

Computational Study of Proteus, Lestrygonians, and Penelope in *Ulysses* : an Intermediate Report

Kazuo Nagahara

Vocabulary studies of Joyce have been executed studiously since Miles Hanley published his *Word Index to James Joyce's Ulysses* in 1937, and at present every work of Joyce has its word index. Without their aids it is almost impossible to read Joyce, for his work enforces formidable cross-references to the reader. But the most careful and full examination of Joyce's vocabulary is that of Erwin R. Steinberg. In his *Stream of Consciousness and Beyond in Ulysses* (1973) he compares the streams of consciousness of Stephen, Bloom, and Molly as Joyce presents them in three chapters of *Ulysses*: Proteus, Lestrygonians, and Penelope. By analyzing meticulously sentence length and structure; word length and frequency; and type-token ratios as well as poetic images and sound correspondences, he shows to what extent Joyce used the stream-of-consciousness technique to differentiate his characters and project their personalities. It is not only the most comprehensive study of the stream-of-consciousness technique of James Joyce, but also an important contribution of stylostatistics and psycholinguistics to literature. But his analyses are based to a large extent on the six samples (approximately 175 words each) collected from each chapter, and they are not exempt from calculation errors inevitable to manual work. I carried out re-examination of Steinberg's result by using a computer, and found significant differences in the data. Just the figures reveal a lot. So I decided to publish them as an intermediate report of the computational study of Proteus, Lestrygonians, and Penelope in *Ulysses*.

The machine readable text of the three chapters was made from the 1984 Gabler edition. Most computer based studies of vocabulary analysis take the graphic form surrounded by blank spaces on a page for a word, but in this study words are identified by their spelling. Without regard to their case and form of letters, when any three given words have the same spelling they are counted as the same type of word, and its token is 3. If necessary, we put special markers to distinguish between them.

It is my pleasure to add here sincere gratitude to Professor Hisanori Kimira of Otaru University of Commerce who kindly helped me process the data. Without his assistance this study would not have been possible.

1. COMPUTATIONAL ANALYSIS OF SETENCE AND WORD

1. A. TOKEN (TOTAL WORDS), NUMBER OF SENTENCES, AVERAGE LENGTH OF A SENTENCE, AND THE LONGEST SENTENCE

	PROTEUS	LESTRYGONIANS	PENELOPE
TOKEN	5,677	12,552	24,161
SENTENCES	716	1,981	1
AV. SENTENCE	7.93	6.34	24,161
LONGEST SENTENCE	39	73	24,161

The length of sentences is scored by words.

1. B. DISTRIBUTION OF SENTENCE LENGTH

LENGTH	PROTEUS	LESTRYGONIANS
1-5	320(44.69)	1,100(55.53)
6-10	206(28.77)	582(29.38)
11-15	108(15.08)	190(9.59)
16-20	41(5.73)	69(3.48)
21-25	27(3.77)	23(1.16)
26-30	8(1.12)	10(0.50)
31-35	3(0.42)	3(0.15)
36-40	3(0.42)	2(0.10)
41-45		1(0.05)
46-75		1(0.05)
TOTAL	716(100)	1,981(99.99)

Parentheses are percentages of total.

1. C. FREQUENCY AND DISTANCE OF PUNCTUATIONS

PUNCTUATIONS	PROTEUS		LESTRYGONIANS	
	FREQ.	DISTANCE	FREQ.	DISTANCE
!	38	149.39	58	216.41
,	552	10.28	485	25.88
.	604	9.40	1,791	7.01
:	57	99.60	95	132.13
;	6	946.17	1	12,552.00
?	74	76.72	134	93.67
—	20	283.85	149	84.24

The distance of punctuations is scored by words.

1. D. DISTRIBUTION OF THE PUNCTUATION DISTANCE

DISTANCE	PROTEUS	LESTRYONIANS
1	307(22.78)	527(19.75)
2	235(17.45)	369(13.82)
3	196(14.55)	379(14.20)
4	136(10.10)	353(13.23)
5	106(07.87)	255(09.55)
6-10	284(21.08)	583(21.84)
11-15	66(04.90)	140(05.25)
16-20	11(00.82)	38(01.42)
21-25	6(00.45)	14(00.52)
26-30		7(00.26)
31-75		4(00.15)
TOTAL	1,347(100)	2,669(99.99)

1. E. TYPE (DIFFERENT WORDS), TOKEN (TOTAL WORDS), TYPE-TOKEN RATIOS, AND AVERAGE LENGTH OF A WORD

	PROTEUS	LESTRYONIANS	PENELOPE
TYPE	2,321	3,629	3,644
TOKEN	5,677	12,552	24,161
TTR	.409	.307	.203
AV. WORD	4.48	4.41	3.91

The sample of TTR is 10,000 words.

The length of words is scored by letters.

1. F. DISTRIBUTION OF WORD LENGTH

LENGTH	PROTEUS	LESTRYGOINIANS	PENELOPE
1	255(04.49)	448(03.57)	1,440(05.95)
2	863(15.20)	1,883(15.00)	4,578(18.95)
3	1,165(20.52)	2,723(21.69)	5,999(24.83)
4	1,083(19.08)	2,485(19.80)	4,883(29.23)
5	746(13.14)	1,805(14.38)	2,745(11.36)
6	522(09.19)	1,203(09.58)	1,766(07.31)
7	431(07.59)	864(06.88)	1,387(05.74)
8	269(04.74)	519(04.13)	621(02.57)
9	160(02.82)	312(02.49)	371(01.54)
10	95(01.67)	145(01.16)	209(00.87)
11	38(00.67)	76(00.61)	100(00.41)
12	23(00.41)	48(00.38)	38(00.16)
13	9(00.16)	22(00.18)	12(00.05)
14	8(00.14)	11(00.09)	6(00.02)
15	3(00.05)	1(00.01)	3(00.01)
16-	7(00.12)	7(00.06)	3(00.01)
TOTAL	5,677(99.99)	12,552(100)	24,161(100.01)

2. FREQUENCY AND LEVEL OF WORDS IN PROTEUS, LESTRYGOINIANS AND PENELOPE

2. A. WORD FREQUENCY

FREQ.	PROTEUS	LESTRYGOINIANS	PENELOPE
51-	10(00.43)	30(00.83)	71(02.11)
31-50	8(00.34)	18(00.50)	44(01.21)
21-30	11(00.47)	29(00.80)	57(01.56)
16-20	9(00.39)	25(00.69)	32(00.88)
11-15	21(00.90)	54(01.49)	82(02.25)
10	5(00.22)	16(00.44)	12(00.33)
9	5(00.22)	23(00.63)	32(00.88)
8	8(00.34)	29(00.80)	33(00.91)
7	15(00.65)	37(01.02)	49(01.34)
6	26(01.12)	48(01.32)	70(01.92)
5	27(01.16)	77(02.12)	90(02.47)
4	65(02.80)	108(02.98)	127(03.49)
3	109(04.70)	198(05.46)	209(05.74)
2	277(11.93)	504(13.89)	518(14.22)
1	1,725(74.32)	2,433(67.04)	2,212(60.70)
TOTAL	2,321(100)	3,629(100)	3,644(100)

2. B. VERBAL RELATIONSHIP OF THE THREE EPISODES

APPEAR IN 3 EPISODES	633 types	
APPEAR IN 2 EPISODES	PROTEUS AND LESTRYGONIANS	196 types
	PROTEUS AND PENELOPE	200 types
	LESTRYGONIANS AND PENELOPE	782 types
APPEAR IN 1 EPISODE ONLY	PROTEUS	1,313 (56.79) types
	LESTRYGONIANS	2,948 (56.43) types
	PENELOPE	2,065 (56.67) types

2. C. FUNCTION WORDS (319) AND CONTENT WORDS

	FUNCTION WORDS		CONTENT WORDS	
	TYPE	TOKEN	TYPE	TOKEN
PROTEUS	147(06.33)	2,428(42.77)	2,174(93.67)	3,249(57.23)
LESTRYGONIANS	190(05.24)	5,541(44.14)	3,439(94.76)	7,011(55.86)
PENELOPE	216(05.93)	13,991(57.91)	3,428(94.07)	10,170(42.09)

2. D. LONGMAN DEFINING VOCABULARY (2,000) AND COBUILD ESSENTIAL WORDS (3,000) OF THE EPISODES

	PROTEUS	LESTRYGONIANS	PENELOPE
TYPE	799(34.42)	1,331(36.68)	1,482(40.67)
TOKEN	3,827(67.41)	9,222(73.47)	20,335(84.16)

Refer to the defining vocabulary of *Longman Dictionary of Contemporary English* (new edition, 1987) and the essential wordlist of *Collins Cobuild Essential English Dictionary* (1988).

2. E. SPECIAL WORDS OF THE EPISODES

	PROTEUS		LESTRYGONIANS		PENELOPE	
	TYPE	TOKEN	TYPE	TOKEN	TYPE	TOKEN
PERSONAL NAMES	105 (04.52)	139 (02.40)	252 (06.94)	500 (03.98)	223 (06.12)	323 (01.34)
PLACE WORDS	46 (01.98)	57 (01.00)	99 (02.73)	127 (01.01)	117 (03.21)	167 (00.69)
AGGLUTINATIONS	138 (05.95)	141 (02.48)	155 (04.27)	161 (01.28)	158 (04.34)	171 (00.71)
FOREIGN WORDS	159 (06.85)	198 (03.48)	19 (00.52)	29 (00.23)		
SYMBOLS	8 (00.34)	11 (00.19)	23 (00.63)	27 (00.22)	69 (01.89)	165 (00.68)
ONOMATOPOEIAS	36 (01.55)	54 (00.95)	38 (01.05)	76 (00.61)	28 (00.77)	86 (00.36)
TOTAL	492 (21.20)	600 (10.57)	586 (16.15)	920 (07.33)	595 (16.33)	912 (03.77)

Unitalicized foreign words are not counted.

3. EXPRESSIVE BEHAVIOR

3. A. REFERENCES TO SELF

WORD	PROTEUS	LESTRYGONIANS	PENELOPE
I	87	130	772
me	39	58	268
my	33	28	193
Ill			56
Im			49
myself		5	35
Id			38
mine	6	2	7
I'm	1	12	
I'll	3	6	
Ive			4
I'd		2	
TOTAL	169(02.98)	243(01.94)	1,422(05.33)

3. B. REFERENCES TO SELF AND OTHERS

WORD	PROTEUS	LESTRYGONIANS	PENELOPE
we	6	21	64
wed			5
we'll		2	
weve			1
our	9	9	15
us	3	9	46
ourselves	1	1	1
TOTAL	19(00.33)	42(00.33)	132(00.49)

3. C. NEGATIVES

WORD	PROTEUS	LESTRYGONIANS	PENELOPE
not	29	40	100
no	20	46	61
never	1	19	53
didnt			37
wouldnt			36
dont			31
couldnt			27

WORD	PROTEUS	LESTRYGONIANS	PENELOPE
nothing	3	6	13
cant			20
wasnt			19
don't	2	13	
doesnt			10
didn't	1	9	
can't	1	9	
hadnt			9
no-one	4	5	
won't	3	6	
nobody			8
nor	1		7
couldn't	1	7	
wouldn't	1	6	
doesn't	1	6	
arent			6
none	2		3
isnt			5
havent			5
neither			4
isn't		4	
werent			3
weren't	2	1	
oughtnt			3
wasn't		2	
nought	2		
nonsense			2
hasnt			2
darent			1
TOTAL	74(01.30)	179(01.43)	465(01.92)

3. D. COMPARISON OF NEGATIVES AND YES

WORD	PROTEUS	LESTRYGONIANS	PENELOPE
NEGATIVES	74(01.30)	179(01.43)	465(01.92)
YES	9(00.16)	20(00.16)	91(00.34)

3. E. VERBAL BEHAVIOR AS A MANIFESTATION OF IDIOSYNCRATIC PSYCHOLOGICAL DEFENSE MECHANISMS

CATEGORY : WORD	PROTEUS	LESTRYGONIANS	PENELOPE
UNCERTAINTY :	8(00.14)	32(00.25)	117(00.44)
believe		4	12
hope		4	14
perhaps	3	11	7
suppose		6	55
wonder	5	7	29
RETRACTION :	8(00.14)	31(00.25)	138(00.52)
but	7	17	57
(of) course		12	48
except			6
though	1	1	21
unless		1	6
FEELING :	5(00.09)	18(00.14)	63(00.24)
alone	2		5
hate		3	15
happy		5	
love	3	7	32
pity		1	8
sorry		2	3
EXPLAINING :	11(00.19)	33(00.26)	111(00.42)
(on) account (of)		1	14
because		5	48
how	4	18	30
why	7	9	19
EVALUATION :	12(00.21)	69(00.55)	131(00.49)
bad	1	10	11
damn	1	3	9
good	3	26	20
lovely		4	20
nice		6	28
poor	5	15	21
queer	1	1	4
silly	1	3	1
stupid		1	6
wrong			11

The words less than frequency level 5 of total words of the three episodes are neglected.