

# 愛、名誉、それとも栄光？

——コンピュータ処理による

フランス古典主義演劇テキスト研究の試み 1——

永井典克

この小論は、コンピュータ処理によるフランス古典主義演劇テキスト研究へのアプローチを試みたものである。文学研究に何故コンピュータ処理を持ち出すのかという問いかけには、一人の研究者が一生かかっても読み切ることができない文献をコンピュータは瞬時に処理することができるからと返答することができるだろう。例えば、ある作家、ある作品にある単語が特徴的であるという結論を出すためには、他の作家、他の作品との比較が不可欠である。今回、Frantext というテキスト・データベースを利用して 5000 冊以上のテキストに現れる単語を調査した。これが Google ブックスを利用すれば、利用可能なテキストは 3000 万冊以上に増える。このような大量のテキストの処理はコンピュータが最も得意とするものである。コンピュータによる大量のテキストの処理により、論の客観性はより確かなものとなるであろう。

インターネット、コンピュータの世界は日進月歩で変化しているため、この小論は 2017 年の時点で、インターネット、コンピュータを用いた文学研究は何ができるのかの記録でしかない。

## 1 語彙索引 (Concordance) 作成の夢

ある文学者、もしくはある文学テキストの研究は、その文学者が用いた、もしくはそのテキスト内に登場する単語の分析と切り離すことが出来ない。多くの文学研究は、単語分析から始まると言っても過言ではない。

コンピュータを用い、1968 年に初めて 17 世紀フランスの演劇作家

ジャン・ラシーヌの『語彙索引』<sup>1</sup>を作成したフリーマンは、「作家にはそれぞれ偏愛する単語というものがある。その単語は、言説の中で繰り返し現れ、それを使う者の隠された願望や弱点を暴露してしまう」というサント＝ブーヴの言葉を引用し<sup>2</sup>、単語分析の重要性を強調する。そして、この単語分析は単なる印象から生じるものであってはならない。フリーマンは「ある作家のある表現のすべての用法を見つけ、それについて確信をもって話す」ためには、正確な語彙索引（Concordance）の作成が必要であると続ける。

しかし、この作業は簡単なものではなかった。語彙索引は、通常、興味深い単語があった時に、手書きで各行をカードに書き写し、参照番号を振り、その結果生じた無数のカードをアルファベット順に分類することで作られるものであった<sup>3</sup>。

この方法では、ある作家、もしくは——比較的短い作品を除き——作品の包括的な語彙索引作成は不可能に近い。包括的な語彙索引作成は文学研究者にとっては願望・夢でしかなかった。語彙索引作成が可能だった比較的短い作品としては、詩作品や演劇作品が挙げられる。古典主義期の演劇作品の単語数は1万9千程度でしかない。実際、フリーマンによるラシーヌの『語彙索引』出版の2年前、フランスでケマダによりラシーヌの悲劇『フェードル』のコンピュータを利用せずに作成した『語彙索引』<sup>4</sup>が出版されている。しかし、単語数は1万9千程度でしかないと言っても、全てを人力で行うには無理がある。そのため、ケマダは極めてユニークな仕掛けを作り出していた。テキストはまずタイプにより穿孔テープ（パンチテープ）化され、このテープから機械的に2種類の異なる穿孔されたカードが作り出される。そして、自動分類装置が機械的にこの穴あきカードを文法上のカテゴリーに沿って分類したのである<sup>5</sup>。

このような状況を変え、包括的な語彙索引作成を可能にしたのがコンピュータの到来であった。フリーマンは、コンピュータにより語彙索引作成のための労力は大いに軽減された<sup>6</sup>とするが、彼が初めて作成に取

り掛かった時の苦労はケマダのそれと同じくらい大変なものであった。彼はまず、ケマダと同じようにラシーヌのテキストをタイプにより鑽孔テープ（パンチテープ）化した。だが、ケマダとは異なり、彼はその鑽孔テープのデータをコンピュータが解析することができる磁気テープに移したのである。当時、コンピュータの情報は磁気テープに保存されていたが、ラシーヌのテキストは磁気テープにして 2000 メートルの長さにとんだという。この膨大な準備作業が終わってしまえば、ここからの作業はすべてコンピュータ上で行われた。今日では（後述するようにテキストデータが存在するならば）、どの単語がどれだけ使われたかは、コンピュータを使えば一瞬で分析が終わる。では、今現在、コンピュータ処理により、文学研究は何ができるようになっているのであろうか。

## 2 Frantext を用いた分析

勿論、今日では語彙索引作成以上のことがコンピュータ上で行える。

この章では、フランテキスト Frantext<sup>7</sup> というテキスト・データベースを用いて、どのような分析が可能なのかを見ることにしよう。

Frantext は ATILF-CNRS (Analyse et Traitement Informatique de la Langue Française - Centre national de la recherche scientifique フランス語の解析とコンピュータ処理・フランス国立科学研究センター) 旧 INaLF (Institut National de la langue française 国立フランス語研究所) により開発されたテキスト・データベースである。70 年代に『国語大辞典』TLF (*Trésor de la langue française*) の編纂のために約 1000 のテキストを集めることから始められたこのデータベースは、辞書編纂後は CD を媒体とした Discotext、1998 年からはウェブで参照できるデータベース Frantext へと姿を変えていく。

もともと辞書編纂が目的だったこのデータベースは中規模のデータベースであり、包括的であることを目標としていない。あくまでフランス語の適切なサンプルを提供することを目標としている。しかし、そうは言っても、2016 年 12 月時点で 5116 点のテキストを有する Frantext は、

フランス語文献に関しては現時点で世界最大級のテキスト・データベースであろう。

Frantext では、まず、ある作家、もしくはある作品の中に、ある単語（表現）がどのくらい出現するか「生起」（occurrence）のリストを作成することができる。が、それ以上に文学研究に役に立つのが、ある単語（表現）が、別のある単語（表現）と関係があるかないかを、その単語（表現）間の距離を測り確かめることができる「共起」（cooccurrence）のリストを作成することができる点である。このようなリスト作成は極めて困難なことであったため、過去においては、こうした索引を作ることで自体が研究として認められ、評価を受けていたのである<sup>8</sup>。

## 2-1 「愛」、「名誉」、それとも「栄光」？

Frantext を用いた分析対象として、17 世紀の劇作家 4 名、悲劇作家ピエール・コルネイユ、同じく悲劇作家のジャン・ラシーヌ、喜劇作家のモリエール、オペラ台本家のキノーらの作品と、18 世紀の哲学者で演劇も著しているヴォルテールの演劇作品を取り上げることにする。

コルネイユは 17 世紀前半の 1630 年代から世紀の半ばにかけて、ラシーヌとモリエールは 17 世紀の後半に、キノーは 17 世紀末に活躍した作家である。

更に、彼らのテキストに現れる特徴が時代的要因によるものかどうかを調べるために 17 世紀、18 世紀、19 世紀、20 世紀の Frantext に登録されている全テキストを同時に参照することにする。調査対象のテキスト数はそれぞれコルネイユ 34 本、ラシーヌ 11 本、モリエール 33 本、キノー 11 本、ヴォルテール 11 本、参照するテキスト数は 17 世紀 605 本、18 世紀 585 本、19 世紀 1096 本、20 世紀 2102 本である（2016 年 12 月 28 日参照）。

まず、テキスト・データベースを用いた彼らの作品の分析を単語の生起リスト作成から行う。

ここでは有名なコルネイユのディレンマ *dilemme cornélien* を出発点と

しよう。

文学史的に定義するならば、コルネイユのディレンマは、主人公たちが「理性に由来する「名誉」 *honneur* と、非論理的な情念に由来する「愛」 *amour* の二つの衝動によって悩まされること<sup>9)</sup>」である。コルネイユの悲劇では、このディレンマの解決策は一つしかない。「愛を得るためには、名誉を無傷のままで保っておく必要がある」ということだ。

一方、17世紀末に国王ルイ14世のための音楽付き悲劇（オペラ）を音楽家リュリと作り続け、人気を博したキノーの作品では、「愛」と「栄光」 *gloire* が対立軸として現れる。国王のための音楽付き悲劇では、「愛」と「栄光」に引き裂かれるが最後には「栄光」を選ぶ国王が主役として登場する。「愛」か「栄光」かという議論は、ルイ14世の宮廷において最も好まれたものの一つであった<sup>10)</sup>。

つまり、17世紀前半のコルネイユの作品では「愛」か「名誉」かが問題になり、17世紀末のキノーの作品では「愛」か「栄光」かが問題にされていたというのが、文学史上の定説である。この二つのディレンマの関係はどのようなものであったのだろうかというのが、今回の調査で明らかにしたいことである。

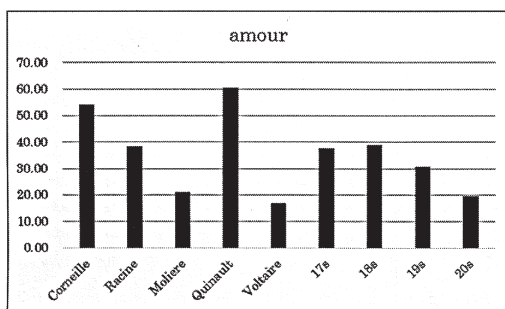
まず、それぞれの作家達と、17世紀、18世紀、19世紀、20世紀のFrantextに現時点で登録されている全テキストにおいて、「愛」、「名誉」、「栄光」が生起した総数と、作品ごとの平均生起回数を調べてみた。

	Corneille	Racine	Moliere	Quinault	Voltaire		17s	18s	19s	20s
総単語数	666878	210729	514588	122023	192775		24203239	36178109	76007436	125343556
作品数	34	11	33	11	11		605	585	1096	2102
amour	1844	422	694	666	185		22842	22832	33849	41242
honneur	465	80	313	22	98		9626	10589	13024	10421
gloire	579	213	146	180	115		9343	7341	8154	6705

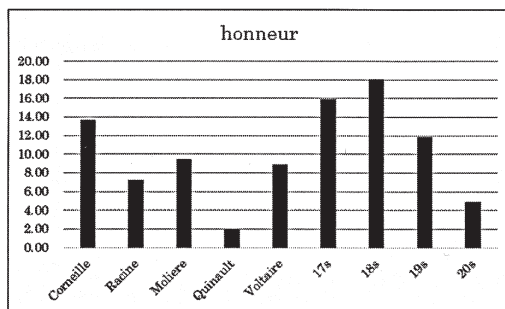
表1 「愛」「名誉」「栄光」生起回数

	Corneille	Racine	Moliere	Quinault	Voltaire		17s	18s	19s	20s
amour	54.24	38.36	21.03	60.55	16.82		37.76	39.03	30.88	19.62
honneur	13.68	7.27	9.48	2.00	8.91		15.91	18.10	11.88	4.96
gloire	17.03	19.36	4.42	16.36	10.45		15.44	12.55	7.44	3.19

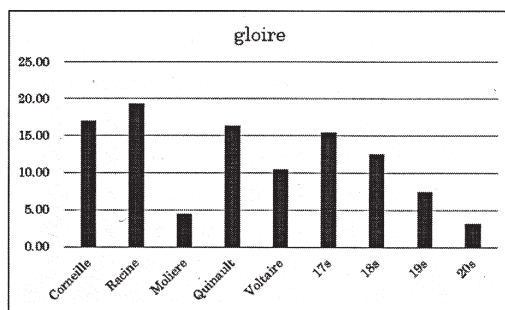
表2 「愛」「名誉」「栄光」平均生起回数



グラフ 1 「愛」 amour の作品当たりの登場頻度



グラフ 2 「名誉」 honneur の作品当たりの登場頻度



グラフ 3 「栄光」 gloire の作品当たりの登場頻度

この生起回数の調査結果から、以下のことが読み取れる。

まず「愛」という単語については、コルネイユ、キノーともに平均よりも使用回数が多いが、喜劇作家モリエールの作品ではそれほど使用されていない。「愛」という単語は喜劇よりも悲劇で使われたと考えられる。18 世紀啓蒙思想家ヴォルテールの演劇作品でも使用回数が減っている点も興味深い。

次に「名誉」という単語に関して、コルネイユ、ラシーヌ、モリエール、キノー、ヴォルテールの 5 者では、1 作品あたりの使用回数平均は確かにコルネイユが一番多い。対して、キノーは「名誉」に全く注意を払っていない。だが、コルネイユの「名誉」の使用回数を、17 世紀、18 世紀の平均が上回っている点に注意したい。コルネイユと言えば「愛と名誉」のディレンマがすぐに思い出される作家だが、実は、他の作家と比較すると、それほどまでに「名誉」という単語を使っていないことが明らかになる。

最後に「栄光」という単語に関しては、コルネイユ、ラシーヌ、キノーでは使用回数にあまり差は見られない。3 者とも 17 世紀平均の数値とほぼ同じであり、この時代の特徴を表していると言える。しかし、モリエール作品ではほとんど使われておらず、この喜劇作家の舞台では「栄光」は重要でなかったことが分かる。

また「名誉」も「栄光」も 19 世紀から 20 世紀にかけて使用頻度は激減していることが見て取れる。「名誉」、「栄光」は 17 世紀から 18 世紀、つまりはフランス絶対王政期に特有の概念であったのではないかという仮説がここでは立てられるだろう。

## 2-2 「愛」、「名誉」、「栄光」の共起回数の調査

「愛」、「名誉」、「栄光」というそれぞれの単語の生起回数の調査では、オペラ作家キノーの作品において「名誉」が重要でなかったことが判明しただけで、他の単語に関してはそれほど情報が得られなかった。しかし、「愛」と「名誉」、「愛」と「栄光」の共起回数を調べてみると、異

なる姿が見えてくる。

次の表はどちらも、二つの単語が前後を問わずに 20 単語以内に現れる回数を、Frantext を用いて出力し、まとめたものである。

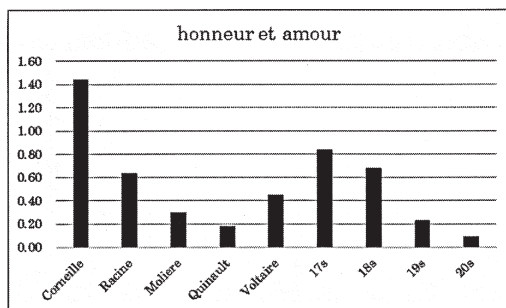
	Cornelle	Racine	Moliere	Quinault	Voltaire		17s	18s	19s	20s
amour et honneur	49	7	10	2	5		508	399	252	194
amour et gloire	67	23	8	49	7		688	609	441	269

表3 「愛」と「名誉」、「愛」と「栄光」共起回数

	Cornelle	Racine	Moliere	Quinault	Voltaire		17s	18s	19s	20s
amour et honneur	1.44	0.64	0.30	0.18	0.45	-	0.84	0.68	0.23	0.09
amour et gloire	1.97	2.09	0.24	4.45	0.64		1.14	1.04	0.40	0.13

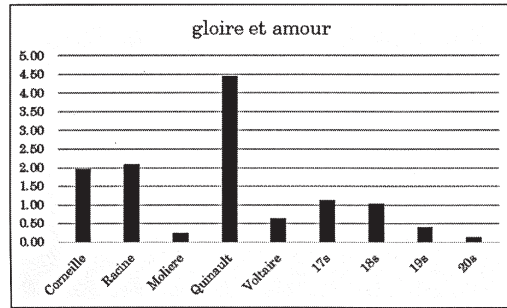
表4 「愛」と「名誉」、「愛」と「栄光」平均共起回数

グラフにしてみると分かりやすい（グラフ4、グラフ5）。「名誉」、「愛」の使用頻度は他の作家と変わらなかったコルネイユだが、「名誉」と「愛」の2単語が20単語以内に共起する回数は群を抜いて高くなっている。対して17世紀末のオペラ作家キノーのほうは、「名誉」と「愛」の共起回数は低く、「栄光」と「愛」の共起回数が他の作家と比べても極めて高い。



グラフ4 名誉と愛が共起する頻度





グラフ5 栄光と愛が共起する頻度

この結果からコルネイユ作品では「愛」と「名誉」が、キノー作品では「愛」と「栄光」が結び付けられて捉えられていることが多いということ、すなわち文学史上の定説の正しさが改めて示されたと言える。

ところで、コルネイユ作品と一口で言っても、彼は悲劇しか書かなかったわけではない。彼の全ての作品において、「愛」と「名誉」が問題となっているのであろうか。ジョルジュ・フォレストイエは、この「愛か名誉か」という「ディレンマはコルネイユの悲劇（少なくとも『ル・シッド』 *Le Cid* から『ポリュクト』 *Polyeucte* までの4大悲劇）に特徴的であるとされているが、コルネイユ悲劇の本質的なディレンマとずれが生じている」と指摘している<sup>11</sup>。フォレストイエによれば、コルネイユのディレンマの本質は、英雄に選択肢がないような状況を指すものであり、ル・シッドに見られる有名なディレンマは悲劇ではなく、悲喜劇に特有のものでしかない。ここで悲喜劇 *tragi-comédie* とは、17世紀前半に流行した、悲劇的なテーマであるが結末は幸福なものとなる劇のことで、代表作はまさにコルネイユの『ル・シッド』とされる。

そこでコルネイユの作品毎に、「愛」と「名誉」が共起する回数を調べてみると（表5）、悲喜劇『ル・シッド』の10回、英雄喜劇『アラゴンのドン・サンシュ』 *Don Sanche d'Aragon* の6回という出現回数が他と比べて確かに多い。英雄喜劇 *comédie héroïque* は、身分の高い登場人物たちが恋に悩むが、幸せな結末を迎える筋書きを持つ劇作品で、これ

を悲喜劇と区別することは難しい。『アラゴンのドン・サンシュ』も悲喜劇と分類して良いだろう。従って、コルネイユの作品群では悲喜劇において「愛」と「名誉」という単語が共起する率が高いと言える。但し、悲喜劇ならば「愛」と「名誉」が必ず共起するとは言えない。例えば1631年の悲喜劇『クリタンドル』*Clitandre*では「愛」と「名誉」の共起回数は1回に過ぎない。また、この結果は直ちにコルネイユの悲喜劇において、「愛」か「名誉」かという二者択一が提起されていることを示すものではない。共起回数による分析はあくまで「愛」という単語が「名誉」という単語と極めて近い位置で多く発現していることを示しているのに過ぎない。「愛」と「名誉」がどのような関係にあるか、「名誉」と「愛」が対立する選択肢なのか、「名誉」があれば「愛」を得ることができるのか、「愛」イコール「名誉」なのか、など二つの関係はさまざまに考えられる。共起回数の分析だけでは、どのような文脈で使われているかという意味の解釈まで到達することは難しいのである。コンピュータ処理による分析で、どこまでこのような意味の解釈に踏み込むことができるのかは次の稿以降で扱うこととして、今回は共起回数の分析に留めたい。

Mélite	1629	2
Clitandre	1631	1
La Veuve	1632	4
La Galerie du Palais	1633	1
La Suivante	1634	1
L'Illusion comique	1635	1
Le Cid	1637	10
Horace	1639	2
Cinna	1639	1
La Suite du Menteur	1644	3
Théodore	1645	1
Héraclius	1647	1
Don Sanche d'Aragon	1649	6
Andromède	1650	1

Pertharite	1651	1
Œdipe	1659	4
Sophonisbe	1663	1
Othon	1664	1
Agésilas	1666	4
Tite et Bérénice	1670	1
Psyché	1671	1
Suréna	1674	1

表5 コルネイユ作品中で「愛」と「名誉」が共起する回数

一方、キノー作品の中で「愛」と「栄光」が共起する回数が多い作品を調べてみると（表6）、『テゼ』と『ロラン』がそれぞれ13回、11回と回数が多い。どちらもリュリが音楽をつけた音楽付き悲劇（オペラ）である。フランスでは、音楽付き悲劇はルイ14世の栄光を歌うために作られたため、悲劇と名前は付けられているが、結末は予定調和的に幸福なものとならなければならなかった。キノーとリュリの音楽付き悲劇では、「愛」を捨て、「栄光」を選ぶ国王＝ルイ14世の姿が描かれる。実際、『テゼ』ではアテネ国王が、『ロラン』ではシャルルマーニュ大帝の甥ロラン（オルランド）が、自らの愛を諦め、栄光を選択する場面で幕は閉じる。このようにキノーの音楽付き悲劇においては「愛」と「栄光」という単語が共起する率が高いと言える。但し、先ほどのコルネイユの場合と同様に、キノーの音楽付き悲劇ならば「愛」と「栄光」が必ず共起するわけではないということは言うまでもない。

Astrate, roy de Tyr	1664	7
Bellérophon	1671	8
Alceste	1674	5
Thésée	1675	13
Atys	1676	2
Proserpine	1680	1
Le Triomphe de l'amour	1681	2
Roland	1685	11

表6 キノー作品中で amour と gloire が共起する回数

従って、先ほどのコルネイユ作品では「愛」と「名誉」が、キノー作品では「愛」と「栄光」が結び付けられて捉えられていることが多いという分析は書き換えられる必要がある。コルネイユ作品では、悲劇ではなく、悲喜劇において「愛」と「名誉」が問題となる率が高い（そこでは主人公は「名誉」を保つことで「愛」を得ることができる）。一方、キノーの作品では、音楽付き悲劇において「愛」と「栄光」が問題となる率が高い（そこでは主人公＝国王は「愛」を捨て、「栄光」を選択する）。どちらも幸福な結末を迎える点で純粋な悲劇とは異なり、よく似た存在であったと言える悲喜劇と音楽付き悲劇において、「愛」と「名誉」、「愛」と「栄光」が問題にされたことになる。

ここで、コンピュータ処理で何ができるのかという最初の問いかけに一旦戻ろう。

共起回数の分析では「愛」と「名誉」、「愛」と「栄光」のような対立項を設けて分析を行ったが、このような対立項の存在そのものをコンピュータ処理による分析で示すことはできるものなのだろうか？

そこで、「愛」という単語の前後1語（フランス語では形容詞は後置するものが殆どで、前置するものは少ないため）に付く上位10単語を調べてみたものが次の表7である。コルネイユにおいて「真の」véritable、ラシーヌにおいては「罪深い」criminel、モリエール、キノーにおいては「甚だしい」extrême (extresme) という単語が多いことがわかるが、平均回数にするとどれも1回を割っている。これらの作家において、「愛」という単語に常に同じ形容詞が使われていたのではないと明らかになる。

Corneille	1844 平均	Racine	422 平均	Moliere	694 平均	Quinault	666 平均	
ton	35	1.03 ton	8	0.57 extrême	6	0.18 vostre	22	2.00
véritable	15	0.44 criminel	4	0.29 parfait	5	0.15 extresme	10	0.91
seul	14	0.41 chaste	3	0.21 ardente	4	0.12 tendre	9	0.82
paternel	8	0.24 extrême	3	0.21 veut	4	0.12 fidelle	8	0.73
digne	7	0.21 aveugle	2	0.14 pur	3	0.09 cét	6	0.55
extrême	7	0.21 fatal	2	0.14 sait	3	0.09 malheureux	6	0.55
parfait	7	0.21 fol	2	0.14 savent	3	0.09 extrême	5	0.45
conjugal	6	0.18 folle	2	0.14 tendre	3	0.09 seul	5	0.45
doit	6	0.18 funeste	2	0.14 ton	3	0.09 ton	5	0.45
vertueux	6	0.18 malheureux	2	0.14 ardent	2	0.06 veut	5	0.45

表7 「愛」Amour につく形容詞

直接付く形容詞では傾向がはっきりとしないために、更に、「愛」という単語前後 20 語以内に共起する単語上位 20 を調べ、まとめたものが次の表 8 である。

Cornelle	2068 平均	Racine	470 平均	Molière	825 平均	Quinault	834 平均				
cœur	209	6.15	cœur	60	4.29	cœur	103	3.12	cœur	139	12.64
jour	171	5.03	jour	50	3.57	jour	80	2.42	vostre	107	9.73
vos	143	4.21	vos	50	3.57	psyché	68	2.06	may	75	6.82
faut	119	3.50	madame	36	2.57	vos	68	2.06	rais	70	6.36
ton	116	3.41	yeux	36	2.57	faut	53	1.61	jour	67	6.09
dime	115	3.38	père	34	2.43	yeux	47	1.42	dime	65	5.91
voir	113	3.32	seigneur	34	2.43	voir	45	1.36	ay	62	5.64
madame	102	3.00	ch	28	2.00	yeux	44	1.33	faut	61	5.55
grand	97	2.85	filis	28	2.00	monseigneur	39	1.18	cœurs	57	5.18
seigneur	96	2.82	quai	27	1.93	ch	35	1.06	doux	56	5.09
yeux	96	2.82	glaise	23	1.64	dime	35	1.06	choeur	49	4.45
haine	87	2.56	haine	23	1.64	belle	34	1.03	glaise	49	4.45
veux	86	2.53	dime	22	1.57	scène	33	1.00	angelique	46	4.18
quai	83	2.44	faut	22	1.57	veut	33	1.00	heureux	43	3.91
doit	80	2.35	veut	22	1.57	valère	32	0.97	doit	41	3.73
veut	80	2.35	vie	21	1.50	doux	31	0.94	vos	40	3.64
encor	78	2.29	voir	21	1.50	fort	31	0.94	cybele	38	3.45
scène	73	2.15	foi	20	1.43	vœux	31	0.94	s'engarde	38	3.45
dimer	71	2.09	s'ang	20	1.43	quai	30	0.91	medar	37	3.36
lou	71	2.09	funeste	19	1.36	dai	29	0.88	scène	36	3.27

表 8 「愛」Amour の前後 20 語以内に共起する単語

4 人の作家共通で上位 5 位までに「心」cœur、「日」jour という単語が入ってくる。しかし、コルネイユに特徴的とされる「名誉」は 20 位に入らず、キノーにおける「栄光」もようやく 11 位でしかない。つまり、コンピュータ処理により単に共起回数を調査するだけでは、今のところ、コルネイユの「愛」と「栄光」のような、ある作家を特徴づける対立項が何かを機械的に導き出すことは難しいようである。

## 2-3 「名誉」と「栄光」

前章において、17 世紀前半に流行した悲喜劇では「愛」と「名誉」が、17 世紀末に流行した音楽付き悲劇では「愛」と「栄光」が問題にされる率が高いという結果が得られたが、そこから 17 世紀前半から後半にかけて、「「愛」か「名誉」か」から「「愛」か「栄光」か」という価値観の変化があったのではないかという仮説を立てることにする。

まず明らかにしておきたいことは、「名誉」と「栄光」の違いである。そこで、コルネイユ、ラシーヌ、モリエール、キノー、ヴォルテールの作家のそれぞれで「名誉」と「栄光」が 20 単語以内に共起する回数を

確認してみると、次の表9のようになる。

honneur et gloire	Corneille	Racine	Moliere	Quinault	Voltaire
共起回数	43	7	6	1	9
平均共起回数	1.26	0.64	0.18	0.09	0.82

表9 「名誉」と「栄光」の共起回数

ここから、確かにコルネイユは明らかに「名誉」と「栄光」の関係を意識していたが、キノーは二つの単語の間に関係性を見出していなかったと判明する。実際、コルネイユは、この二つは全く違うものと認識していたようだ。

メデ：

私は私の義務をなします。

あなたがあなたの義務をなすように。

私には名誉が重要なのです。あなたに栄光が重要なように。

MÉDÉE：

Je ferai mon devoir, comme tu fais le tien.

L'honneur doit m'être cher, si la gloire t'est chère：

Pierre CORNEILLE, *La Toison d'or*, ACTE II, SCÈNE II.

そして、この二つの違いはコルネイユによると「名誉」は「栄光」の前提条件、つまり「栄光」は「名誉」の後に来るものであったようだ。

クリタンドル：

あなたにお仕えする名誉が、私に栄光をもたらすのです。

CLITANDRE：

L'honneur de vous servir m'apporte assez de gloire,

Pierre CORNEILLE, *Clitandre*, ACTE V, SCÈNE IV.

次に、17世紀前半から17世紀後半にかけて価値観の変化が起きたの

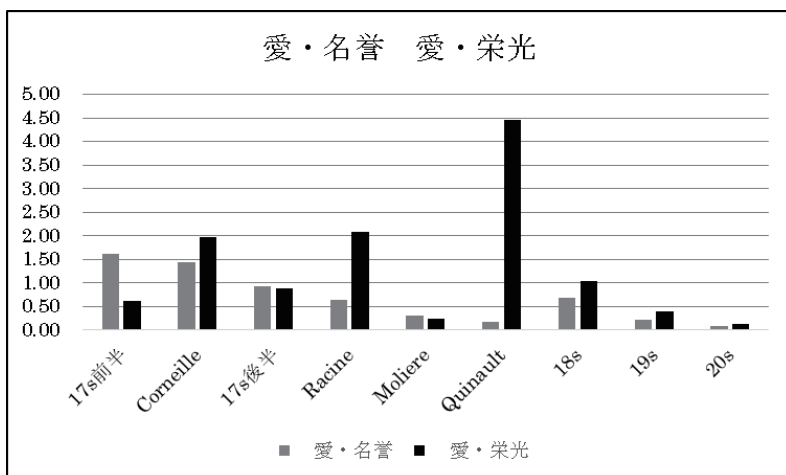
かを調べるために、先ほどの表4（「愛」と「名誉」、「愛」と「栄光」平均共起回数）のデータ中、17世紀分を17世紀前半と後半に分割して調べなおしてみることにする。重複を避けるために、17世紀前半の作品からはピエール・コルネイユ、17世紀後半の作品からはコルネイユ、キノーの作品を除き、それぞれ181作品、291作品における「愛」と「名誉」、「愛」と「栄光」という単語が20単語以内に共起する回数、平均を表4に追加し、作家の時代順に並び替えたものが次の表10、グラフ6である。

	17s前半	Cornelle	17s後半	Racine	Moliere	Quinault	18s	19s	20s
総単語数	10485676	666878	12313704	210729	514588	122023	36178109	76007436	125343556
作品数	181	34	291	11	33	11	585	1096	2102
amour et honneur	292	49	270	7	10	2	399	252	194
平均	1.61	1.44	0.93	0.64	0.30	0.18	0.68	0.23	0.09
amour et gloire	113	67	259	23	8	49	609	441	269
平均	0.62	1.97	0.89	2.09	0.24	4.45	1.04	0.40	0.13

表10 「愛」と「名誉」、「愛」と「栄光」共起回数変化

すると17世紀前半の作品群及びコルネイユの作品群における「愛」と「名誉」という単語が共起する平均回数は、後の時代のそれに比べると明らかに高いことが分かる。そして、「愛」と「名誉」の共起回数は、17世紀後半以降減っている。

一方、「愛」と「栄光」という単語が共起する平均回数に関しては、キノー作品における共起回数平均が同時代の他の作家、他の時代の作品に比べて2倍以上高いものになっていることが見て取れる。キノーほどではないが、コルネイユもラシーヌも実は「愛」と「栄光」を関連付けていたことも明らかになる。また、こちらの共起回数は19世紀から20世紀にかけて減っていくものの、17世紀から18世紀にかけては変化していない。



グラフ6 「愛」と「名誉」、「愛」と「栄光」共起回数変化

以上から、次の傾向を読み取ることができるだろう。

1: 「愛」と「名誉」に関しては17世紀前半には、結び付けられて考えられることが多く、後半からは減っていく。コルネイユの作品の「愛」と「名誉」の使われ方には、時代的な影響があるようだ。

2: 「愛」と「栄光」に関しては、時代的な特徴というよりも、キノーの作品に特有のテーマを構成していた。

しかし、実はこのような結論に飛びつくことはこの段階ではできないのである。

以上はあくまでもFrantextというテキスト・データベースを用いた分析結果からの仮説に過ぎない。Frantextという最大級のフランス語文献のデータベースの現時点での限界がそこにはあるからだ。

#### 2-4 Frantextの限界、他のデータベースの可能性

2016年時点で10世紀から21世紀までの5000点を超えるフランス語



によるテキストを有するデータベース Frantext は、もともとフランス語大辞典を作るために作られたデータベースであり、フランス語のサンプルを提供するにすぎない。Frantext はフランス語・フランス文学の包括的なテキスト・データベースではない。

例えば 17 世紀前半ではヨーロッパ中で大流行した田園小説『アストレ』*L'Astrée* はデータベースに含まれているが、演劇作家でもあったラ・カルプルネード *La Calprenède* の演劇作品や大河小説『カッサンドル』*Cassandra*、『クレオパートル』*Cleopatre* は含まれていないし、後半でもキノー以外の音楽付き悲劇台本作家、例えばラ・グランジュ＝シャンセル *La Grange-Chancel* などは含まれていない。Frantext は包括的なデータベースではないため、ある種のヒエラルキーにより選ばれた作家が優先的に登録されているのである。しかし、文学史上ではマイナーであり、研究対象とはされることは少ないが、当時は人気もあったような作家の作品は、時代精神の変化を調べるためには不可欠の要素となる。

従って、これらのマイナー作家のデータを現段階では欠く Frantext というデータベースを用いて試みた分析は、文学史上に名を残し、データベースに登録されているような作家の特性については正確なものではあるものの、時代精神の変化についての分析は正確なものにはなりえない。そのような時代精神の変化の調査を可能にするものは、包括的な大規模なデータベースの存在でしかない。包括的なデータベースを目指しているのはフランス国内ではフランス国立図書館 BNF (*Bibliothèque nationale de France*) のウェブサイト (<http://gallica.bnf.fr/>)、それ以外では世界最大級の検索エンジン Google のサービスの一つである Google ブックス (<https://books.google.co.jp/>) がよく知られている。

前者はもともと 1988 年にフランソワ・ミッテラン大統領により建物をもたないバーチャルな図書館として構想されたものである。当初は図書館内の端末から 10 万冊の画像データにアクセスできることを目的としたが、ウェブの台頭により、ウェブ上で誰もが閲覧できるデジタル図書館の構築へと軌道修正を行った<sup>12</sup>。

後者は、1996 年スタンフォード大学の大学院生が始めた図書館の蔵書を自由に検索するための「スタンフォード・デジタル図書館技術プロジェクト」が、後に巨大な検索エンジンへと変化していったものである。2004 年になると Google は当初の使命に立ち返り、自社で本の内容を読み取ってデータ化する「Google ブックス」プロジェクトに乗り出した。9 年後には 3000 万冊を超える本がデジタル化されたが、これはこれまでに出版された本のほぼ 4 冊に 1 冊にあたる<sup>13</sup>。

Frantext との違いは、Frantext がテキスト・データベースだったのに対して、BNF も Google もデータを画像で持つデータベースとなっている点である。このままではテキスト処理を行えないため、両者とも、デジタル化された画像データを OCR（Optical Character Recognition、光学的文字認識）処理し、テキスト・ベースのデータとしても持っている。しかし、この OCR 処理の精度に現時点では問題があるようだ。

今回は Google ブックスが提供している画像データで検証を行うことにする。次の図 1 は Google ブックスに登録されている 17 世紀末の劇作家ラ・グランジュ＝シャンセル（先述したように Frantext にはこの作家は登録されていない）の全集<sup>14</sup>の画像の一部である。

**L'empereur aujourd'hui voit la fête éclatante ,  
Où l'hymen se prépare à remplir son attente ,  
Et de tout l'univers attirant les regards ,  
Ma fille va monter au trône des Césars.  
J'attendois ce moment avec impatience ,  
Non pour être plus près de la toute-puissance ,  
Ni pour voir l'empereur me charger d'un em-  
ploi  
Dont ta sœur Pulcherie est plus digne que  
moi ;  
Mais pour jouir , ami , de la douceur extrême  
De voir fixer le sort d'une fille que j'aime ,  
Et loin de cette cour , précipitant mes pas ,  
Eviter des grandeurs qui ne me touchent pas .**

図 1 ラ・グランジュ＝シャンセル全集から<sup>15</sup>

17 世紀にフランスで出版された書籍はすでに見やすい活字を採用しており、現在の書籍と大差がない。この箇所をタイプで打ち直してみると以下ようになる。

L'empereur aujourd'hui voit la fête éclatante  
Où l'hymen se prépare à remplir son attente  
Et de tout l'univers attirant les regards,  
Ma fille va monter au trône des Césars.  
J'attendois ce moment avec impatience,  
Non pour être plus près de la toute-puissance,  
Ni pour voir l'empereur me charger d'un emploi  
Dont la soeur Pulcherie est plus digne que moi ;  
Mais pour jouir, ami, de la douceur extrême  
De voir fixer le sort d'une fille que j'aime,  
Et loin de cette cour, précipitant mes pas,  
Eviter des grandeurs qui ne me touchent pas.

この画像を Google が OCR 処理し、裏に張り付けたテキストデータは以下のようになっている。

L empereur aujourd hui voit la fête éclatante  
Où l hymen se prépare à remplir son attente  
t Et de tout l univers attirant les regards  
Ma fille va monter au trône des Césars  
J attendois ce moment avec impatience  
Non pour être plus près de la toute puissance  
Ni pour voir l empereur me charger d un en ploi  
Dont fa soeurr Pulcherie est plus digne moi  
Mais pour jouir ami de la douceur extrême  
Devoir fixer le sort d une fille que j aime  
Et loin de cette cour précipitant mes pas  
Eviter des grandeurs qui ne me touchent pas

フランス 17 世紀の活字の特徴として « s » が縦長で « f » と混同しやすいというものがあり、これは全て誤認識している。92 単語中 10 単語を誤認識しているが、« s » と « f » を修正すれば、« m » と « n » の取り違いなどの純粋な誤認識は 4 単語になる。それでも確率としては 4.3 パーセントの誤認識があることになる。古典主義期演劇テキストは 1 作あたり 2 万字とすると、この率では OCR 処理された演劇テキスト 1 作あたりに約 870 字のミスがあることになる。

正	誤
Et	t Et
puissance	puiflance
emploi	en ploi
sœur	foettr

更に、この時代のテキストは活字が滲んでいるなど判読に悩むことが多い。ラ・グランジュ＝シャンセルの同じテキストだが数ページ後に次のような箇所がある。

**Hé bien ! dans ce récit que trouvez-vous à crain-**  
**dre ?**  
**Qui sera donc heureux, si vous êtes à plaindre ?**  
**L'hymen de votre fille est prêt à s'achever ;**  
**L'ambassadeur persan va bientôt arriver :**  
**Si le jeune empereur attendoit la présence,**  
**Vous savez les raisons de cette déférence,**  
**Et que malgré les feux & son empressément,**  
**Son devoir l'obligeoit à ce retardement.**

図2 ラ・グランジュ＝シャンセル全集から<sup>16</sup>

このようなテキストは人間でも読むのが難しいが、機械処理が今のところ苦手とするようだ。タイプで打ち直したものと、Google による OCR 処理されたデータを見比べみよう。タイプで打ち直したものは次の

ようになる。

Hé bien dans ce récit que trouvez-vous à craindre ?  
 Qui fera donc heureux, si vous êtes à plaindre  
 L'hymen de votre fille est prêt à s'achever  
 L'ambassadeur persan va bientôt arriver  
 Si le jeune empereur attendoit sa présence  
 Vous savez les raisons de cette déférence  
 Et que malgré ses feux & son empressement  
 Son devoir l'obligeoit à ce retardement

Hé bien dans ce céeit que trouvez vous à çrai dre  
 Qui fera dwpfesreux fi vous êtes àplaind ef  
 L hymen de votre filleeft prêt à s achever  
 L ambafladeur perfan va bientôt arriver  
 Si le e ne empereur tcendoit fa préjfence  
 Vous fevez les raifons de cette défér eae  
 Et que malgré fes feux & jbn empreflemeot  
 Son devoir l obligeoit à ce retardement

Google が提供するテキストデータは次のようなものだ。

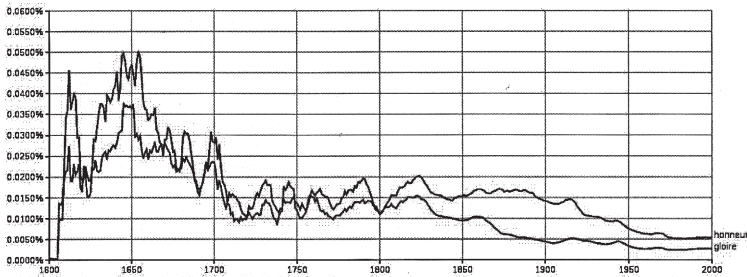
先ほどと同じように純粋な誤認識だけを拾っていくと、次のリストのようになる。

正	誤
récit	céeit
craindre	çrai dre
à plaindre	àplaind ef
ambassadeur	ambafladeur
jeune	e ne
attendoit	tcendoit
présence	préjfence
savez	fevez

déférence	défer eace
son	jbn
empressement	empreflemeot

今回は 59 単語中 11 単語の純粋な誤認識があり、18.6 パーセントの誤認識となる。これだけ滲みがあるページでこれだけの誤認識で済んでいるのは立派な成績だと言えなくもないが、やはり高い誤認識率だろう。このようなページはそれほど多いとは言えないが、この時代のテキストには必ず存在する。従って、17 世紀から 18 世紀のフランス語のテキストを現段階で単純に機械的に OCR 処理した場合、5 パーセントから 20 パーセント程度の誤認識はあると考えておくべきであろう。但しこれはあくまで 17 世紀、18 世紀など活字が現在と異なるか、滲んでいるなどの問題がある場合の数字であり、それ以降のテキストでは誤認識率は低いのではないと思われる。いずれにしても、この数字は今後、処理性能が向上することで改善されることが期待される。

さて、17 世紀から 18 世紀のフランス語のテキストに 5 パーセント以上の誤認識の可能性があるものの、母数で 3000 万冊を超える Google ブックスのデータベースは、利用することができるのであれば、極めて強力な分析ツールの材料となる。そして、このデータベースから、単語の使用頻度の変化を調べることができるようにした分析ツールが、Google Ngram Viewer (<https://books.google.com/ngrams>) というウェブサービスである。実際に、1600 年から 2000 年までの期間でフランス語の本に限り、「名誉」、「栄光」という単語が時代ごとにどの程度使われているかを Google Ngram Viewer を用いて、出力したものが、次のグラフである。



グラフ7 Google Ngram Viewer による「名誉」と「栄光」の使用頻度変遷<sup>17</sup>

この結果には5パーセント強の誤差が含まれていると推定されるとはいえ、明らかにフランス語のテキストにおいて「名誉」と「栄光」という単語の使用率が17世紀中ごろをピークとして、減少していく様子が見て取れる。「愛」か「名誉」か、「愛」か「栄光」かというディレンマは17世紀絶対王政期のフランスに特有の現象であったと言える。

このNgram Viewerでは生のデータをダウンロードすることも出来るため、このデータを使ったどのような分析が可能なのかについては、次回以降調査することにした。

### 3 まとめ、今後の課題

コンピュータ処理により、フランス古典主義演劇テキスト研究は何ができるのかを調査するにあたり、本論では2016年12月時点で5116点のフランス語テキストを有するFrantextを用いて、単語の生起・共起回数による分析を行った。その結果、17世紀悲劇作家のコルネイユの悲喜劇において「愛」と「名誉」が、17世紀末の音楽付き悲劇台本作家キノーの作品では「愛」と「栄光」が問題とされる率が高いという文学史上の定説を確認することができた。特に、「愛」と「名誉」に関しては17世紀前半には結び付けられて考えられることが多く、後半からは減っていくという時代的要素が強いことも判明した。対して、「愛」と「栄光」に関しては、時代的な特徴というよりも、キノーの作品に特有の

テーマを構成していたのではないかと考えられる。しかし、このような時代的要素に関する結論は、あくまでも Frantext というテキスト・データベースを用いた分析結果からの仮説に過ぎない。Frantext は信頼できるテキストデータを 5000 点超提供してくれているが、その数は時代の変遷の分析には不足する。一方、Google ブックスが提供するデータは 3000 万点を超えるため極めて母数が大きく、時代の変化を十分に映し出すものと考えられるが、フランス古典主義時代の文献などに関しては、OCR 処理における誤認識率が高いという欠点を持つ。

従って、コンピュータ処理により、フランス古典主義演劇研究の可能性を追求していくにあたって、今後はいかに 17 世紀、18 世紀フランス演劇全体の信頼できるテキストデータを得るかが課題となってくる。勿論、一人の研究者が全文献をスキャンし、OCR をかけ、修正していくなどということは不可能である。そのため、Google、BNF などによる OCR データを修正し、使うことができないかを検討することになるだろう。本来的には、Google、BNF による OCR 精度の向上が望まれるというのは言うまでもない。

Google などによりすでに提供されているデータを修正し得られたデータをもとに今回行ったような共起回数による分析を行うとどのような結果が得られるのだろうか。

また、テキストデータから意味の分析は可能なのだろうか。今回は「愛」と「名誉」、「愛」と「栄光」という単語の共起回数による分析を行ったが、今回の共起回数分析ではこれらの単語の意味関係（「愛」か「名誉」かなのか、「愛」と「名誉」かなのか、「名誉」によって「愛」を得るのか、など）を明らかにすることはできていない。また、「愛」か「名誉」か、「愛」か「栄光」か」という対立軸の存在も、今のところコンピュータ処理による単語の単純な分析からは簡単に出てくるものではないと分かった。コンピュータ処理により、意味内容の分析がどこまで可能なのか。もしコンピュータ処理により意味内容の分析が可能ならば、このような対立軸が存在するのか、存在するのであれば何処に存在



するのか、つまりはある作家、ある作品の特徴がどのようなものであるかを、コンピュータ処理が明らかにすることすら可能になるのではないだろうか。

次回以降は演劇テキストのデータ化の可能性、および、テキストデータをもとに意味の分析が可能なのかどうかを検証することにしたい。

本稿は平成 28 年度成城大学特別研究助成「自然言語処理によるフランス古典主義演劇研究の可能性」の研究成果である。

注

- 1 Bryant C. FREEMAN, Alan BATSON (Programmeur), *Concordance du Théâtre et des Poésies de Jean Racine*, 2 v., New York, Cornell University Press, 1968.
- 2 *Ibid.*, p. vii.
- 3 *Ibid.*, p. viii-ix.
- 4 B. QUEMADA, *J. Racine Phèdre Concordances Index et Relevés statistiques*, Librairie Larousse, 1966.
- 5 *Ibid.*, p. b.
- 6 Freeman, *Op. cit.*, p. ix.
- 7 <http://www.frantext.fr/> (2016 年 12 月 28 日参照)
- 8 沢田肇「研究ノート：<Frantext> について」『上智大学仏語・仏文学論集』34、上智大学、2000 年、150 頁。
- 9 Jacques Lindecker, *Les grands écrivains français*, Editions l'Etudiant, 2002, p. 43.
- 10 Norman Buford, *Quinault, librettiste de Lully: le poète des grâces*, Editions Mardaga, 2009, p. 144.
- 11 Georges Forestier, *Essai de génétique théâtrale : Corneille à l'œuvre*, Librairie Droz, 1996, p. 352.
- 12 <http://gallica.bnf.fr/html/und/a-propos> (2017 年 1 月 4 日参照)
- 13 エレツ・エイデン&ジャン＝バティースト・ミシェル、阪本芳久訳『カルチャロミクス 文化をビッグデータで計測する』草思社、2016 年、30 頁。
- 14 [https://books.google.co.jp/books?id=qVEGAAAAQAAJ&dq=la+grange-chancel&hl=ja&source=gbs\\_navlinks\\_s](https://books.google.co.jp/books?id=qVEGAAAAQAAJ&dq=la+grange-chancel&hl=ja&source=gbs_navlinks_s) (2016 年 12 月 28 日参照)
- 15 Joseph de La Grange-Chancel, *Œuvres de Monsieur de la Grange-Chancel. Tome II*, Chez les librairies associés, 1758, p. 10.
- 16 *Ibid.*, p. 15.
- 17 <https://books.google.com/ngrams> (2017 年 1 月 5 日参照)

