Computers and the Humanities **22** (1988) 1–10.

The Stylo-Statistical Method of Literary Analysis

Robert F. Allen

Dept. of French, Rutgers — the State University, New Brunswick, NJ 08903, U.S.A.

Abstract: The stylo-statistical approach to literary analysis provides scholars with a method for extracting pertinent data of manageable size from the overall mass of data provided by the computer in the analyses of large bodies of texts. How this is done is explained in the article that follows, beginning with a brief description of the programs that provide the data for the statistical analysis, followed by a description of one of the statistical methods currently used, with examples drawn from a stylo-statistical analysis of Flaubert's *Madame Bovary.*

Key Words: Stylo-statistics, French literature, Flaubert, Constant.

When this author presented a paper on stylo-statistics at a past MLA Convention, members of the audience were asked if they were familiar with the term stylo-statistics. Only four or five of the nearly fifty persons in attendance fully understood the term. If one may extrapolate from these figures, it would appear that few humanists understand this approach to literary analysis. The aim of this paper, then, is first to describe the stylo-statistic method of literary analysis, in particular, the key-word approach [1], and secondly to explain what literary scholars can hope to accomplish by using this method.

There can be little doubt that scholars who conduct content analysis of texts would like to be able to identify the vocabulary that the author stresses, the vocabulary that he returns to time and again, in other words the vocabulary that is characteristic of the text. Stylo-statistics allows one to do this by identifying the key-words of the work under study. One must be careful, however, for the term "key-word" has a special meaning in stylo-statistics. A meaning that one usually encounters for the term is, "one of the most frequently used words of the text". In stylo-statistics it has the more restricted meaning of: "one of the most frequently used words of the text *that deviates significantly in a positive sense from the norm* [2]."

Norm frequencies for vocabularies have been established for most of the western European languages. Norms for other linguistic structures, such as phrases, clauses, verbal tenses, etc. have not yet been created, or if they have been, are not easily accessible, or do not cover all time periods. Frequency lists for twentieth-century French verbal tenses have been established in an unpublished doctoral dissertation at the University of Strasbourg. To my knowledge, there are no such lists for other periods.

In the area of French vocabulary, norm frequencies have been established by the Institut National de la Langue Française (formerly the Trésor de la Langue Française) at Nancy, France, based upon a corpus of more than 70,000,000 words, drawn from 19th and 20th century literature. The norm frequencies are broken down into time periods: by century, half century, and intervals ranging from 4 to 26 years with an average interval of 10.7 years. For the century and half centuries, norm frequencies are provided for the various genres: prose, poetry, soliloquy, prose poetry, and dialogue.

The concept that a norm exists for the linguistic structures of a language can be traced back to Ferdinand de Saussure and his theory of *langue et*

Robert F. Allen is in the French Department of Rutgers, the State University, where he teaches courses related to computer applications for language and literature. His principal field of research is the stylo-statistical analysis of the 19th century French novel.

parole[3]. *La langue* represents the grammar of a language — its fixed structure, i.e., the norm — while *la parole* represents the individual usage of the language, including slight but permissible variations from the norm. Stylo-statistics has carried de Saussure's theory one step further by declaring that the vocabulary, as well as other linguistic structures, have a fixed frequency of use or norm frequency. It follows from this that an author's style can be identified by the structures that deviate significantly from the norm. Any reference to a style that is considered characteristic of a writer should not be to linguistic structures that occurred only once or even a few times; the structure should occur often enough to be felt as typical of the writer. For it to be felt as typical of the writer, it has to have a frequency of use significantly higher than the norm.

There are ways of determining whether or not the deviation from that norm is a significant one, and they will be explained below. However, it is necessary to realize first that the norm is a measure of the probability of the word occurring in the text. The act of picking words out of one's head is comparable to picking different colored marbles out of an urn. If the urn contains, for example, 25 black marbles, 20 white, 15 blue, 12 orange, 10 green, 10 brown, and 8 red marbles (a total of one hundred), there is one chance out of four of picking a black marble, one chance out of five of picking a white one, fifteen chances out of a hundred of picking a blue one, etc. If the marbles are replaced by words and the numbers thought of as norm frequencies, we have some idea of how stylo-statisticians believe that the choice of words is influenced by the laws of chance or probability.

Of course, when a marble is picked out of an urn, it is a completely random choice, while a writer's choice of a word is not. The subject matter, the literary genre, and the conscious and unconscious preferences for particular linguistic structures exclude complete random selection. Nevertheless, stylo-statisticians have observed that the end result is similar to a random situation. If reasonable account is taken of the subject matter and of the literary genre, any resulting significant deviation from the norm can be explained by the author's preference for particular linguistic structures.[4]

The measure of deviation from the norm is referred to as the "standard deviate" and is calculated by using the central limit theorem (Z):

The following terms are used in the formula:

X = Absolute or observed frequency
ANF = Adjusted norm frequency, i.e., the norm frequency calculated relative to the size of the text.

$$Z = \frac{(X - ANF)}{\sqrt{ANF}}$$

In the *Dictionnaire des fréquences* the norm frequencies are relative to one hundred million words, i.e., calculated for a text of one hundred million words. If we are working, for instance, with a text of 118,745 words (the total number of words in *Madame Bovary*), the norm frequencies will have to be calculated so that they are proportional to that size text. If the word deviates by more than two and a half standard deviates, it is considered that the deviation is a significant one.

If one considers the example of tossing a coin in the air, the concept of normal versus significant deviation from the norm may be better understood. If a coin is tossed in the air one hundred times, theoretically one would obtain 50 heads and 50 tails, with a norm frequency of 50 for either heads or tails. A result of 50/50 would be surprising. In stochastic distributions results usually group around a projected figure, so that one expects a slight deviation from the norm. If the proportion were 55/45 we would find that quite normal and results obtained from the central limit theorem would support that belief.

$$Z = \frac{(55 - 50)}{\sqrt{50}} = 0.7 \text{ Standard Deviates (SD)}$$

If the proportion were 60/40 (1.4 SD) we would not find that abnormal either while with 65/35 (2.1 SD), we might be surprised at our degree of good or bad luck. With a 70/30 ratio (2.8 SD) we would most certainly feel that the toss was rigged, and at 80/20 (4.2 SD) there would be no question about it. Although there are many more factors to take into account when dealing with linguistic structures, significant deviation from the norm can be felt in much the same way as with a game of heads or tails. We are sensitive to repetition of nouns, adjectives, verbs, and adverbs, while

we accept more frequent repetition of other words, such as articles, prepositions, and conjunctions which have higher norm frequencies than the former group.

As the central limit theorem is not accurate when adjusted norm frequencies are below "5", for these frequencies the deviation from the norm must be calculated using Poisson's formula:

$$pf(X) = ANF^{x} * e^{-anf}/X!$$

"X" is the word's frequency of use.

Pf(X) is the measure of probability of the word occurring X number of times in the text.

"e" is equal to 2.71828.

The exclamation mark (!) is the symbol for "factorial". If the word's frequency is 4, X! would be the factorial of 4, i.e., $4 \times (4 - 1) \times (4 - 2) \times (4 - 3) = 24$.

Statisticians have taken the value of pf(X) and converted it to a probability factor. A pf(X) value of 10^{-2} or .02 means that the frequency of occurrence has two chances in a hundred of occurring *by chance*, which is so slight that one may conclude that the author repeated the structure that frequently by choice, whether on a conscious or unconscious level. In the tables that follow, the Poisson values are abbreviated. 8^{-5} is equal to $8 * 10^{-5}$ or .00008. Translated, this means that the frequency of occurrence has 8 possibilities in 100,000 (1 in 12,500) of occurring by chance. One must remember when reading the Poisson values that

$$A^{-B} = A * 10^{-B}.$$

The degree of probability for each standard deviate and Poisson value follows:

TABLE I
Degree of Probability for Standard Deviates and Poisson Values

STANDARD DEVIATES	POISSON VALUE	DEGREE OF PROBABILITY
1.0	0.317	About one chance in three
2.0	0.046	About one chance in twenty
2.5	0.012	About one chance in a hundred
3.0	0.0027	Less than three chances in a thousand
4.0	0.00006	Six chances in one hundred thousand
4.5	0.000006	Six chances in a million

To understand stylo-statistics, it helps to see how the method works. Examples drawn from my own research in the stylo-statistical anlaysis of the nineteenth-century French novel will serve to demonstrate the method. At this moment, I have four novels in machine-readable form and have analyzed them stylo-statistically. They are: *Adolphe, Le Rouge et le noir, Le Père Goriot*, and *Madame Bovary*. Two other novels, one from the seventeenth-century (*La Princesse de Clèves*), the other from the eighteenth-century (*Manon Lescaut*), have had their adjectives counted by hand.

The computer program that was used in this research, and which was written in the SNOBOL4 computer language, identifies and then lemmatizes the word, i.e., reduces the inflected word to its base form, keeps track of its frequency of use, notes where the word occurs in the text, and calculates its relative frequency — relative to 100,000 words. The program does essentially what a concordance program does with one important difference — it lemmatizes the vocabulary. This is done in three ways: (1) programmed routines within the program, (2) tables, (3) annotation of the text.

Programmed Routines

An example of programmed routines can be seen in the one that distinguishes between words that have a hyphen or apostrophe as part of their structure, e.g., *aujourd'hui* and words that are joined by a hyphen or apostrophe, e.g., *j'ai*. In computational linguistics, the word is usually defined as a series of characters preceded and followed by a blank space or punctuation. However, this definition would never identify such words as *vis-à-vis, aujourd'hui, garde-manger, chef-d'oeuvre*, or *coq-à-l'âne*, since the apostrophe and hyphen are punctuations. On the other hand, if we were to include the apostrophe and hyphen with the alphabetical characters we would find *j'ai, dit-il, moi-même, va-t'en, parle-t-il*, and *allez-vous-en* classified as words. Consequently, the definition has to take into account these two structural groupings. Let us set the two groups side by side and take a closer look at them.

In the group of words joined by an apostrophe or hyphen, certain elements within the structure

can be classified, e.g., the "*j*" of "*j'ai*" can be classified with the group of words whose final "e" can be dropped when followed by an immediate constituent that begins with a vowel or unaspirated "h", while "*il*" of "*dit-il*" is one of a number of subject or object pronouns which can be suffixed to the verb with a hyphen, e.g., the inverted form of the verb. Then there is the use of *-ci, -là, -même*, and what Grevisse refers to as the analogical "t" of *parle-t-il.* In other words there is a relatively small number of easily classifiable words that are joined to other words either by an apostrophe or a hyphen. This small number of words is put into a table. The program checks to see if the word is in the table and if it is, the two structures that are joined by the hyphen or apostrophe are treated as separate words, otherwise as one word.

TABLE II
Words Formed or Joined by " ' " or "-"

Words Formed or Joined by	One Word	Several Words
'	aujourd'hui	j'ai
-	garde-manger	dit-il
		garçon-ci
- -	vis-à-vis	parle-t-il
		allez-vous-en
- '	chef-d'oeuvre	va-t'en
- - '	coq-à-l'âne	
- ' -	sot-l'y-laisse	

Tables

The tables contain, in addition to the small number of words mentioned above that are joined to other words by the hyphen or apostrophe, lists of:

- Function words, e.g., prepositions, conjunctions, pronouns, noun determiners
- Irregular forms of verbs, adjectives, nouns, and adverbs
- Stems of regular verbs
- Words that are exceptions to the rules, e.g., to lemmatize the plural form of the noun one removes the final "s" as in the case of *maisons.* On the other hand, we do not want to remove the "s" in words such as *bras, dos, bois* which are the exceptions.

Once a string of letters has been identified as a possible word (WORD), the tables are checked to see if the string of characters matches the structure that is in the table. Each structure in the table has a value assigned to it. For example, if the string of characters happened to be *bénigne*, upon checking the table of irregular adjectives the computer would find ⟨'bénigne'⟩ = 'bénin', and the program would assign to WORD the new value (*bénin*) as well as a marker that would identify the word's grammatical category.

Annotation

The text is annotated to identify the grammatical category of homographs. This is done by either prefixing or suffixing a code of fixed length to the word.

Most homographs can be identified through programming. *Bois* can be either noun (woods) or verb (first or second person of "to drink.") However, if the preceding word is a noun determiner, *bois* is a noun. If the grammatical category of homographs that may be either noun or verb can be determined by programming, it is more difficult to distinguish between homographs such as *vieux, petit, grand* that may be either noun or adjective, between *que* as conjunction, adverb, or relative pronoun, or the various categories of *tout*, etc. There is the probability that with hundreds of hours of additional programming such homographs can be identified, and the attempt is being made to do so. Until the grammatical category of every homograph has been identified through programming, and there is very little possibility that every homograph can be, those that are not identified have to be annotated[5].

* * *

Separate programs sort the data and format the output into fields and columns so that the presentation of the data is clear and easily understood. An example of this is seen in Table III. Next, the alphabetized list is fed back into the computer, the different parts of speech are separated, and each grammatical category is printed out in a descending order of frequency of use. One can thus immediately identify the most frequently used words of each category. Examples of the most frequently used adjectives of *Adolphe*[6] can be seen in Table IV. The "Z" scores represent the

TABLE III
A Page from Word Index of ADOLPHE
ALPHABETIZED WORD INDEX: ADOLPHE
TOTAL NUMBER OF WORDS = 29,180

LEMMA	ABS FREQ	REL FREQ	PAGE/LINE REFERENCES					
1 à (prep)	(567	1945)						
2 abandon (noun)	(6	21)	02202	05421	06012	07008	07222	09214
3 abandonner (vrb)	(9	31)	02322	04402	09322	09817	10602	11301
			11618	11924	13312			
4 abattement (noun)	(1	3)	04425					
5 abattu (adj)	(1	3)	09218					
6 abîme (noun)	(1	3)	10505					
7 abîmer (vrb)	(1	3)	14010					
8 abjurer (vrb)	(3	10)	04403	08003	13601			
9 abondamment (adv)	(1	3)	13528					
10 aborder (vrb)	(1	3)	09606					
11 abri (noun)	(3	10)	02708	07205	14101			
12 absence (noun)	(15	51)	02202	03513	03825	04210	05713	07402
			07621	07910	07910	09101	11111	12205
			12321	12707	14618			
13 absent (adj)	(2	7)	03908	12614				
14 absenter (vrb)	(1	3)	05802					
15 absorber (vrb)	(2	7)	04802	10615				
16 abuser (vrb)	(2	7)	07705	11616				
17 accablé (adj)	(1	3)	10521					
18 accabler (vrb)	(4	14)	05214	05226	08231	10312		
19 accepter (vrb)	(6	21)	02627	06613	06901	07306	09113	10317
20 accès (noun)	(3	10)	06028	08617	10217			
21 accompagner (vrb)	(2	7)	08117	09014				
22 accorder (vrb)	(10	34)	01918	04003	04503	04729	05625	06323
			07524	10119	10315	11715		
23 accourir (vrb)	(1	3)	13224					
24 accoutumé (adj)	(1	3)	10803					
25 accoutumer (vrb)	(3	10)	02118	02709	05616			
26 accroître (vrb)	(1	3)	05805					
27 accueillir (vrb)	(2	7)	02420	12016				
28 accumuler (vrb)	(1	3)	02328					
29 accusation (noun)	(1	3)	11902					
30 accuser (vrb)	(9	31)	02617	06503	09322	10313	11013	11525
			11616	11913	12722			
31 acéré (adj)	(1	3)	14429					
32 acharnement (noun)	(1	3)	14207					
33 acharner (vrb)	(3	10)	08011	08409	14401			
34 acheter (vrb)	(2	7)	06620	10130				

number of standard deviates derived from the central limit theorem.

Upon looking at the list below, one will see that some of the words are used less frequently than the norm, while others are used at about the same frequency. Since the frequency of use of these words is less than normal, or at least close to it, the words are *not* considered characteristic of the text. When these words are deleted from the list, only the key-words remain (see Table V).

* * *

An example of what stylo-statistics reveals about an author's style can be seen in an analysis of

TABLE IV
Most Frequently Used Adjectives of ADOLPHE

	ADJECTIVES	ANF	ABS FREQ	RELV FREQ	NORM FREQ	Z	P(x)
1	même	78.6	56	192	269262	−2.53	
2	seul	29.7	43	147	101858	2.42	
3	grand	60.8	25	86	208212	−4.60	
4	nouveau	21.8	24	82	74792	0.45	
5	heureux	10.7	21	72	36537	3.10	
6	propre	8.3	21	72	28348	4.48	
7	premier	32.1	20	69	110117	−2.13	
8	malheureux	5.6	29	65	19214	9.92	
9	long	9.7	18	62	33242	2.69	
10	libre	4.7	15	51	16041		8^{-5}#
11	doux	8.8	13	45	30315	1.43	
12	tel	6.2	12	41	21083	2.37	
13	triste	6.3	12	41	21748	2.20	
14	violent	2.3	12	41	7904		5^{-6}
15	bizarre	1.5	11	38	5179		5^{-7}
16	dernier	17.6	11	38	60430	−1.58	
17	entier	6.7	11	38	22909	1.65	
18	profond	6.9	11	38	23591	1.51	
19	vif	5.3	11	38	18145	2.50	
20	content	2.4	10	34	8238		2^{-4}
21	froid	5.7	10	34	19516	1.72	
22	général	13.2	10	34	45169	−0.86	
23	jeune	23.4	10	34	80185	−2.76	
24	faible	3.9	9	31	13231		1^{-2}
25	commun	5.4	8	27	18417	1.11	
26	naturel	6.8	8	27	23387	0.49	
27	nécessaire	5.7	8	27	19588	0.89	
28	petit	38.7	8	27	132626	−4.94	
29	sûr	4.1	8	27	14000		3^{-2}
30	bon	20.1	7	24	68721	−2.92	
31	funeste	1.4	7	24	4940		5^{-4}
32	immobile	1.8	7	24	6186		2^{-3}
33	possible	6.2	7	24	21165	0.35	
34	sombre	3.2	7	24	11054		3^{-2}
35	tendre	3.4	7	24	11503		3^{-2}
36	tranquille	3.0	7	24	10208		2^{-2}
37	vague	2.2	7	24	7533		5^{-3}
38	amer	1.2	6	21	4025		2^{-3}
39	cher	9.3	6	21	31780	−1.09	
40	dur	2.3	6	21	7802		2^{-2}
41	étranger	6.1	6	21	20860	−0.05	
42	généreux	1.9	6	21	6455		2^{-2}
43	importun	2.9	6	21	10013		5^{-2}
44	inquiet	1.6	6	21	5396		5^{-3}
45	noble	4.7	6	21	16254	0.61	
46	particulier	6.3	6	21	21614	−0.16	
47	passager	0.9	6	21	3143		3^{-4}
48	passionné	1.1	6	21	3724		8^{-4}
49	reconnaissant	0.8	6	21	2743		2^{-4}
50	subit	0.8	6	21	2616		2^{-4}

$A^{-B} = A * 10^{-B}$

TABLE V
Key-Adjectives of ADOLPHE

ADJECTIVES	ABS	ANF	Z	P(x)
1 malheureux	29	5.6	9.92	
2 bizarre	11	1.5		5^{-7}
3 violent	12	2.3		5^{-6}
4 propre	21	8.3	4.48	
5 libre	15	4.7		8^{-5}
6 subit	6	0.8		2^{-4}
7 reconnaissant	6	0.8		2^{-4}
8 content	10	2.4		2^{-4}
9 timide	6	0.9		3^{-4}
10 passager	6	0.9		3^{-4}
11 funeste	7	1.4		5^{-4}
12 passionné	6	1.1		8^{-4}
13 heureux	21	10.7	3.10	
14 amer	6	1.2		2^{-3}
15 immobile	7	1.8		2^{-3}
16 inquiet	6	1.6		5^{-3}
17 vague	7	2.2		5^{-3}
18 long	18	9.7	2.69	
19 généreux	6	1.9		1^{-2}
20 faible	9	3.9		1^{-2}
21 vif	11	5.3	2.50	

Flaubert's use of the key-adjectives that occur in *Madame Bovary*. The first step in the analysis is to examine the key-adjectives as they occur out of context as seen in Table VI.

The difference between the key-adjectives of Constant and Flaubert is striking. Those of Constant are abstract, and relate to emotions and impressions, while those of Flaubert are heavily loaded with color, and adjectives that relate to what can be experienced with the senses, in other words his adjectives are concrete. This is what one would expect since Flaubert is considered a representative of the school of realism, of which one of the principal characteristics was the description of the environment in all its concrete details. Analysis of the adjectives from the novels of the 17th, 18th, and the first half of the 19th century confirm our findings as indicated in Table VII.

TABLE VI
The Key-adjectives of MADAME BOVARY

KEY-WORD	Z	KEY-WORD	Z
1. noir	11.4	11. rouge	6.7
2. pâle	9.8	12. brun	6.3
3. blanc	9.4	13. fermé	6.0
4. vert	9.0	14. nu	5.5
5. jaune	9.0	15. large	5.1
6. bleu	8.7	16. gros	4.3
7. long	8.3	17. perdu	4.0
8. immobile	8.0	18. gris	3.8
9. vide	7.0	19. lourd	3.2
10. suspendu	6.9	20. fin	3.2

It should be understood that the sole purpose of statistics in stylo-statistical studies is to identify vocabulary that *may be* of interest for stylistic analysis. Very little can be learned about a writer's style from a list of words until they have been studied within their context. What can be learned from a contextual analysis of Flaubert's key-adjectives?

Flaubert has given us a clue as to what to look for when examining *Madame Bovary*. He has claimed that in writing the novel, he wanted to create a tonal quality the color of mold, an atmosphere similar to that of the wood louse[7]. His intent was to create what he considered to be the depressingly oppressive atmosphere typical of the bourgeois society of provincial France during the 1850s.

At another time he wrote in one of his letters to Louise Colet, "The reader will not notice, I hope, all the psychological work hidden under the form, but he will feel the effects[8]." This is a startling statement, for he is saying that he has done something with the language, which will act on the reader's unconscious, in such a way that it will not be noticed — or at least he hopes so.

How did Flaubert plan to establish the atmosphere? What psychological effects was he referring to? A study of the key-words provides the answer. When the adjectives are examined within their context, one notes that they have a figurative or symbolic value that is either inherent to the word itself or acquired through usage in the text. Further analysis reveals that Flaubert has taken these adjectives and used them as a musician would a leitmotif; as the story unfolds, the figurative and symbolic values of the adjectives act on the unconscious of the reader, establishing the moods of the story.

The adjective *vide* provides a example of this. We first encounter it in a description of Charles

TABLE VII
The Key-adjectives of:

La Princesse de Clèves	Manon Lescaut	Le Rouge et le noir	Le Père Gorlot
1. affligé	1. perfide	1. grossier	1. jeune
2. surpris	2. tendre	2. jeune	2. parisien
3. galant	3. honnête	3. sévère	3. pauvre
4. extrême	4. sage	4. affreux	4. riche
5. violent	5. surpris	5. extrême	5. heureux
6. aimable	6. charmant	6. riche	6. élégant
7. extraordinaire	7. extraordinaire	7. singulier	7. épouvantable
8. particulier	8. commode	8. magnifique	8. social
9. agréable	9. médiocre	9. brillant	9. joli
10. embarrassé	10. fidèle	10. cruel	10. faux
11. digne	11. malheureux	11. joli	11. horrible
12. touché	12. unique	12. noble	12. beau
13. blessé	13. considérable	13. immense	13. cher
14. fâché	14. vif	14. sage	14. honnête
15. insupportable	15. satisfait	15. terrible	15. mauvais
16. heureux	16. cher	16. étonné	16. bon
17. intime	17. doux	17. tremblant	17. vieux
18. grand	18. cruel	18. pâle	18. fou
19. véritable	19. capable	19. profond	19. haut
20. malheureux	20. heureux	20. fier	20. pur

Bovary as he is walking across a field with his bride, Emma:

> . . . elle [Emma] enlevait les herbes rudes avec les petits dards de chardons pendant que Charles, les mains *vides*, attendait qu'elle eût fini. (026/14[9])

When we are out walking, or pausing during a walk, our hands are usually empty. So why does Flaubert mention this obvious detail about Charles? It certainly isn't to let us know that his hands are empty. It is rather that by making the reader conscious of Charles' hands, we feel that Charles also is conscious of them, that he does not know what to do with them. An image is thus created of someone who is rather awkward and not very sophisticated. More important, the emptiness of the hands appears symbolic of the emptiness of his being, of the life that Emma will have with him now that they are married.

In another example of Flaubert's use of *vide*, Emma hopes to be invited once again to the annual ball held by a local nobleman. The hoped for invitation fails to arrive: "*Après l'ennui de cette déception, son coeur de nouveau resta vide.*" (059/14) Here we encounter the figurative sense of *vide* which, when modifying the word heart, evidently means void of hope and joy.

In another instance, at the end of a boring evening for Emma, we read: "*Le feu se mourait dans les cendres; la théière était vide.*" (092/36) It is evident that the fire and the teapot are symbols for life which not only is empty of vitality, but so empty as to be virtually dead.

In the next example, Emma has just received a letter from her lover telling her that he is leaving her. She runs up into the attic, and looks out the window: "*En bas, sous elle, la place du village était vide, les cailloux du trottoir scintillaient, les girouettes des maisons se tenaient immobiles.*" (191/33)

In establishing the mood, Flaubert describes the emptiness of the scene from several viewpoints. The village square is so empty of life that something so small as a pebble — the sparkling of the pebbles — strikes one's attention. There is a deathly stillness that is evidenced by the stillness of the weathervanes.

* * *

In another series of examples, the key-adjective

immobile conveys the sense of inertness, of the lack of movement and vitality. In this sense, *immobile*, like *vide*, evokes the depressingly dreary French provincial life of the 1850s.

Emma's first lover is about to move away and consequently leave her. She is in her room, feeling quite depressed. Flaubert does not explicitely describe Emma's emotional state — his description of the room conveys her emotion:

> . . . il [the lover] se cacha derrière un pilier, afin de contempler une dernière fois cette maison blanche . . . Il crut voir une ombre derrière la fenêtre, dans la chambre, mais le rideau, se décrochant de la patère comme si personne n'y touchait, remua lentement ses longs plis obliques, qui d'un seul bond s'étalèrent tous, et il resta droit, plus *immobile* qu'un mur de plâtre. (112/20)

At another point in the story, Emma takes on another lover, Rodolphe, who in his turn abandons her. When she discovers that he has left her, she faints. It is the beginning of a nervous breakdown (*fièvre cérébrale*) that leaves her in a debilitated state for months: "*Elle restait étendue, la bouche ouverte, les paupières fermées, les mains à plat, immobile et blanche comme une statue de cire.*" (194/11) It is worth noting here that two other key-adjectives: *blanc* and *fermé* occur in this sentence, and both reinforce the feeling of utter loss.

There is then a description that sets the atmosphere: "*Cependant, la neige sur le toit des halles jetait dans la chambre un reflet blanc, immobile; ensuite, ce fut la pluie qui tombait.*" (198/18)

Flaubert finally associates immobility with death: "*Quant au souvenir de Rodolphe, elle l'avait descendu tout au fond du coeur; et il restait là, plus solennel et plus immobile qu'une momie de roi dans un souterrain.*" (200/25)

When Flaubert characterizes an inanimate noun as immobile, it is obvious that he is doing something more than qualifying the noun. He is describing the atmosphere that the character feels. Flaubert's description of the ball at la Vaubyessard makes it quite clear that he is describing a family and a class that are dying out. When he describes the statue, which is incapable of movement, as "immobile", it is evident that the adjective is being used more to create an atmosphere than to characterize the statue: "*Sur le grand poêle de porcelaine à baguettes de cuivre, une statue de femme drapée jusqu'au menton regardait immobile la salle pleine de monde.*" (045/40)

How we see our environment often depends upon our emotional state. When we are happy, everything about us seems bright, lively and cheerful; when we are depressed, the environment appears drab and lifeless. Flaubert does not tell us what the emotional condition of his characters is, he describes how they see their environment. For example, Emma feels that she is at the breaking point — psychologically; she seeks advice and moral support from her priest, who is more concerned with the mundane affairs of his church than with her, and consequently is of no help. She returns, quite dejected, to her home and her bedroom:

> Le jour blanchâtre des carreaux s'abaissait doucement avec des ondulations. Les meubles à leur place semblaient devenus plus *immobiles* et se perdre dans l'ombre comme dans un océan ténébreux. La cheminée était éteinte . . . (107/25).

We again note the use of *immobile* to characterize an inanimate object. At the same time, *blanc* or its derivatives has been used in several examples to convey a feeling of despair. As Flaubert uses the word, it acquires the underlying meaning associated with sickly or deathly whiteness. When the word is used outside that context, it still retains, on an unconscious level, the atmosphere associated with sickness or death. Note also reference to the fire that is extinguished or dead.

Virtually all of Flaubert's key-adjectives (a few such as *bleu, vert, long* are exceptions) operate in this way, i.e., they are used in a context where the figurative or symbolic meaning is evident. When the context no longer calls forth the figurative or symbolic value, it is felt on an unconscious level. The following examples are drawn, one each, from the greater part of the remaining key-adjectives: (The effect created by *blanc* and *fermé* have been described above.) "*Le lendemain fut, pour Emma, une journée funèbre. Tout lui parut enveloppé par une atmosphère noire. . . .*" (115/01)

Both the adjectives *brun* and *rouge* are associated with everything that is rustic, that lacks refinement, that lacks attraction for Emma:

> Tout en lançant contre la borne un long jet de salive *brune*, il soulevait du genou son instrument, . . . (061/18)

> Quoiqu'il [Charles] ne fût pas large des épaules, son habit veste . . . laissait voir, par la fente des parements, des poignets *rouges* . . . (003/17)

The atmosphere can be felt weighing down on the characters: "*L'air du bal était lourd, les lampes pâlissaient.*" (048/36)

Gros and *large* produce much the same effect. Within their context they characterize persons and things as being "big, stout, coarse," while *large*, at the same time, appears to magnify the disagreable aspect (disagreable to Emma) inherent in the noun it modifies.

> Il *Charles* avait sa casquette enfoncée sur les sourcils et ses deux *grosses* lèvres tremblotaient. (095/06)

> Le long des bâtiments s'étendait un *large* fumier. (013/33)

Thus we have here, exposed to us through a stylo-statistical study of certain key-adjectives of *Madame Bovary*, one of the means that Flaubert used to (1) establish the moods of the novel and (2) create particular psychological effects upon the reader.

* * *

To summarize, now that humanists have access to computers, data in machine-readable form, and programs to operate on the data they also have new avenues of research available to them, among which is the stylo-statistical method. This method permits scholars to examine texts not only with more accuracy, but in greater depth, and it permits them to obtain information about a writer's style, e.g., that obtained from an analysis of the author's use of key-words or key-structures, that otherwise would be unattainable.

Notes

[1] Other aspects of stylo-statistics are described in my article, "Computer-aided Stylistic Analysis," to appear in *Computational Linguistics, An International Handbook on Computer Oriented Language Research and Applications*, I. Bátori, W. Lenders, W. Putschke, eds., (Berlin: Walter de Gruyter).

[2] See Pierre Guiraud. *Les Caractères Statistiques du Vocabulaire* (Paris: Presses Universitaires de France, 1954).

[3] *Cours de Linguistique Générale*, (Paris: Payot, 1964), p. 37.

[4] For a thorough understanding of the statistical basis of stylo-statistics, there are two excellent texts that treat the subject, one by Anthony Kenny, *The Computation of Style* (Oxford: Pergamon Press, 1982), the other, Charles Muller's *Initiation aux méthodes de la statistique linguistique* (Paris: Librairie Hachette, 1973). I shall give only a brief explanation here of the statistical basis of stylo-statistics.

[5] For a more complete explanation of lemmatization, see chapter 2 of my book, *A Stylo-Statistical Study of "Adolphe"*, (Genève — Paris: Slatkine — Champion, 1984).

[6] Examples are drawn from my book: *A Stylo-Statistical Study of "Adolphe"*, (Genève — Paris, Slatkine — Champion, 1984).

[7] "Je n'ai eu que l'idée de rendre un ton, cette couleur de moisissure de l'existence des cloportes. . . .", cited by Pierre Monier, "Flaubert Coloriste", *Mercure de France*, MVCLXIII (December 1921), p. 409.

[8] "Le lecteur ne s'apercevra pas, je l'espère, de tout le travail psychologigue caché sous la forme, mais il en ressentira l'effet." Letter of Jan. 1, 1854 to Louise Colet.

[9] The first three numbers at the end of each quote refer to the page, the last two numbers refer to the line. References are to the Classiques Garnier edition, 1968.