A Stylo–Statistical Analysis of W. Somerset Maugham's Short Stories (6)

——Ashenden or the British Agent (1928)——

WADA Koumei · YOSHIOKA Ken'ichi

Introduction

In 1928 *Ashenden or the British Agent* was published (hereafter referred to as *Ashenden*). This book comprises sixteen pieces of spy fiction. Maugham describes the book as based on his experiences in the Intelligence Department during the war (the First World War), but rearranged for the purpose of fiction.[1] However, when *the Complete Short Stories* (3 vols) were published in 1951, the 15 Ashenden stories were conflated into six respectively under a single title. The only one Maugham omitted is Section XIII *The Flip of a Coin*. But this story could have been included as a connected narrative of *His Excellency* in the collected edition.[2] The conflation was made as follows (The marks and the numerals shown after each title are respectively the abbreviations and the tokens of the 16 stories) :

Ashenden	Tokens	The Complete Short Stories
1. *R*.(R.)	899	
2. *A Domiciliary Visit* (DOM)	5,080	*Miss King*
3. *Miss King* (KIN)	7,130	
4. *The Hairless Mexican* (MEX)	6,010	
5. *The Dark Woman* (DAR)	3,660	*The Hairless Mexican*
6. *The Greek* (GRE)	4,868	
7. *A Trip to Paris* (TRI)	6,435	*Giulia Lazzari*
8. *Giulia Lazzari* (GIU)	77,62	
9. *Gustav* (GUS)	2,175	*The Traitor*
10. *The Traitor* (TRA)	11,937	
11. *Behind the Scenes* (SCE)	2,269	*His Excellency*
12. *His Excellency* (EXC)	10,936	
13. *The Flip of a Coin* (FLI)	1,633	omitted
14. *A Chance Acquaintance* (CHA)	5,208	
15. *Love and Russian Literature* (LOV)	4,131	*Mr Harrington's Washing*
16. *Mr Harrington's Washing* (HAR)	5,660	

R, the first story, is exceptionally short (899 words), but serves as the introductory or the beginning part of the whole stories.

The aim of this paper is to find out stylistic features of Maugham's short stories through a stylo-statistic or stylometric analysis of *Ashenden*. The text used for our analysis is *Ashenden or the British Agent* (London : Heinemann, 1967).

1. Word–length and TTR

Types, tokens, TTR, the mean word–length, and the number of letters in the longest word in each of the 16 stories are shown in Table 1(a).

The mean of word–length ranges from 4.0 to 4.5 letters with SD between 2.1 and 2.6. This is almost true of *The Mixture as Before* (1940) and *Creatures of Circumstance* (1947).[3] The mode of word–length in each story, which is simply the peak or the highest point of the frequency polygon, is three in letters.

As a general tendency, TTR is liable to be under the influence of the tokens ; the types increase as the tokens increases, but the increasing rate of the types comes down gradually. The column of TTR in Table 1(a) clarifies this tendency, though TRA, the longest story, is second to the bottom when arranged in the rank order with the text having the highest TTR at the top. In case we conflate the 16 episodes into a single connected story, the TTR drops in proportion to the increment of the texts (Table 1(b)). The mean of word–length is 4.3 letters (SD=2.3).

The longest words in *Ashenden* are such hyphenated ones as 'aristocratic–looking' (TRA) and 'fifty–million–dollar' (HAR).[4] Incidentally, the longest single word that does not employ hyphens is

Table 1 (a)   The Measurements of Word–length

Short Stories	Mean(SD)	Max Length	Type	Token	TTR
R.	4.26 (2.37)	16	386	899	0.43
DOM	4.29 (2.36)	15	1,323	5,080	0.26
KIN	4.33 (2.33)	17	1,648	7,130	0.23
MEX	4.15 (2.28)	18	1,465	6,010	0.24
DAR	4.04 (2.23)	15	1,022	3,660	0.28
GRE	4.18 (2.17)	15	1,294	4,868	0.27
TRI	4.22 (2.32)	17	1,641	6,435	0.26
GIU	4.06 (2.18)	16	1,460	7,762	0.19
GUS	4.28 (2.40)	19	701	2,175	0.32
TRA	4.30 (2.34)	20	2,346	11,937	0.20
SCE	4.46 (2.59)	16	782	2,269	0.34
EXC	4.27 (2.32)	20	2,251	10,936	0.21
FLI	4.31 (2.42)	16	608	1,633	0.37
CHA	4.41 (2.48)	17	1,483	5,208	0.28
LOV	4.43 (2.44)	16	1,173	4,131	0.28
HAR	4.32 (2.43)	20	1,286	5,660	0.23

Table 1 (b) TTR of the Cumulative Texts

Short Stories	Mean(SD)	Max Length	Type	Token	TTR
01–01	4.26 (2.37)	16	386	899	0.43
01–02	4.29 (2.36)	16	1,466	5,979	0.25
01–03	4.31 (2.35)	17	2,441	13,109	0.19
01–04	4.26 (2.33)	18	3,134	19,119	0.16
01–05	4.22 (2.31)	18	3,488	22,779	0.15
01–06	4.22 (2.29)	18	3,923	27,647	0.14
01–07	4.22 (2.29)	18	4,486	34,082	0.13
01–08	4.19 (2.27)	18	4,861	41,844	0.12
01–09	4.19 (2.28)	19	4,983	44,019	0.11
01–10	4.22 (2.30)	20	5,776	55,956	0.10
01–11	4.23 (2.31)	20	5,959	58,225	0.10
01–12	4.23 (2.31)	20	6,636	69,161	0.10
01–13	4.23 (2.31)	20	6,723	70,794	0.09
01–14	4.25 (2.32)	20	7,129	76,002	0.09
01–15	4.26 (2.33)	20	7,375	80,133	0.09
01–16	4.26 (2.34)	20	7,598	85,793	0.09

Table 2 The Frequency Distribution of Hyphenated Words

Text	No. of Types	No. of such Types Frequency	%	Tokens
R.	11	13	1.45	899
DOM	22	23	0.45	5,080
KIN	51	66	0.93	7,130
MEX	35	43	0.72	6,010
DAR	12	12	0.33	3,660
GRE	22	29	0.60	4,868
TRI	32	35	0.54	6,435
GIU	30	48	0.62	7,762
GUS	10	10	0.46	2,175
TRA	60	82	0.69	11,937
SCE	17	18	0.79	2,269
EXC	45	58	0.53	10,936
FLI	11	11	0.67	1,633
CHA	28	36	0.69	5,208
LOV	15	17	0.41	4,131
HAR	19	20	0.35	5,660

'straightforwardness' (GUS).

A chi–square test indicates that the differences of frequency among the texts are statistically significant.

2. Hapax Legomena and HTR

The numbers of the hapax legomena (HL), the ratio of the number of hapax legomena to the types (HTR), and HL / Tokens are shown in Table 3.

Table 3 Hapax Legomena, HTR and HL / Tokens

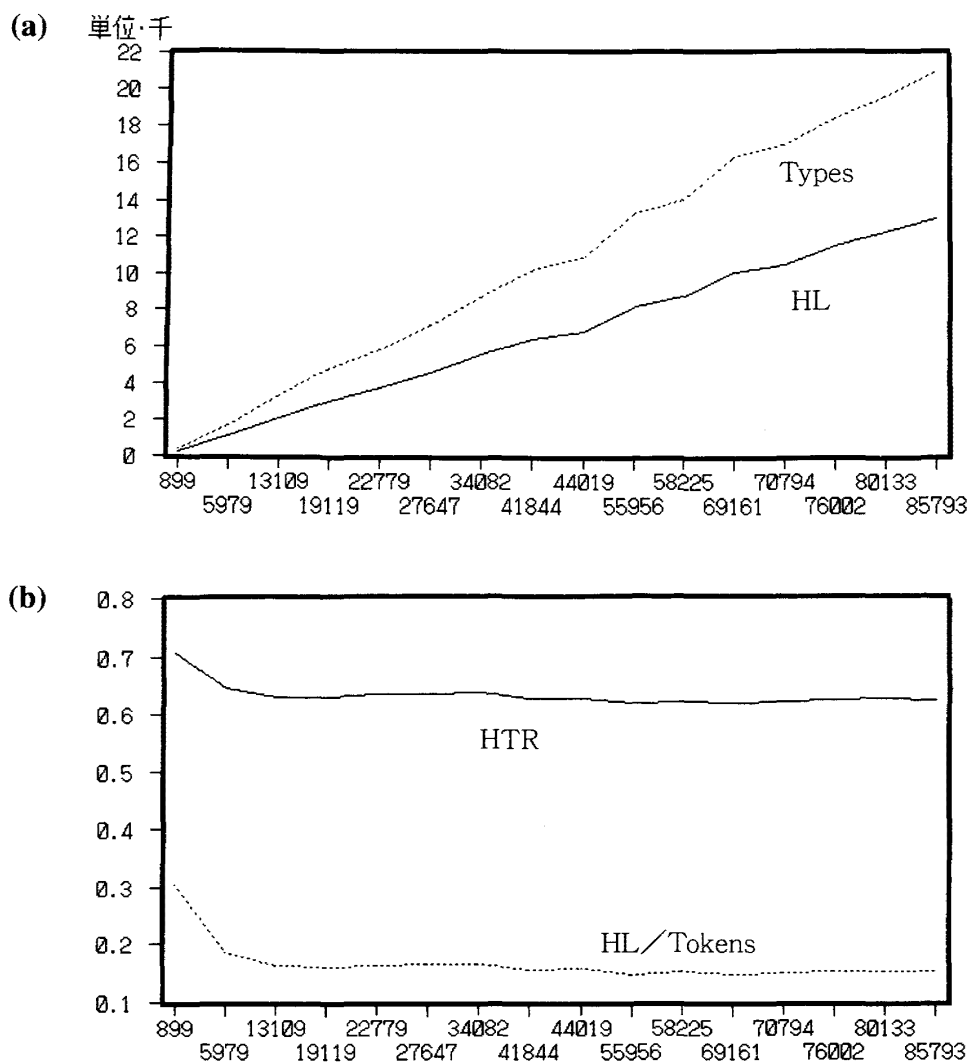| Text	HL	Types	HTR	HL / Tokens
R.	273	386	0.71	0.30
DOM	830	1,323	0.63	0.16
KIN	1.011	1,648	0.61	0.14
MEX	919	1,465	0.63	0.15
DAR	675	1,022	0.66	0.18
GRE	827	1,294	0.64	0.17
TRI	1,059	1,641	0.65	0.16
GIU	815	1,460	0.56	0.10
GUS	446	701	0.64	0.21
TRA	1,375	2,346	0.59	0.12
SCE	529	782	0.68	0.23
EXC	1,337	2,251	0.59	0.12
FLI	410	608	0.67	0.25
CHA	1,000	1,483	0.67	0.19
LOV	748	1,173	0.64	0.18
HAR	753	1,286	0.59	0.13

(a) 単位·千

(b)

Fig. 1 Polygon of Cumulative HTR and the Tokens

The number of hapax legomena increases with the increment of the tokens, but its increasing ratio drops. If we assume Ashenden is a single connected story, the inclement of Hapax legomena and that of the tokens can be graphed. (Fig. 1)

3. Richness of Vocabulary

We applied the statistics gained from the 16 texts to the Dugast's formula for calculating 'richness of vocabulary' (Table 4(a)).

Dugast's formula is

$$R(d) = \frac{\log^2 N}{\log N - \log V}$$

where N is the tokens, and V is the types.[5]

Table 4 (a)   Richness of Vocabulary

Text	R(d)	Text	R(d)
R.	23.76	GUS	22.65
DOM	23.50	TRA	23.52
KIN	23.34	SCE	24.34
MEX	23.29	EXC	23.76
DAR	22.92	FLI	24.06
GRE	23.63	CHA	25.32
TRI	24.44	LOV	23.92
GIU	20.85	HAR	21.88

According to Dugast, an index of 18 indicates a limited vocabulary, 20 an average vocabulary, and 24 a rich vocabulary. The indices of the 16 texts ranges between 20.85 and 25. 32.

As for the cumulative data, we got the following results (Table 4 (b)).

Table 4 (b)   Richness of Vocabulary

Text	R(d)	Text	R(d)
R.	23.76	GUS	22.79
DOM	23.36	TRA	22.86
KIN	23.23	SCE	22.94
MEX	23.34	EXC	23.01
DAR	22.30	FLI	24.01
GRE	23.26	CHA	25.18
TRI	24.33	LOV	23.21
GIU	22.85	HAR	23.12

When we regard *Ashenden* as a single connected story, we can safely say it has an average and rich vocabulary.

4. Repetitiveness of Vocabulary (Rep)

Rep is the ratio of the total number of the fifty most frequently used words to the tokens. The

formula for calculating Rep is

$$Rep = \frac{\sum\limits_{i=1}^{50} f_i}{N}$$

where N is the tokens.[6]

Table 5 shows the indices of Rep and the ratio of the fifty words to the types in each text.

Table 5 Repetitiveness of Vocabulary

Text	Rep	50/types	Text	Rep	50/types
R.	0.538	0.12	GUS	0.516	0.07
DOM	0.523	0.04	TRA	0.502	0.02
KIN	0.501	0.03	SCE	0.506	0.06
MEX	0.498	0.03	EXC	0.511	0.02
DAR	0.528	0.05	FLI	0.509	0.08
GRE	0.508	0.04	CHA	0.506	0.03
TRI	0.499	0.03	LOV	0.496	0.04
GIU	0.527	0.03	HAR	0.503	0.04

The total number of these fifty words occupies fifty percent in each text, though the ratio of these fifty words to the types is only small, ranging between 2 and 12 percent. The reason why R. takes an exceptionally high percentage may be that its types are small in number.

5. Sentence–length

Table 6 shows statistical features of sentence–length.

Table 6 Sentence–length

Text	Mean (SD)	MJ	MSJ	Text	Mean (SD)	MJ	MSJ
R.	16.6 (11.9)	−0.5	211.2	GUS	14.2 (11.2)	0	196.5
DOM	18.9 (14.4)	0.1	301.8	TRA	16.9 (13.9)	0	323.2
KIN	16.7 (13.4)	0	225.5	SCE	20.4 (14.8)	0.1	329.5
MEX	15.1 (12.6)	0.1	223.9	EXC	17.6 (14.1)	0	355.4
DAR	16.1 (13.6)	−0.1	315.6	FLI	13.0 (9.9)	0	172.9
GRE	13.3 (12.3)	0	239.4	CHA	22.2 (16.9)	−0.1	497.8
TRI	16.0 (14.8)	0	335.3	LOV	16.8 (13.9)	0	304.3
GIU	11.6 (9.7)	0	152.3	HAR	12.7 (9.6)	0	193.7

The sentence–length, contrary to the word–length, has a high degree of dispersion, in that the difference of the values of SD is greater (Table 7). In order to analyze the alternation of sentence–length, we calculated the mean of the difference between the length of each sentence and its immediate preceding sentence, which we refer to as the Mean Jump (MJ) after Schils and Haan (1993).[7]

The MJs having a minus sign like R., DAR, and CHA indicate that short sentences tend to come after long ones, and vice versa with the ones having a plus sign.

The longest sentence contains 153 words. Maugham prolongs the sentence using conjunctions

and semicolons.

It might be, he mused, as he rode along the lake on a dappled horse with a great rump and a short neck, like one of those prancing steeds that you see in old pictures, but this horse never pranced and he needed a firm jab with the spur to break even into a smart trot—it might be, he mused, that the great chiefs of the secret service in their London offices, their hands on the throttle of this great machine, led a life full of excitement ; they moved their pieces here and there, they saw the pattern woven by the multitudinous threads (Ashenden was lavish with his metaphors), they made a picture out of the various pieces of the jigsaw puzzle ; but it must be confessed that for the small fry like himself to be a member of the secret service was not as adventurous an affair as the public thought. (TRI)

Table 7 shows the frequencies of such conjunctions as 'and', 'but', colon, semicolon, and comma.

Table 7 Frequency of Conjunctions

Text	'and'	'but'	colon	semicolon	comma
R.	27	2	3	1	44
DOM	149	42	6	14	254
KIN	209	57	9	17	293
MEX	184	44	6	36	284
DAR	114	24	1	14	150
GRE	178	36	5	32	194
TRI	200	52	6	47	233
GIU	221	42	4	10	277
GUS	56	14	0	3	77
TRA	376	104	16	77	559
SCE	60	24	2	6	90
EXC	350	99	13	45	429
FLI	39	11	1	3	57
CHA	180	43	5	22	208
LOV	132	34	6	12	171
HAR	178	37	3	20	232
χ^2	✕	✕	✕	◯	◯

The mark ◯ indicates that the frequency in each text is stastistically significant and the mark ✕ does not.

6. Resemblance of the Stories

Based upon the correlation coefficient worked out from the frequencies of 1) content words, 2) function words,[8] and 3) the types that occur in each pair of the 16 texts, we constructed a dendrogram to see if some clusters appear (Fig. 2(a), (b), and (c)).
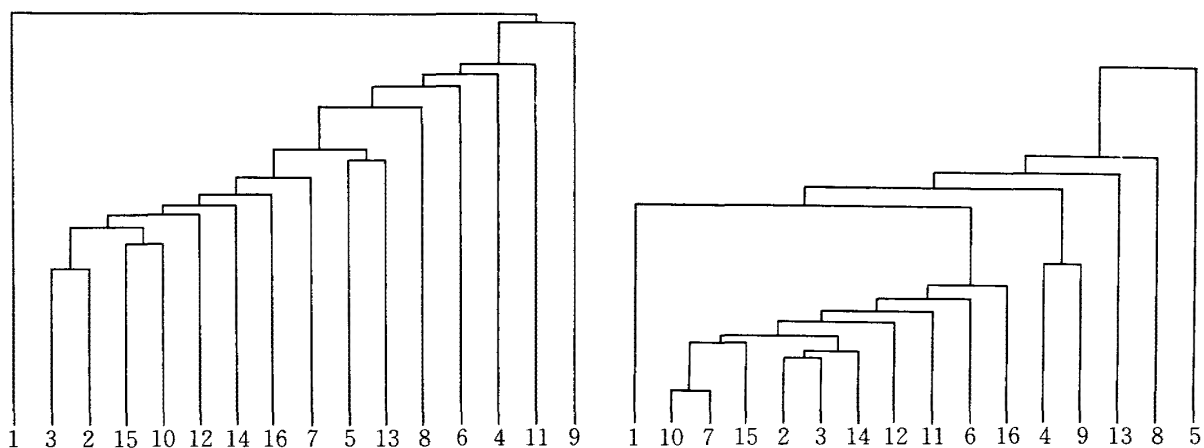
Fig. 2 (a) Content Words
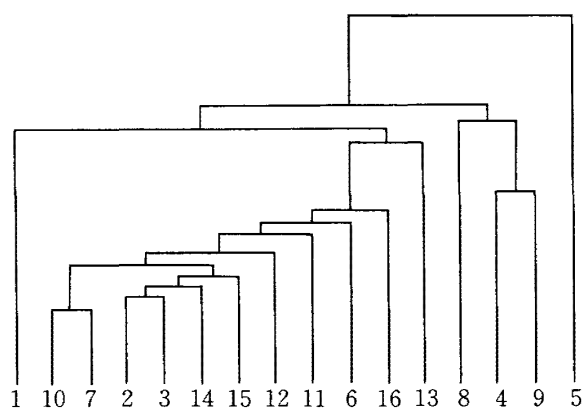
Fig. 2 (b) Function Words

Fig. 2 (c) Types

In the dendrograms there are some clusters of stories which have become grouped together ;

1) In the case of content words, DOM and KIN seem to be closest together, and TRA and LOV come next, and then DAR and FLI. That R. has the largest dissimilarities may be due to R.'s part it plays in the whole stories as an introduction.

2) In the case of function words, TRI and TRA have the closest resemblance and DOM and KIN the second, and MEX and GUS the third.

3) In the case of the types, the shape of the dendrogram is almost the same with 2).

7. Summary and Conclusion

Let us summarise our findings :

1) Maugham uses four–letter words most often in *Ashenden*. This is true of his other collections of short stories.

2) There is a remarkable occurrence of hyphnated words in such stories as KIN, TRA, and EXC.

3) HTR is about 60 percent on average (except R.), while the ratio of hapax legomena to the tokens is about 20 percent on average. Incidentally, HTR is about 57 percent and HL / tokens is about 15 percent in *Creatures of Circumstance*.

(4) According to Dugast's norm, the vocabulary in *Ashenden* is on an average level or rich.

(5) The ratio of the fifty most frequently used words to the types ranges from 2 to 12 percent. Their total number, however, occupies no less than fifty percent in each of the stories.

(6) The mean sentence lengths vary rather remarkably in Ashenden, compared with those of his other collections of short stories.

We can know the degree of jumps of sentence–length from the MSJ, while we can know the patterns of jumps from MJ ; plus jump and minus jump. The former indicates that long sentences follow short ones and the latter the reverse order through the values gained from MJ ; we have too many zeroes.

(7) The dendrogram based on function words and that on the tokens have very similar shapes.

Notes

1) William Somerset Maugham, *Ashenden or the British Agent* (London : Heinemann, 1928), Preface vii

The episodes first appeared in magazines as six short stories, each of between 12,000 and 15,000 words in length. John Whitehead, *Maugham––A Reappraisal* (London : Vision and Barnes & Noble, 1987), p.139

As for the name Ashenden, T. Morgan says : the name Ashenden, which Maugham used again in *Cakes and Ale*, has its origin in his years at the King's School, where one of his classmates, perhaps one whom he admired or envied, had the name, which he now adopted as his own. In 1954 the daughter–in–law of his school friend wrote to ask him about the choice of the name. "I chose the name Ashenden," he replied, "because like Gann and Driffield, it is a common surname in the neighborhood of Canterbury, where I spent many years of my youth. The first syllable had to me a peculiar connotation which I found suggestive." Ted Morgan, *Maugham—A Biography* (N.Y. : Simon and Schuster, 1980), p. 206–7

2) This short tale with an ambiguous conclusion might be thought of as 'best forgotten' by Maugham.
Stanley Archer, *W. Somerset maugham—A Study of the Short Fiction* (N.Y. : Twayne, 1993), p.43

3) See our papers, *Journal of Tezukayama College*, No. 31 (1994) and No. 32 (1995).

4) We included hyphens in counting letters.

5) Daniel Dugast, *Vocabulaire et Discours* (Geneva : Editions Slatkine,1979), p.67

6) Robert F. Allen, *A Stylo–Statistical Study of* ⟨*Adolphe*⟩ (Geneve : Slatkine–Champion, 1984), p.164–166

7) Erik Schils and Pieter de Haan, 'Characteristics of Sentence Length in Running Text', *Literary and Linguistic Computing*, Vol. 8, No. 1, p.22, 1993

They propose a formula to yield a measure which they refer to as MSJ, the mean squared jump. It is expressed in the following way :

$$\text{MSJ} = \frac{1}{n-1} \underset{i=1}{\overset{n-1}{SUM}} (L_{i+1} - L_i)^2.$$

8) Following Fries (1952), we designate 203 words as 'function words'.