

# Elyot、Marlowe、Shakespeareの語彙分析と比較

—コンピュータ利用による—

君 羅 久 則

## はじめに

近年のコンピュータのハードウェア、ソフトウェア両面における技術の進歩と普及にはめざましいものがあり、処理の高速化、メモリ装置の大容量化等が実現され、その結果、文学の研究者個人でさえも必要なコンピュータ機器を揃え、単にワープロとしてばかりでなく、テキスト分析などの研究に役立てることが可能になってきた。とはいえ、一面では、テキストの文章や語彙の分析など研究者個人々々が考える目的や必要性に十分応えてくれる、公表されたソフトウェアは余り多くはない。わずかに Oxford Computing Service が公開している Micro-OCP (Oxford Concordance Program) が現在のところ、パーソナルコンピュータで走らせることが出来、しかも信頼するに足る文章解析用プログラムであるといえる<sup>(1)</sup>。しかし、この Micro-OCP にしても、その主な機能は、Word list、Word index、それに Concordance を作成することであり、例えば複数の作品や作家間の比較分析などはできない。従って、研究者個人が自分の思い通りの目的を達成しようとするなら、勢い、独自のプログラムを組まなくてはならなくなる。幸いなことに、ある程度の訓練さえ積めば、門外漢のわれわれ文学や語学の研究者でも自分の目的を達

---

(1) Micro-OCP の他にも、例えば、WordCruncher (Electronic Text Corporation) や CLOC (University of Birmingham) などのテキスト分析用のプログラムがあるが、残念ながらどの機種でも使えるというわけではない。CLOC については、Christopher Butler, *Computers in Linguistics* (Oxford, 1965) に OCP との簡単な比較が記されている。

成し得るプログラムをパーソナルコンピュータ上で組めるような簡易言語やデータベース等のアプリケーションソフトウェアが普及しており、また、大きなテキストを扱うにはどうしても大きな容量のメモリ装置が必要になるが、大容量のハードディスクの普及によってこの支障も緩和されつつある。

文学・語学の分野におけるコンピュータ専門のジャーナルである *Literary and Linguistic Computing*<sup>(2)</sup> を見ても分かる通り、現在、文学の分野においても多くの研究者がコンピュータを利用して文体や語彙の分析に取り組んでいる。しかし、残念なことに、その研究結果は公表されても、その分析のために使用されたプログラムまでは公開されないのが通例である。数多くのプログラムが公開され、その中から研究者の目的に応じてプログラムを選択して使用することが可能になるなら、文学の分野においても、コンピュータを利用した研究が飛躍的に発展するであろう。少なくともある作家のコンコーダンスやグロサリーを独自に作成し、特定の語彙やセンテンスを拾い出したり、あるいはそれらの作品における分布を調べるといったように、文学作品の研究に強力な武器となる道具を提供してくれると考えられる。

小論では、筆者が独自に開発したテキスト分析用のプログラム、TAS (Text Analyzing System)<sup>(3)</sup> を利用して、Sir Thomas Elyot の *The Boke*

(2) Oxford University Press, 1986 年以後。それ以前の *Bulletin of the Association for Literary and Linguistic Computing* (ALLC Bulletin) と *Journal of the Association for Literary and Linguistic Computing* という ALLC の 2 つの機関誌が統合されたもので、文学・語学のコンピュータ利用による調査研究をカバーしているジャーナルである。

(3) 元々は dBASE II (Ashton-Tate) で組まれていたが、その後 dBASE III、dBASE III PLUS に変更し、更に現在では、この簡易言語の文法に従って組まれた後、QuickSilver (WordTech) によってコンパイルされ、MS-DOS (3.30) のリンカーによりリンクされている。同プログラムを利用して行った、永原和夫氏と共著になる、高等学校英語教科書の分析、「高等学校用英語教科書 Unicorn I、II、IIB の語彙分析」小樽女子短期大学「研究紀要」第 19 号 (北村正司学長追悼論文集、1990 年 1 月) がある。ただし、今回の分析に利用した TAS は劇作品や詩をも分析できるように改訂されているので、TAS に関する記述は小論の記述とは異なるところがある。

*Named The Governour* を中心に、Christopher Marlowe、Shakespeare の作品に用いられている語彙の分析と比較を主として行い、一部文体に関わる分析も試みる。さらに、現代イギリスの散文とも比較するために Muriel Spark の作品数編も加えた。その中で、TAS プログラムそのものにも若干触れながら、語彙分析に限らずこの種のテキスト分析にコンピュータを利用する上でのいくつかの問題点も明らかにしたい。

## 1. 調査の対象と方法

イギリスのルネッサンスのヒューマニストである Sir Thomas Elyot (1490? -1546) が著した *The Boke Named The Governour* (以下 *The Governour* と記す) は英語によって書かれた最初の教育論<sup>(4)</sup>であり、その中で彼は母国語による教育を主張し、更には *The Dictionary* (1538) を出版するなど、Elyot は Roger Ascham 等とともに母国語の擁護者の一人である。*The Governour* は、後に成長して統治者になるべき子弟の教育について論じたものであるが、スポーツや音楽など広範囲の話題について触れており、Elyot の主張の裏付けとして古今を通じて豊富なエピソードを語っている。特に、アレクサンダー大王やジュリアス・シーザーなどギリシア、ローマからの例が多く、この書物は当代ばかりでなく、後の詩人、劇作家など、特に Shakespeare に大きな影響を与えたことが明らかにされている<sup>(5)</sup>。

*The Governour* の中で Elyot が使用している語彙や文章を他の劇作家のそれと比較してみることは無意味なことではないし、その Word index ない

---

(4) F. Watson (ed.), *The Boke Named The Governour* (London, 1907), pp. xi-xii 参照。

(5) D. T. Starnes は *Henry IV* や *The Merchant of Venice* などの Shakespeare の作品に Elyot の影響があることを明らかにしている。("Shakespeare and Elyot's *Governour*," *Univ. of Texax Studies in English*, No.7 (Austin, 1927), 112-32)。Ralph M. Sargent は特に *The Two Gentlemen of Verona* に対する影響を論じている。("Sir Thomas Elyot and the Integrity of *The Two Gentlemen of Verona*," *PMLA* LXV (1950), 1169-70) .

しは Concordance を用意することは、Shakespeare をはじめ、ルネッサンス期の詩人や劇作家のことばを正確に知る上でも重要な意義を有すると思われる。比較の対象としては、Christopher Marlowe (1564-1593) と William Shakespeare (1564-1616) の劇作品からそれぞれ数編ずつ取り上げることにする。さらに現代イギリスの散文の例として、Muriel Spark (1918-) の数編の作品も加えた。

*The Governour* は 1531 年にロンドンで Thomas Berthlet の印刷所で初めて出版され、その後、16 世紀だけでも 7 回の改訂版が出されており<sup>(6)</sup>、かなり人気のあった書物の一つと考えられる。この初版 (1531) に基づいて H. H. S. Croft が 1883 年に編集出版したものをここでは基本テキストとして用いている<sup>(7)</sup>。

Marlowe の作品としては、Fredson Bowers 編 Cambridge 版<sup>(8)</sup>から、*Dido, the Queen of Carthage* (1594 年出版、以下 *Dido* と略記)、*Tamburlaine* 第一部 (1590、*1 Tamburlaine*)、同第二部 (1590、*2 Tamburlaine*)、*The Jew of Malta* (1633、*The Jew*)、*Doctor Faustus* (1604、*Faustus*)、それに *Edward the Second* (1594、*Edward*) の 6 編の悲劇を取り上げる。詩と翻訳の作品を除けば、この 6 編で Marlowe の劇作品の殆ど全てが含まれることになる。この他にはわずかに 1 編、*The Massacre at Paris* を残すだけである。

Shakespeare からは、*The Two Gentlemen of Verona* (以下 *Gentlemen* と略記)、*The Merchant of Venice* (*Merchant*)、*Twelfth Night*、*Much Ado about Nothing* (*Ado*)、*King Henry IV* 第一部 (*1 Henry IV*) の 5 編、主に喜劇を選んだ。いずれも Shakespeare の死後 1623 年に出版された二つ

(6) Facsimile 版 *The Booke Named the Governour* (Menston, 1970) のはしがきを参照。

(7) H. H. S. Croft (ed.), *The Booke Named the Governour* (London, 1883; rptntd. New York, 1967), 2 vols.

(8) Fredson Bowers (ed.), *The Complete Works of Christopher Marlowe* (Cambridge University Press, 1973; 1981), 2nd ed., 2 vols.

折り本初版 (First Folio)<sup>(9)</sup>を基本テキストとして用いた。Shakespeare の場合には、喜劇的な作品を選んだことから、人文主義的興味に富むことと、悲劇に比べて散文が多く使われている点は、多少なりとも Elyot に近づける意図がなかったわけではないが、それよりも作品の選択に関しては、Marlowe の場合も同様であるが、その長さ (使用されている語の延べ数) を最大の根拠としている。*The Governour* の使用語数は延べ 101,883 語であり、約 10 万語を目処として Marlowe と Shakespeare からそれぞれ 6 編の悲劇と 5 編の喜劇及び喜劇的作品を選び出した。

Spark からは *The Prime of Miss Jean Brodie*、*The Girls of Slender Means*、*The Go-away Bird*、*The Portobello Road*、*The Dark Glasses* の長編、短編の物語合わせて 5 編を選んだ。前 2 者は Penguin Books 版を、残り 3 編については *Collected Stories I* に収録されたものを使用した<sup>(10)</sup>。

尚、上記の基本テキストとして使用した版の他に、*The Governour* に関しては、Facsimile 版、並びに Watson 編の Everyman's Library 版、Marlowe の悲劇については J. B. Steane 編の Penguin Classics 版、また Shakespeare の作品に関しては First Folio の Facsimile 版と Arden 版<sup>(11)</sup>をそれぞれ参照し、後で触れるように若干の修正を行う際に利用している。

これらのテキストをそれぞれ作品毎にテキストファイル (SDF) として入

(9) Charlton Hinman (ed.), *The Norton Facsimile. The First Folio of Shakespeare* (New York, 1968).

(10) *The Girls of Slender Means* (Harmondsworth, Middlesex, 1963; rprntd. 1966), *The Prime of Miss Jean Brodie* (Harmondsworth, 1961; rprnted. 1965) を使用し、*The Go-away Bird*, *The Portobello Road*, *The Dark Glasses* の 3 編は *Collected Stories I* (London, 1967) から採用した。

(11) 参照した Arden 版テキストは次の通りである。John Russell Brown (ed.), *The Merchant of Venice* (London, 1955; rprntd. 1989); Clifford Leech (ed.), *The Two Gentlemen of Verona* (London, 1969; rprntd. 1986), A. R. Humphreys (ed.), *Much Ado about Nothing* (London, 1981; rprntd. 1989), J. M. Lothian and T. W. Craik (edd.), *Twelfth Night* (London, 1975; rprntd. 1988), and A. R. Humphreys (ed.), *The First Part of King Henry IV* (London, 1960; rprntd. 1989).

力した。このようなテキスト分析を行う場合にはテキストファイルをいかに正確に作成するかが重要であり、コンピュータを利用する上で、機械可読の形式のテキストを準備する難しさが大きな障害になっていると思われる。例えば 40 編近い劇作品と 2 編の長詩、ソネット、それに数編の詩作品を含む Shakespeare の全作品をワープロ等でタイプ入力し、プリンタで印刷し校正を行うとすれば、考えただけで利用を断念しかねない難作業になる。文章分析用のプログラム同様、テキストも機械可読の形で公開され、提供されるようになれば、この種の研究も大きな発展を遂げることが可能になると考えられる。その意味で、現在多くのコンピュータ可読のテキストを集積しつつある Oxford Computing Service の Oxford Text Archive (OTA)<sup>(12)</sup> は貴重な存在である。幸いに筆者は First Folio についてはこの OTA から公表されているテキストファイルを利用することが出来た<sup>(13)</sup>。しかし、このような場合でも、個別のプログラムに合わせて文字ばかりでなく、例えば Act や Scene の指示、ト書きと台詞の区別等をコンピュータが認識できるようにいわばテキストを加工しなければならない (encoding) という問題が残されている。Oxford Text Archive の First Folio には独自の encoding が施されているが、この調査では、後で述べるように、TAS で識別することができる記号に変更した上で利用している。

*The Governour* はタイトルページと目次を除いて巻頭の 'Proem' から最後の 'TALOS' まで全て入力した。ただし、'CHAPTER I,' 'CHAPTER II' 等の章番号は除外した。しかし、各章の見出しは、時として 3 行に及ぶ文から成っている場合もあるので、テキストとして取り扱っている。またページ欄外の小見出しは、人名や書名 (しかも略されていることが多い) や番号が

---

(12) Oxford Text Archive には英語ばかりでなく、ギリシア語、フランス語など様々な言語のテキストが集められている。Oxford Computing Service が出している OTA List (May 1991) を参照。

(13) すなわち OTA List, U-119-E。このテキストファイルの利用に関しては Oxford University Computing Service に深く感謝する。

主なので、削除した。

Marlowe と Shakespeare の劇作品については、Dramatis Personae を除いて、タイトル名、ト書き、Prologue や Epilogue も含めて入力されている。ただし、Speech Prefix は入力されているが、特殊な記号をつけてテキストとは別の取扱いをした。従って、Speech Prefix の登場人物名は使用語数には含まれることはない。*Faustus* のように Marlowe の劇には、劇全体が通し番号の行表示だけで Act や Scene に分かれていない劇がある。この場合には Bowers が [ ] つきで、Act と Scene を表示しているのので、その表示に従って Act、Scene に分けている。同様に、First Folio の中には Scene の表示がない劇が見られる。この場合には、Arden 版に従って Scene の指定を行った。

今回の調査に使用したテキストファイルにはコンピュータが的確に識別できるように次のような原文には印刷されていない記号を施した。

- <C ...> ‘<C’&‘>’とで囲まれた部分は Speech Prefix であることを示し、テキスト本文とは別個の取扱いとなる。
- <D ...> ‘<D’&‘>’とで囲まれた部分は劇のト書きのように本文と区別する必要がある時に使用する。テキストとして分析の対象となるが、この部分に属する単語にはデータベースファイルではDの記号がつけられているので、ト書きとして識別することが出来るようになっている。*The Governour* の各章の見出しにもこの記号がつけられている。
- <H> 行の先頭につけてその行を行数のカウントから除外するために使用する。劇では詩としての1行を複数の登場人物に割り当てることがよくある。このような場合に複数の行を1行として数えるために2行目以下にこの<H>をつける。また、*The Governour* の各章の見出しにもこの印をつけ、行数を数えないようにしている。従って、ページの先頭から章が始まっている時はそのページの0行目の表示になる。
- 。 このカタカナ用のマルは文の終わりを指示するための記号で、例えばト書きで‘Exit’とあってピリオドのないときにこの印をつけて文の終わりで

あることを示す。さもなければこの語が次の文の先頭に属することになる。この印はデータベースファイルでは削除される。

- ・ 省略符号としてのピリオドが語または文字の末尾についている場合に終止符としてのピリオドと区別するために利用する。
- 人名などで一人の名前が空白を間に挟んで2語からなるときに、それを1語として取り扱う、いわゆるワードラップをかけるために使用する。この調査では *The Governour* の South\_hampton と、First Folio の中で、同一人物が、例えば Gads Hill と Gads-Hill のように空白とハイフンとが混用されている時に限って使用した。データベースファイルでは空白に変わる。
- ハイフンのついた人名などをハイフンで分断するのを避けるために通常のハイフンとは区別してカナのハイフンを使う。
- ^ アポストロフィと区別するためにシングルクォーテーションの代わりに用いる。

その他、テキスト名を登録するための記号 ‘~’ やテキストの章や劇の Act を示す ‘#’、ページや Scene を指定するための ‘\*’ の記号を用いている。又、1語がハイフンを用いて2行にまたがって印刷されている場合、TAS は自動的にハイフンを取り除き、その語は次の行の先頭に位置するものとして処理する。このような語が本来ハイフンが入るべき語であるときは、これを避けるために、語全体を前の行に繰り上げるか、次の行に送って、入力した<sup>(14)</sup>。

テキストファイルの記述の最後に、ここで次の点についてそれぞれのテキストを修正したことを付記しなければならない。Croft 編の *The Governour* には、明らかに脱字と見られる部分と活字の判別できない部分とが3箇所ある。これらは Facsimile 版と Everyman's Library 版とを参照して訂正した。

---

(14) 通常は語全体を前行に繰り上げるが、その場合に次行が空白になってしまうときは次の行に送って入力した。



又、同書で、これは Facsimile 版も同じであるが、縮約された語のうち次の 10 語、延べ 21 語は *th aige* のように空白を入れて 2 語とした。尚、その際に アポストロフィはつけていない、又各語のうち括弧書きの数字のあるものは 延べ語数を示す。

*thaige*, *thauncient*, *thentent* (5), *theffected*, *thone* (2), *thother* (2), *thaccomplysshement*, *themperour*, *thepublike*<sup>(15)</sup>, *thaccompt*.

First Folio の中で ‘Hercules’ (MAN, 666)、‘Toby’ (TN, 2340)、‘Sir M.’ (1H4, 2612) の 3 箇所が Speech Prefix として取り扱われているが、Facsimile 版に従って、台詞の一部とした<sup>(16)</sup>他、次のような訂正を施した。左側が誤り、右側が訂正、括弧内の作品名の後の数字は OTA のテキストファイルで作品ごとに付された行番号である。

last	lost	(MV 660)	cozening	for cozening	(1H4 229)
Salanio	Slanio	(MV 793)	made	made with	(1H4 1055)
Sommer	Sommmer	(MV 1275)	fat	fatch	(1H4 1103)
let vs	let vs vs	(MV 2449)	most	most most	(1H4 1442)
sola,	sola, sola.	(MV 2594)	of of	of	(1H4 2025)
In faith	Infaith	(MAN 711)	out all	out of all	(1H4 2025)
Helpe me	Help	(MAN 1593)	a Lazarus	as Lazarus	(1H4 2400)

(15) この語だけは他の語と縮約の性質が異なるので恐らく誤植と思われるが、他のテキストでも同様であるが、このような場合にでも原文通りとした。

(16) ただし、実際に参照した Facsimile 版は Charlton Hinman 版ではなく、Sidney Lee 編になる Facsimile 版である (Sidney Lee (ed.), *Shakespeare's Comedies, Histories, and Tragedies* (Oxford, 1902))。その他、First Folio のテキストファイルでは、オリジナルにはないが、付加された記号のうち 10 数カ所訂正したことも付記しておかなければならない。尚、作品名の略号については、次節の冒頭参照。

Art thou	Art thou thou(MAN 2346)	Step	Steps	(1H4 2544)
a nayword	an ayword (TN 829)	builke	bulke	(1H4 2699)
Master	M. (TN 1998)	yea	gea	(1H4 2744)
pauyn	panyn (TN 2363)			

ここには主として語形、語数に影響を与えるものだけを挙げたが、この他に句読点等についても修正を加えた。

尚、次のような種類の語には以下に示す特別な記号をつけ特殊語として区別した。

- \* P 人名、国名、地名等の固有名詞とその形容詞
- \* L 明らかにラテン語として意識して使用されている語<sup>(17)</sup>
- \* G 同様なギリシア語
- \* M ラテン語、ギリシア語以外の外国語
- \* S 数字、記号、略語等

以上のようにして作成されたテキストファイルは、例えば *Merchant* の冒頭の7行を見ると次のようになっている。

~MV The Merchant of Venice.

# 01

\* 101

<H><D Actus \* L primus \* L.>

<H><D Enter Anthonio \* P, Salarino \* P, and Salanio \* P.>

(17) 例えば、*Faustus*, I. i. 41に見られるようなラテン語のセンテンスに含まれる語など。但し、劇のト書きによくある‘Exit’や‘Exeunt’等は特殊語としては取り扱っていない。

<H><C Anthonio \* P.>

In sooth I know not why I am so sad,  
 It wearies me: you say it wearies you;  
 But how I caught it, found it, or came by it,  
 What stufte 'tis made of, whereof it is borne,  
 I am to learne: and such a Want-wit sadnesse makes of

<H>mee,

That I haue much ado to know my selfe.

このようにして準備されたテキストファイルをデータベースファイル(DBF)に読み込み、分析に必要な一種のデータベースを作成するのであるが、このDBF作成には次のようなコンピュータ機器とソフトウェアを利用した。

コンピュータ NEC PC 9801 RA  
 プリンタ EPSON VP 1600  
 TAS (私家製テキスト分析プログラム)  
 dBASE III PLUS (Ashton-Tate)  
 QuickSilver (WordTech) (コンパイラ)<sup>(18)</sup>

*The Governour* であれば各巻毎に、Marlowe と Shakespeare の劇、それに Spark の物語の場合には作品毎にテキストファイルを読み込み、単語単位に分解し、テキスト名、ページ(劇の場合には Act と Scene)、それに行番号をつけたデータベースファイルを中間データとして作成する。この中間デー

(18) dBASE III PLUS と QuickSilver は TAS を運用する上では必要ない。TAS 作成に利用したソフトウェアである。ただし、前者はデータの検証並びに一部の分析にダイレクト・コマンドで利用している。

タでは、テキストファイルでは単語に付随するクォーテーション・マークやピリオドや記号等は単語から分離され、他のファイルにまとめられる。この中間データを基に、*The Governour* であれば 3 巻を、劇作品の場合には Marlowe 6 編、Shakespeare 5 編の作品をひとまとめにして、それぞれ最終的なデータベースが作られる。このデータベースは主として、アルファベット順に並べられ、使用頻度とコードのついた、単語リストのファイルと、単語のコードと出現位置を登録した位置ファイルとから構成される。その他、センテンスのリスト、ハイフン語リスト、句読点リスト、Speech Prefix のリストなどの DBF もこのとき同時に生成される。

分析や参照のために利用される Word list、Word index、Concordance の作成に必要なデータベースは以上のようにテキストファイルから中間データ作成、データベース構築のように 2 段階に分けて作られるが、これらの処理は TAS ではテキストファイル名と目的とするデータベース名を指定するだけで、後はすべて自動的に実行される。

## 2. 使用テキスト概観

細かい分析に入る前に分析・比較の対象として使用したテキストの規模などを概観するために、ここで、作品それぞれの出版の体裁とページ数、テキストの行数、総語数、異語数についてまとめてみた。次の表 1 のようになる。

テキスト名には *The Governour* (ELT)、Book I (GV1)、Book II (GV 2)、Book III (GV 3) のように省略して記し、Marlowe (MRL) の作品は概ね最初の 3 文字を省略形とし、Shakespeare (SHK) は C. T. Onions の省略形を利用した<sup>(19)</sup>。Spark (SPK) は、*The Prime of Miss Jean Brodie* (PJB)、*The Girls of Slender Means* (GSM)、*The Go-away Bird* (GAB)、*The Portobello Road* (PBR)、*The Dark Glasses* (DGL) のように略記し

---

(19) C. T. Onions, *A Shakespeare Glossary* (Revised by Robert D. Eagleton, Oxford, 1986; rptd. 1989) を参照。

表1 テキストの規模

Texts	体裁	pp.	ll.	T.W.	D.W.
ELT (1531)	Quarto	515	10,411	101,883	9,552
GV1		201	4,081	39,791	5,552
GV2		137	2,717	26,932	4,266
GV3		177	3,613	35,160	5,053
MRL	Quarto Quarto Octavo Octavo Quarto Quarto	418	14,330	107,416	10,756
DID (1594)		52	1,814	13,938	2,842
JEW (1633)		76	2,583	19,009	3,206
1TA (1590)		70	2,422	17,856	3,609
2TA (1590)		70	2,444	18,259	3,573
EDW (1594)		82	2,853	21,437	3,400
FAU (1604)		68	2,214	16,917	3,263
SHK (1623)	Folio	109	13,501	104,324	9,584
TGV		19	2,315	17,223	1,853
TN		21	2,585	19,894	3,361
MAN		21	2,676	21,150	3,227
MV		22	2,740	21,376	3,501
1H4		26	3,185	24,681	4,143
SPK	18cm 19cm 20cm 20cm 20cm	363	11,636	103,034	9,194
PJB (1961)		124	4,232	39,744	5,081
GSM (1963)		136	1,105	33,173	4,901
GAB (1958)		58	1,876	17,016	2,930
PBR (1958)		28	891	8,213	1,811
DGL (1961)		17	532	4,888	1,114

た。テキスト名の後の括弧書きの数字は出版年を表し、Marlowe の作品の出版年は Fredson Bowers による<sup>(20)</sup>。また、ページ数は、*The Governour* については Facsimile 版 (1970)、Shakespeare も Facsimile 版 (1902) によるが、Marlowe だけは Bowers の校訂が施されているので Cambridge 版のページ数を入れてある。これらの本の体裁とページ数によって印刷されたテキスト

(20) 注 (8) を参照。

の大体の大きさが知られるであろう。*The Governour* は四つ折り本で延べ515 ページ、Shakespeare の5編の劇は二つ折り本で109 ページになる<sup>(21)</sup>。Spark の場合は本の高さで表示したが、合わせて363 ページになる。しかし、この表に示した行数は印刷のそれではなく、テキストファイル上で純粋にテキストとして数えられる行数を表し、空白行や Speech Prefix だけからなる行などは含まない。総語数 (T. W.) はテキストに用いられている全ての語の延べ数 (token) で、異語数 (D. W.) は同形の語が何回出現しても1としてしか数えない場合の異なり語 (type) の数である。

前にも触れたように、Marlowe と Shakespeare の作品は *The Governour* の総語数 101,883 語を目処に選ばれているが、結果として Shakespeare で約 2,500 語、Marlowe では約 5,500 語、*The Governour* よりも多くなっている。しかし、興味深いことには、これらのテキストをその使用文字数によって比べると、*The Governour* は 476,275 文字を使用しており、Shakespeare の 422,526、Marlowe の 462,144、さらに Spark の 442,678 よりも多いことになる<sup>(22)</sup>。従って、総語数では他の2人に劣るが、われわれが書物として接するときのテキストの大きさとしては十分比肩しうる規模を有しているといえる。

### 3. 文の長さと語

各作家あるいは作品の特徴を把握するために文の長さと語の長さについて調べてみた。文の長さは1文に含まれる語の数で、又語の長さは語中の文字数によって測る。平均語数だけでは、劇作品の場合にはト書きの 'Exit' や

(21) 面白いことに First Folio の *1 Henry IV* には記されたページ番号に2ページ分欠けているところがあるが、ここに上げた数字は実質のページ数である。同様に他の作品にも番号を飛ばしているところがあるが、重複しているところもあるので最終的には見かけのページと実質のページ数は一致する。

(22) これらの文字数はデータベースファイルとして作成された後の単語リストから算出しているのでハイフンやアポストロフィ、句読点や記号は含まない。

‘Aside’なども1文を構成していることになり、しかもその頻度も高いので、この調査のように特に散文の作品と比較する場合には正確な指標とはならない。そこで、平均の語数ばかりでなく、それぞれの長さの文の頻度も調べてみた(表2 a)。語の長さについてもそれぞれの文字数の語の頻度を調べてみることにする。

表2 a 文の長さとは頻度

length	ELT	MRL	SHK	SPK
1- 5	49 ( 1.84)	1,312 (21.35)	1,760 (27.51)	1,424 (20.26)
6- 10	150 ( 5.62)	1,849 (30.08)	1,606 (25.10)	2,004 (28.52)
11- 15	242 ( 9.07)	617 (10.04)	766 (11.97)	1,219 (17.35)
16- 20	292 (10.94)	638 (10.38)	640 (10.00)	811 (11.54)
21- 25	259 ( 9.70)	420 ( 6.83)	418 ( 6.53)	534 ( 7.60)
26- 30	287 (10.75)	278 ( 4.52)	286 ( 4.47)	313 ( 4.45)
31- 35	229 ( 8.58)	282 ( 4.59)	218 ( 3.41)	232 ( 3.30)
36- 40	212 ( 7.94)	154 ( 2.51)	162 ( 2.53)	171 ( 2.43)
41- 45	167 ( 6.26)	125 ( 2.03)	135 ( 2.11)	110 ( 1.57)
46- 50	153 ( 5.73)	113 ( 1.84)	82 ( 1.28)	66 ( 0.94)
51- 55	120 ( 4.50)	78 ( 1.27)	66 ( 1.03)	46 ( 0.65)
56- 60	87 ( 3.26)	57 ( 0.93)	53 ( 0.83)	29 ( 0.41)
61- 65	81 ( 3.03)	51 ( 0.83)	31 ( 0.48)	17 ( 0.24)
66- 70	60 ( 2.25)	33 ( 0.54)	39 ( 0.61)	16 ( 0.23)
71- 75	39 ( 1.46)	26 ( 0.42)	27 ( 0.42)	8 ( 0.11)
76- 80	48 ( 1.80)	17 ( 0.28)	18 ( 0.28)	4 ( 0.06)
81- 85	38 ( 1.42)	21 ( 0.34)	18 ( 0.28)	9 ( 0.13)
86- 90	29 ( 1.09)	23 ( 0.37)	16 ( 0.25)	6 ( 0.09)
91- 95	16 ( 0.60)	8 ( 0.13)	8 ( 0.13)	3 ( 0.04)
96-100	23 ( 0.86)	5 ( 0.08)	10 ( 0.16)	1 ( 0.01)
101-105	13 ( 0.49)	3 ( 0.05)	5 ( 0.08)	1 ( 0.01)
106-110	13 ( 0.49)	5 ( 0.08)	4 ( 0.06)	0 ( 0.00)
111-115	11 ( 0.41)	5 ( 0.08)	6 ( 0.09)	2 ( 0.03)
116-120	7 ( 0.26)	4 ( 0.07)	3 ( 0.05)	0 ( 0.00)
121-	44 ( 1.65)	22 ( 0.36)	21 ( 0.33)	1 ( 0.01)
Total	2,669(100.01)	6,146(100.00)	6,398 (99.99)	7,027 (99.98)
T. W.	101,883	107,416	104,324	103,034
Average	38.17	17.48	16.31	14.66

括弧内の数字は各テキストの文の合計から見た百分率を表す。

文の長さは A. Q. Morton が *Literary Detection* の中で利用した文の長さの集計法<sup>(23)</sup>に基づいて、語数 1 から 5 まで、6 から 10 までのように 5 語刻みでまとめてある。ただし、121 語以上の長さの文は一まとめにして表示した。この表から分かるように、平均語数の長さとしては Spark が 14.66 で一番短く、Elyot は平均 38.17 語の長さの文を使用しており、Marlowe の 17.48 語、Shakespeare の 16.31 語よりも長く、他の 3 者の 2.2~2.6 倍の長さである。また、他の 3 人は 1-5、6-10 の一番短い方の文を約 50% の割合で使用しているのに対し、Elyot は 16-20 の文を最も多く使用し、前 2 者同様に過半数を越えるところまで頻度の高い順に選んで行くと、11-15 から 36-40 までになり、150 語を越える特に長い文も 19 回見られる。この傾向は、興味深いことに作品毎に (Elyot は各巻毎に) 見た場合も変わらない (Appendix I の表 A を参照)。従って作品の傾向というよりは作者の特徴を表すものと考えられる。ちなみに Elyot の最も長い文は 275 語で 2 回使用されている。Shakespeare も 271 語という Elyot に匹敵するような長さの文を使っている。しかし、Elyot の文が他の 2 人に比べて長いのは、散文で書かれ、しかも論文であることを考慮しなければならない。しかも、劇には比較的短いト書きが多く使われていることも考え併せるべきであろう。そこで、もう少し細かく見るために 3 人の作者が最も多く使っている文を頻度の高い方から順番に 10 位まで拾い出してみると、Elyot は 27 語から成る文を最も多く用い、最も短い文は 14 語である。それに対し、Marlowe は 16 語が一番長く、Shakespeare はすべて 10 語以下の文である (表 2 b)。

Marlowe、Shakespeare に共通して語数 1、2、3、4 の文が入っているが、これは短いト書きと関係があるかも知れない。ト書きに多く用いられ、しかも頻度も高いと考えられる 'exit'、'exeunt'、'enter'、'aside' の 4 語に限ってその頻度を調べると下記の表 3 a のようになる。勿論これらの語が全てト

---

(23) *Literary Detection* (New York, 1978)、pp.46-47.



表 2 b 使用頻度の高い文の長さ

ELT		MRL		SHK		SPK	
Freq.	Len.	Freq.	Len.	Freq.	Len.	Freq.	Len.
70( 2.62)	27	488( 7.94)	8	497( 7.77)	4	447( 6.36)	6
67( 2.51)	26	457( 7.44)	9	360( 5.63)	2	432( 6.15)	7
63( 2.36)	23	400( 6.51)	7	355( 5.55)	8	418( 5.95)	5
62( 2.32)	17	332( 5.40)	4	348( 5.44)	9	403( 5.74)	8
62( 2.32)	18	294( 4.78)	6	345( 5.39)	6	395( 5.62)	9
60( 2.25)	16	286( 4.65)	2	342( 5.35)	5	381( 5.42)	4
59( 2.21)	19	272( 4.43)	1	306( 4.78)	7	327( 4.65)	10
58( 2.17)	21	222( 3.61)	3	293( 4.58)	3	291( 4.14)	11
56( 2.10)	15	210( 3.42)	10	268( 4.19)	1	286( 4.07)	3
55( 2.06)	14	203( 3.30)	16	252( 3.94)	10	253( 3.60)	12
612(22.92)		3,164(51.48)		3,366(52.62)		3,633(51.70)	

書きに用いられて短い文を作っているわけではないが、そのかなりの部分は短文を増加させていると思われる。念のためにこのうち一番頻度の高い‘enter’は、文の長さもいろいろな例が考えられるので、この語が含まれる文の長さを調べてみた。TASではある特定の語についてそれを含む文を画面上で見ながら、その語の性質、機能などを入力し、性質、機能別に集計することができる機能を備えている。この機能を利用するとこの種の調査は比較的容易である。それによれば、MarloweとShakespeareでは、‘enter’を含む文の長さは表3bのようになり、Marloweで約半数が、またShakespeareでは

表 3 a ト書きに多い語の頻度

Words	MRL	SHK
exit	111	86
exeunt	163	85
aside	338	234
Total	692	415

表 3 b enterを含む文の長さ

Len.	MRL	SHK
1	0	1
2	53	64
3	46	24
4	40	39
5	28	23
6	171	83
Total	338	234

3分の2が5語以下の短い文に含まれている。

各作品毎に最も多く使われている文の長さを見してみる (Appendix I 表 B) と興味深いことが分かる。Marlowe と Shakespeare では上位の 10 番以内は殆ど変わらない。特に Shakespeare では順位は多少変わるが 5 編全てにおいて 1 語から 10 語までの文が占めている。Marlowe では総合テキストで最も頻度の高い 7、8、9 語の文が順番が入れ替わりながらも 6 編中 FAU を除いた<sup>(23)</sup> 5 編でやはり上 3 位を占め (ただし、2 TA では 7 語文が 15 語文と同位)、最も多く使われる文の長さは 8 語である。Shakespeare では 4 語の文ということになり、作品によってあまり変わらないところから作品というよりは作者の特徴を表すものと考えてよいのではなかろうか。しかし、Elyot ではそれほどはっきりした特徴は現れていない。総合テキストで上位の 26、27 語文、それに 18 語文が各テキストでも比較的上位に入っているのが分かる程度である。Elyot の場合長い文を多く使っているために語数は他の 2 人とほぼ同じでも、文の数は半分以下にしかならない。その結果頻度の差が余り大きくない (総合テキストでも 1 位と 10 位の差はわずかに 19 回である) ので、はっきりした特徴が出ないとも思われる。もしそうだとすれば、もっと多くのテキストを追加すれば差が出て来るかも知れない。しかしながら、*The Governour* に関する限りでは、Elyot はいろいろな長さの文を、しかも比較的長い文を使うところから、いうなれば丘原型の分布を示し、他の 3 者はいずれも、比較的短い文を集中して使用し、長い文は極端に少ないので、懸崖型にその分布の特徴があるといえる。長さから見た文の分布のこの傾向は図 1 に示したグラフに顕著に現れている。

次に、Elyot、Marlowe、Shakespeare、Spark の 4 人が使用している語の

---

(24) FAU は Marlowe の作品の中でも最も本文の異同の激しいテキストである (Bowers の FAU の 'Textual Introduction' を参照) ことと関係があるかも知れないが今はその検討の余裕はないし、上位 10 位に入っている文の種類は総合テキストと大きく変わってはいないのでそこまで考える必要はないのかも知れない。

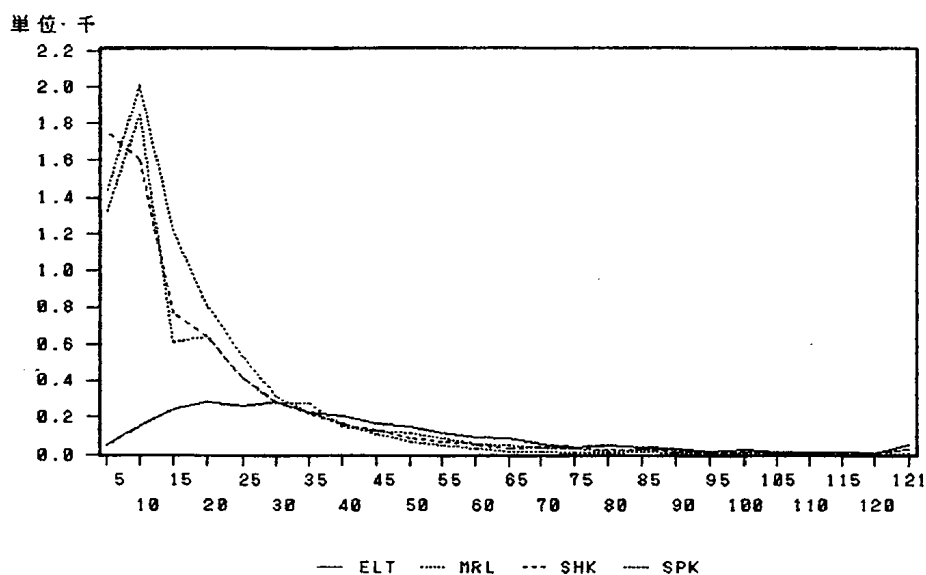


図1 長さから見た文の分布

長さについて比較してみることにする。語の長さは語に含まれる文字数で測るが、アポストロフィやハイフン、それにワードラップの空白は文字数として算入していない。例えば、'em、Bilia-borza、Gads Hill等はそれぞれ2文字、10文字、8文字の長さとして計算される。文の長さと同様に平均文字数だけでなく、それぞれの長さの語の使用頻度を異語数と総語数の両方の場合についてその分布も調べてみると表4のようになり、この表に基づいてグラフを描くと図2のようになる。

Elyotは異語数で平均7.58文字、総語数で4.67文字の長さの語を使い、4人の中では最も長く、Marloweの6.66、4.30文字、Shakespeareの6.64、4.05文字と比較すると、それぞれMarloweより0.98、0.27、Shakespeareより0.96、0.62文字も長いことになる。同じ散文のSparkの平均の長さ(7.03、4.30)と比較してみても0.55、0.27の差が見られる。各作家の最も多く用いる語の長さは、異語数で較べると、Elyotは7文字語、ShakespeareとMarloweは6文字語であり、Sparkは7文字語と6文字語がほぼ同じ位使用している。総語数で較べると、他の3人が3文字語が一番多いのに対して、Elyotは2文字語を最も多く使用している。

表4 語の長さと頻度

Len.	ELT		MRL	
	D.W.	T.W.	D.W.	T.W.
1	7 ( 0.07)	2,229 ( 2.19)	10 ( 0.09)	3,757 ( 3.50)
2	50 ( 0.52)	21,127 ( 20.74)	63 ( 0.59)	17,500 ( 16.29)
3	124 ( 1.30)	18,970 ( 18.62)	332 ( 3.09)	22,942 ( 21.36)
4	470 ( 4.92)	14,348 ( 14.08)	992 ( 9.22)	21,658 ( 20.16)
5	1,059 ( 11.09)	11,789 ( 11.57)	1,801 ( 16.74)	15,085 ( 14.04)
6	1,508 ( 15.79)	10,735 ( 10.54)	2,212 ( 20.57)	10,125 ( 9.43)
7	1,628 ( 17.04)	6,766 ( 6.64)	1,985 ( 18.45)	7,112 ( 6.62)
8	1,551 ( 16.24)	6,169 ( 6.05)	1,479 ( 13.75)	4,515 ( 4.20)
9	1,290 ( 13.51)	4,366 ( 4.29)	945 ( 8.79)	2,716 ( 2.53)
10	911 ( 9.54)	2,753 ( 2.70)	547 ( 5.09)	1,201 ( 1.12)
11	524 ( 5.49)	1,518 ( 1.49)	232 ( 2.16)	505 ( 0.47)
12	254 ( 2.66)	731 ( 0.72)	105 ( 0.98)	151 ( 0.14)
13	112 ( 1.17)	258 ( 0.25)	41 ( 0.38)	61 ( 0.06)
14	47 ( 0.49)	97 ( 0.10)	8 ( 0.07)	84 ( 0.08)
15	12 ( 0.13)	17 ( 0.02)	3 ( 0.03)	3 ( 0.00)
16	5 ( 0.05)	10 ( 0.01)	1 ( 0.01)	1 ( 0.00)
Total	9,552 (100.01)	101,883 (100.01)	10,756 (100.02)	107,416 (100.00)
Av.	7.58	4.67	6.66	4.30

Len.	SHK		SPK	
	D.W.	T.W.	D.W.	T.W.
1	14 ( 0.15)	5,712 ( 5.48)	17 ( 0.18)	3,460 ( 3.36)
2	65 ( 0.68)	18,069 ( 17.32)	74 ( 0.80)	16,568 ( 16.08)
3	316 ( 3.30)	23,734 ( 22.75)	276 ( 3.00)	25,313 ( 24.57)
4	918 ( 9.58)	21,672 ( 20.77)	916 ( 9.96)	20,381 ( 19.78)
5	1,692 ( 17.65)	13,794 ( 13.22)	1,260 ( 13.70)	11,846 ( 11.50)
6	1,919 ( 20.02)	8,777 ( 8.41)	1,515 ( 16.48)	9,043 ( 8.78)
7	1,718 ( 17.93)	5,634 ( 5.40)	1,551 ( 16.87)	6,792 ( 6.59)
8	1,203 ( 12.55)	3,284 ( 3.15)	1,271 ( 13.82)	4,221 ( 4.10)
9	799 ( 8.34)	2,081 ( 1.99)	938 ( 10.20)	2,564 ( 2.49)
10	497 ( 5.19)	890 ( 0.85)	632 ( 6.87)	1,535 ( 1.49)
11	275 ( 2.87)	437 ( 0.42)	373 ( 4.06)	695 ( 0.67)
12	107 ( 1.12)	154 ( 0.15)	211 ( 2.29)	369 ( 0.36)
13	42 ( 0.44)	55 ( 0.05)	102 ( 1.11)	169 ( 0.16)
14	12 ( 0.13)	21 ( 0.02)	42 ( 0.46)	54 ( 0.05)
15	6 ( 0.06)	9 ( 0.01)	10 ( 0.11)	17 ( 0.02)
16	1 ( 0.01)	1 ( 0.00)	6 ( 0.07)	7 ( 0.01)
Total	9,584 (100.02)	104,324 ( 99.99)	9,194 ( 99.98)	103,034 (100.01)
Av.	6.64	4.05	7.03	4.30

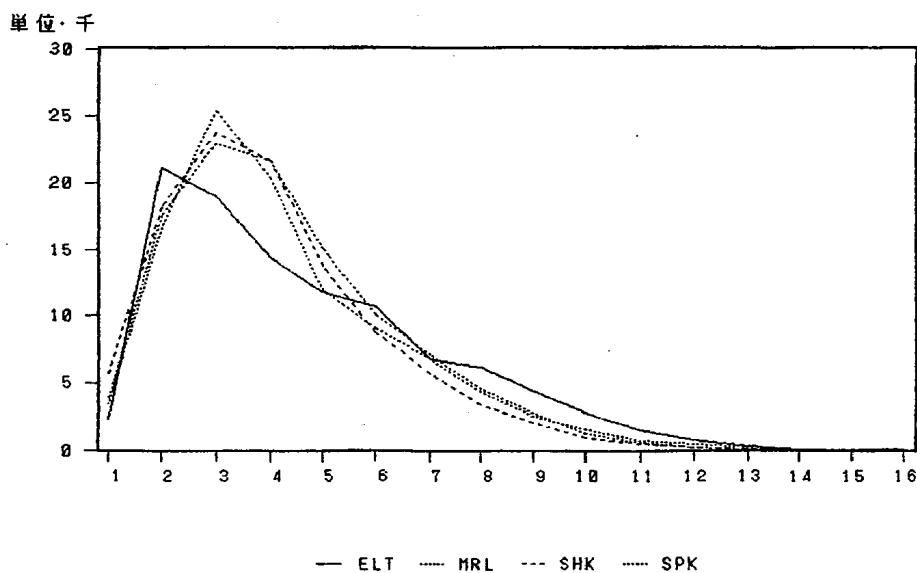


図2 長さ別に見た語の分布

作品別に語の長さを較べてみると、異語数ではまちまちの値を示すのに対し、総語数では類似した傾向を示す。例えば、Marloweでは平均文字数が、最も短い4.05 (JEW) から最も長い4.49 (2 TA) までその幅は0.44であり、Shakespeareではもっと幅がせばまって、3.99から4.07の、0.08の間にすべての作品が入る。(Appendix II表A参照)。Elyotでも各巻毎に見た場合、4.64から4.70の間に3巻とも入る。最も好んで使われる語の長さについても総語数で見ると、Elyotは各巻とも2文字なのに対して、他の3人は各作品とも3文字で、全体としてみた場合と変わらない。Elyotは他の3人と異なり、2文字語を最も多く使用し、平均文字数では4.67と最も長いことになる。

このように総語数からみた平均文字数、最も多く使用される語の長さとともに、作家の特色を示す指標として有効であることがわかるのである。そこで、表4のうち、総語数から見た語の長さの分布をグラフにしてみたのが図2である。グラフからも他の三者が3文字語で高い頂点をなし、後長くなるにつれ急勾配で減少するのに対して、Elyotは2文字語を頂点としてなだらかに段を形成し、対照的である。前者を鋭鋒型とするなら、Elyotは段丘型の分布を見せている。

## 4. 語彙の比較

Elyot、Marlowe、Shakespeare、Spark の4テキスト間で語彙がどの程度共通に用いられているかを調べた。Elyot の語彙を基本として出現頻度101以上、51以上100までのように11以上の頻度をもつ語はいくつかの頻度のグループにまとめてある。

表5 頻度別に見た語彙の比較

Freq.	ELT	MRL	SHK	SPK
101-	109 ( 1.25)	89 ( 0.93)	87 ( 0.97)	78 ( 0.91)
51-100	102 ( 1.17)	64 ( 0.67)	65 ( 0.72)	49 ( 0.57)
31-50	141 ( 1.61)	63 ( 0.66)	61 ( 0.68)	44 ( 0.52)
21-30	145 ( 1.66)	75 ( 0.79)	80 ( 0.89)	48 ( 0.56)
16-20	173 ( 1.98)	63 ( 0.66)	63 ( 0.70)	42 ( 0.49)
11-15	296 ( 3.39)	112 ( 1.17)	125 ( 1.39)	86 ( 1.01)
10	80 ( 0.92)	29 ( 0.30)	33 ( 0.37)	22 ( 0.26)
9	100 ( 1.15)	41 ( 0.43)	39 ( 0.43)	27 ( 0.32)
8	136 ( 1.56)	52 ( 0.55)	53 ( 0.59)	34 ( 0.40)
7	164 ( 1.88)	41 ( 0.43)	57 ( 0.63)	29 ( 0.34)
6	226 ( 2.59)	81 ( 0.85)	84 ( 0.93)	51 ( 0.60)
5	307 ( 3.52)	86 ( 0.90)	91 ( 1.01)	53 ( 0.62)
4	423 ( 4.84)	130 ( 1.36)	119 ( 1.32)	67 ( 0.79)
3	708 ( 8.11)	207 ( 2.17)	221 ( 2.45)	110 ( 1.29)
2	1,352 ( 15.48)	280 ( 2.94)	273 ( 3.03)	182 ( 2.13)
1	4,270 ( 48.90)	580 ( 6.08)	586 ( 6.51)	392 ( 4.59)
Total	8,732 (100.00)	1,993 ( 20.89)	2,037 ( 22.62)	1,314 ( 15.40)
Vocab.	9,552	10,756	9,584	9,194
Excl.SPC	8,732 (100.00)	9,539 (100.00)	9,005 (100.00)	8,533 (100.00)
SPC	820	1,217	579	661

ここでの比較にはあらかじめ特別な記号を付して区別した特殊語を除いた語を対象に行っている。特殊語はSPCとして参考までにそれぞれの合計を表4に記した。Excl.SPCが特殊語を除いた語の総数を表し、表中の括弧書きの数字はそれぞれのテキストのExcl.SPCから割り出した百分率である。

る。

Elyot 以外の 3 テキストの中では Shakespeare が最も高く、22.62 パーセント、約 5 分の 1 の語を Elyot と共有していることになる。しかし、頻度 101 以上の高い頻度の語は他の 3 テキストでも比較的多く含まれているが、出現頻度が下がるにつれて他のテキストに含まれる語は少なくなり、全体的には Marlowe で 20.89、Shakespeare でも 22.62 パーセントの低率でしか出現しないことになる。特に、現代の散文例である Spark では 15.40 の低率である。TAS では頻度のレベルを指定するとこれらの包含関係は自動的に算出され、各テキストの使用頻度を付した一覧表 Word List を印刷することが可能である。その語彙リストの一ページ目を Appendix III に掲げておいた。その表を見ても分かるとおおり、例えば、'suche' や 'shulde' のように高い頻度を持つ語が他のテキストに現れないのは、主として古い綴りの語の場合に限られている。この傾向は頻度の低い語についても当てはまり、全体として低率の共有関係しか見られない。また、現代英語の散文とは最も少なくしか一致しないことになるのもうなずけるのである。Elyot では、アリストテレスの綴りは同一人物にも関わらず、'Aristotel' (3 回)、'Aristotell' (10)、'Aristotelle' (4)、'Aristotle' (20)、'Aristotell' (1) の 5 通りにつづられる。括弧書きの頻度を見ても分かるとおおり、これは決して入力ミスではない。このようにいわばスペリングの「揺れ」の大きいルネサンス期のオリジナル・スペリングにかかわる問題を検討する場合には、後でみる句読法のそれと同様、単に作家の書き方ばかりでなく、印刷の段階、特に植字工の問題も含んでおり、結局はテキスト研究の成果をまつしかないと思われる。逆に、全てのスペリングや句読点を網羅することの可能な、コンピュータを利用したテキスト分析は、印刷の違いや植字工の癖を割り出したりすることが比較的容易であるので、テキスト研究に少なからず貢献することも可能である。しかし、ここではそこまでの余裕はないので、検討することは不可能であり、例えばスペリングを現代英語のそれに置き換えて比較するというような便法も考えられるが、今は将来の課題とするしかない。

## 5. 句読法について

最後に、Elyot、Marlowe、Shakespeare、Spark の4テキストに見られる句読点について調べてみることにする。各テキストでどの句読点がどんな割合で使用されているかを知るために、いくつかの句読点の使用頻度を算出するとともに、句読点間の距離もコンピュータによって測定した。ここでいう距離は、ある句読点から次の句読点までの間に存在する語の数によって測ることにする。こうすることによって、文の中で、コンマやピリオドをどんな割合で使用しているかが知られ、各作家に標準的な文の姿を描き出すことができ、それぞれの作家がどのくらい「息の長い」文、もしくは簡潔な文を使用しているかが分かるであろう。

コンマ、ピリオド、コロン、セミコロン、感嘆符、疑問符、引用符（シングルとダブルの2種）、括弧（‘ ’と’ ’の2種）、ダッシュ、ハイフンの12種類の句読点のほか、アポストロフィも含めてその頻度を調べた。表5には、編集者が挿入したと考えられる bracket (‘ [] ’) と、Spark だけに見られる省

表5a 句読点の使用頻度

PUN.	ELT			MRL			SHK			SPK		
	FREQ.	SENT.	WORD	FREQ.	SENT.	WORD	FREQ.	SENT.	WORD	FREQ.	SENT.	WORD
!	3	0.00	0.00	56	0.01	0.05	40	0.01	0.04	80	0.01	0.08
”	0	0.00	0.00	0	0.00	0.00	0	0.00	0.00	92	0.01	0.09
’	8	0.00	0.01	1,337	0.22	1.24	1,763	0.28	1.69	1,886	0.27	1.83
(	302	0.11	0.30	81	0.01	0.08	234	0.04	0.22	9	0.00	0.01
)	301	0.11	0.30	82	0.01	0.08	232	0.04	0.22	9	0.00	0.01
,	9,436	3.54	9.26	12,020	1.96	11.19	10,943	1.71	10.49	7,008	1.00	6.80
-	2	0.00	0.00	380	0.06	0.35	1,036	0.16	0.99	1,636	0.23	1.59
.	2,480	0.93	2.43	4,861	0.79	4.53	4,890	0.76	4.69	6,249	0.89	6.06
—	5	0.00	0.00	70	0.01	0.07	8	0.00	0.01	184	0.03	0.18
:	274	0.10	0.27	1,446	0.24	1.35	2,196	0.34	2.10	154	0.02	0.15
;	466	0.17	0.46	363	0.06	0.34	468	0.07	0.45	203	0.03	0.20
?	184	0.07	0.18	1,193	0.19	1.11	1,391	0.22	1.33	564	0.08	0.55
、	7	0.00	0.01	0	0.00	0.00	0	0.00	0.00	4,987	0.71	4.84



略記号‘…’を除いて全ての句読点とアポストロフィとを掲げる。表中でSENT. は1文当たりの数、WORDは同じく総語数で除した値であるが、後者は100倍してから除してあるので、100語当たりの数を表す<sup>(25)</sup>。

Elyotはコンマを総数でみると、Sparkに次いで少ない9,436回しか用いていないが、元々文が長いので、1文当たりでは、3.54個と最も多く、Shakespeareの2倍になり、総語数との割合では、Shakespeareの10.49、Marloweの11.19に近づいているといえる。ShakespeareとMarloweでは大体9語に1回の割合でコンマを用いているのに対し、Elyotは11語に1回の割合で用いている。

しかし、Marlowe、Shakespeareの2人はコロン、セミコロンを多く使っているので、やはり「息の長さ」はElyotのそれには較べられない。そこで、文の長さに関係するピリオド、疑問符、感嘆符に、4種の句読点(‘-’、‘:’、‘;’、‘,’)を加えて、これら7種の句読点間の距離を語数によって測ってみることにする(表5b)。

距離は1から5語までのように、5語刻みでグループにしてある。同時に2種類以上の句読点が出現したときは、1つしか現れないのと同じく計算しているので、表5bのTOTALは表5aの7種の句読点の頻度を単純に加算した値とは異なる。ElyotではSparkの最高49語に次いで長い、46語の長さも見られるが、「距離」の平均で見ると、Sparkの7.14に近い7.93であり、約8語に1回は上述の句読点のいずれかには出会うことになる。Shakespeare、Marloweの約5語に1回よりは長いですが、文の長さの違いから考えると、「息の長さ」では両者の差は相当に緩和されていると考えられる。

このように、Elyotは他の4人の作家に較べて2.2~2.6倍の長さの文を使っているが、コンマを相当の割合で使用し、バランスを保っているのでは

---

(25) ‘ ’は引用符(シングル)を表わす。表中で左右が対になるべき括弧(‘(’&‘)’)が異なる頻度を示しているが、オリジナルでも必ずしも左右の対が一致していないことがある。

表5b 句読点間の距離

LENGTH	ELT	MRL	SHK	SPK
1-5	4,412	10,797	11,966	6,867
6-10	5,253	8,361	6,784	4,492
11-15	2,244	670	1,029	1,834
16-20	688	137	136	771
21-25	180	19	18	296
26-30	49	3	3	101
31-35	13	0	0	43
36-40	4	0	0	16
41-45	2	0	0	5
46-50	1	0	0	2
TOTAL	12,846	19,987	19,936	14,427
MEAN	7.93	5.37	5.23	7.14

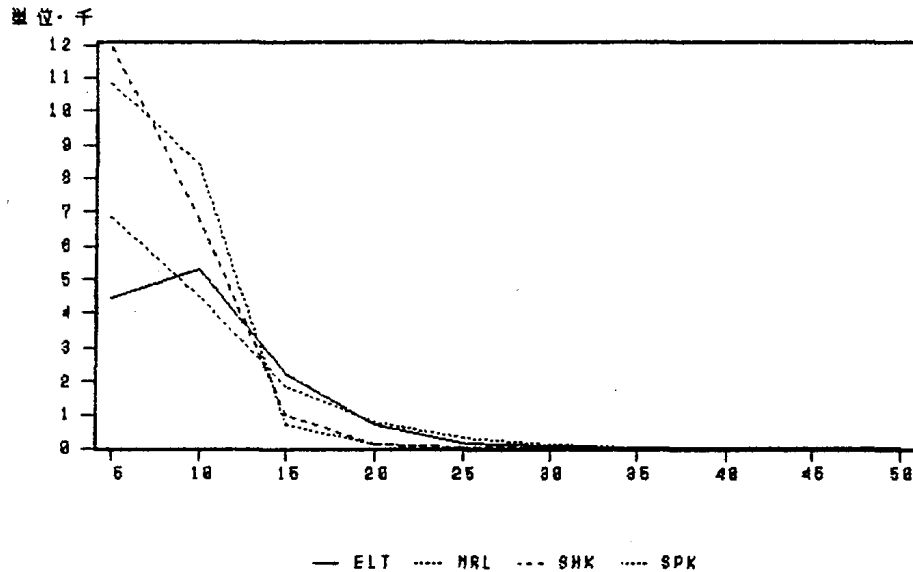


図3 句読点の分布

る。

句読点の分布をグラフで見ると、Elyot は他の3人とは異なる分布を示している(図3)が、彼が他の3人と顕著に異なる点は、Elyot はアポストロフィとハイフンをほとんど用いていないことである。ハイフンについては、'pre-eminence' を2度使用しているだけである。Shakespeare は Marlowe

と較べても格段に多くのハイフンを使用している。Elyot はアポストロフィをほとんど用いない。しかし、*The Governour* が出版されてから Shakespeare の First Folio が出版されるまでにはほぼ1世紀近くの時間の隔りがあり、印刷所も異なる。表記法の違いもあれば、印刷所、特に植字工の癖なども当然変化しているので、一概に論ずることはできない。例えば、Elyot では所有格でもアポストロフィなしで使うことが、Shakespeare などより、多い。また、Marlowe もそうであるが、Shakespeare では、韻文であることから、語を縮約することがしばしば起こるためにアポストロフィの数が目だつて多くなる。

劇作家に感嘆符と疑問符が多いのも、対話を主体とする劇作品の特徴である。シングルクォーテーション・マークを約5,000回も使っていることから分かるように、短い会話を多用する Spark に、散文の割には、疑問符と感嘆符が比較的多く用いられているのも理解できる。

## 結論

Elyot、Marlowe、Shakespeare、それに Spark の4種類のテキストの Word list、Concordance を用意し、それらの分析と比較を試みた。各テキスト間で類似よりは相違が目だつ結果に終わったが、しかし、それでも、他の3つの作品と比較することによって、Elyot の特色を幾分なりとも把握することができたと思われる。Elyot の最も好んで用いる語の長さは2文字語であるが、平均では4.67文字の語を、38.17語連ねる、非常に長い文を使いながらも、コンマなどの句読点を約8語に1回の割合で利用し、うまくバランスを保っている。文の長さの分布も、Shakespeare と Marlowe とは異なり、さまざまの長さの文を使い分けており、丘原型にその特徴がある。長さ別に見た語の分布では2文字語を頂点とする段丘型の分布を示し、長い語を比較的多く使っている。他方、語形の一致による語彙の直接的な比較はスペリングの違いや「揺れ」などのためにあまり明瞭な意味を持たないが、それでも、Marlowe よりも Shakespeare の方がより多くの語を共有していることがわ

かる。しかし、それも Elyot の本文校訂研究の成果を待ってテキスト研究を徹底するならば、もっと精密な比較が可能になると考えられる。テキスト研究を徹底的に行った上で、漢字に対する「振りがな」のように、全ての語に現代英語の綴りを添えるならば、直接的な語形一致による比較にも有効な手段を提供することになると思われる。この「振りがな」を一段と進めて、例えば、動詞や形容詞などの活用語などは原形に還元し、同形異語を区別できるようにすれば、更に精密な比較も可能になるとと思われる。語形から判断する一般的なコンピュータの語彙処理とは異なったものとなり、しかも手間暇のかかる作業を必要とするが、文学作品のテキスト分析には重要な要素であろう。なぜなら、文学においては語の形ばかりではなく、その語が有する含蓄や更には語が醸し出すイメージも劣らず重要なものであり、コンピュータを利用して機械的に語形だけで処理するのではなく、それを補完する意味で、この「振りがな」と同形異語の区別、原形への還元の活用は不可欠であると考えられる。将来の課題としたい。

## Appendix I

表A 文の長さと頻度 (作品別)

Length	GV1	GV2	GV3	DID	JEW
1- 5	14( 1.52)	24( 3.09)	11( 1.14)	91( 14.72)	435( 30.04)
6- 10	40( 4.33)	46( 5.93)	64( 6.60)	211( 34.14)	497( 34.32)
11- 15	78( 8.44)	71( 9.15)	93( 9.60)	46( 7.44)	122( 8.43)
16- 20	90( 9.74)	100( 12.89)	102( 10.53)	55( 8.90)	128( 8.84)
21- 25	69( 7.47)	87( 11.21)	103( 10.63)	35( 5.66)	81( 5.59)
26- 30	96( 10.39)	76( 9.79)	115( 11.87)	26( 4.21)	39( 2.69)
31- 35	77( 8.33)	60( 7.73)	92( 9.49)	36( 5.83)	49( 3.38)
36- 40	70( 7.58)	59( 7.60)	83( 8.57)	16( 2.59)	23( 1.59)
41- 45	60( 6.49)	59( 7.60)	48( 4.95)	23( 3.72)	17( 1.17)
46- 50	48( 5.19)	46( 5.93)	59( 6.09)	13( 2.10)	12( 0.83)
51- 55	39( 4.22)	37( 4.77)	44( 4.54)	10( 1.62)	11( 0.76)
56- 60	36( 3.90)	22( 2.84)	29( 2.99)	8( 1.29)	9( 0.62)
61- 65	38( 4.11)	19( 2.45)	24( 2.48)	4( 0.65)	7( 0.48)
66- 70	28( 3.03)	8( 1.03)	24( 2.48)	5( 0.81)	3( 0.21)
71- 75	21( 2.27)	8( 1.03)	10( 1.03)	4( 0.65)	4( 0.28)
76- 80	19( 2.06)	19( 2.45)	10( 1.03)	3( 0.49)	0( 0.00)
81- 85	21( 2.27)	3( 0.39)	14( 1.44)	4( 0.65)	2( 0.14)
86- 90	16( 1.73)	5( 0.64)	8( 0.83)	10( 1.62)	3( 0.21)
91- 95	8( 0.87)	1( 0.13)	7( 0.72)	2( 0.32)	2( 0.14)
96-100	11( 1.19)	7( 0.90)	5( 0.52)	2( 0.32)	0( 0.00)
101-105	9( 0.97)	2( 0.26)	2( 0.21)	2( 0.32)	0( 0.00)
106-110	8( 0.87)	4( 0.52)	1( 0.10)	3( 0.49)	0( 0.00)
111-115	2( 0.22)	5( 0.64)	4( 0.41)	2( 0.32)	0( 0.00)
116-120	3( 0.32)	2( 0.26)	2( 0.21)	1( 0.16)	0( 0.00)
121-	23( 2.49)	6( 0.77)	15( 1.55)	6( 0.97)	4( 0.28)
TOTAL	924(100.00)	776(100.00)	969(100.01)	618( 99.99)	1,448(100.00)
Average	43.06	34.71	36.28	22.55	13.13

Length	1TA	2TA	EDW	FAU
1- 5	109( 12.78)	98( 14.00)	289( 20.31)	290( 26.27)
6- 10	204( 23.92)	119( 17.00)	529( 37.17)	289( 26.18)
11- 15	114( 13.36)	61( 8.71)	136( 9.56)	138( 12.50)
16- 20	103( 12.08)	62( 8.86)	176( 12.37)	114( 10.33)
21- 25	78( 9.14)	68( 9.71)	74( 5.20)	84( 7.61)
26- 30	59( 6.92)	50( 7.14)	48( 3.37)	56( 5.07)
31- 35	51( 5.98)	58( 8.29)	54( 3.79)	34( 3.08)
36- 40	34( 3.99)	36( 5.14)	21( 1.48)	24( 2.17)
41- 45	26( 3.05)	28( 4.00)	21( 1.48)	10( 0.91)
46- 50	19( 2.23)	35( 5.00)	17( 1.19)	17( 1.54)
51- 55	13( 1.52)	17( 2.43)	15( 1.05)	12( 1.09)
56- 60	7( 0.82)	19( 2.71)	6( 0.42)	8( 0.72)
61- 65	6( 0.70)	14( 2.00)	10( 0.70)	10( 0.91)
66- 70	9( 1.06)	8( 1.14)	6( 0.42)	2( 0.18)
71- 75	4( 0.47)	5( 0.71)	4( 0.28)	5( 0.45)
76- 80	2( 0.23)	4( 0.57)	3( 0.21)	5( 0.45)
81- 85	5( 0.59)	3( 0.43)	5( 0.35)	2( 0.18)
86- 90	1( 0.12)	4( 0.57)	3( 0.21)	2( 0.18)
91- 95	0( 0.00)	2( 0.29)	1( 0.07)	1( 0.09)
96-100	1( 0.12)	1( 0.14)	1( 0.07)	0( 0.00)
101-105	0( 0.00)	0( 0.00)	1( 0.07)	0( 0.00)
106-110	1( 0.12)	1( 0.14)	0( 0.00)	0( 0.00)
111-115	2( 0.23)	0( 0.00)	1( 0.07)	0( 0.00)
116-120	0( 0.00)	3( 0.43)	0( 0.00)	0( 0.00)
121-	5( 0.59)	4( 0.57)	2( 0.14)	1( 0.09)
Total	853(100.02)	700( 99.98)	1,423( 99.98)	1,104(100.00)
Average	20.93	26.08	15.06	15.32

Length	TGV	TN	MAN	MV	1H4
1- 5	325( 27.92)	454( 32.24)	288( 23.36)	221( 21.48)	472( 30.18)
6- 10	355( 30.50)	318( 22.59)	362( 29.36)	235( 22.84)	336( 21.48)
11- 15	114( 9.79)	189( 13.42)	171( 13.87)	95( 9.23)	197( 12.60)
16- 20	135( 11.60)	127( 9.02)	103( 8.35)	111( 10.79)	164( 10.49)
21- 25	48( 4.12)	94( 6.68)	76( 6.16)	79( 7.68)	121( 7.74)
26- 30	40( 3.44)	68( 4.83)	50( 4.06)	60( 5.83)	68( 4.35)
31- 35	40( 3.44)	45( 3.20)	38( 3.08)	50( 4.86)	45( 2.88)
36- 40	27( 2.32)	29( 2.06)	30( 2.43)	42( 4.08)	34( 2.17)
41- 45	22( 1.89)	30( 2.13)	27( 2.19)	21( 2.04)	35( 2.24)
46- 50	15( 1.29)	17( 1.21)	18( 1.46)	20( 1.94)	12( 0.77)
51- 55	8( 0.69)	8( 0.57)	12( 0.97)	22( 2.14)	16( 1.02)
56- 60	7( 0.60)	7( 0.50)	10( 0.81)	16( 1.55)	13( 0.83)
61- 65	3( 0.26)	3( 0.21)	3( 0.24)	9( 0.87)	13( 0.83)
66- 70	6( 0.52)	8( 0.57)	5( 0.41)	10( 0.97)	10( 0.64)
71- 75	5( 0.43)	3( 0.21)	6( 0.49)	6( 0.58)	7( 0.45)
76- 80	1( 0.09)	4( 0.28)	4( 0.32)	6( 0.58)	3( 0.19)
81- 85	1( 0.09)	1( 0.07)	6( 0.49)	5( 0.49)	5( 0.32)
86- 90	4( 0.34)	0( 0.00)	4( 0.32)	5( 0.49)	3( 0.19)
91- 95	0( 0.00)	0( 0.00)	1( 0.08)	4( 0.39)	3( 0.19)
96-100	2( 0.17)	1( 0.07)	1( 0.08)	2( 0.19)	4( 0.26)
101-105	0( 0.00)	0( 0.00)	3( 0.24)	2( 0.19)	0( 0.00)
106-110	2( 0.17)	0( 0.00)	2( 0.16)	0( 0.00)	0( 0.00)
111-115	0( 0.00)	1( 0.07)	2( 0.16)	2( 0.19)	1( 0.06)
116-120	0( 0.00)	0( 0.00)	2( 0.16)	1( 0.10)	0( 0.00)
121-	4( 0.34)	1( 0.07)	9( 0.73)	5( 0.49)	2( 0.13)
Total	1,164(100.02)	1,408(100.00)	1,233(99.98)	1,029(99.99)	1,564(100.01)
Average	14.80	14.13	17.15	20.77	15.78

Length	PJB	GSM	GAB	PBR	DGL
1- 5	347( 15.10)	458( 21.02)	343( 23.77)	150( 23.11)	126( 27.51)
6- 10	576( 25.07)	577( 26.48)	490( 33.96)	206( 31.74)	155( 33.84)
11- 15	391( 17.01)	360( 16.52)	269( 18.64)	109( 16.80)	90( 19.65)
16- 20	299( 13.01)	264( 12.12)	139( 9.63)	71( 10.94)	38( 8.30)
21- 25	210( 9.14)	172( 7.89)	83( 5.75)	45( 6.93)	24( 5.24)
26- 30	125( 5.44)	104( 4.77)	48( 3.33)	26( 4.01)	10( 2.18)
31- 35	99( 4.31)	75( 3.44)	29( 2.01)	19( 2.93)	10( 2.18)
36- 40	86( 3.74)	56( 2.57)	18( 1.25)	9( 1.39)	2( 0.44)
41- 45	57( 2.48)	32( 1.47)	14( 0.97)	7( 1.08)	0( 0.00)
46- 50	38( 1.65)	22( 1.01)	3( 0.21)	2( 0.31)	1( 0.22)
51- 55	26( 1.13)	14( 0.64)	4( 0.28)	1( 0.15)	1( 0.22)
56- 60	14( 0.61)	12( 0.55)	2( 0.14)	1( 0.15)	0( 0.00)
61- 65	8( 0.35)	9( 0.41)	0( 0.00)	0( 0.00)	0( 0.00)
66- 70	5( 0.22)	10( 0.46)	0( 0.00)	1( 0.15)	0( 0.00)
71- 75	3( 0.13)	3( 0.14)	0( 0.00)	1( 0.15)	1( 0.22)
76- 80	2( 0.09)	2( 0.09)	0( 0.00)	0( 0.00)	0( 0.00)
81- 85	3( 0.13)	5( 0.23)	1( 0.07)	0( 0.00)	0( 0.00)
86- 90	3( 0.13)	2( 0.09)	0( 0.00)	1( 0.15)	0( 0.00)
91- 95	2( 0.09)	1( 0.05)	0( 0.00)	0( 0.00)	0( 0.00)
96-100	1( 0.04)	0( 0.00)	0( 0.00)	0( 0.00)	0( 0.00)
101-105	1( 0.04)	0( 0.00)	0( 0.00)	0( 0.00)	0( 0.00)
106-110	0( 0.00)	0( 0.00)	0( 0.00)	0( 0.00)	0( 0.00)
111-115	2( 0.09)	0( 0.00)	0( 0.00)	0( 0.00)	0( 0.00)
116-120	0( 0.00)	0( 0.00)	0( 0.00)	0( 0.00)	0( 0.00)
121-	0( 0.00)	1( 0.05)	0( 0.00)	0( 0.00)	0( 0.00)
Total	2,298(100.00)	2,179(100.00)	1,443(100.01)	649( 99.99)	458(100.00)
Average	17.30	15.22	11.79	12.65	10.67



表B 頻度の高い文(作品別)

GV1		GV2		GV3		DID		JEW	
Freq.	Len.	Freq.	Len.	Freq.	Len.	Freq.	Len.	Freq.	Len.
23(2.49)	16	25(3.22)	17	30(3.10)	26	58(9.39)	8	128(8.84)	8
22(2.38)	13	24(3.09)	27	30(3.10)	37	48(7.77)	7	119(8.22)	9
22(2.38)	26	21(2.71)	18	29(2.99)	23	44(7.12)	9	108(7.46)	7
22(2.38)	29	20(2.58)	21	26(2.68)	15	39(6.31)	6	101(6.98)	4
22(2.38)	31	19(2.45)	16	26(2.68)	27	25(4.05)	2	100(6.91)	1
20(2.16)	27	19(2.45)	19	25(2.58)	21	24(3.88)	1	95(6.56)	2
20(2.16)	30	19(2.45)	23	25(2.58)	32	22(3.56)	10	83(5.73)	6
20(2.16)	36	19(2.45)	28	23(2.37)	14	19(3.07)	16	71(4.90)	3
18(1.95)	18	19(2.45)	36	23(2.37)	18	19(3.07)	17	68(4.70)	5
18(1.95)	19	18(2.32)	25	22(2.27)	19	17(2.75)	4	59(4.07)	10
18(1.95)	24					17(2.75)	5		

1TA		2TA		EDW		FAU	
Freq.	Len.	Freq.	Len.	Freq.	Len.	Freq.	Len.
59(6.92)	8	30(4.29)	9	154(10.82)	9	86(7.79)	8
47(5.51)	7	29(4.14)	8	128(9.00)	8	78(7.07)	4
44(5.16)	9	28(4.00)	7	120(8.43)	7	72(6.52)	2
41(4.81)	15	28(4.00)	15	86(6.04)	4	66(5.98)	9
32(3.75)	6	26(3.71)	1	67(4.71)	16	53(4.80)	6
32(3.75)	14	23(3.29)	3	66(4.64)	6	51(4.62)	1
31(3.63)	16	22(3.14)	4	64(4.50)	2	51(4.62)	3
28(3.28)	1	22(3.14)	16	61(4.29)	10	49(4.44)	7
28(3.28)	4	21(3.00)	6	52(3.65)	3	40(3.62)	14
27(3.17)	17	18(2.57)	17	51(3.58)	15	38(3.44)	5

TGV		TN		MAN		MV		1 H 4	
Freq.	Len.	Freq.	Len.	Freq.	Len.	Freq.	Len.	Freq.	Len.
86(7.39)	4	131(9.30)	4	86(6.97)	6	69(6.71)	4	128(8.18)	4
76(6.53)	8	96(6.82)	3	83(6.73)	4	69(6.71)	8	107(6.84)	2
76(6.53)	9	88(6.25)	2	80(6.49)	9	56(5.44)	9	94(6.01)	5
75(6.44)	7	87(6.18)	5	78(6.33)	7	45(4.37)	2	80(5.12)	6
70(6.01)	6	74(5.26)	6	69(5.60)	8	42(4.08)	5	76(4.86)	9
68(5.84)	2	71(5.04)	8	57(4.62)	5	39(3.79)	7	74(4.73)	3
62(5.33)	5	60(4.26)	7	53(4.30)	1	37(3.60)	1	70(4.48)	8
58(4.98)	10	60(4.26)	9	52(4.22)	2	36(3.50)	10	69(4.41)	1
57(4.90)	1	53(3.76)	10	49(3.97)	10	35(3.40)	6	56(3.58)	10
52(4.47)	3	52(3.69)	1	43(3.49)	3	28(2.72)	3	54(3.45)	7

PJB		GSM		GAB		PBR		DGL	
Freq.	Len.	Freq.	Len.	Freq.	Len.	Frq.	Len.	Freq.	Len.
130(5.66)	7	131(6.01)	5	125(8.66)	6	49(7.55)	7	39(8.52)	4
129(5.61)	6	128(5.87)	9	101(7.00)	5	48(7.40)	8	36(7.86)	5
116(5.05)	5	123(5.64)	7	95(6.58)	10	45(6.93)	4	36(7.86)	7
116(5.05)	8	117(5.37)	6	94(6.51)	7	42(6.47)	6	34(7.42)	6
114(4.96)	9	115(5.28)	8	91(6.31)	4	41(6.32)	9	34(7.42)	8
93(4.05)	4	113(5.19)	4	90(6.24)	8	34(5.24)	5	27(5.90)	11
87(3.79)	10	95(4.36)	3	86(5.96)	9	30(4.62)	2	26(5.68)	9
86(3.74)	11	94(4.31)	10	73(5.06)	3	27(4.16)	3	25(5.46)	2
85(3.70)	13	87(3.99)	11	64(4.44)	11	27(4.16)	11	25(5.46)	10
76(3.31)	12	84(3.85)	12	57(3.95)	2	26(4.01)	10	22(4.80)	3

## Appendix II

表 A 語の長さ と 頻度 (作品別)

Len.	GV1 D.W.	GV1 T.W.	GV2 D.W.	GV2 T.W.	GV3 D.W.	GV3 T.W.
1	6( 0.11)	861( 2.16)	3( 0.07)	647( 2.40)	5( 0.10)	721( 2.05)
2	42( 0.76)	8,373( 21.04)	32( 0.75)	5,406( 20.07)	37( 0.73)	7,348( 20.90)
3	88( 1.59)	7,147( 17.96)	80( 1.88)	5,295( 19.66)	83( 1.64)	6,528( 18.57)
4	334( 6.02)	5,561( 13.98)	299( 7.01)	3,921( 14.56)	292( 5.78)	4,866( 13.84)
5	674( 12.14)	4,611( 11.59)	563( 13.20)	3,088( 11.47)	621( 12.29)	4,090( 11.63)
6	926( 16.68)	4,373( 10.99)	705( 16.53)	2,717( 10.09)	796( 15.75)	3,645( 10.37)
7	968( 17.44)	2,712( 6.82)	671( 15.73)	1,701( 6.32)	812( 16.07)	2,353( 6.69)
8	867( 15.62)	2,413( 6.06)	694( 16.27)	1,652( 6.13)	787( 15.57)	2,104( 5.98)
9	704( 12.68)	1,750( 4.40)	521( 12.21)	1,140( 4.23)	642( 12.71)	1,476( 4.20)
10	485( 8.74)	1,051( 2.64)	348( 8.16)	647( 2.40)	466( 9.22)	1,055( 3.00)
11	254( 4.57)	555( 1.39)	202( 4.74)	436( 1.62)	272( 5.38)	527( 1.50)
12	127( 2.29)	241( 0.61)	95( 2.23)	202( 0.75)	137( 2.71)	288( 0.82)
13	48( 0.86)	94( 0.24)	38( 0.89)	59( 0.22)	65( 1.29)	105( 0.30)
14	21( 0.38)	39( 0.10)	14( 0.33)	18( 0.07)	25( 0.49)	40( 0.11)
15	6( 0.11)	7( 0.02)	0( 0.00)	0( 0.00)	9( 0.18)	10( 0.03)
16	2( 0.04)	3( 0.01)	1( 0.02)	3( 0.01)	4( 0.08)	4( 0.01)
Tot. Av.	5,552(100.03) 7.36	39,791(100.01) 4.68	4,266(100.02) 7.27	26,932(100.00) 4.64	5,053( 99.99) 7.49	35,160(100.00) 4.70

Len.	DID D.W.	DID T.W.	JEW D.W.	JEW T.W.	1 TA D.W.	2 TA T.W.
1	4( 0.14)	537( 3.85)	8( 0.25)	823( 4.33)	4( 0.11)	434( 2.43)
2	31( 1.09)	2,058( 14.77)	41( 1.28)	3,259( 17.14)	34( 0.94)	2,967( 16.62)
3	133( 4.68)	2,915( 20.91)	171( 5.33)	4,402( 23.16)	151( 4.18)	3,651( 20.45)
4	361( 12.70)	2,957( 21.22)	429( 13.38)	3,900( 20.52)	415( 11.50)	3,336( 18.68)
5	592( 20.83)	2,148( 15.41)	693( 21.62)	2,635( 13.86)	642( 17.79)	2,460( 13.78)
6	606( 21.32)	1,528( 10.96)	698( 21.77)	1,512( 7.95)	750( 20.78)	1,765( 9.88)
7	487( 17.14)	827( 5.93)	473( 14.75)	1,148( 6.04)	638( 17.68)	1,261( 7.06)
8	347( 12.21)	579( 4.15)	300( 9.36)	630( 3.31)	446( 12.36)	880( 4.93)
9	165( 5.81)	255( 1.83)	216( 6.74)	432( 2.27)	260( 7.20)	573( 3.21)
10	85( 2.99)	99( 0.71)	110( 3.43)	172( 0.90)	162( 4.49)	301( 1.69)
11	21( 0.74)	22( 0.16)	43( 1.34)	60( 0.32)	59( 1.63)	172( 0.96)
12	5( 0.18)	8( 0.06)	16( 0.50)	19( 0.10)	37( 1.03)	41( 0.23)
13	3( 0.11)	3( 0.02)	6( 0.19)	10( 0.05)	9( 0.25)	13( 0.07)
14	1( 0.04)	1( 0.01)	2( 0.06)	7( 0.04)	2( 0.06)	2( 0.01)
15	1( 0.04)	1( 0.01)	0( 0.00)	0( 0.00)	0( 0.00)	0( 0.00)
16	0( 0.00)	0( 0.00)	0( 0.00)	0( 0.00)	0( 0.00)	0( 0.00)
Tot. Av.	2,842(100.02) 6.12	13,938(100.00) 4.25	3,206(100.00) 6.08	19,009( 99.99) 4.11	3,609(100.00) 6.41	17,856(100.00) 4.49

Len.	2 TA D.W.	2 TA T.W.	EDW D.W.	EDW T.W.	FAU D.W.	FAU T.W.
1	4( 0.11)	501( 2.74)	5( 0.15)	770( 3.59)	6( 0.18)	692( 4.09)
2	35( 0.98)	2,885( 15.80)	36( 1.06)	3,661( 17.08)	44( 1.35)	2,670( 15.78)
3	152( 4.25)	3,857( 21.12)	148( 4.35)	4,499( 20.99)	180( 5.52)	3,618( 21.39)
4	432( 12.09)	3,416( 18.71)	403( 11.85)	4,524( 21.10)	446( 13.67)	3,525( 20.84)
5	662( 18.53)	2,693( 14.75)	647( 19.03)	2,885( 13.46)	634( 19.43)	2,264( 13.38)
6	727( 20.35)	1,771( 9.70)	753( 22.15)	2,026( 9.45)	639( 19.58)	1,523( 9.00)
7	612( 17.13)	1,360( 7.45)	574( 16.88)	1,278( 5.96)	531( 16.27)	1,238( 7.32)
8	430( 12.03)	759( 4.16)	384( 11.29)	1,014( 4.73)	366( 11.22)	653( 3.86)
9	291( 8.14)	584( 3.20)	234( 6.88)	481( 2.24)	219( 6.71)	391( 2.31)
10	138( 3.86)	269( 1.47)	136( 4.00)	193( 0.90)	109( 3.34)	167( 0.99)
11	55( 1.54)	124( 0.68)	53( 1.56)	68( 0.32)	50( 1.53)	59( 0.35)
12	24( 0.67)	27( 0.15)	18( 0.53)	29( 0.14)	22( 0.67)	27( 0.16)
13	9( 0.25)	11( 0.06)	8( 0.24)	8( 0.04)	13( 0.40)	16( 0.09)
14	0( 0.00)	0( 0.00)	1( 0.03)	1( 0.00)	3( 0.09)	73( 0.43)
15	1( 0.03)	1( 0.01)	0( 0.00)	0( 0.00)	1( 0.03)	1( 0.01)
16	1( 0.03)	1( 0.01)	0( 0.00)	0( 0.00)	0( 0.00)	0( 0.00)
Tot.	3,573( 99.99)	18,259(100.01)	3,400(100.00)	21,437(100.00)	3,263( 99.99)	16,917(100.00)
Av.	6.35	4.44	6.28	4.24	6.18	4.30

Len.	TGV D.W.	TGV T.W.	TN D.W.	TN T.W.	MAN D.W.	MAN T.W.
1	7( 0.25)	948( 5.50)	12( 0.36)	1,118( 5.62)	6( 0.19)	1,244( 5.88)
2	35( 1.23)	2,867( 16.65)	50( 1.49)	3,602( 18.11)	33( 1.02)	3,618( 17.11)
3	158( 5.54)	3,902( 22.66)	175( 5.21)	4,589( 23.07)	163( 5.05)	5,004( 23.66)
4	381( 13.35)	3,785( 21.98)	421( 12.53)	4,097( 20.59)	437( 13.54)	4,477( 21.17)
5	606( 21.24)	2,215( 12.86)	699( 20.80)	2,577( 12.95)	679( 21.04)	2,611( 12.35)
6	560( 19.63)	1,486( 8.63)	672( 19.99)	1,684( 8.46)	616( 19.09)	1,609( 7.61)
7	457( 16.02)	889( 5.16)	546( 16.25)	996( 5.01)	508( 15.74)	1,183( 5.59)
8	281( 9.85)	505( 2.93)	347( 10.32)	578( 2.91)	325( 10.07)	685( 3.24)
9	184( 6.45)	389( 2.26)	200( 5.95)	346( 1.74)	223( 6.91)	417( 1.97)
10	114( 4.00)	151( 0.88)	119( 3.54)	154( 0.77)	135( 4.18)	181( 0.86)
11	46( 1.61)	56( 0.33)	78( 2.32)	103( 0.52)	62( 1.92)	75( 0.35)
12	15( 0.53)	21( 0.12)	26( 0.77)	33( 0.17)	30( 0.93)	34( 0.16)
13	5( 0.18)	5( 0.03)	13( 0.39)	14( 0.07)	9( 0.28)	11( 0.05)
14	2( 0.07)	2( 0.01)	1( 0.03)	1( 0.01)	0( 0.00)	0( 0.00)
15	2( 0.07)	2( 0.01)	2( 0.06)	2( 0.01)	1( 0.03)	1( 0.00)
16	0( 0.00)	0( 0.00)	0( 0.00)	0( 0.00)	0( 0.00)	0( 0.00)
Tot.	2,853(100.02)	17,223(100.01)	3,361(100.01)	19,894(100.01)	3,227( 99.99)	21,150(100.00)
Av.	6.14	4.05	6.19	3.99	6.22	4.01

Len.	MV D.W.	MV T.W.	1H4 D.W.	1H4 T.W.
1	6( 0.17)	1,144( 5.35)	7( 0.17)	1,258( 5.10)
2	35( 1.00)	3,742( 17.51)	38( 0.92)	4,240( 17.18)
3	172( 4.91)	4,824( 22.57)	181( 4.37)	5,415( 21.94)
4	432( 12.34)	4,262( 19.94)	508( 12.26)	5,051( 20.47)
5	683( 19.51)	2,847( 13.32)	840( 20.28)	3,544( 14.36)
6	748( 21.37)	1,868( 8.74)	868( 20.95)	2,130( 8.63)
7	577( 16.48)	1,186( 5.55)	685( 16.53)	1,380( 5.59)
8	396( 11.31)	800( 3.74)	432( 10.43)	716( 2.90)
9	225( 6.43)	410( 1.92)	281( 6.78)	519( 2.10)
10	131( 3.74)	187( 0.87)	161( 3.89)	217( 0.88)
11	59( 1.69)	67( 0.31)	90( 2.17)	136( 0.55)
12	29( 0.83)	30( 0.14)	29( 0.70)	36( 0.15)
13	6( 0.17)	6( 0.03)	12( 0.29)	19( 0.08)
14	1( 0.03)	1( 0.00)	8( 0.19)	17( 0.07)
15	1( 0.03)	2( 0.01)	2( 0.05)	2( 0.01)
16	0( 0.00)	0( 0.00)	1( 0.02)	1( 0.00)
Tot. Av.	3,501(100.01) 6.25	21,376(100.00) 4.07	4,143(100.00) 6.64	24,681(100.00) 4.05

Len.	PJB D.W.	PJB T.W.	GSM D.W.	GSM T.W.	GAB D.W.	GAB T.W.
1	11( 0.22)	1,143( 2.88)	7( 0.14)	1,000( 3.01)	2( 0.07)	564( 3.31)
2	47( 0.93)	6,182( 15.55)	49( 1.00)	5,415( 16.32)	40( 1.37)	2,689( 15.80)
3	202( 3.98)	9,808( 24.68)	188( 3.84)	8,076( 24.35)	145( 4.95)	4,422( 25.99)
4	579( 11.40)	8,072( 20.31)	576( 11.75)	6,428( 19.38)	447( 15.26)	3,322( 19.52)
5	776( 15.27)	4,861( 12.23)	699( 14.26)	3,625( 10.93)	485( 16.55)	1,907( 11.21)
6	868( 17.08)	3,608( 9.08)	786( 16.04)	2,825( 8.52)	496( 16.93)	1,517( 8.92)
7	841( 16.55)	2,529( 6.36)	809( 16.51)	2,237( 6.74)	469( 16.01)	1,218( 7.16)
8	656( 12.91)	1,505( 3.79)	649( 13.24)	1,518( 4.58)	337( 11.50)	615( 3.61)
9	464( 9.13)	991( 2.49)	475( 9.69)	944( 2.85)	239( 8.16)	387( 2.27)
10	318( 6.26)	550( 1.38)	310( 6.33)	593( 1.79)	138( 4.71)	218( 1.28)
11	169( 3.33)	269( 0.68)	177( 3.61)	258( 0.78)	68( 2.32)	86( 0.51)
12	95( 1.87)	145( 0.36)	92( 1.88)	146( 0.44)	43( 1.47)	49( 0.29)
13	39( 0.77)	56( 0.14)	58( 1.18)	77( 0.23)	15( 0.51)	16( 0.09)
14	13( 0.26)	18( 0.05)	18( 0.37)	21( 0.06)	5( 0.17)	5( 0.03)
15	3( 0.06)	7( 0.02)	5( 0.10)	6( 0.02)	1( 0.03)	1( 0.01)
16	0( 0.00)	0( 0.00)	3( 0.06)	4( 0.01)	0( 0.00)	0( 0.00)
Tot. Av.	5,081(100.02) 6.76	39,744(100.00) 4.31	4,901(100.00) 6.85	33,173(100.01) 4.37	2,930(100.01) 6.40	17,016(100.00) 4.24

Len.	PBR D.W.	PBR T.W.	DGL D.W.	DGL T.W.
1	3( 0.17)	468( 5.70)	3( 0.27)	285( 5.83)
2	30( 1.66)	1,482( 18.04)	30( 2.69)	800( 16.37)
3	113( 6.24)	1,764( 21.48)	86( 7.72)	1,243( 25.43)
4	303( 16.73)	1,586( 19.31)	213( 19.12)	973( 19.91)
5	316( 17.45)	953( 11.60)	182( 16.34)	500( 10.23)
6	290( 16.01)	748( 9.11)	173( 15.53)	345( 7.06)
7	273( 15.07)	473( 5.76)	163( 14.63)	335( 6.85)
8	210( 11.60)	388( 4.72)	111( 9.96)	195( 3.99)
9	122( 6.74)	173( 2.11)	53( 4.76)	69( 1.41)
10	81( 4.47)	104( 1.27)	47( 4.22)	70( 1.43)
11	42( 2.32)	44( 0.54)	26( 2.33)	38( 0.78)
12	11( 0.61)	11( 0.13)	14( 1.26)	18( 0.37)
13	6( 0.33)	8( 0.10)	8( 0.72)	12( 0.25)
14	8( 0.44)	8( 0.10)	2( 0.18)	2( 0.04)
15	2( 0.11)	2( 0.02)	1( 0.09)	1( 0.02)
16	1( 0.06)	1( 0.01)	2( 0.18)	2( 0.04)
Tot.	1,811(100.01)	8,213(100.00)	1,114(100.00)	4,888(100.01)
Av.	6.22	4.17	6.04	4.13

## Appendix III Word List

Word	ELT	MRL	SHK	SPK	Word	ELT	MRL	SHK	SPK
the	4908	3865	3150	5958	tyme	239	0	0	0
and	4345	4162	2964	2544	there	239	160	257	252
of	4050	2437	2087	2660	some	236	116	189	149
to	3087	2395	2226	2902	suche	234	0	0	0
in	2488	1366	1398	1871	shulde	231	0	0	0
that	2204	1378	1351	937	unto	225	141	0	1
be	1812	888	857	458	vertue	224	13	25	0
a	1549	1354	2213	2221	an	224	113	244	299
is	1494	788	1246	410	moste	211	0	0	0
he	1340	657	709	1072	well	208	133	330	156
his	1233	985	667	577	whan	206	0	0	0
or	1192	341	243	230	great	205	122	78	57
as	1139	736	793	635	beinge	203	0	0	0
it	938	664	1094	1145	onely	197	31	42	0
whiche	913	0	0	0	nowe	188	1	0	0
by	841	409	475	437	wolde	183	0	0	0
for	831	1006	1071	841	who	182	112	147	319
with	816	1244	907	904	which	182	204	229	317
they	769	306	263	519	people	182	11	13	74
all	760	545	403	358	god	167	110	107	15
nat	754	0	0	0	after	153	36	53	171
but	724	790	844	514	on	144	360	376	851
I	649	2170	3185	1218	kynge	143	0	0	0
was	616	196	249	1909	without	139	25	54	45
them	585	279	239	237	selfe	137	119	202	0
other	551	68	74	131	weale	136	2	0	0
haue	495	0	758	0	mought	135	1	0	0
their	493	417	133	266	Wherfore	132	2	3	0
also	447	1	6	30	out	132	126	183	264
man	420	130	324	129	excellent	127	8	32	3
this	415	813	770	453	parte	126	0	0	0
hym	406	0	0	0	euery	126	0	58	0
so	405	549	652	274	made	125	128	82	99
more	372	225	272	193	wise	124	14	35	5
men	367	153	106	62	my	124	1904	1682	356
were	359	156	211	506	howe	124	0	0	0

any	350	75	134	96	will	122	619	705	91
noble	345	35	46	3	publike	122	1	1	0
than	338	94	21	121	yet	120	222	195	46
him	331	653	669	319	citie	118	23	1	0
do	331	138	234	178	thinge	116	0	0	0
moche	324	0	0	0	me	116	1032	1103	253
if	321	337	554	241	wyll	115	0	0	0
may	317	285	196	26	owne	114	39	80	0
shall	306	589	432	39	can	114	144	176	64
at	306	326	313	729	before	114	81	76	130
one	286	147	249	341	amonge	114	0	0	0
had	286	135	180	1165	ye	112	90	39	7
called	265	1	1	65	whom	110	65	38	42
where	262	190	131	116	many	110	51	67	53
what	260	484	575	278	longe	110	0	0	0
no	242	361	544	258	wherin	104	4	1	0
good	241	128	370	87	we	104	537	265	216
from	241	422	269	471	thou	103	725	705	5