

品詞分析から見る夏目漱石の前期作品の文体の特異性

杉浦清人

1. 夏目漱石の特異性

今日、情報技術の発展に伴って情報技術と人文学の融合を目指すデジタル・ヒューマニティーズの機運が高まっている。デジタル・ヒューマニティーズは多様な可能性を持つが、文学研究におけるテーマの一つとして、数的データによって作品の解釈を試みる計量文体分析がある。計量文体分析には、大量の作品を比較できることや、人間の読解とは異なった角度からの知見を得ることができるという利点がある。¹

計量文体分析の方法・対象は多様でありうるが、ここでは主に夏目漱石の作品を題材として用いる。その理由として、漱石は日本近代文学を代表する大作家であること、インターネット上で公開されている青空文庫²においてすべての主要作品が電子化されていること、短編小説よりも揺らぎが少ないと考えられる長編小説をいくつも執筆しているために扱いやすいことがある。

さらに、より大きな理由は、近代日本文学の文体史において漱石は独特の興味深い位置を占めているからである。

日本における計量文体分析の先駆者である心理学者の波多野完治はこう書いている。

鴎外や漱石をいかに処理すべきであろうか。実をいえば私にはこれを「文章史」として処理すべき案がないのである。鴎外の文章は現代にあっても一つの古典であり模範であって、それは歴史の中で時間的に整理することがむずかしい。また漱石の変貌するスタイルは歴史家を幻惑せしめるばかりである。むしろこのような作家は文章史の「潮流」というものからはいちおう切り離して考えるべきもののように思われるのである。³

ここで波多野は、漱石の文体は「変貌する」ため、理解・説明しにくいとしている。

すなわち、漱石はその作家生活の中で、作品によって大きく異なった文体を用いて執筆しているということである。

批評家の柄谷行人は漱石の特異性について次のように書いている。

¹ 計量文体分析のより詳しい歴史的・理論的説明については、杉浦清人(2017)「文学研究におけるデジタル・ヒューマニティーズの可能性：文章心理学・計量文献学・マクロ分析」、『れにくさ:現代文芸論研究室論集』7,pp80-96.を参照。

² <https://www.aozora.gr.jp/>

³ 波多野完治(1988)『文章心理学入門』小学館,pp146-147.

夏目漱石は、初期の『吾輩は猫である』や『坊っちゃん』、また『漾虚集』から『明暗』に至る小説、さらに俳句や漢詩を書いている。つまり、多種多様な“ジャンル”に及んでいる。こういう作家は日本だけでなく、たぶん外国にもいないだろう。⁴

柄谷の漱石とジャンルについての議論をやや単純化して要約すれば、漱石は自然主義に代表される典型的な近代小説とは異なる、前近代的な要素を含んだ作品を書いていたのであり、そのために文体も近代小説とは異なっていたのである。このような特殊な立ち位置にいる作家であるから、漱石の文体の分析は他の作家以上に日本近代文学史において重要なのである。

伝記的事実として、漱石の作家生活において転換点となったのは明らかに、それまで盟友高浜虚子の雑誌『ホトトギス』などに作品を掲載していた彼が、東京帝国大学の職を辞して朝日新聞社に入社し職業作家となったことである。

だが、内容・文体の面から考えれば、朝日新聞入社後最初の作品である『虞美人草』を後期の始まりとするよりも、『三四郎』から後のものを漱石の後期作品とみなしたほうが妥当と考えられる。

作家の水村美苗はこの転換について次のように書いている。

漱石のように異様に多様な作品を残した作家は、ひとつの作品ごとにはげしく変わり、ひとつの作品ごとに何かちがうものへ向かう必然のようなものを感じさせる。だが『虞美人草』を境に漱石の作品が決定的に変わったのは、『虞美人草』以降、まさにその多様性自身の振幅がきわめて尋常になったことにある。『虞美人草』のあと漱石は「美文」を書くのをやめ、多様なジャンルにわたることもなくなった。⁵

『三四郎』以後の漱石は比較的似通った作品を、執筆中に彼が死亡したため未完の作品となった『明暗』まで書き続けた。後期の漱石作品はほぼ一貫して男女の三角関係というテーマを扱っており、その恋愛関係の中で男性が近代的な自意識に苦しむという図式を持っている。また、どれも現代から見ても読みやすい文体の近代小説といえる。

つまり、漱石は『虞美人草』までの（正確には『虞美人草』の後のルポルタージュ的な異色作である『坑夫』までの）前期においては多様な、前近代的な要素を含む作品を書いていたが、『三四郎』以後は多様な作品を書かなくなり、近代的な小説を書くようになり、文体も変化したと図式化することが出来る。そこで、このテキスト読解・文学理論的な漱石の前期と後期の差異が、計量的文体分析の結果と一致するかの検証を試みる。

⁴ 柄谷行人(2001)『増補 漱石論集成』平凡社,pp127.

⁵ 水村美苗(2009)『日本語で書くということ』筑摩書房,pp121.

2. 品詞分析という方法について

ここでは計量文体分析の方法として、KHCoder⁶を用いた品詞分析を行うこととする。

日本における人文的な計量文体分析の先行研究として、波多野完治や安本美典によるものがある。これらは当時の情報技術が未発達であったため、作品の一部を切り出して、人力で品詞などの要素をカウントするという方法をとっていた。現在においては、電子化されたテキストがあればコンピュータによる計量が可能であり、そちらのほうがテキスト全体を扱うことができるため（テキストの一部を切り出して分析する場合、それがテキスト全体とずれた文体であったなら、テキストの誤った解釈に至るだろう）、そして時間がかからないため、また他人によって検証しやすいために優れている。日本語の分析には品詞の解析が必要であるが、Chasen⁷や MeCab⁸といった解析器が開発されているので利用することができる（本発表では KHCoder を用いて間接的に MeCab を使用している）。さらに、KHCoder によって長いテキストの品詞解析の結果を利用することが容易になる。

ただ、日本語の品詞をどのように分類するか自体が言語学的にも統一的な見解を得られていない難解な問題であり、加えて形態素解析の精度が完璧ではないという問題があるため、品詞分析はかなり繊細かつ誤りやすい作業となることは留意しておかなければならない。

3. 品詞分析の実例

ここでは漱石の長編作品のみを扱う。短編作品では少ない語数の中でどのような場面を描いているかなどによって集計結果がかなり揺らぐと考えられるが、長編作品ならばより安定した結果が得られるからである。

	名詞	動詞	形容詞	副詞	連体詞	接頭詞	接続詞	感動詞	助詞	助動詞
吾輩は猫である	31.60%	15.69%	1.95%	3.57%	1.16%	0.82%	0.76%	0.52%	33.15%	10.76%
坊っちゃん	30.35%	17.14%	2.02%	3.36%	0.99%	0.65%	0.63%	0.36%	33.08%	11.42%
草枕	31.91%	16.22%	2.25%	3.30%	1.26%	0.68%	0.66%	0.55%	33.62%	9.53%
野分	31.38%	16.30%	2.09%	3.08%	1.03%	0.65%	0.89%	0.42%	32.43%	11.73%
虞美人草	31.69%	16.45%	2.37%	3.19%	0.73%	0.90%	0.75%	0.50%	33.05%	10.37%
坑夫	28.78%	17.66%	2.29%	3.98%	1.31%	0.41%	1.25%	0.41%	32.84%	11.07%
三四郎	30.28%	16.67%	2.06%	3.66%	1.38%	0.33%	1.07%	0.31%	33.05%	11.19%
それから	30.79%	15.89%	1.83%	3.05%	1.21%	1.88%	1.19%	0.16%	32.50%	11.51%
門	30.71%	15.89%	2.04%	3.47%	1.18%	1.26%	1.05%	0.15%	33.21%	11.05%
彼岸過迄	31.12%	15.72%	2.04%	3.50%	1.32%	0.45%	1.03%	0.15%	33.25%	11.42%
行人	30.17%	15.48%	1.90%	3.63%	1.46%	0.70%	1.24%	0.19%	32.83%	12.38%
ころ	29.17%	15.68%	1.72%	3.40%	1.48%	0.20%	1.21%	0.09%	32.50%	14.55%
道草	31.03%	15.22%	1.99%	3.30%	1.65%	0.69%	1.24%	0.18%	32.22%	12.48%
明暗	30.13%	15.49%	1.78%	3.82%	1.32%	1.20%	1.39%	0.30%	31.90%	12.67%

図 1

漱石の作品を MeCab の品詞体型の大分類のみで品詞分析し、文体をわかりやすくするために記号を取り除いた結果が図 1 である。作品の順番は上から執筆順になっている。すなわち『吾輩は猫である』から『坑夫』までが前期作品であり、『三四郎』から『明暗』までが後期作品である。

⁶ <http://khc.sourceforge.net/>

⁷ <http://chasen-legacy.osdn.jp/>

⁸ <https://www.mlab.im.dendai.ac.jp/~yamada/ir/MorphologicalAnalyzer/MeCab.html>

まず、品詞の割合にそれほど極端な差異はないということがわかる。どれも多くの人間に読めるように書かれた近代日本語の物語文なのだから不思議ではない。しかし、その中でもある程度の差異は存在する。

全体的に前期作品のほうが名詞や動詞や形容詞や感動詞が多い傾向にあり、一方で助動詞や連体詞や接続詞は後期作品のほうが多い傾向にあるということがわかる。

このデータに対して R^2 を用いて主成分分析を行うと、次の図のような結果が得られた。

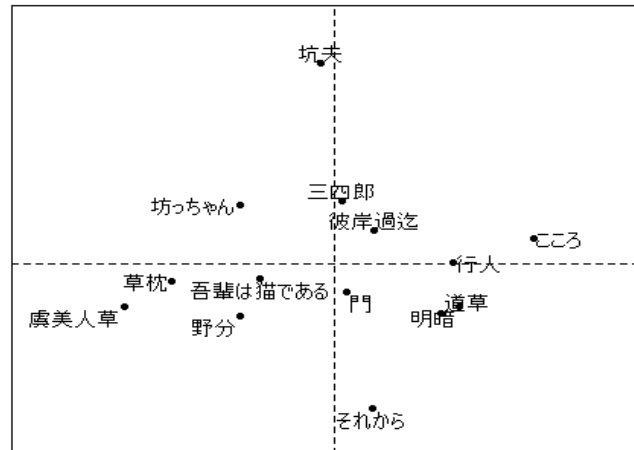


図 2

この結果を見ると、全体として前期作品が図の左側に位置し、後期作品は図の右側に位置している。つまり、大まかな品詞分類による分析で、漱石の前期作品と後期作品に文体の変化があることが認められる。より詳しく見れば、図の中央に近い『三四郎』『それから』『門』『彼岸過迄』は後期の前半の作品で、後期の後半の作品である『行人』『ころ』『道草』『明暗』の4作品はさらに前期と離れた文体になったということもできるだろう。よって、漱石後期をさらに二つの時期に分けて考えるほうが適切かもしれない。

4.美文の分析

品詞分析においては、品詞をより細かい分類を用いて分けることもできる。そのほうが大まかな分類よりもテキストの特定の性質を捉えるためには適切と考えられる。しかし、小説の品詞分析は、先述したように割合の差はそれほど大きくなく、品詞を細かく分けてデータを出してみてもかえってそのデータが何を意味しているのかわからなくなる恐れがある。

また、本発表では純粋な統計だけではなく、文学作品の解釈と統計を組み合わせることを目指す。

そこで、ここでは「美文」というテーマを用いてさらに詳しい分析を行う。先に引用した文章において水村が述べているように、『虞美人草』は美文で書かれている。また、『草枕』も『虞美人草』に近い美文で書かれている。

⁹ <https://cran.r-project.org/>

ここで言う美文とは、たとえば以下のようなものである。

しかもこの姿は普通の裸体のごとく露骨に、余が眼の前に突きつけられてはおらぬ。すべてのものを幽玄に化する一種の靈気のなかに髣髴として、十分の美を奥床しくもほのめかしているに過ぎぬ。片鱗を澆墨淋漓の間に点じて、虬竜の怪を、楮毫のほかにも想像せしむるがごとく、芸術的に観じて申し分のない、空気と、あたたかみと、冥邈なる調子とを具えている。六々三十六鱗を丁寧に描きたる竜の、滑稽に落つるが事実ならば、赤裸々の肉を浄洒々に眺めぬうちに神往の余韻はある。余はこの輪廓の眼に落ちた時、桂の都を逃れた月界の嫦娥が、彩虹の追手に取り囲まれて、しばらく躊躇する姿と眺めた。¹⁰

これは『草枕』中の、語り手が女性の裸体を目撃した場面であるが、この文章は女性について説明しているような印象を与えない。内容よりもむしろ難解な漢語が多く使われている文章表現のほうに印象的である。また、言葉を次々と連ねていく密度の濃さが目を引く。漱石前期の作品である『草枕』や『虞美人草』はこのような美文で書かれており、後期作品の読みやすい文体とは異なっている。図2においてもこの2作品は後期作品と最も距離が離れていた。

では、このような美文の性質は、統計にはどのように表れているだろうか。それを明らかにするために、名詞の中で詳しく分類を見ることにする。

	名詞一般	サ変名詞	形容動詞語幹	固有名詞	代名詞	非自立名詞	名詞接尾	名詞その他
吾輩は猫である	14.91%	2.64%	1.65%	1.33%	2.12%	4.27%	2.53%	2.15%
坊っちゃん	13.52%	2.53%	1.61%	1.07%	2.74%	4.03%	2.38%	2.47%
草枕	17.63%	1.83%	1.51%	1.28%	1.84%	3.46%	2.19%	2.18%
野分	14.48%	2.63%	1.63%	1.63%	1.98%	4.34%	2.49%	2.21%
虞美人草	15.74%	1.75%	1.63%	2.51%	1.68%	3.65%	2.58%	2.15%
坑夫	14.05%	2.12%	1.55%	0.82%	1.19%	4.38%	2.45%	2.22%
三四郎	12.87%	2.14%	1.40%	3.29%	1.53%	4.03%	2.18%	2.86%
それから	14.48%	2.65%	1.50%	1.43%	1.72%	3.85%	2.19%	2.98%
門	13.62%	2.11%	1.25%	3.45%	1.56%	3.97%	1.97%	2.79%
彼岸過迄	13.00%	2.61%	1.58%	2.38%	3.10%	4.19%	1.89%	2.37%
行人	14.20%	2.29%	1.51%	1.10%	2.64%	3.99%	2.02%	2.41%
ころ	11.68%	2.46%	1.54%	0.35%	4.78%	5.14%	1.24%	1.98%
道草	13.58%	2.38%	1.58%	1.85%	3.12%	4.17%	1.85%	2.51%
明暗	12.25%	2.64%	1.68%	1.65%	3.68%	4.10%	1.83%	2.31%

図3

これは、先ほどと同じ記号を取り除いた品詞分析の結果から、名詞のより詳しい分類のみ取り出したものである。このように詳しい分類を見ると、先ほどの大まかな分類では見えなかったことが見えてくる。美文で書かれている作品、『草枕』と『虞美人草』は、名詞一般の割合が漱石作品の中で最も高く、サ行変格活用に接続する名詞と非自立名詞の割合が最も少ないという特徴を持っている。

これが美文の性質であると考えることができる。『草枕』『虞美人草』で用いられているような美

¹⁰ 夏目漱石(1929)『草枕』岩波書店,pp96-97.

文で用いられる名詞はより漢文的で難解であり、サ行変格活用の「～する」という表現はあまり用いずに名詞や動詞を連ねていく文章表現を多用する。また、非自立名詞とは「もの」「よう」「事」などの形式名詞が主である。名詞以外では、この2作品は具体的な内容を表す動詞や形容詞も多く、一方で助動詞や非自立動詞といったニュアンスをつける言葉の割合が低い。このような語の選択によって、これらの美文は密度が濃い文章として読者に感じられるのである。

このように統計分析と人間による文学の読解を組み合わせることで、人間の印象を統計によって検証したり、逆に統計の解釈を読解によってすることが可能となる。

5. 漱石と自然主義

ここまで漱石作品のうちでの文体差を見てきたが、文体分析は漱石以外の作家の作品と混合した場合でも妥当するような一般性を持つべきである。特に計量的分析は客観的に異なる作家の作品を比較できるのであるから。

そこで、次に漱石作品をそれと近い時期の他の作家の小説作品、それも自然主義の作品とともに分析する。文学史的には、明治時代の後半において、田山花袋『蒲団』の影響などがあり自然主義が日本近代文学の主流となっていくが、漱石は「余裕派」と呼ばれ、自然主義とは異なる潮流と見られていた。その漱石も後期には自然主義に近づいていくのだが、前期作品は自然主義とは大きく性質が異なっている。

では、漱石の作品の文体が自然主義と異なっているのか、品詞分析によって統計的に検証してみる。漱石の作品と同じく青空文庫から入手した文学史的に自然主義の作品として位置づけられる作品を、細かい品詞分類で集計し記号を除いた上で漱石の作品とともに主成分分析を行うと次の図のような結果となった。

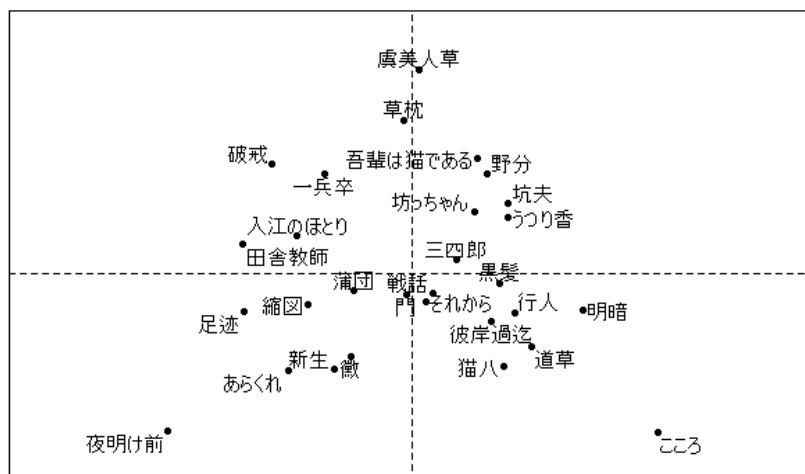


図 4

まず、全体として漱石の作品と他の作家の自然主義の作品が混ざりあっていることがわかる。つまり、前期の漱石作品すべてが自然主義と大きく異なっているとは、少なくともこの品詞分析からは言

えない。しかし、『草枕』と『虞美人草』が図の最も上に来ている。つまり、この美文で書かれた2作品は同時代の自然主義の作品と比較して、ある種の傾向において最も極端な文体であると考えられる。この主成分に寄与している成分に名詞一般や動詞や形容詞や形容動詞などがあり、逆方向に寄与している成分に非自立動詞や接尾動詞や非自立名詞や助動詞や連体詞などがあること、そして『草枕』や『虞美人草』に近い自然主義の作品である『破戒』や『一兵卒』が特に美文ではないことを考えると、ここに表れているのは美文に限らず補助的な役割を果たす語が少なく密度の濃い文体というほうが適切と考えられる。

つまり、漱石の美文的な2作品『草枕』『虞美人草』は、同時代の自然主義の作品と比べて、最も密度が濃い特徴的な文体を持っていることが品詞の統計分析によって示されたといえる。

6.終わりに

本発表では、夏目漱石の前期作品の特異性というテーマを、品詞割合の統計分析という方法で検証することを試みた。その結果、漱石の前期作品と後期作品には文体上の差異があること、その中で『草枕』『虞美人草』の2作品は美文で書かれているために特異な密度の濃い文体であること、そしてそれらは漱石作品だけではなく、同時代の自然主義の作品と比べても特徴的であることが示された。

この品詞分析だけでは理論的出発点である漱石の前期作品の特異性を十分に示すことはできていないが、文学的読解と計量的分析を組み合わせることで分析を行い知見を得ることにはある程度成功したといえるだろう。