
TANT
TARD
TRES
VOLONTIERS
Y
PEU
TOT

ADV COMPOSED

AUSSI-TOT
 AUSSI-BIEN
 LA*-DESSUS
 DESSOUS
 LA*-DEDANS
 POURQUOI
 JAMAIS
 CI-DESSUS
 BIENTOT
 DEDANS
 DESSUS
 DEPUIS
 DEBOUT
 AUTOUR
 DAVANTAGE
 AUTREFOIS
 LONGTEMPS
 BEAUCOUP
 TOUJOURS
 CEPENDANT

ADV MANNER

AINSI
 BIEN
 COMME
 DEBOUT
 COMMENT
 EXPRES
 ENSEMBLE
 GRATIS
 MAL
 MIEUX
 PIS
 PLUTOT
 QUASI
 VOLONTIERS

ADV QUANTITY

ASSEZ
 AUSSI
 AUTANT
 AUTREMENT
 BEAUCOUP
 COMBIEN
 COMME
 COMMENT
 DAVANTAGE
 ENVIRON
 FORT
 GUERE
 MOINS
 PAS
 PEU
 PLUS
 PLUSIEURS
 PRESQUE
 QUELQUE
 SI
 TANT
 TELLEMENT
 TOUT
 TOUT-A*-FAIT
 TRES
 TROP

ADVERBS TIME

ALORS
 APRES
 AUJOURD'HUI
 AUPARAVANT
 AUSSI-TOT
 AUTREFOIS
 AVANT
 BIENTOT
 CEPENDANT
 DEJA
 DEMAIN
 DEPUIS
 DESORMAIS
 ENCORE
 ENFIN
 ENSUITE
 JAMAIS
 LONGTEMPS
 LORS
 MAINTENANT
 PUIS
 QUAND
 QUELQUEFOIS
 SOUVENT
 TANTOT
 TARD
 TOT
 TOUJOURS

ADVERBS PLACE

ALENTOUR
 ARRIERE
 ATTENANT
 AUTOUR
 AVANT
 BAS
 CI
 CONTRE
 DEDANS
 DEHORS
 DERRIERE
 DESSOUS
 DESSUS
 DEVANT
 ICI
 LA*
 LOIN
 OU*
 OUTRE
 PRFS
 PROCHE
 LA*-DEDANS

ADV AFFIRM

ASSUREMENT
 AUSSI
 CERTAINEMENT
 BIEN
 CERTES
 OUI
 PRECISEMENT
 SI
 VOLONTIERS
 VRAIMENT

ADV NEG

NON
 NE
 AUCUNEMENT
 NULLEMENT
 NE-EST-CE-PSA
 PAS

ADVERBS DOUBT

PEUTETRE

PRONOUNS

AUTRE
 CEUX
 CELUI
 CHACUN
 DONT
 JE
 QUELCONQUE
 QUICONQUE
 TU
 IL
 ELLE
 ON
 CE
 NOUS
 VOUS
 UNS
 ILS
 ELLES
 L'ON
 LEUR
 CECI
 CELA
 NOTRE
 PERSONNE
 QUELQUES-UNS
 VOTRE
 QUI
 LEQUEL
 LAQUELLE
 OU*
 QUOI
 SE

ADJECTIVES

AUCUN
 AUTRE
 QUELCONQUE
 QUELQUE
 TOUT
 NOTRE
 VOTRE
 LEUR
 CE
 UN
 CERTAIN
 CHAQUE
 MEME
 NUL
 QUEL
 SON
 TEL
 MON

ARTICLES

LE
UN
DES

VERBS GENERAL

ALLER
AVOIR
CROIRE
DEMANDER
DEVOIR
DIRE
DONNER
ENTENDRE
ETRE
FAIRE
FALLOIR
METTRE
PARLER
PASSER
POUVOIR
PRENDRE
REGARDER
REPONDRE
SAVOIR
TROUVER
VENIR
VOULOIR
FORCE
RESTE
VOIR

CONJUNCTIONS ALL

COORD CONJ EXP
COORD CONJ ALT
COORD CONJ OP
COORD CONJ TRAN
COORD CONJ CON
COORD CONJ CAUSE
COORD CONJ L O
COORD CONJ LIA

PREPOSITIONS ALL

PREP LATIN ORIG
PREP GENERAL
PREP AS ADVERBS

ADVERBS ALL

ADV LATIN ORIGIN
ADV COMPOSED
ADV MANNER
ADV QUANTITY
ADV TIME
ADV PLACE
ADV AFFIRM
ADV NEG
ADV DOUBT

NOUNS

CHOSE
ENFANT
FEMME
FOIS
HOMME
JOUR
MAIN
MONSIEUR
POINT
TEMPS
CAS
CAUSE
COTE
DOMMAGE
FACE
MOIEN
MORAL
PERSONNES
PROPOS
TAS

APPENDIX R

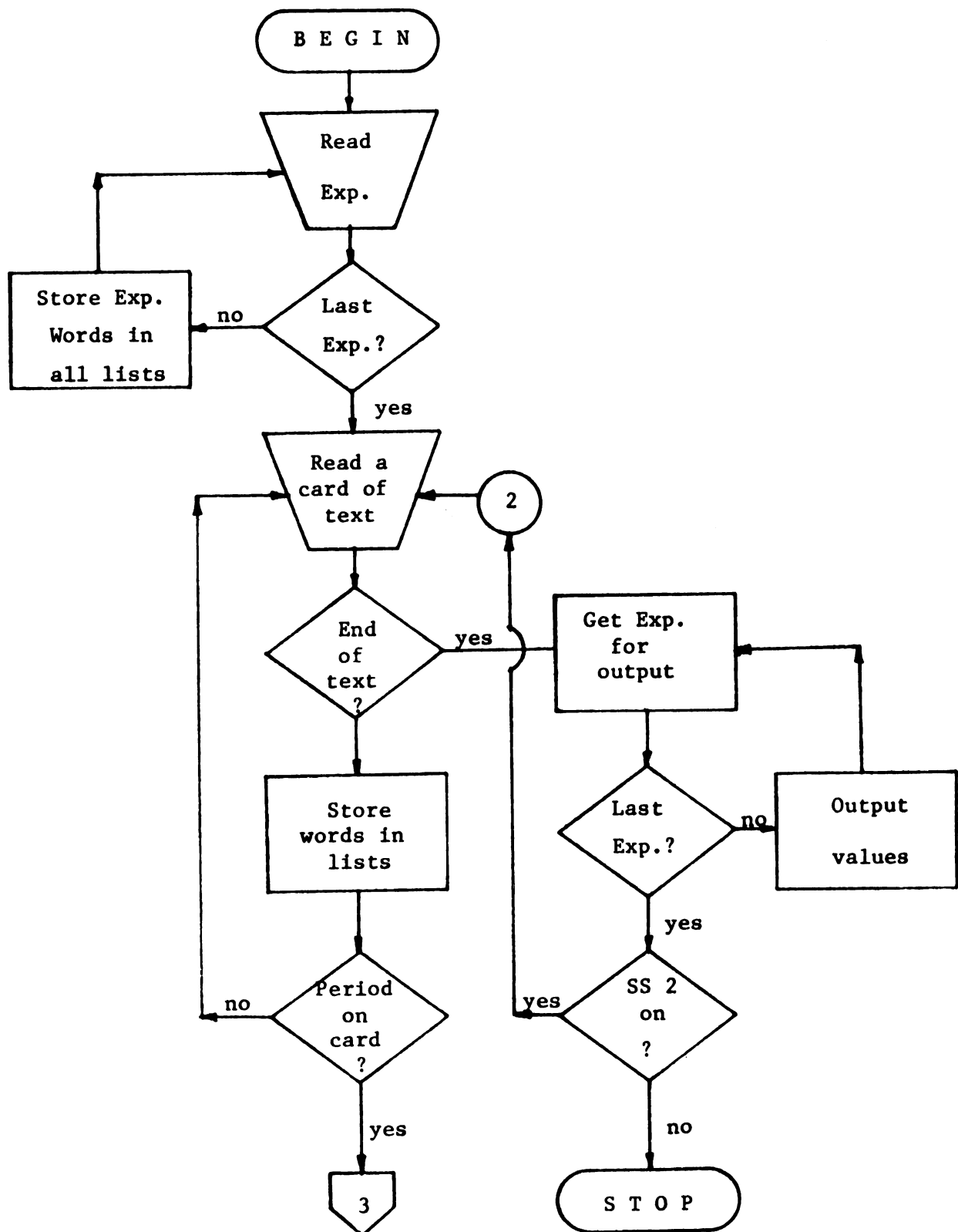
APPENDIX R

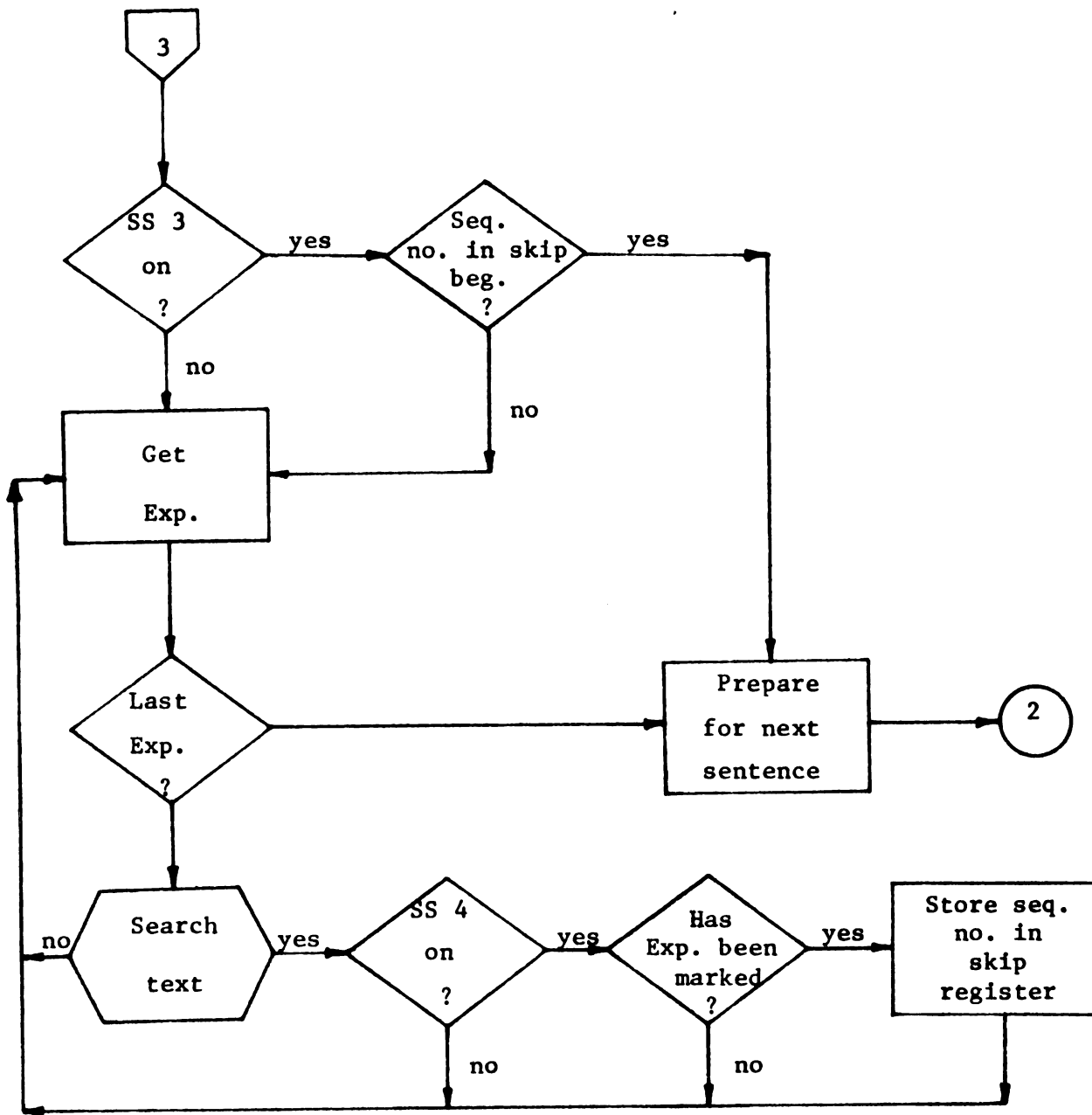
TABLE 4:41D. FREQFUN GROUPED DATA, ABSOLUTE FREQUENCIES

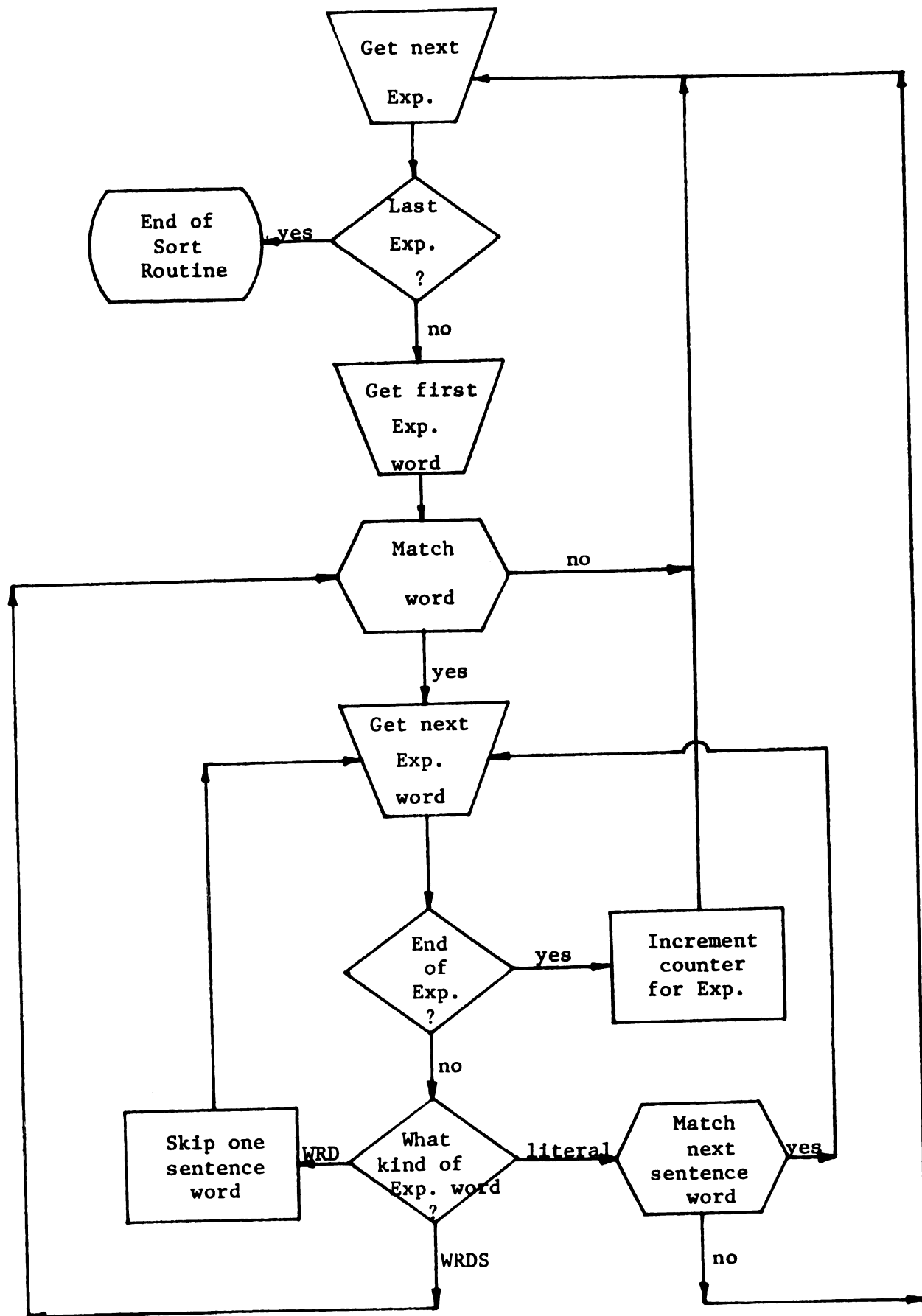
Group	P	X	B1	B2	L1
Coord. conj.-Latin origin	317	4985	5409	1343	3674
Coord. conj.-alternation	16	243	159	50	93
Coord. conj.-cause	7	253	315	72	218
Coord. conj.-consequence	6	267	211	63	170
Coord. conj.-explication	1	119	99	12	55
Coord. conj.-liaison	126	1833	1804	511	1272
Coord. conj.-opposition	116	1665	1548	427	1107
Coord. conj.-transition	0	23	23	4	25
Coord. conj.-all	342	5566	5902	1495	4064
Subord. conj.-all	7	269	149	80	179
Conjunctions-all	349	5835	6051	1575	4243
Prepositions-Latin origin	522	8690	8687	2116	4786
Prepositions-adverbial	164	2841	3908	712	1786
Prepositions-all	568	9372	9357	2307	5318
Adverbs-Latin origin	174	2472	2586	619	1434
Adverbs-affirmation	29	529	729	155	480
Adverbs-composed	15	277	385	71	219
Adverbs-doubt	1	11	17	2	19
Adverbs-manner	58	853	864	234	737
Adverbs-negation	78	1682	1792	379	1212
Adverbs-place	30	451	453	113	216
Adverbs-quantity	156	2573	2644	634	1590
Adverbs-time	30	589	703	195	508
Adverbs-all	391	6580	8058	1696	1668
Adjectives-(poss., demon., indef.)	262	4265	4169	1045	2694
Articles-(def., indef.)	330	5825	5704	1636	3423
Nouns	45	654	1151	232	665
Pronouns	482	7403	8054	1975	5192
Verbs	381	5011	6226	1431	3923

APPENDIX S

APPENDIX S
EXSOR FLOWCHART







APPENDIX T

APPENDIX T

LIST OF EXSOR EXPRESSIONS

AFIN DE	GRACE AUX
AFIN QUE	GRACE A*
AINSI QUE	HOMME DE BIEN
ALORS QUE	HORS DE
AMOUR DE VOUS	HORS DE DOUTE
APPAREMENT QUE	ICY
APRES LE TOUT	IL EST
APRES QUE	IL EST A*
A* BON DROIT	IL EST BESOIN
A* BOUT	IL EST QUESTION
A* CAUSE DE	IL EST SI VRAI
A* CAUSE QUE	IL EST VRAI
A* CE QUE	IL EST WRDS A*
A* CENT MILLE	IL EST WRDS QUE
A* CETTE HEURE	IL ETOIT
A* CONDITION QUE	IL FALLOIT
A* COTE	IL FAUDRA
A* COTE DE	IL FAUDRA QUE
A* COUP DE EPEES	IL FAUDROIT
A* COUPS DE PLUME	IL FAUT
A* COUVERT	IL NE EST PAS
A* DOUBLE POIDS	IL NE EN EST PAS DE MEME
A* FORCE DE	IL NE EST PAS JUSQUE
A* JAMAIS	IL NE EST POINT DE
A* LA BESOGNE	IL NE ETOIT PAS
A* LA CONSCIENCE	IL NE FALLOIT PAS
A* LA FAVEUR	IL NE FAUDRA PAS
A* LA FAVEUR DE	IL NE FAUDROIT PAS
A* LA MANIERE	IL NE FAUT PAS
A* LA MERCI	IL NE RESTE WRDS QUE
A* LA VERITE	IL NE SEMBLE PAS
A* LAQUELLE	IL NE SERA PAS
A* LEUR DROIT	IL NE SEROIT PAS
A* LONGS TRAITS	IL NE Y A PAS
A* L@ APPETIT	IL NE Y A PAS LOIN
A* L@ EGARD DE	IL NE Y AIT
A* L@ EGARD DES	IL SE EN FAUT BIEN
A* L@ EGARD DU	IL SE ENSUIT QUE
A* L@ OCCASION	IL SE PASSE
A* MESURE QUE	IL SE PEUT
A* MOINS DE CELA	IL SERA
A* MOINS QUE	IL SEROIT
A* MOINS QUE WRDS NF	IL VOUS WRDS MIEUX
A* NEANT	IL WRDS AGI
A* PEINE	IL WRDS AGIRA

A* PEU DE FRAIS
 A* PEU PRES
 A* PLUS FORTE RAISON
 A* PRESENT QUE
 A* PROPORTION QUE
 A* PROPOS
 A* PROPOS DE
 A* PUR ET A* PLEIN
 A* QUOI
 A* QUOI BON
 A* SA CHARGE
 A* SON GOUT
 A* TACHE
 A* TOUT LE MOINS
 A* TOUT PERDRE
 A* TOUTE HEURE
 A* TRAVERS
 A* VOTRE TOUR
 A* WRDS EGARD
 A*-PEU PRES
 A*-PROPOS
 ASSEZ BONNE FOI
 ATTENDU QUE
 AU BOUT DU COMPTE
 AU BOUT DU MONDE
 AU CONTRAIRE
 AU DELA
 AU DELA DE
 AU DESSOUS
 AU DESSOUS DE
 AU DESSUS
 AU DESSUS DE
 AU DEVANT
 AU DEVANT DE
 AU JUSTE JUGE
 AU LARGE
 AU LIEU
 AU LIEU DE
 AU LIEU QUE
 AU MILIEU
 AU MILIEU DE
 AU MOIEN
 AU MOINS
 AU PARTI
 AU PIED DE LA LETTRE
 AU PIS ALLER
 AU PLUTOT
 AU PRES
 AU PRIX
 AU PRIX DE
 AU RESTE
 AU TEMPS
 AU TRAVERS
 AU TRAVERS DE

IL WRDS AGIT
 IL WRDS SEMBLE
 IL Y AIT
 JAMAIS WRDS NE
 JOINT QUE
 JUSQUE AU
 JUSQUE AUX
 JUSQUE A*
 JUSQUE A* CE QUE
 JUSQUES DANS
 JUSQUE LA* QUE
 LA JUSTE CRAINTE
 LA JUSTE IDEE
 LA MINE
 LA MOINDRE CAUSE
 LA VOIE DES REMONTRANCES
 LE BIEN COMMUN
 LE COMBLE
 LE DEVOIR
 LE DROIT DE EXAMINER
 LE DROIT DE ELECTION
 LE LONG DE
 LE PUBLIC
 LE REPOS PUBLIC
 LES ARMES A* LA MAIN
 LES AUTRES
 LES UNS
 LOIN DE
 LORS MEME QUE
 LORS QUE
 LORSQUE OUTRE
 LUMIERE DE LA EXAMEN
 LA ADRESSE DE
 LA AN
 LA AUTRE
 LA EGALITE
 LA OEIL SUR
 LON
 MAINS DU BOURREAU
 MAL AISE
 MAL ASSORTI
 MAL A* PROPOS
 MAL FONDEZ
 MEILLEUR WRDS QUE WRDS NE
 MEILLEURE WRDS QUE WRDS NE
 MENACE DE
 MENU PEUPLE
 MOINDRE QUE
 MOINS
 MOINS DE SUJET
 MOINS WRDS QUE
 MOINS WRDS QUE WRDS NE
 MOINS WRD WRDS QUE
 NE EST-CE PAS

AUCUN WRDS NE
 AUCUNE WRDS NE
 AUPRES DE
 AUSSI WRD WRDS QUE
 AUSSITOT QUE
 AUSSI-TOT QUE
 AUTANT DE
 AUTANT QUE
 AUTANT WRDS AUTANT
 AUTANT WRDS QUE
 AUTOUR DE
 AUTREMENT QUE
 AUX DEPENS
 AUX PIEDS
 AVANT QUE
 AVANT QUE DE
 AVANT QUE WRDS NF
 AVOIR SUSPECT
 AVOIR SUSPECTE
 BANQUE OUVERTE
 BESOIN QUE
 BIEN COMMUN
 BIEN DE AUTRES
 BIEN ENTENDU
 BIEN ENTENDU QUE
 BIEN LOIN
 BIEN PUBLIC
 BIEN QUE
 BLANC AU NOIR
 BON A* DIRE
 BON GRE MAL GRF
 BON GRE MALGRE
 BON LUI SEMBLE
 BON SENS
 BONNE FOI
 C'EST-A*-DIRE
 CAR ALORS
 CAR COMME
 CAR DE
 CAR EN CAS QUE
 CAR PUIS QUE
 CAR QUAND
 CAR QUE
 CAR WRDS SI
 CAS OU*
 CAUSE QUE
 CE A* QUOI
 CE DONT
 CE EST
 CE EST AINSI
 CE EST AUSSI
 CE EST CE QUE
 CE EST CE QUI
 CE EST DE QUOI

NE EST-CE POINT
 NE EST-IL PAS
 NE FUT-CE QUE
 NE GUERE
 NE IMPORTE QUI
 NE IMPORTE QUOI
 NE JAMAIS
 NE PAS
 NE PEUT-ON PAS
 NE PLUS
 NE POINT
 NE SONT-CE PAS
 NE WRD GUERE
 NE WRD JAMAIS
 NE WRD PAS
 NE WRD PLUS
 NE WRD POINT
 NE WRD WRD WRDS GUERE
 NE WRD WRD WRDS JAMAIS
 NE WRD WRD WRDS PAS
 NE WRD WRD WRDS PLUS
 NE WRD WRD WRDS POINT
 NE WRD WRDS JAMAIS
 NE WRD WRDS NI
 NE WRD WRDS PAS
 NE WRD WRDS POINT
 NE WRD WRDS QUE
 NE WRD WRDS QUE EN CE QUE
 NE WRDS AUCUN
 NE WRDS AUCUNE
 NE WRDS GUERES
 NE WRDS JAMAIS PERSONNE
 NE WRDS JAMAIS RIEN
 NE WRDS NUL
 NE WRDS NULLE
 NE WRDS PAS NI WRDS NE
 NE WRDS PERSONNE
 NE WRDS PLUS JAMAIS
 NE WRDS PLUS PERSONNE
 NE WRDS PLUS QUE
 NE WRDS PLUS RIEN
 NE WRDS PLUS WRDS QUE
 NE WRDS QUE DE
 NE WRDS QUE PARCE QUE
 NE WRDS RIEN
 NE WRDS RIEN WRDS NI WRDS PERSO
 NE WRDS RIEN WRDS QUE
 NI MEME QUE
 NI WRD WRDS NI WRD WRDS NI WRD
 NON COMPRIS
 NON GUERE
 NON PARCE QUE WRDS MAIS PARCE (C
 NON PAS
 NON PAS QUE

CE EST LA*
 CE EST QUE
 CE EST SELON
 CE EST WRDS CE QUE
 CE EST WRDS CE QUI
 CE EST WRD WRDS QUE
 CE EST-A*-DIRE
 CE ETOIT
 CE FUT ETE
 CE NE EST PAS
 CE NE EST PAS LA PEINE
 CE NE EST PLUS
 CE NE ETOIT PAS
 CE NE SERA PAS
 CE NE SEROIT PAS
 CE NE SEROIENT PAS
 CE NE SOIT QUE
 CE NE SONT PAS
 CE QUE
 CE QUE CE EST QUE
 CE QUI
 CE QUI EST
 CE SERA
 CE SEROIENT
 CE SEROIT
 CE SONT
 COMBIEN DE
 COMBIEN PEU
 COMME
 COMME AUTANT
 COMME PAR EXEMPLE
 COMME SI
 COMME TRES-PROPRE
 COMMENT LE MONDE EST FAIT
 CONSIDERE A* PART
 COUP DE ESSAI
 COUPS DE FOUDRE
 COURIR SUS
 CRAINS QUE WRDS NE
 CRAINT QUE WRDS NE
 DANS LE CAS
 DANS LE MOMENT QUE
 DANS LE TEMS QUE
 DE AFFAIRE
 DE AILLEURS
 DE AILLEURS QUE
 DE AUCUN
 DE AUCUNS
 DE AUTANT DE
 DE AUTANT PLUS QUE
 DE CE TEMS-LA*
 DE CE QUE
 DE CE QUI
 DE DENT

NON PLUS
 NON POINT
 NON SEULEMENT
 NON WRDS MAIS
 NONOBSANT QUE
 NUL WRDS NE
 NULLE WRDS NE
 OR ALORS
 OR DE MEME QUE
 OR SI
 OU BIEN
 OU MEME QUE
 OU QUE
 OU WRD WRDS OU
 OUTRE CE QUE
 OUTRE CELA
 OUTRE QUE
 OU* L'ON
 OU* ON
 PAIX PUBLIQUE
 PAR AVANCE
 PAR BONHEUR
 PAR CONSEQUENT
 PAR DESSUS
 PAR LA VOIE
 PAR LE FER ET PAR LE FEU
 PAR PROVISION
 PAR RAPPORT
 PAR TOUT
 PAR TOUT AILLEURS
 PAR TOUT LE RESTE
 PARCE QUE
 PAS DES
 PAS MEME
 PENDANT QUE
 PETIT SENS
 PEU DE EGARD
 PEU IMPORTE
 PEU WRD WRDS IMPORTE
 PEUR WRDS NE
 PIRE WRDS QUE WRDS NE
 PIS ENCORE
 PLUS AVANT
 PLUS DE PART
 PLUS DE SUJET
 PLUS DE UN
 PLUS DE UNE FOIS
 PLUS ENCORE
 PLUS QUE
 PLUS WRD WRDS QUE
 PLUS WRDS PLUS
 PLUS WRDS QUE WRDS NE
 PLUT A* DIEU QUE
 PLUTOT QUE DE

DE ENTRE VOUS
 DE FOND EN COMBLE
 DE HAUT EN BAS
 DE HEURE EN HEURE
 DE LA MANIERE
 DE LA SORTE
 DE LAQUELLE
 DE LA* VIENT QUE
 DE MEME
 DE MEME QUE
 DE OU*
 DE OU* EST SORTI
 DE OU* IL SE ENSUIT
 DE OU* VIENT QUE
 DE PAIR
 DE PART ET DE AUTRE
 DE PEUR DE
 DE PEUR QUE
 DE PLUS
 DE PLUS EN PLUS
 DE RIEN
 DE SORTE QUE
 DE SUITE
 DE TROP
 DE UN COTE
 DELA LA MER
 DEPUIS CE TEMS-LA*
 DEPUIS DE
 DEPUIS PEU
 DEPUIS PLUS DE
 DEPUIS QUE
 DE*S AVANT QUE
 DE*S LE MOMENT QUE
 DE*S QUE
 DES UNS WRDS DES AUTRES
 DEVANT CELA
 DIEU AIDANT
 DIEU ME EN EST TEMOIN
 DIEU SOIT LOUE
 DONT
 DOUBLE MESURE
 DOUTE QUE WRDS NE
 DOUTOIT QUE WRDS NE
 DOUX COMME DU LAIT
 DROIT AUX BIENS
 DROITS DU PEUPLE
 DU PEU DE CAS QUE
 DU PEU DE SOIN QUE
 DU MOINS
 DU TOUT
 DURANT MEME
 EMOTION DE COEUR
 EN AMI
 EN ATTENDANT QUE

PLUTOT WRD WRDS QUE
 POINT DE
 POINT DU TOUT
 POUR CAUSE
 POUR CE QUI EST
 POUR CE QUE
 POUR DE
 POUR LA CAUSE
 POUR LE MOINS
 POURCE QUE
 POURVU QUE
 PREMIER VENU
 PRES DE
 PRINCE DES APOTRES
 PRINCE LEGITIME
 PRISE DE ARMES
 PUIS QUE
 QUAND MEME
 QUANT A*
 QUE
 QUE A* CE QUE
 QUE CE SOIT
 QUE COMME
 QUE COMME POUR
 QUE DE
 QUE DEVOIR QUE
 QUE EST-IL BESOIN
 QUE PAR
 QUEL DOMMAGE
 QUEL QUE
 QUEL QUE SOIT
 QUELLE QUE
 QUELLES QUE
 QUELQUE CHOSE DE
 QUELQUE ENTREE
 QUELQUE UN
 QUELQUE WRDS QUE
 QUELS QUE
 QUI EST-CE QUI
 QUI PLUS EST
 QUI QUE CE SOIT
 QUOI DE
 QUOI QUE
 QUOI QUE POURTANT
 QUOI QUE WRDS QUE
 REPOS PUBLIC
 RIEN DE
 RIEN NE
 RIEN NI PERSONNE NE
 RIEN WRD NE
 RIEN WRD WRDS NE
 RIEN WRDS DE
 RIEN WRDS NE
 SANS CARACTERE

EN AVANT
 EN CAS
 EN CAS DE
 EN CAS DE BESOIN
 EN CAS QUE
 EN CE QUE
 EN CE TEMPS-LA*
 EN COMPTE
 EN CONFIRMATION
 EN CONSCIENCE
 EN DEPIT
 EN DEPIT DE
 EN DROIT
 EN EFFET
 EN EFFORT
 EN ETAT DE
 EN FACE
 EN FACE DE
 EN FACON DU MONDE
 EN FAISANT VOIR
 EN FAVEUR
 EN LA MEME MANIERE QUE
 EN MAIN
 EN MATIERE DE
 EN MEME TEMPS
 EN MEME TEMS
 EN MOINS
 EN PEU DE TEMPS
 EN PROPRE
 EN QUALITE DE
 EN QUEL TEMPS
 EN QUOI
 EN RUINE
 EN SA PUISSANCE
 EN SOUHAITS
 EN TEMPS ET LIEU
 EN UN LIEU
 EN UN MOT
 EN VOUS EST
 EN WRD WRDS QUE
 EN WRD WRDS VENU
 EN WRDS CAS
 EN WRDS LIEU
 ENCORE DE
 ENCORE MIEUX
 ENCORE MOINS
 ENCORE QUE
 ENCORE UN COUP
 ENTANT QUE
 ENTRA EN FAVEUR DE
 ENTRE AUTRES
 ENTRE AUTRES CHOSES
 ESPRIT DE SATYRE
 ET QUE AINSI

SANS DETOUR
 SANS DOUTE
 SANS PEINE
 SANS PREOCCUPATION
 SANS QUE
 SANS RESERVE
 SE WRD TENIR AU
 SE WRD TENIR AUX
 SE WRD TENIR A*
 SECOUER LE JOUG
 SELON QUE
 SERMENS DE FIDELITE
 SERVIR DE BRIDE
 SI BIEN QUE
 SI CE NE EST
 SI CELA EST
 SI WRD WRDS SI
 SI WRD WRDS SI WRD WRDS SI
 SINON QUE
 SINON WRDS QUE
 SOIT QUE WRD WRDS SOIT QUE
 SOIT WRD WRDS OU
 SOIT WRDS OU
 SONNER LE TOCSIN
 SOUFFRIR PARTAGE
 SOUS LA ETENDART
 SOUTENANT QUE
 SUFFIT QUE
 SUIVANT QUE
 SUPOSE QUE
 SUR QUOI
 SUR TOUT
 TANT DE
 TANT DE FOIS
 TANT DE WRD WRDS QUE
 TANT DE WRDS QUE
 TANT PARCE QUE WRDS QUE
 TANT QUE
 TANT SE EN+FAUT
 TANT WRD WRDS QUE
 TANDIS QUE
 TANTOT WRDS TANTOT
 TEL QUE
 TEL QUI
 TOUT AINSI WRDS QUE
 TOUT AU PLUS
 TOUT AUTRE
 TOUT CE QUI
 TOUT CELA
 TOUT COMME MOI
 TOUT DE MEME QUE
 TOUT DE UN COUP
 TOUT DROIT
 TROP SOUVENT

ET WRD WRDS ET
ETANT DONNE
ETANT EN LIEU DE
EU EGARD AU
EU EGARD AUX
EU EGARD A*
FAUTE DE
FAUX PAS
FEU ET FLAMME
FORT BIEN
FORT CLAIREMENT
FORT MAL
GARDE AU
GARDE AUX
GARDE A*
GRACE AU

UN AUTRE
UN BIEN
UN BON PIED
UN CHACUN
UN GRAND COMPTE
UN MEME
UN TAS DE
UN TEL
UN TOUT
UN WRD AUTRE
VALABLE DE
VEU QUE
VOIE DE LA AUTORITE
VOILA-T-IL
VU QUE
Y COMPRIS

APPENDIX U

APPENDIX U

EXPRESSIONS IN GROUPS

ADVERBIAL PHRASES 143

 AUTANT DE
 AU TEMPS
 COMBIEN DE
 DE AUTANT DE
 ENCORE DE
 EN AVANT
 LOIN DE
 TANT DE
 AU DELA
 AU DESSOUS
 AU DESSUS
 AU DEVANT
 ICY
 DURANT MEME
 SANS DOUTE
 APRES LE TOUT
 EN MEME TEMPS
 DEPUIS CE TEMS-LA*
 EN PROPRE
 A* COUVERT
 AU MOIEN
 AU PARTI
 COMME AUTANT
 EN MOINS
 LORSQUE OUTRE
 QUOI QUE POURTANT
 AU PIS ALLER
 A* BON DROIT
 A* TOUTE HEURE
 A* TRAVERS
 DE PART ET DE AUTRE
 DEVANT CELA
 EN UN LIEU
 PAR TOUT
 SUR TOUT
 A* CETTE HEURE
 A* COTE
 A* COUPS DE PLUMF
 A* DOUBLE POUIS
 A* LA CONSCIENCE
 A* LONGS TRAIT
 A* PEU DE FRAIS
 A* TOUTE HEURE
 A* PUR ET A* PLFIN
 AU BOUT DU COMPTF
 AUX PIEDS
 BLANC AU NOIR
 DE ENTRE VOUS
 DE PAIR
 DE PLUS EN PLUS
 DOUBLE MESURE
 EN CE TEMPS-LA*
 EN CONSCIENCE
 EN MAIN
 EN WRDS LIEU
 EN WRDS CAS
 ENCORE MOINS

ADVERBIAL PHRASES 143

 MOINS DE SUJET
 A* COUP DE EPEES
 A* JAMAIS
 A* LA MANIERE
 A* LEUR DROIT
 A* LA OCCASION
 A* LA VERITE
 A* LA BESOGNE
 A* LA APPETIT
 A* LA MERCI
 A* MOINS DE CELA
 A* NEANT
 A* PEU PRES
 A* PLUS FORTE RAISON
 A* TOUT LE MOINS
 A* VOTRE TOUR
 AU BOUT DU MONDE
 AU LARGE
 AU MILIEU
 AU MOINS
 AU PIED DE LA LETTRE
 AU PLUTOT
 BIEN LOIN
 BON GRE MALGRE
 COMBIEN PEU
 COMME TRES-PROPRE
 CONSIDERE A* PART
 DE FOND EN COMBLE
 DE LA MANIERE
 DE LA SORTE
 DE MEME
 DE SUITE
 DE TROP
 DE UN COTE
 DU TOUT
 EN FACON DU MONDE
 EN PEU DE TEMPS
 EN QUEL TEMPS
 EN RUINE
 EN SA PUISSANCE
 EN SOUHAITS
 ENCORE MIEUX
 FORT CLAIREMENT
 FORT MAL
 PIS ENCORE
 A* PROPOS
 A* SON GOUT
 DE CE TEMS-LA*
 DE HEURE EN HEURE
 DELA LA MER
 EN AMI
 EN MEME TEMS
 LES ARMES A* LA MAIN
 PAR AVANCE
 PAR BONHEUR
 PAR CONSEQUENT
 PAR LA VOIE

ADVERBIAL PHRASES -CONT-

FORT BIEN
 PAR LE FER ET PAR LE FEU
 PAR PROVISION
 PAR TOUT AILLEURS
 PAR TOUT LE RESTE
 PLUS AVANT
 PLUS DE UNE FOIS
 POUR LE MOINS
 SANS DETOUR
 SANS PREOCCUPATION
 SANS PEINE
 SANS RESERVE
 TANT DE FOIS
 TOUT AU PLUS
 TOUT COMME MOI
 TOUT DE UN COUP
 TOUT DROIT
 TROP SOUVENT
 DE OU*
 DEPUIS PEU
 A* CENT MILLE
 A* PEINE
 A* WRDS EGARD
 TANTOT WRDS TANTOT
 DANS LE MOMENT QUF
 DE AILLEURS QUE
 TANT PARCE QUE WRDS QUF
 TANT QUE
 TOUT DE MEME QUF
 COMPARATIVES AND SUPERLATIVES 25

AUSSI WRD WRDS QUF
 AUTANT QUE
 AUTANT WRDS QUF
 AUTANT WRDS AUTANT
 DE AUTANT PLUS QUF
 LA MOINDRE CAUSE
 MEILLEUR WRDS QUE WRDS NE
 MEILLEURE WRDS QUF WRDS NE
 MOINDRE QUE
 MOINS
 MOINS WRDS QUE
 MOINS WRDS QUE WRDS NF
 MOINS WRD WRDS QUF
 PIRE WRDS QUE WRDS NF
 PLUS WRDS PLUS
 PLUS QUF
 PLUS WRD WRDS QUF
 PLUS WRDS QUE WRDS NF
 PLUTOT WRD WRDS QUE
 QUE A* CE QUE
 TANT DE WRDS QUF
 TANT DE WRD WRDS QUE
 TANT WRD WRDS QUF
 TEL QUE
 TEL QUI

COORDINATING CONJUNCTIONS 33

COMME
 COMME PAR EXEMPLE
 COMME SI
 DE MEME QUE
 EN LA MEME MANIERE QUE
 ENTANT QUE
 ET WRD WRDS ET
 CAR ALORS
 CAR QUAND
 CAR QUE
 CAR WRDS SI
 EN EFFET
 ET QUE AINSI
 SOUTENANT QUE
 EN UN MOT
 OR ALORS
 OR SI
 AU CONTRAIRE
 AU RESTE
 DE AILLEURS
 DE PLUS
 DU MOINS
 PLUS ENCORE
 SINON QUE
 SINON WRDS QUE
 OU BIEN
 OU MEME QUE
 OU QUE
 OU WRD WRDS OU
 SOIT QUE WRD WRDS SOIT QUE
 SOIT WRDS OU
 SOIT WRD WRDS OU
 C'EST-A*-DIRE

SUBORD. CONJ. -- CAUSE IND. 10

A* CAUSE QUE
 ATTENDU QUE
 CAR COMME
 CAR PUIS QUE
 CAUSE QUE
 PARCE QUE
 POUR CE QUE
 PUIS QUE
 VEU QUE
 VU QUE

SUBORD. CONJ. -- PURPOSE SUB. 02

AFIN QUE
 DE PEUR QUE

SUB. CONJ. - CONSEQ. - SUBJ. 3

 A* CE QUE
 SI BIEN QUE
 DE SORTE QUE

SUB. CONJ. - CONCESSION 17

 AU LIEU QUE
 OUTRE CE QUE
 QUAND MEME
 BIEN QUE
 ENCORE QUE
 NONOBTANT QUE
 OUTRE QUE
 QUELQUE WRDS QUE
 QUEL QUE
 QUELLE QUE
 QUELLES QUE
 QUELS QUE
 QUOI QUE
 QUOI QUE WRDS QUE
 SANS QUE
 SI WRD WRDS SI
 SI WRD WRDS SI WRD WRDS SI

SUB. CONJ. - CONDITION 7

 A* CONDITION QUE
 A* MOINS QUE
 A* MOINS QUE WRDS NE
 CAR EN CAS QUE
 EN CAS QUE
 POURVU QUE
 SUPPOSE QUE

SUB. CONJ. - TIME 16

 A* MESURE QUE
 APRES QUE
 AUSSI-TOT QUE
 DANS LE TEMS QUE
 DEPUIS QUE
 DE*S LE MOMENT QUE
 DE*S QUE
 LORS MEME QUE
 LORS QUE
 PENDANT QUE
 A* PRESENT QUE
 AVANT QUE
 AVANT QUE WRDS NE
 DE*S AVANT QUE
 EN ATTENDANT QUE
 JUSQUE A* CE QUE

SUB. CONJ. - COMP. - IND. 5

 A* PROPORTION QUE
 AUTREMENT QUE
 OR DE MEME QUE
 SELON QUE
 TOUT AINSI WRDS QUE

ADJECTIVAL NEGATIONS 8

 AUCUN WRDS NE
 AUCUNE WRDS NE
 NE WRDS AUCUN
 NE WRDS AUCUNE
 NE WRDS NUL
 NE WRDS NULLE
 NUL WRDS NE
 NULLE WRDS NE

ADVERBIAL NEGATIONS 35

 JAMAIS WRDS NE
 NE WRD GUERE
 NE WRD WRD WRDS GUERE
 NE WRD JAMAIS
 NE WRD WRDS JAMAIS
 NE WRD WRD WRDS JAMAIS
 NE WRDS JAMAIS RIEN
 NE PAS
 NE WRD PAS
 NE WRD WRDS PAS
 NE WRD WRD WRDS PAS
 NE PLUS
 NE WRD PLUS
 NE WRD WRD WRDS PLUS
 NE WRDS PLUS QUE
 NE WRDS PLUS RIEN
 NE WRDS PLUS WRDS QUE
 NE WRDS QUE DE
 NE POINT
 NE WRD POINT
 NE WRD WRDS POINT
 NE WRD WRD WRDS POINT
 NE WRDS QUE PARCE QUE
 NE WRD WRDS QUE EN CE QUE
 NE WRD WRDS QUE
 NON PAS
 NON PAS QUE
 NON PLUS
 NON POINT
 NON SEULEMENT
 PAS DES
 PAS MEME
 PEU WRD WRDS IMPORTE
 POINT DE
 POINT DU TOUT

NOUN PHRASES 51

VOIE DE LA AUTORITE
 UN TOUT
 UN GRAND COMPTE
 UN BON PIED
 UN BIEN
 SERMENS DE FIDFLITE
 SANS CARACTERE
 REPOS PUBLIC
 PRISE DE ARMES
 PRINCE LEGITIME
 PRINCE DES APOTRES
 PREMIER VENU
 PLUS DE SUJET
 PLUS DE PART
 PEU DE EGARD
 PETIT SENS
 MAINS DU BOURREAU
 LA MINE
 MENU PEUPLE
 LE DROIT DE EXAMINER
 LE DROIT DE ELECTION
 LE DEVOIR
 LA VOIE DES REMONTRANCES
 LA JUSTE IDEE
 LA JUSTE CRAINTF
 LA OEIL SUR
 LA EGALITE
 HOMME DE BIEN
 BONNE FOI
 BIEN DE AUTRES
 AMOUR DE VOUS
 MAL A* PROPOS
 LE COMBLE
 LE BIEN COMMUN
 LA AN
 FEU ET FLAMME
 ESPRIT DE SATYRF
 ENCORE UN COUP
 DROITS DU PEUPLE
 COUPS DE Foudre
 COUP DE ESSAI
 BON SENS
 BIEN PUBLIC
 BIEN COMMUN
 AU JUSTE JUGF
 ASSEZ BONNE FOI
 LE REPOS PUBLIC
 LE PUBLIC
 FAUX PAS
 DROIT AUX BIENS
 EN CAS DE BESOIN

INTERJECTIVE PHRASE 4

A* QUOI BON
 DIEU AIDANT
 DIEU SOIT LOUE
 QUEL DOMMAGE

CONJUNCTIVAL NEGATIONS 5

NE WRD WRDS NI
 NI MEME QUE
 NI WRD WRDS NI WRD WRDS NI WRD WRDS NE
 NON WRDS MAIS
 NON PARCE QUE WRDS MAIS PARCE QUE

PRONOMINAL NEGATIONS 11

DE AUCUN
 DE RIEN
 NE WRDS PERSONNE
 NE WRDS RIEN
 NE WRDS RIEN WRDS QUE
 RIEN DE
 RIEN NE
 RIEN WRD NE
 RIEN WRDS NE
 RIEN WRD WRDS NE
 RIEN WRDS DE

MISCELLANEOUS PREP PHRASES 4

EN EFFORT
 ENTRE AUTRES
 HORS DE DOUTE
 OUTRE CELA

PREP. PLUS VERB 14

A* FORCE DE
 AFIN DE
 AU LIEU DE
 AVANT QUE DE
 CAR DE
 CAS OU*
 DE CE QUE
 DE CE QUI
 DE PEUR DE
 DU PEU DE CAS QUE
 DU PEU DE SOIN QUE
 EN CE QUE
 PLUTOT QUE DE
 QUE COMME POUR

PREP PLUS NOUN 44

A* COTE DE
 A* L@ EGARD DE
 A* L@ EGARD DES
 A* L@ EGARD DU
 A* LA FAVEUR
 A* LA FAVEUR DE
 A*-PROPOS
 AU DELA DE
 AU DESSOUS DE
 AU DESSUS DE
 AU DEVANT DE
 AU MILIEU DE
 AUPRES DE
 AUX DEPENS
 DEPUIS PLUS DE
 EN CONFIRMATION
 EN DEPIT
 EN DEPIT DE
 EN DROIT
 EN ETAT DE
 EN FAVEUR
 EN MATIERE DE
 EN QUALITE DE
 ETANT EN LIEU DE
 FU EGARD A*
 FU EGARD AU
 GRACE A*
 GRACE AU
 HORS DE
 JUSQUE A*
 JUSQUE AU
 JUSQUE AUX
 JUSQUES DANS
 L@ ADRESSE DE
 PAR DESSUS
 PAR RAPPORT
 POUR DE
 POUR LA CAUSE
 POUR CAUSE
 PRES DE
 QUANT A*
 QUE COMME
 UN TAS DE
 VALABLE DE
 PREP PLUS NOUN OR VERB 11

A* CAUSE DE
 A* PROPOS DE
 AINSI QUE
 AU LIEU
 DANS LE CAS
 EN CAS
 EN CAS DE
 FAUTE DE
 MENACE DE

QUI DE QUE PAR

VERBS MISCELLANEOUS 20

TANT SE EN FAUT
 A* BOUT
 A* SA CHARGE
 A* TACHE
 A* TOUT PERDRE
 AVOIR SUSPECTE
 BON LUI SEMBLE
 COMMENT LE MONDE EST FAIT
 COURIR SUS
 DE OU* EST SORTI
 DIEU ME EN EST TEMOIN
 EN COMPTE
 EN FAISANT VOIR
 EN WRD WRDS VENU
 GARDE A*
 SECOUR LE JOUG
 SERVIR DE BRIDE
 SE WRD TENIR A*
 SONNER LE TOCSIN
 SOUFFRIR PARTAGE

VERB WITH NEGATION 30

CE NE EST PAS
 CE NE FTOIT PAS
 CE NE EST PAS LA PEINE
 CE NE EST PLUS
 CE NE SERA PAS
 CE NE SEROIT PAS
 CE NE SOIT QUE
 CE NE SONT PAS
 IL NE EN EST PAS DE MEME
 IL NE EST PAS
 IL NE EST PAS JUSQUE
 IL NE EST POINT DE
 IL NE ETOIT PAS
 IL NE FALLOIT PAS
 IL NE FAUDRA PAS
 IL NE FAUT PAS
 IL NE RESTE WRDS QUE
 IL NE SEMBLE PAS
 IL NE SERA PAS
 IL NE SEROIT PAS
 IL NE Y AIT
 IL NE Y A PAS
 IL NE Y A PAS LOIN
 NE EST-CE PAS
 NE EST-CE POINT
 NE EST-IL PAS
 NE FUT-CE QUE
 NE PEUT-ON PAS
 NE SONT-CE PAS
 SI CE NE EST

PRONOUN PHRASES 29

A* LAQUELLE
 A* QUOI
 CE A* QUOI
 CE QUE
 CE QUI
 CE QUE CE EST QUI
 DE LAQUELLE
 DES UNS WRDS DES AUTRES
 DONT
 EN QUOI
 EN WRD WRDS QUI
 ENTRE AUTRES CHOSSES
 LA AUTRE
 LES AUTRES
 LES UNS
 L'ON
 OU* L'ON
 OU* ON
 QUELQUE UN
 QUOI DE
 SUR QUOI
 TOUT AUTRE
 TOUTE AUTRE
 TOUT CE QUI
 TOUT CELA
 UN AUTRE
 UN CHACUN
 UN MEME
 UN TEL

VERB PLUS CONJUNCTION 19

BESOIN QUE
 CE EST CE QUE
 CE EST CE QUI
 CE EST QUI
 CE EST WRD WRDS QUI
 CE EST WRDS CE QUI
 CE EST WRDS CE QUI
 DE LA* VIENT QUI
 DE OU* IL SE ENSUIT
 DE OU* VIENT QUI
 DOUTE QUE WRDS NE
 IL EST WRDS QUE
 IL FAUDRA QUE
 IL SE ENSUIT QUI
 JOINT QUI
 PEUR WRDS NE
 PLUT A* DIEU QUI
 QUE DEVOIR QUI
 SUFFIT QUE

ALL VERBAL PHRASES 04

VERB + CONJ
 VERB + IMPERSON
 VERB MISC ETR AND VERB + NEGATION

VERB PLUS IMP. IL OR CE 43

CE EST
 CE EST AINSI
 CE EST AUSSI
 CE EST DE QUOI
 CE EST LA*
 CE EST SELON
 CE ETOIT
 CE EUT ETE
 CE SERA
 CE SEROIENT
 CE SEROIT
 CE SONT
 CE QUI EST
 EN VOUS EST
 IL EST
 IL EST A*
 IL EST BESOIN
 IL EST QUESTION
 IL EST SI VRAI
 IL EST VRAI
 IL ETOIT
 IL SE EN FAUT BIEN
 IL SERA
 IL SEROIT
 IL VOUS WRDS MIEUX
 IL WRDS AGI
 IL WRDS AGIRA
 IL WRDS AGIT
 IL FALLOIT
 IL FAUDRA
 IL FAUDROIT
 IL FAUT
 IL WRDS SEMBLE
 IL SE PASSE
 IL SE PEUT
 IL Y AIT
 POUR CE QUI EST
 QUE EST-IL BESOIN
 QUE CE SOIT
 QUI PLUS EST
 QUI QUE CE SOIT
 SI CELA EST
 VOILA-T-IL
 ALL NEGATIONS 04

ADJ. NEG. PHRASE
 ADV. NEG. PHRASE
 CONJ. NEG. PHRAS
 PRON. NEG. PHRAS

PRON. NEGATIONS 04

MISC. PREP.
 PREP. + NOUN
 PREP. + VERB
 PREP + N OR V

APPENDIX V

APPENDIX V

TABLE 4:54A. EXSOR GROUPED DATA, ABSOLUTE FREQUENCIES

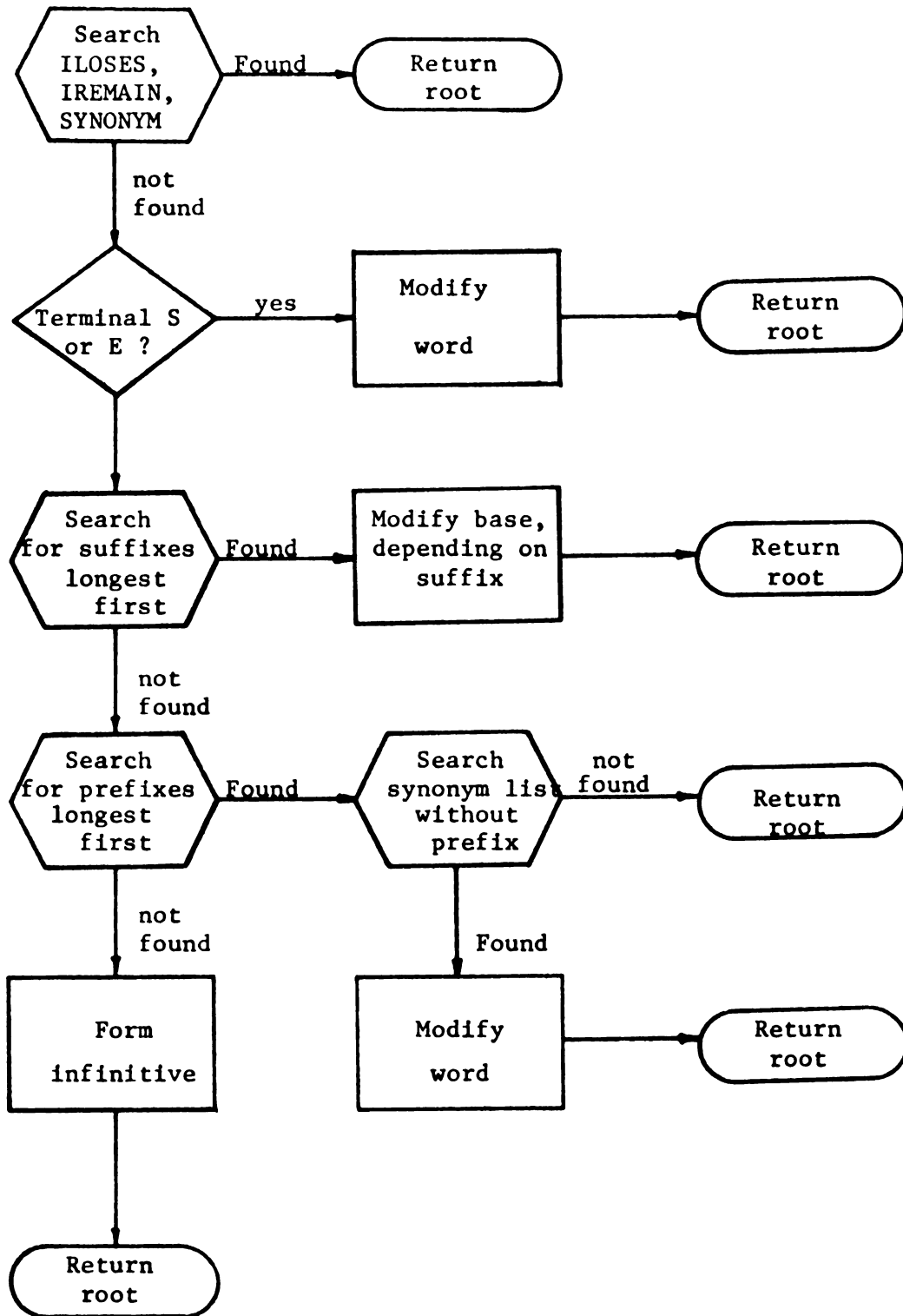
Expression Group	P	X	B1	B2	L1
Adjective phrases	0	14	6	2	2
Adverb phrases	24	459	322	119	260
Interrogative phrases	1	3	1	0	1
Noun phrases	2	113	78	18	22
Pronoun phrases	79	843	976	252	667
Comparatives & superlatives	32	615	503	173	258
Coordinating conjunctions - E*	44	543	496	155	353
Coordinating conjunctions - E & F**	384	6109	6354	1636	4333
Subordinating conjunctions - cause	7	48	109	36	162
Subordinating conjunctions - comparison	0	7	5	1	4
Subordinating conjunctions - concession	12	136	177	43	140
Subordinating conjunctions - condition	1	28	18	0	16
Subordinating conjunctions - consequence	1	29	20	10	30
Subordinating conjunctions - purpose	1	9	13	1	15
Subordinating conjunctions - time	6	42	105	19	42
Subordinating conjunctions - all (E)	27	281	429	103	402
Subordinating conjunctions - all (E & F)	34	550	578	183	578
Miscellaneous prepositionals	0	11	10	1	0
Prepositions & noun	9	143	86	27	54
Prepositions & verb	3	59	79	18	51
Prepositions & noun or verb	7	145	155	31	86
Verb & conjunction	12	194	214	73	172
Verb & impersonal	25	439	484	125	292
Verbs - miscellaneous	3	24	7	1	4
Adjectival negations	3	62	62	5	48
Adverbial negations	90	1827	1876	477	1301
Conjunctival negations	1	66	71	15	42
Pronominal negations	11	118	135	41	65
All negations	105	2073	2144	538	1456

* = EXSOR
 ** = FREQFUN

APPENDIX W

APPENDIX W

ENROOT FLOWCHART



APPENDIX X

APPENDIX X

ENROOT SYNONYM LIST

(A*)=(AU), (AUX)
 (ABONDER)=(ABONDANT)
 (ABSOUORE)=(ABSOU)
 (ABSURD)=(ABSURDITES), (ABSURDUM)
 (ABUSER)=(ABUSIF)
 (ACADEMI)=(ACADEMICIENS), (ACADEMIQUES)
 (ACCEPT)=(ACCEPTER)
 (ACCUSER)=(ACCUSATEUR)
 (ACCIDEN)=(ACCID)
 (ACTUEL)=(ACTUELLE)
 (ADRESSER) = (ADRE) , (ADDRESS)
 (AFFERMIR)=(AFFERM)
 (AFFLIGER)=(AFFLICTE) , (AFFLICTION)
 (AGEN)=(AGENS), (AGENT)
 (AGIR)=(AGISSANT)
 (AIMER)=(AIMENT)
 (AISE)=(AISEZ)
 (ALARMER)=(ALLARMER)
 (ALLEGUER)=(ALLEGUE), (ALLEQUER)
 (ALLEMAN)=(ALLEMAND) , (ALLEMAGN)
 (ALLER)=(VA), (ALL), (VAIS), (VONT), (AILL), (IR), (AILLE)
 (ALLIER)=(ALLIER)
 (AMBASADE)=(AMBASSADE)
 (AMBIION)=(AMBITIEU)
 (AMOUR) = (AMOURFUX), (AMOURER)
 (ANARCHI)=(ANARCHIQUE)
 (ANEANTIR)=(ANFANTISSEMENT)
 (AN)=(ANEE), (ANNEE)
 (ANONYM)=(ANONYME), (ANONIME)
 (APARENC)=(APARENCE), (APARENT), (APAREMMENT)
 (APARENC)=(APPAREMMENT), (APPARENCE)
 (APFELLER)=(APPFEL) , (APPFELL) , (APEL)
 (APFELLER) = (APPFELLER) , (APPFELER) , (APELER) , (APPELLE)
 (APERCEVOIR)=(APERCEVOIR) , (APERCEVOIR)
 (APLAUDIR)=(APLAUDISSEMENT)
 (APLIQUER)=(APPLICATION), (APPLIQUE)
 (APOSTASI)=(APOSTAT)
 (APPRENDRE)=(APRENANT), (APRENANT), (APRENS), (APREND)
 (APPRENDRE)=(APRENDRE-T-IL), (APRENDRE), (APREMENT)
 (APPRENDRE)=(APRENNENT), (APPRIS), (APRENER)
 (APPUYER)=(APPUY)
 (AQUERIR)=(AQUIS), (AQUIERT), (AQUIT)
 (ARDENT)=(ARD) , (ARDEUR)
 (ARGUMEN)=(ARGU)
 (ARMER)=(ARME) , (ARMEMENT)

(ARTIFIC)=(ARTIFICI)
 (ASSURER) = (ASURER)
 (ATHE)=(ATHEISME),(ATHEE),(ATHEES)
 (ATTEINDRE)=(ATTEIGNIS),(ATTEINTE)
 (ATTENDRE) = , (ATTENDR)
 (AUCUN)=(AUCUNE)
 (AVALER)=(AVAUX)
 (AVANTAGE)=(ANYTHING)
 (AVERTIR) = (AVER) , (AVERTIT) , (AVERTISS)
 (AVIDITE)=(AVIDE)
 (AVOIR) = (A) , (AI) , (AV) , (ONT) , (EU) , (EUT) , (E) , (AURAI)
 (AVOIR)=(AIENT),(AVEZ-VOU),(AVONS-NO)
 (AVOIR) = (AURA) , (AURONS) , (AUREZ) , (AURONT) , (AIE)
 (AVOIR) = (AY) , (AURIS) , (AUROIT) , (AURIONS) , (AURIEZ)
 (AVOIR) = (AYE) , (AIT) , (AURAS) , (AUROIENT) , (EUES) , (EUE)
 (AVOIR) = (AUR) , (AVEZ) , (AVONS) , (AYER) , (EUTES) , (EUS)
 (AVOUE) = (AVE) , (AVEU)
 (AZYLE)=(AZILE)
 (BABYLON)=(BABYLONIEN)
 (BARASSER)=(BARR)
 (BARBARE)=(BARBARI)
 (BATTRE)=(BATUS),(BATRE),(BATTU)
 (BEAU)=(BEAL) , (BEL) , (BELLE)
 (BENIR)= (BENEDICT)
 (BILE)=(BILIEUSE)
 (BOULVERSER)=(BOULVERSER)
 (BRASSER)=(BRASSER) , (BRER)
 (BRIGAN)=(BRIGANDAGE)
 (BRITANNI)=(BRITANNIQUE)
 (CABAL)=(CABALIST)
 (CALOMNIE)=(CALOMNIES),(CALOMNI),(CALOMNIATEUR)
 (CALVIN)=(CALVINIST)
 (CANON)=(CANONICI)
 (CANTON)=(CANTONNEZ)
 (CARTESIEN)=(CARTESIANISM)
 (CATHOLIC)=(CATHOLIC)
 (CEDER)=(CESSION)
 (CELA)=(CA)
 (CELLENT)=(CELLER) , (CELLER)
 (CELUI)=(CELL) , (CEUX) , (CELLE)
 (CENT)=(CENT)
 (CEPENDANT)=(CEPENDANT)
 (CERNER)=(CERNANT)
 (CERTAIN)=(CERTAINEMENT),(CERTES),(CERTITUDE)
 (CESSER)=(CESSION)
 (CHANT)=(CHANTS),(CHANCET)
 (CHARME)=(CHARMANT)
 (CHAST)=(CHASTETE)
 (CHAUD)=(CHAUDIMENT)
 (CHEMINER)=(CHEMIN)
 (CHIMER)=(CHIMERIQUE)
 (CHOISIR)=(CHOIX)
 (CINO)=(CINO)
 (CITER)=(CITATION)

(CITOIEN)=(CITOYEN)
 (CLERGE)=(CLERC)
 (COLOMNIER)=(COLOMIEZ)
 (COMBATTRE)=(COMBATIRENT),(COMBATRE)
 (COMENCER) = (COMMENCE)
 (COMMUNIER)= (COMMUNION)
 (COMMUNIQUER)=(COMMUNICATION), (COMMUNIQUASSE)
 (COMISION)=(COMMISSI),(COMMISSIO)
 (COMMOD)=(COMMODEMENT)
 (CONCEVOIR)=(CONCEU),(CONCOI)
 (CONCLUR)=(CONCLURRE),(CONCLUT),(CONCLUANTES),(CONCLURE)
 (CONDAMNER)=(CONDAMNMF)
 (CONDUIRE)=(CONDUIS)
 (CONFERER)=(CONFAIRE)
 (CONFONDRE)=(CONFUSION)
 (CONOITRE) = (CONNOIST) , (CONNOITRE) , (CONNU)
 (CONOITRE) = (CONNIR) , (CONNOISS) , (CONNOIR) , (CONNOSS)
 (CONOITRE) = (CONN) , (CONNA) , (CONNO) , (CONNU) , (CONNOIT)
 (CONSEILLER)=(CONSEILFER),(CONSEIL)
 (CONSENTIR)=(CONSENTFMENT)
 (CONVRTIR) = (CONVER) , (CONVERTI)
 (CONVIER)=(CONVIEZ)
 (CONVAINCRE)=(CONVAINQUOIT)
 (CONVENIR)=(CONVIENT),(CONVIENENT)
 (CONDITIO)=(CONDIRE) , (CONDITON)
 (CONSTANT)=(CONSTANCE) , (CONSTAMMENT)
 (CONTENT)=(CONTENS)
 (CORRECT)=(CORRECTEUR),(CORRECTIF)
 (COUPER)=(COUPNIT),(COUPEURS)
 (COURONER) = (COURONNE)
 (COURIR) = (CO)
 (COURIER)=(COURRIER)
 (COURONER)=(COURONNANT)
 (COUVRIRE)=(COUVERT),(COUVRE)
 (CRAINIRE)=(CRAIN),(CRAINT),(CRAIGN),(CRAIGNER),(CRAIGNIS)
 (CRASSER)=(CRFR)
 (CREEER)=(CREEE)
 (CRETIEN) = (CHRETIEN) , (CHRISTIAN)
 (CRIER)=(CRIS)
 (CRITIQUE)=(CRITQUER)
 (CRIME)=(CRIMINFL)
 (CROIRE)=(CR),(CROY),(CRU),(CRD),(CROYER),(CRUT),(CROIR)
 (CROIRE)=(CROYABLE),(CROYE),(CRFANCE),(CREDULE)
 (CROITRE)=(CROISSANT)
 (CRUE)=(CRIEMENT),(CRUES)
 (CRUEL) = (CRUAUT)
 (CUFILLIR) = (CUFIL)
 (CUIRE)=(CUI SANT)
 (CURER)=(CURAT)
 (CURIEUX)=(CURIOSITE)
 (DAUPHIN)=(DAUPHINE),(DAUHPIN)
 (DE)=(DFS),(DU),(DFLFS)
 (DEBONAIR) = (DEBONNAIR)
 (DECIDER)=(DECI)

(DECOUVRIRE)=(DECOUVRANT),(DECOUVRE),(DECOUVIR),(DECOUVROIT)
 (DECRIRE)=(DECRIROIENT),(DECRIEZ)
 (DEFAIRE)=(DEFAITE)
 (DEFENDRE)=(DEFFENS),(DEFFENDRE),(DEFFENDE)
 (DEFENDRE)=(DEFFENSIVE),(DEFEENS),(DEFENSE),(DEFENSEUR)
 (DEFAULT)=(DEFAULTS)
 (DELIVRER)=(DELIVER)
 (DELAIR)=(DELAIS)
 (DENT)=(DENTS)
 (DEREGLEMENT)=(DEREGLEMENT)
 (DERNIER)=(DERNIERE)
 (DESHONORER)=(DESHONNEUR)
 (DESTINER)=(DESTINA),(DESTENIR)
 (DESIR)=(DESIRS)
 (DEVENIR)=(DEVENOIR),(DEVENOIS),(DEVENOIT),(DEVIENT)
 (DEVANT)=(DEVANS)
 (DEVOIR)=(DOIT),(DU*), (DEVOIT),(DEVONS),(DOIVENT)
 (DEVOIR)=(DEVR),(DUE),(DEV),(DOIVE),(DUT),(DOIV),(DOIS)
 (DIABLE)=(DIABOLIQUE)
 (DIFFAMER)=(DIFFAMATOIR)
 (DIFFICILE)=(DIFFICUL)
 (DIGNITE)=(DIGNIF)
 (DIRE)=(DIRER)
 (DIRE)=(DI),(DIT),(DIS),(DIR),(DISER),(DISAN),(DISE)
 (DISCUTER)=(DISCUS),(DISCUSSI)
 (DISSENTION)=(DISSENSIONS)
 (DISTINCT)=(YES)
 (DIVERTIR)=(DIVERTIS)
 (DIVERS)=(DIVERSIT)
 (DIVINIT)=(DIVINITEZ)
 (DOCILE)=(DOCILITE)
 (DOGME)=(DOGMATIS),(DOGMATIQUE)
 (DONNER)=(DONNEREZ-VOUS),(DON)
 (DOUX)=(DOUCE),(DOUCEUR)
 (DUQUEL)=(DESQUEL)
 (ECHER)=(ECHER)
 (ECLAIRER)=(ECLAIRE)
 (ECRIRE)=(ECRIRE),(EC-IV),(ECRIVANT),(ECRIVIT),(ECRIVOIR)
 (EDIFIER)=(EDIFIANT)
 (ELECTION)=(ELU),(FLECTEUR),(ELECTIF),(ELECTORA)
 (EMBARRASSER)=(EMBARRASSANT)
 (EMPLOIER)=(EMPLOY)
 (EMPIRE)=(EMPIRANT)
 (ENFANT)=(ENFAN)
 (ENORME)=(ENORMITE)
 (ENSUIVRE)=(ENSUIT)
 (ENTREPRENDRE)=(ENTREPREIS)
 (ENTIER)=(GUTS)
 (ENVOIER)=(ENVOIR)
 (EPOUVANTAIL)=(EPOUVENTAIL)
 (EPRUVE)=(EPRUVE)
 (ESCLAVE)=(ESCLAVAG)
 (ESPAGNOL)=(ESPAGNOL)
 (ESPION)=(ESPIONS)

(ESSAYER)=(ESSAIRER)
 (ESSENTIEL)=(ESSENCE)
 (ETABLIR)=(ESTABLIR),(ETABLISS)
 (ETERNEL)=(ETERNEL)
 (ETRE)=(ETAIT),(ETES),(FUT-CE),(FUT-IL),(FUSSENT-ILS)
 (ETRE)=(ETOIT),(ETIONS),(ETIEZ),(ETOIENT),(FUS),(FUT)
 (ETRE)=(ETOIT-IL),(ESTRE),(FUST),(SOMMES-NOUS),(ES)
 (ETRE)=(FUMES),(FUTES),(FURENT),(ETANT),(ESTE),(SERAI)
 (ETRE)=(FUSSIEZ),(FUSSENT),(SOIS),(SOIT),(SOIENT)
 (ETRE)=(SERAS),(SERA),(SERONS),(SEREZ),(SERONT),(SEROIS)
 (ETRE)=(SERAIT),(SERIONS),(SERIEZ),(SEROIENT),(SO)
 (ETRE)=(SOIT-IL),(SONT-ILS),(SONT-ELLES),(SONT-CE)
 (ETRE)=(SOYONS),(SOYEZ),(FUSSE),(FUSSES),(FUSSIONS),(SER)
 (ETRE)=(SUIS),(EST),(SOMMES),(ETE),(SONT),(ETOIS)
 (ETRANGE)=(ETRANGES)
 (ETUDIER)=(ETUDIE),(ETUDE)
 (EUX-MEME)=(EUXMEME)
 (EVANGIL)=(EVANGELIQUES),(EVANGELI)
 (EVENEMENT)=(EVENEMENT)
 (EVEQUE)=(EVEQUE)
 (EVIDENT)=(EVIDENT),(EVIDENCE)
 (EXCEPTER)=(EXCEPTION)
 (EXCES)=(EXCESSIF)
 (EXEMPTER)=(EXEMPTION),(EXEMPT),(EXEMPT)
 (EXERCICE)=(EXERCICES)
 (EXPEDIER)=(EXPEDIER)
 (EXPLIQUER)=(EXPLICATION)
 (EXPRES)=(EXPRE)
 (EXPRIMER)=(EXPRESSION)
 (FACILE)=(FACILITE)
 (FAIRE)=(F),(FA),(FAIR),(FASSER),(FER),(FERA),(FI)
 (FAIRE)=(FAIS),(FAITES),(FAISONS),(FIMES),(FAISEUR)
 (FAIRE)=(FIRER),(FAITE),(FAISER),(FE),(FERER),(FIR)
 (FALLOIR)=(FALOIR),(FAUT),(FAILL),(FAUDR),(FALLU),(FALL)
 (FALLOIR)=(FALU),(FAUDRE),(FAL)
 (FAMILLE)=(FAMILIE)
 (FANATISME)=(FANATIQUE)
 (FANTAISI)=(FANTAISES)
 (FAUTE)=(FAUTES)
 (FAUX)=(FAUX),(FAUSSE),(FAUSSETTE)
 (FAVORISER)=(FAVORABLE)
 (FEINDRE)=(FEINS),(FEINT),(FEINTS)
 (FEROCE)=(FEROC),(FEROCIT)
 (FEUILLETER)=(FEUILLES)
 (FERIR)=(FERIR)
 (FIER)=(FIER),(FIERES)
 (FINIR)=(FINI),(FINIS),(FINIE)
 (FIDELITE)=(FIDEL),(FIDELE)
 (FISQUER)=(FISC)
 (FLATTER)=(FLATE),(FLATERIE)
 (FLECHIR)=(FLECHISSE)
 (FLETRIR)=(FLETRI),(FLETRISSANTES)
 (FLEURIR)=(FLEUR),(FLORISS),(FLORISSANT)

(FLEXION) = (FLFX)
 (FONDER) = (FONDAT), (FONDEUR), (FONDOIER), (FONDANT), (FONDOIT)
 (FONDATION) = (FONDEMENT), (FONDEMENT)
 (FORMER) = (FORMIST)
 (FORMALITE) = (FORMELS), (FORMELLES)
 (FOU) = (FOLF), (FOLLE)
 (FRAINDRE) = (FRAINT)
 (FRAPPER) = (FRAPER)
 (FRAIS) = (FRAICH), (FRAICHE)
 (FRAUD) = (FRAUDE), (FRAUDULEUSE)
 (FRUIT) = (FRUITS)
 (FURIFUX) = (FURFUR)
 (GAZETTE) = (GAZETTIER), (GAZETIER)
 (GIRQUET) = (GIRQUETTE)
 (GLORIFIER) = (GLOIRE)
 (GOUTER) = (GOUT), (GOUST)
 (GOUVERNER) = (GOUVERNE)
 (GRELE) = (GRFSLF)
 (GRIEF) = (GRIEVF), (GRIEVES)
 (GROSSIR) = (GROS), (GROSSE), (GROSSES)
 (GROSSIERE) = (GROSSIER)
 (GUERE) = (GUERES)
 (HABILET) = (HABILE), (HABILES)
 (HABITANT) = (HABITANT)
 (HAIR) = (HAIS)
 (HAILLON) = (HAILLER)
 (HARDI) = (HARDIF)
 (HERITER) = (HERITIER)
 (HISTOIRE) = (HISTOR), (HISTOIR), (HISTORIO)
 (HONNORER) = (HONORABLE)
 (HONNET) = (HONNETET)
 (HONTE) = (HONTEUX), (HONTEUSEMENT)
 (HOSTILE) = (HOSTILIT)
 (HUMAIN) = (HUMANITE), (HUMAINE), (HUMAINE)
 (ICELUI) = (ICFLL), (ICEUX)
 (IDOLE) = (IDOLATRE), (IDOLATRI)
 (IL) = (SIEN), (SIENNE), (SIENS), (LUI)
 (IMMEDIAT) = (IMMEDIAT)
 (IMPOSTURE) = (IMPOSTEURS)
 (INDEPEND) = (THIS)
 (INDIFFER) = (THAT)
 (INDIGNE) = (INDIGN)
 (INFALLIBLE) = (INFALLIBILE)
 (INFIDELITE) = (INFIDELLE)
 (INFINI) = (INFINIMENT), (INFIT)
 (INFIRM) = (INFIRMITÉ)
 (INJUR) = (INJURIFUX), (INJURE)
 (INNOCENT) = (INNOCENCEMENT), (INNOCENT), (INNOCENTES)
 (INSCRIRE) = (INSCRIT)
 (INSULTER) = (INSULTANT)
 (INSOLENT) = (INSOLER), (INSOLEN), (INSOLENC)
 (INSU) = (INSCU)
 (INTENTER) = (INTENTION)
 (ITER) = (ITATION), (ITATEUR)

(JALOU)=(JALOUSI),(JALOUSIE)
 (JESU)=(JESUS-CHRIST),(MESSIE)
 (JEUNE)=(JEUN)
 (JOINDRE)=(JOIGNANT),(JOIGNIS)
 (JOURNEE)=(JOURNEFL)
 (JOURNAL)=(JOURNAAL)
 (JUSTIFIE)=(JUSTIC),(JUST),(JUSTIFIC),(JUSTE)
 (LACHE)=(LACHETF)
 (LAISSER)=(LAIR),(LAISSONS)
 (LANGUE)=(LANGAGE)
 (LF)=(L@),(LES),(LA)
 (LECON)=(LECONS)
 (LEGEND)=(LEGENDAIRE)
 (LEGER)=(LEGEREMENT)
 (LEMENT)=(LEMENTAI)
 (LEQUEL)=(LESQUEL),(LAQUELL),(LESQUELL)
 (LEUR)=(LEURS)
 (LIBELL)=(LIBELLI),(LIBELLAT)
 (LIBERAL)=(LIBERALMENT),(LIBERALITE)
 (LIBERTE)=(LIBERT),(LIBERTRE)
 (LIBRAIRE)=(LIBRAIRIE)
 (LIBRE)=(LIBERTE)
 (LIFU)=(LIFUX)
 (LIRE)=(LISER),(LU),(LIR),(LISANT),(LISE),(LUE),(LUS)
 (LITERATUR)=(LITERAIRES),(LITTERATURE)
 (LIVRE)=(LIVRER),(LIVREZ)
 (LIVRE)=(LIVRET),(LIVRETS)
 (LOGER)=(LOGE),(LOGEOIS)
 (LOGI)=(LOGICIEU),(LOGIQUE)
 (LOISIR)=(LOISIR)
 (LONGTEMP)=(LONG-TEM)
 (LONG)=(LONGUE),(LONGUES),(LONGUEUR)
 (LORSQUE)=(LORQUE)
 (LOUER)=(LOUANGE),(LOUANGES)
 (MAGISTRA)=(FOOL)
 (MAITRE)=(MAITRISE)
 (MAJESTE)=(MAJEST),(MAJESTAT)
 (MALHONET)=(MALHONNE)
 (MALIC)=(MALICIEUSEMENT)
 (MALIGNE)=(MALIGN),(MALIGNIT)
 (MANIER)=(MANIERE),(MANIERES)
 (MANIFEST)=(MANIFES)
 (MARCHER)=(MARCHION)
 (MARQUER)=(MARQUABLE),(MARQUANT)
 (MARCHAND)=(GIRL)
 (MARTYR)=(MATYR)
 (MAUDIRE)=(MAUDIR)
 (MAUVAIS)=(MAUVAISE)
 (MECHANT)=(MECHAN),(MECHAMMENT),(MECHANTS)
 (MECONTEN)=(SOSO)
 (MEDIRE)=(MEDISANTE)
 (MEDIATE)=(MEDIATEURS)
 (MEILLIEUR)=(MEILLIEU)
 (MEME)=(MESME),(MEMES)

(MEMOIRE)=(MEMOIRFS)
 (MENER)=(MENANT),(MENFMT),(MENEUR),(MENOIT)
 (MEPRISER)=(MEPRIS),(MEPRI),(MEPRISE),(MEPRISANS)
 (MERVEILLE)=(JUNK)
 (METTRE)=(MET),(METT),(MIRER),(MIS),(METTION),(MIR),(METTE),(MISE)
 (METTRE) = (METTANT)
 (MIFUX)=(MIEU)
 (MINENT)=(INFNT),(INFNC)
 (MINUT)=(MINUTFF),(MINUTIES)
 (MIRACLE)=(MIRACL),(MIRACULE)
 (MISSION)=(MISSIONNAIRE)
 (MOIEN)=(MOIFNNAN),(MOYEN)
 (MOMENT) = (MOMFN)
 (MONTER)=(MONTEZ)
 (MON)=(MA),(MFS),(MOI),(MOI-MEME),(MIEN),(MIENNE),(MIENS),(ME)
 (MONARCHI) = (MONARCH),(MONARCHO),(MONARQUE),(MONACHAL),(MONARQU)
 (MONSIEUR) = (SIFUR),(MESSIFUR),(MONSEUR)
 (MONSTRE)=(MONSTREUX),(MONSTRUFUX)
 (MONT) = (MONTS)
 (MORAL)=(MORALITE),(MORALISER)
 (MORTELL)=(MORTELLF)
 (MOURIR)=(MEURT),(MORT),(MORTS)
 (MYSTERE)=(MYSTERIF)
 (NAIF)=(NAIVEMENT)
 (NATURE) = (NATUR)
 (NCOTRER) = (NCOTREF),(CONTRER)
 (NEAMMOIN)=(NEAMMOINS),(NEANMOIN)
 (NECESSAIRE)=(NECCFSSITE)
 (NEGOCIER)=(NEGOTIATEUR)
 (NET)=(NETTE)
 (NEUTRE)=(NEUTRALI)
 (NIER)=(NIAIS),(N),(NIANT)
 (NOIRCIR)=(NOIR),(NOIRE)
 (NOMMER)=(NOMME),(NOMS),(NOM)
 (NOMBRE)=(NOMBREUX)
 (NONCER)=(NOCER)
 (NOTOIRE)=(NOTORIET)
 (NOURRIR)=(NOURRISSANT)
 (NOUS)=(NOS),(NOTRE),(NOTRES)
 (NOUVEAU)=(NOUVAUX),(NOUVELLEMENT)
 (NOUVEAU)=(NOUVEAUT),(NOUVEL),(NOUVELL),(NOUVEAL)
 (NUIRE)=(NUIRA),(NUIT),(NUISENT)
 (NUL)=(NULLF),(NULLAM),(NULS),(NULLIT)
 (NVERSER)=(NVERSE)
 (OBFIR)=(OBFISS),(OBFISSAN),(OBFISSE),(OBIIE),(OBEI)
 (OBLIQ)=(OBLIQF),(OBLIQUES),(OBLIQUITEZ)
 (OBLIGER)=(OBLIEZ)
 (OBSCUR)=(OBSCURCI)
 (OBTENIR)=(OBTIENNE)
 (OCCASION)=(OCCO)
 (ODIEUX)=(ODIEUSE)
 (OEUVRE)=(OEUVRES)
 (OFFENSER)=(OFFENCANT)
 (OFFRIR)=(OFFERT)

(OMETTRE)=(OMISSIONS),(OMMISSION),(OBMETTRE),(OMIS)
 (ON)=(SOI),(SOI-MEME),(L'ON)
 (OPINER)=(OPINION),(OPENIATR),(OPINATRE)
 (OPPOSER)=(OPOSER),(OPPOSEZ),(OPPOSOIENT),(OPPOFER),(OPPOSITI)
 (OPPRES)=(OPPRESSE)
 (OPULEN)=(OPULENT)
 (ORFILLE)=(OREIL)
 (ORIGINAL)=(ORIGINEL)
 (ORIGINE)=(ORIGINAIR)
 (OTER)=(OTANT)
 (OUTRAG)=(OUTRAGEA),(OUTRAGEU)
 (OUVRIR)=(OUVERT),(OUVERTEMENT),(OUVERTS),(OUVRANT),(OUVRE)
 (OUVRAGE)=(OUVRAGEZ)
 (PACIFI)=(PACIFICATION),(PACIFIEROIS),(PACIFIQUES)
 (PAPE)=(PAPIST),(PAP),(PAPAL),(PAPISTIQ)
 (PARCOURIR)=(PARCOURU)
 (PARFAIT)=(PARFAITEMENT)
 (PAREIL)=(PAREILLEMENT),(PAREILL)
 (PAREN)=(PARENTE),(PARENT)
 (PARENCE)=(PARAT)
 (PARER)=(PARABLE),(PARAISER),(PARAISON)
 (PAROITRE)=(PARUSSENT)
 (PARTICUL)=(PARTICULIERS),(PARTICLUIERS)
 (PARTIE)=(PARTIFS)
 (PARTIR)=(PARTANT)
 (PASSION)=(PASSIONN)
 (PATIBLE)=(PATIBILIT)
 (PATIENT)=(PATIEN),(PATIENC),(PATIER),(PATI)
 (PATRI)=(PATRIMoine)
 (PAUVRE)=(PAUVRETE)
 (PAYS)=(PAIS)
 (PCONER)=(PCER),(PCONN)
 (PEINDRE)=(PEINT)
 (PERCER)=(PERCANT)
 (PERE)=(PERES)
 (PERIL)=(PERILLEU)
 (PERNIC)=(PERNICIFUSFS),(PERNICIEUX)
 (PERSONNE)=(PERSON),(PERSONNA)
 (PETIT)=(PITITESSE),(PETITESSE),(PETIRE),(PETIR)
 (PEUTETRE)=(PEUT-ETRE),(PEUR-ETRE),(PEUT-ESTRE)
 (PHILOSOP)=(PHILOSOPH)
 (PIRE)=(PIS)
 (PLAINDRE)=(PLAIN),(PLAINT),(PLAIN),(PLAINDR),(PLAINAN)
 (PLAINDRE)=(PLAINIR),(PLAINIR),(PLAINIR)
 (PLUR)=(PLURALITE),(PLURIEL)
 (POLIC)=(POLICE)
 (POLI)=(POLIE),(POLITESSE)
 (POSER)=(POSANT)
 (POSSEDER)=(POSSES),(POSSESSE)
 (POUSSER)=(POUSSENT),(PO)
 (POUVOIR)=(PEUT),(PEUVER),(POUVER),(PUIR),(POURRER),(POUVANT)
 (POUVOIR)=(POURR)
 (POURVU)=(POURVUE)
 (PRATI)=(PRATICABLE),(PRATIQUE)

(PRECIS)=(PRECISE)
 (PREDIRE)=(PREDICAT),(PREDICTI)
 (PREFERER)=(PREFFEREN)
 (PREJUGER) = (PREJUDIC)
 (PRELAT) = (PRELATUR)
 (PREMIER) = (PREMIERF)
 (PRENDRE)=(PRIS),(PRISE),(PRENNER),(PRIRENT),(PRIT),(PRISES)
 (PRESENC)=(PRESEN),(PRESENT),(PRESENTEMENT)
 (PRESSER)=(PRESSANT)
 (PRET)=(PRETE),(PRETES),(PRETS)
 (PRETER)=(PRETATIF),(PRET),(PRETENT),(PRETEZ)
 (PRIER)=(PRI),(PRIA),(PRE)
 (PRIMER)=(PRESSION)
 (PRINC)=(PRINCFS)
 (PRISER)=(PRISABLE)
 (PRISON)=(PRISONNIERS)
 (PRIVER)=(PRIVAT)
 (PROCHER) = (PROUCHER)
 (PRODUIRE)=(PRODUCTION),(PRODUISIT),(PRODUITE)
 (PRODUIRE)=(PRODUIS),(PRODUISE),(PRODUIR)
 (PROTEGER)=(PROTECTE),(PROTECTI)
 (PROBABLE)=(PROBABILITE)
 (PROCE)=(PROCEZ),(PROCES)
 (PRODIGE) = (PRODIG),(PRODIGIE)
 (PROGRE)=(PROGREZ),(PROGRES)
 (PROMT)=(PROMPTEMENT),(PROMTEMENT)
 (PROPHET)=(PROPHETE),(PROPHETIQUE),(PROPHETI)
 (PROPRE)=(PROPR),(PROPREMENT),(PROPRES)
 (PROVINC)=(PROVINCIAL)
 (PUBLIER) = (PUBLIASS),(PUBLIFZ)
 (PUBLIC) = (PUBLIQUE)
 (PUISER)=(PUISONIEN)
 (PUNI)=(PUNITF),(PUNEMENT),(PUNE)
 (PURIFIER)=(PURET),(PURE),(PUR)
 (QUALITE)=(QUALITRE)
 (QUERIR)=(QUERAN),(QUET),(QUIS),(QUER),(QUISE)
 (QUEL) = (QUELL),(QUELF)
 (QUITER)=(QUITTER)
 (QUINQUE)=(QUIN-QUE)
 (RADIOUIR)=(ADIOUCIFS),(ADIOUCIES)
 (RAILL)=(RAILLIURS)
 (RAISON) = (RAISER)
 (RAISONER) = (RAISONNE),(RAISONNA)
 (RAPORTER)=(RAPPORTS)
 (RAPPELER)=(RAPPELLER),(RAPFL),(RAPPELLE)
 (REBATTRE) = (REBATU)
 (REBELLE)=(REBELLIN)
 (RECEVOIR) = (RECEVER),(RECUE),(RECOIVER)
 (RECOMMENCER)=(RECOMMENCAI)
 (RECOURIR)=(RECOURRA),(RECOURER)
 (RECUEILLIR)=(RECUEILS),(RECUEIL)
 (REDEMPT)=(REDEMPTEUR),(REDEMPTION)
 (REFUGIE)=(REFUGIEZ),(REFUGIEZ)
 (RELIRE)=(RELUF)

(RELIGION)=(RELIG)
 (RENDRE) = (RENDIRER)
 (RENONCER) = (RENONCI)
 (RENOIER) = (RENOIER)
 (REPANDRE)=(REPAND), (REPANDENT), (REPANDIT)
 (REPOIRE)=(REPOIRIS), (REPOIS)
 (REPUT)=(REPUTATION)
 (REQUIRIR) = (REQUER), (REQUIS)
 (RESSUSCITER)=(RESSUCTITE)
 (RESPECTE)=(RESPECT)
 (RETABLIR)=(RETABLISSF)
 (REUSSIR)=(REUSSI), (REUSSIT)
 (REVENIR)=(REVIENT)
 (REVENIR)=(REVIR)
 (RHETORIQ)=(RHETHORI)
 (RIRE)=(RIROIT), (RIRFZ)
 (ROI) = (ROIS), (ROY), (ROIAL), (ROIAUM), (ROIAUT)
 (ROMPRE)=(RUPTION), (RUPTIBLE)
 (ROULLER)=(ROULF)
 (SACCAGER) = (SAC)
 (SACRE)=(SACREMFMS)
 (SAINT)=(SAINTFTE)
 (SALE)=(SALETEZ)
 (SANG)=(SANGLAN), (SANGLANT)
 (SATISFRE)=(SATISFAC), (SATISFAI)
 (SATYRSER)=(FATYRF), (FATYR), (FATYRIQUES), (SATYISE)
 (SATYRSER)=(SAYTRFS), (SATIRE), (SATYRIQU), (SATYRE)
 (SAUVER) = (SALUT), (SALUS), (SAUVEUR), (SALV)
 (SAVOIR) = (SAIS), (SAIT), (SCAVOIR), (SCAVANS),
 (SAVOIR)=(SAVANT), (SCAV), (SCAVAN), (SACH), (SAVANTE), (SAUR)
 (SAVOIR)=(SCAIR), (SCAUR), (SAI), (SAV), (SCAI), (SCU)
 (SCANDAL) = (SCANDALE), (SCANDALI)
 (SCRUPUL)=(SCRUPULE)
 (SECOURIR)=(SECOURUS)
 (SECUTER) = (SECTEUR), (SECUTRIC)
 (SECUTER) = (SECT), (SECUTEUR), (SECUTRE), (SECUTION)
 (SECOND)=(SECONDEM), (SECONDER), (SECONDE)
 (SECRET)=(SECRETAIRE), (SECRETEMENT), (SECRETE)
 (SECT)=(SECTAIRE), (SECTATEURS), (SECTES)
 (SEILLER)=(SEIL)
 (SEJOURNER)=(SEJOUR)
 (SEMBLER)=(SENBLEFROIT), (SEMBL), (SSEMBLAT)
 (SEMER)=(SEMAT)
 (SENTIR)=(SENT)
 (SETER) = (SENTA), (SENTAT), (SETEREN), (SENTE)
 (SEIR)=(SEANT), (ASSIS)
 (SEQUENT) = (SEQUEMMENT)
 (SEUL)=(SEULE)
 (SEVER)=(SEVERITE)
 (SIGN)=(SIGNF), (SIGNFS)
 (SIMPLE)=(SIMPLEMENT), (SIMPLES), (SIMPLICITE)
 (SINGUL)=(SINGULARIFZ), (SINGULIER)
 (SOIN)=(SOIGNEUSE), (SOIGNEUX)
 (SOLID)=(SOLIDAIREMENT), (SOLIDEMENR), (SOLIDITE)

(SON) = (SA) , (SFS)
 (SONNE) = (SONNELLE) , (SONNELL)
 (SOUDRE) = (SOL) , (SOLU) , (SOLUTION) , (SOLUTION)
 (SOUFFRIR)=(SOUFFRABLE)
 (SOUS)=(SOUS-GOUVERNER)
 (SPECTER) = (SPECT) , (SPECTUEUX)
 (SPECTEUR)=(SPECTION)
 (SPIRITUEL)=(RAG)
 (SUADER) = (SUA) , (SUADER)
 (SUCCEDER)=(SUCCEP),(SUCCEP)
 (SUFFIR)=(SUFFISANT),(SUFFIRE)
 (SUIVRE)=(SUIT),(SUIVE),(SUIVR),(SUIVI),(SUIVER),(SUITE),(SUIVANT)
 (SUPERIOR) = (SUPERIEUR)
 (SUIVRE)=(SUIVRE)
 (SUPPOSER)=(SUPPOSITION), (SUPPOSIT), (SUPPOSER)
 (SUPPRIMER) = (SUPPRIMER),(SUPPRES),(SUPPRES)
 (SURPRENDRE)=(SUPRENANT),(SURPRISE)
 (SURE)=(SURETE)
 (SUREMENT) = (SURETY)
 (SUSCITER) = (SUSCITER)
 (SUSPENDRE)=(SUSPENS),(SUSPENDRE)
 (TABLIR) = (TABLISSE)
 (TAIRE)=(TU)
 (TEL) = (TELLE)
 (TEMPOR)=(TEMPORAL)
 (TENDRE)=(TEND),(TENDRE),(TENS),(TENDU),(TENTION),(TENSION),(TENIR)
 (TENIR)=(TIENS),(TIENNE),(TEN),(TIENNE),(TIENNER),(TIENNE)
 (TENTER) = (TENTE)
 (TESTANT) = (TESTAN)
 (TESTABLE)=(BOY)
 (TICULIER) = (TICULIER),(TICULIER)
 (TINGUER)=(TINGUIT),(TINGITION),(TINGUANT),(TINGUO),(TINGUEZ)
 (TIRER) = (TIRE)
 (TORRENT)=(TORENT)
 (TOTAL)=(TOTALMENT)
 (TOUT) = (TOUTE) , (TOUS) , (TOUTES)
 (TRACTER)=(TRACT) , (TR)
 (TRADUIRE)=(TRADUCTEUR), (TRADUCTION)
 (TRAHIR)=(TRAHI),(TRAITRE)
 (TRAITER) = (TRAITEMENT)
 (TRAINDRE)=(TRAINT) ,(TRAIGNIR)
 (TRANCHER) = (TRANCHAS)
 (TRAVAIL)=(TRAVAIL) ,(TRAVAIL)
 (TRAVAGAN)=(GOODIES)
 (TRES-VRAI)=(TRES-VRAI)
 (TRIBUER) = (TRIBU) , (TRIBUTION)
 (TRIOMPHER)=(TRIOMPHER)
 (TROMPER) = (TROMP) , (TROMPRE)
 (TRONER) =(THRONE)
 (TROUVER) = (TROUVER)
 (TUER)=(TUOIT)
 (UN)=(UNE)
 (UTIL)=(UTILISE),(UTILISEMENT),(UTILITE)
 (VAINCRE)=(VAINCREMENT),(VAINQUEUR),(VAINQUEUR)

(VANTR) = (VENT)
 (VARIFR)=(VARIF),(VARIENT),(VARIETE)
 (VENDRE)=(VEND),(VENDF),(VENDIT),(VENTE)
 (VENIR)=(VENU),(VIENT),(VENDIT),(VIENS) , (VEN) , (VIENN),(VIENNER)
 (VERSFR)=(VERSE)
 (VERIT)=(VERITABLE),(VFRITE),(VERITE-CI),(VERITEZ)
 (VERTU) = (VERTUEUX)
 (VICTOIRE) = (VICTORIFUX)
 (VIOLEN)=(VIOLEMMENT),(VIOLENTS)
 (VIOLEBLE)=(VOILABLE)
 (VOEU)=(VOEUX)
 (VISIB)=(VISIBILITE),(VISIBLE),(VISIBLEMENT),(VISIBLES)
 (VOILA)=(THERE)
 (VOICI) = (VOICF)
 (VOUFR) = (VOU), (VOCATIF), (VOCATION)
 (VOIAGER)=(VOYAGE)
 (VOIR)=(VERK),(VU),(VOYE),(VO),(VOYER),(VUF),(VUS),(VERRER)
 (VOIR) = (VOIS),(VOIT),(VOY),(VOI)
 (VORISFR)=(VORI)
 (VOULOIR) = (VEUILL) , (VEUILLER) , (VOUDRE) , (VEULENT) , (VE)
 (VOULOIR) = (VEUX) , (VEUT) , (VOUL) , (VOULU) , (VOULUST) , (VOUDR) ,
 (VOULOIR)=(VOULOIS),(VOUDROI) , (VOULR)
 (VOUS)=(VOS) , (VOTRE)
 (ALLER) = (AILLENT),(ALLF),(ALLFES)
 (CEVOIR) = (COIT),(COIS),(CU),(COI),(CEU)
 (PRENDRE) = (PRENDR)
 (POUVOIR) = (PUIT),(POUV)
 (PRENDRE) = (PREN) , (PRENDRE), (PRI)
 (FLATERIF) = (TRIXXX)
 (CITER) = (CITEF)
 (METTRE) = (MI)
 (NAITRE) = (NEF)

APPENDIX Y

APPENDIX Y

GLOSSARY OF WORD LISTS FOR ENROOT

LBAS LSYN	These two lists make up the synonym list. LBAS contains the base or root words, and LSYN contains the synonyms to these roots.
KSUF	A list of all suffixes.
LPRE	A list of all prefixes.
IDROPCHR	These words drop a certain number of letters from the end of the word.
IER	These words drop from 0-6 letters and add ER.
LAD	These words have a number of letters added to the end of the word. The letters are obtained from the corresponding entry in JAD.
LR	These words drop 0-6 letters and add an R.
LRE	These words drop 0-6 letters and add RE.
LIR	These words drop 0-3 letters and add IR.
LAD1	These words have one (1) letter added to them. The letter is obtained from the corresponding entry in JAD1.
IETOER	These words change the final E to an R.
ILOSES	These words lose their final S and have no other changes made.
IREMAIN	These words have no changes made at all.
IS34	These are three and four letter words which lose their final S.
IS5678	These are five, six, seven, and eight letter words which lose their final S.

APPENDIX Z

APPENDIX Z

ENROOT SAMPLE OUTPUT

KEYWORD	FIELD DUPS	CONTEXT
ABAISSER	2	ABAISSSE
ABANDONNER	2	ABANDON
	2	ABANDONNE
		ABANDONNEE
	2	ABANDONNER
ABATTRE		ABATTRE
ABIMER		ABIME
		ABIMER
		ABIMES
ABIRAM		ABIRAM
ABJURER		ABJURER
ABOLIR	2	ABOLIR
		ABOLIRENT
		ABOLISSENT
		ABOLIT
ABOMINABLE	2	ABOMINABLE
	2	ABOMINABLES
		ABOMINATION
ABONDER		ABONDANCE
		ABONDANTE
		ABONDEROIT
ABORD	3	ABORD
ABOUTIR		ABOUTI
		ABOUTIRONT
		ABOUTISSENT
ABRI		ABRI
ABSALON		ABSALON
ABSOLU	3	ABSOLU
	3	ABSOLUE
	6	ABSOLUMENT
		ABSOLUTION
ABSOUFRE		ABSOUFRE
		ABSOUS
		ABSUUS
ABSTENIR		ABSTENIR
		ABSTENU
	2	ABSTINT
ABSTRAIT		ABSTRAITES
ABSURD	7	ABSURDE
		ABSURDUM
ABUSER		ABUS
	2	ABUSANT
	3	ABUSE
		ABUSENT
		ABUSER

APPENDIX AA

APPENDIX AA

PIP FLOWCHART*

