人文科学とコンピュータ　36－6
（1997．11．15）

パスカル全データベース作成と言語解析 (I)

白石修二[†]　輪田裕[‡]　藤村丞[†]

福岡大学 理学部 応用数学科[†]
福岡大学 人文学部 フランス語学科[‡]

メナール版パスカル全集を底本として第2巻と第3巻をデータベース化した。言語解析の基礎のために、またテキストデータベースの堅牢さの向上のために作品ごとの単語の頻度データブックを作成した。頻度データブックの作成を受けてはじめて、パスカル研究が始まったところであり、結論を急ぐべきではないので決定的なことは言えないが、パスカルの書簡にみる人称代名詞に関して若干の発見があり、コメントを付け加えた。

Construction of the Database on the Complete Works of Blaise Pascal and Linguistic Analysis (I)

SHIRAISHI Shuji[†]　WADA Yutaka[‡]　FUJIMURA Shou[†]

Department of Applied Mathematics, Faculty of Science, Fukuoka University[†]
Department of French Language, Faculty of Humanities, Fukuoka University[‡]

We constructed a database of the complete works of Blaise Pascal, on the second and third volume of the Jean Mesnard version as definitive version. We wrote up word frequency books of each work as a solid basis for linguistic analysis and to improve the completenesse of text database. Finally, we present a few discoveries and comments on personal pronoun use in Letters by Pascal, although it is too early to jump to decissive conclusions because we have just started a study of Pascal being based on word frequency books.

1　はじめに

フランス文学という分野は、世界の文学史の中でも確固たる位置を築き、その思想、形態、もろもろにおいてもまた格別の多大な功績をあげてきた文学の一つである。当然わが国においても文学研究者のうち、フランス文学・哲学研究者の数は相当人数にのぼっている。その中でも、<パスカル>その人とその作品群は研究対象として圧倒的な興味を持たれ、研究されている。しかしながらパスカルの作品の特徴ともいえる、文学のみにとどまらない科学的興味（例えば数学、物理、エトセトラ）までをカバーする研究というのはその多岐に渡る知識の多彩さ故に現在まで部分研究というものはあっても包括的なものはないといってもいい状況である。その全作品をデータベース化することこそはこの研究の最大最優先の目的である。このことはパスカル研究者に取っては大変な意義のあることであると思われる。

2　パスカルデータベースの作成

メナール版パスカル全集 [1] を底本とし、パスカル自身が書いた（あるいは書いたと推定されている部分）の第2巻と第3巻を電子テキスト化しテキストデータベース管理システム SIGMA で言語解析等を行っている。SIGMA を用いた種々の言語解析により貴重なデータが得られつつある。1997年度中には現在刊行されている第1巻から4巻までを全て電子化する予定である。電子化する上で生じてくる手順や技法上の問題点及び解決法については報告集で詳しくまとめることにしている。

3　頻度データブック作成

第3巻までのパスカルデータベースを作成し、作品ごとの頻度データブックや各単語の末尾からひくいわゆる逆引き辞書を含む、パスカルデータ資料を作り終えたところである [2]。頻度データブック等は言語解析の基礎であるのみならず、テキストデータベースの堅牢性をより確かなものとするために不可欠のものである。頻度データブックの内容として、以下のものを作成中である。

1. 作品ごとの単語の頻度データ (高頻度順, アルファベット順等)
2. 全作品中の単語の頻度データ (高頻度順, アルファベット順等)
3. 単語の逆引き辞書作成　（作品ごと、もしくは全体で）
4. 熟語及び慣用句の頻度データ（作品ごと、もしくは全体で）
5. その他

4　頻度データブックから得られる若干のコメント

－パスカルの書簡にみる人称代名詞－

長期に亘るデータベース作成によって、現在ようやく頻度データブックの回収が可能となり、現段階ではようやく研究が始まったばかりで何も決定的なことは言えないが、以下の観点から推測できる、注目すべき点についての若干のコメントをもって、報告に代えることにしたい。

4.1　研究の意図

パスカルにおける書簡の位置、あるいは価値について言えば、その主著である『パンセ』の執筆形態の決定と関係がある。彼がその作品をどのような形態で発表しようとしていたのか、つまり、論文形式なのか、手紙形式なのか、あるいは談話形式なのか、実はまだ定説がない。

他方、『パンセ』以外のテキストの多くは上記の三つの形式のどれかである。その中には、もう一つの主著である『プロヴァンシャル』が含まれているが、これは書簡体文学の古典主義における傑作との評価が定着している。しかし、文体論的観点からは、それに先だって執筆された『恩寵文書』が示唆的であろう。これは、人間の自由意志と神の意志である恩寵との関係について書かれたものであるが、その形式は三つに分かれており、手紙、談話、論文の３種類の文体が追求されている。しがって、少くとも手紙形式が『パンセ』の形式の一つとして考慮されていた可能性がある、と考えても不自然ではないだろう。

そこで、手紙形式で書かれた全体を対象として、その文体的特徴を捉え、それをもって、『パンセ』と比較することが一つの方法だと、言えるだろう。以上が、書簡を対象とする理由である。

次ぎに、人称代名詞についてであるが、これにはそれほどしっかりした理由がないのだが、しいて挙げるなら、仮に『パンセ』が書簡体であるならば、そこに散見される「私」という代名詞が誰を指すのか、これは作品解釈の要となるだろう。また、『パンセ』の人称代名詞の研究がすでに先行の研究者によって、ある程度行われており、その成果と、『パンセ』以外のテキストとの比較が容易であることも挙げておきたい。

以上のような観点から、若干の分析とコメントをしたい。ただし、まだ始めたばかりであるため、資料体も十分とは言えず、また、人称代名詞といっても、せいぜい、je(j'), me(m'), moi, / tu, te, toi,/ nous, / on, / vous に限定している。

4.2 若干の分析

パスカルの書簡を分類すると、おおむね５つのグループに分類できる。

1. 家族 (姉、姉夫婦、義兄) あての手紙。
2. シャルロットへの手紙。
3. 論争的手紙。
4. 学問上の手紙。
5. 献辞の手紙。

この姉あるいは姉夫婦にあてた手紙における人称代名詞を見てみると、姉のみにあてて書かれた手紙と、そうでない手紙では顕著な差が見られる。(ただし、1648.4.1. の手紙はその特殊性を直ちには説明できないので対象から除外している。つまり je, tu が極端に少く、また、vous は 0 で、その代わり、nous が 49 という、かなり特異な数字なので、これについては、いずれ考察することにしたい。)

1. 姉にのみ書かれた手紙では、tu, te, t' が使用され、
   もう一方は vous が使用されている。
2. 姉にのみ書かれた手紙では、je : tu = 3 : 1
   義兄＋姉夫婦では、je : vous = 2 : 1
3. 姉にのみ書かれた手紙では、je : nous = 10 : 1
   義兄＋姉夫婦では、je : nous = 1 : 2
4. 姉にのみ書かれた手紙では, on の使用は少い (3)
   義兄＋姉夫婦では、on の使用は比較的多い (13)

さて、このごく大雑把な数字を次ぎの二つの手紙ないしテキストと比較する。

- 『通知』。これは計算機を大法官セギエに献呈する献辞に付された、読者にその機能を説明するテキスト。
- シャルロットへの手紙。

1. 『通知』では、ほぼ tu, te, t' が使用され、もう一方は vous が使用されている。
2. 『通知』では、je : tu = 3 : 1 シャルロットへの手紙では、je : vous = 2 : 1
3. 『通知』では、je : nous = 132 : 1 シャルロットへの手紙では、je : nous = 3 : 1
4. 『通知』では、on の使用は少い (10) シャルロットへの手紙では、on の使用は比較的多い (36)

こうした数字から、姉のみに宛てた手紙と『通知』に共通性が指摘できる。同様に、義兄＋姉夫婦に宛てた手紙とシャルロットへの手紙には共通性が指摘できる。

4.3 若干の考察

1. 姉のみに宛てた手紙には、パスカルが姉に寄せる家族的親密性を指摘できる。それは単に je/tu ばかりではなく、婉曲な「私」を意味する on の少さからも指摘できるだろう。この特徴を『通知』も共有している。パスカルは計算機の使用方法を読むであろう読者にあてて書いているのだが、彼は、その読者を<< ami lecteur >> ;(「友人なる読者よ」) と呼びかけ、親しい友人に擬しているのである。こうした語り口は読者ひとりひとりに直接呼びかけ、その注意を喚起していると言えるだろう。敢えて言えば、呼び込みの口上、あるいは広告の手法と言えるのではなかろうか。計算機の使用方法とその機能の素晴らしさを訴え、多くの人々がこれに関心を寄せ、それを手に入れるように誘う手法である、と言えるだろう。

2. これと逆のことが、もう一方に指摘できるだろう。すなわち、率直というより、遠慮。個人的というより、一般的。シャルロットは、パスカル研究者から、パスカルの恋人と目されたこともあった。メナールは、身分の差からそうした推測を否定しているが、少くとも、義理の兄に対する程度の親密さと、それと裏腹の遠慮が介在するような関係であったのではないだろうか。

5 おわりに

1997年度重点領域「人文コンピュータ」に採用されたことによって、メナール版パスカル全集を底本とし、パスカル自身が書いた（あるいは書いたと推定されている部分）の第4巻までのデータベース化が強力に推進され、今なお現在形で進行中である。それと平行してSIGMAシステムを用いた種々の言語解析、あらゆる数量化調査も行っており、このことによって得られる様々なデータは、貴重であり、その様々なデータを読みとっていくことによって　、新たな研究分野の開拓、解釈を促進する事にもつながるものと予想される。このようにパスカル全集をデータベース化することは、その研究において飛躍的な結果をもたらす可能性を多分に秘めているといえる。またそのことと同時に、データベース化を進めていく過程において生じると予想される様々な SIGMA システムにおけるテクニカルな問題は、データベース管理システム SIGMA の改良、改善にとってもその一助となるものであると期待される。

今回すでに貴重なデータと解析結果が得られたことによってこれまでの研究者がまさしく手作業でこなしていた部分、あるいは手作業では不可能であった部分の諸問題が発見でき、さらなる研究上の可能性が広がることは、必須である。この貴重なデータおよび解析結果は学術図書または CD-ROM の形で出版し、フランス文学研究者等に供したい。

6 補足: SIGMA システム概要と電子化テキスト例

SIGMA システムは1981年より、九州大学大型計算機センターに公開され、多くの研究者から利用されているテキストデータベース管理システムである。公開されているフ

ルテキストデータベースの例として、「トーマスマンファイル」、「ゲーテファイル」、「昆虫学データベース」等がある [4]。SIGMA はデータ (ファイル) を一次元の文字列として管理し、ファイル全体を一度だけ先頭から一字づつ走査する一方向逐次処理に基本をおいているシステムで、大規模なテキストを高速に検索することができる。SIGMA システムの詳しい概要や使用法に関しては文献 [3] を参照されたい。

次は、パスカルデータベースの電子テキストサンプルである。電子テキストには文頭に、区切り語 # と同時に原著での位置が分かるようにボリューム番号 (巻数) とページ数、行数が記されている。例えば、#2023107 は第 2 巻 231 頁の 7 行目を意味している。また、パラグラフの区切りとして、記号 @ を入れているので、文単位以外にパラグラフ単位での検索ができるようになっている。

```
@#2023101 de'finition premie're
@#2023102 quand plusieurs lignes droites concourent a' me^me point,
ou sont toutes paralle'les entre elles,toutes ces lignes sont dites
de me^me ordre ou de me^me ordonnance,et la multitude de ces
lignes est dite ordre de lignes,ou ordonnance de lignes.
@#2023106 de'finition 2
#2023107 par le mot de section de co^ne nous entendons la circonfe'rence
du cercle,l'ellipse,l'hyperbole,la parabole et l'angle rectiligne,
d'autant qu'un co^ne coupe' paralle'lement a' sa base,ou par son
sommet,ou des trois autres sens qui engendrent l'ellipse,l'hyperbole
et la parabole,engendre dans la superficie conique,ou la
circonfe'rence d'un cercle,ou un angle,ou l'ellipse,ou l'hyperbole,
ou la parabole.
@#2023114 de'finition 3
#2023115 par le mot de droite mis seul,nous entendons ligne droite.
```

参考文献

[1] Blaise Pascal, Œuvres complète, DDB.

[2] 白石, 輪田, 柴田, パスカルデータブック 1 (近刊)

[3] 有川ほか, テキストデータベース管理システム SIGMA 第 2 版について, 九大大型計算機センター広報, Vol.20, No.6 (1987)

[4] 樋口, 篠原, 「テキストデータベーストーマスマンファイル」の完成と再編成について, 九大大型計算機センター広報, Vol.20, No.6 (1987)