



Title	テキストデータベースによる色彩表現の研究：芥川龍之介作品への適用
Author(s)	上村, 和美
Citation	大阪大学, 1994, 博士論文
Version Type	
URL	https://hdl.handle.net/11094/29166
rights	
Note	著者からインターネット公開の許諾が得られていないため、論文の要旨のみを公開しています。全文のご利用をご希望の場合は、大阪大学の博士論文についてをご参照ください。

The University of Osaka Institutional Knowledge Archive : OUKA

<https://ir.library.osaka-u.ac.jp/>

The University of Osaka

氏 名	うえむらかずみ 上村和 美
博士の専攻分野の名称	博 士（言語文化学）
学 位 記 番 号	第 1 1 4 1 8 号
学 位 授 与 年 月 日	平成 6 年 3 月 25 日
学 位 授 与 の 要 件	学位規則第 4 条第 1 項該当 言語文化研究科言語文化学専攻
学 位 論 文 名	テキストデータベースによる色彩表現の研究 — 芥川龍之介作品への適用 —
論 文 審 査 委 員	(主査) 教 授 波田 節夫 (副査) 教 授 中西 暉 教 授 今井 光規

論 文 内 容 の 要 旨

文学作品における色彩表現の研究は、従来も波多野完治の創設した文章心理学の分野や、文学研究の分野でも行われてきた。しかし、手作業による色彩表現の抽出には、限界があった。対象作品を多くすれば、抽出作業そのものが目的化してしまい、十分な考察まで行えないし、反対に、対象作品を少なくすれば、用例数が少なくなり、やはり十分な考察が行えなかったからである。

また、そもそも「色彩表現」と言っても、どのような表現を「色彩表現」と呼ぶのかを、明示することもなかった。たとえば、「赤い靴」を「色彩表現」だと認定することに問題はないだろう。しかし、「白鳥」や「赤信号」といった用例を「色彩表現」と認定するかどうかは、論議の分かれるところである。だが、そのような用例を「色彩表現」とするかどうかの基準を明確に示さなければ、用例数そのものにまで影響する。「色彩表現」を考察する上では、まず「色彩表現」とは何か、という根本的な問題から考えていく必要がある。

本論文では、従来の「抽出作業の困難さ」と「色彩表現とは何か」という問題点を克服することを第一の目的とした。第一番目の問題点の解決には、まず、パソコンの利用に注目した。パソコンを利用することにより、**大量のデータ**を正確に扱うことが可能になるのである。

第 1 章では、まず以上のような従来の問題点について述べ、「そもそも文学作品における色彩表現に意味があるのか」についても述べた。

第 2 章においては「日本語における色彩表現」と題し、日本語の中にみる色彩表現語彙について、事例を交えながら考察した。「赤」「青」「白」「黒」が、日本語の中に浸透しているという事実を「基本色名」（赤・青・白・黒・黄・緑・紫）の語彙の特徴や、日本人の「姓」に見られる「赤」「青」「白」「黒」を考察することにより、確認した。同時に、日本語における「青」と「緑」の問題についても、英語との比較を視点のひとつに加え、言及した。

第 3 章では、波多野完治に代表される先行研究とその問題点について論じ、本論文作成上の方法論を導き出した。

第 4 章「芥川龍之介作品のテキストデータ化」では、データ抽出の素材となるテキストデータの作成について論じた。ここでは、書籍をどのように機械可読化状態にするのかという方法と、日本語を扱う上での問題点について述べた。テキストデータ化に関しては、『筑摩全集類聚・芥川龍之介全集』（全 8 巻別巻 1）を原典とし、小説のジャンルにあたる 148 作品と、『侏儒の言葉』を含めた 149 作品のテキストデータ化を行った。

欧米語のような分かち書きをしない日本語を処理しようとするれば、まず「語」の認定という問題から始まり、

KWIC (Key Word In Context) の作成すらも容易ではないことがわかる。また、近年利用が増加しつつある OCR も使用し、マニュアル入力との作業比較も行った。結果は、英語 OCR ほど精度は高くなく、マニュアル入力ではおよそ起きないであろう入力ミスも、かなり発見できた。入力に使用する書籍の状態にもかなり左右されるが、現時点での日本語 OCR の利用は、必ずしも効率的ではない。そして、この章の最後には、テキストデータの将来性についても触れた。

第5章「色彩表現の抽出と分類」では、3つの「抽出基準」を設けた。本論文の「色彩表現」がどのようなものであるのかを、ここで明示する。抽出した後には、用例の分類が次の手順となる。従来の方法にしたがった「対象」別分類と、新たに「機能」別分類を設けた。これにより、抽出した色彩表現を、「対象」と文中での「機能」の両面から考察することが可能となる。

第6章では、作成したテキストデータベースを利用して色彩表現の研究を行う。ここでは、すでに作成している「芥川龍之介作品のテキストデータベース」を利用する。抽出対象とするのは、日本の基本色(きほんしきめい)である赤・青・白・黒・黄・緑・紫とする。色彩表現の抽出に際しては、プログラミング言語 PASCAL により、抽出プログラムを作成した。用例からある程度の文脈を読み取ることができるよう、色彩表現を含む行と、その前後行の合計3行を抽出するようにした。このようにして一旦抽出した色彩表現は、第5章で規定した「抽出基準」により「色彩表現」のみを選択し、さらに「機能別分類」と「対象別分類」を行った。

考察にあたっては、まず、「数量的傾向」から行った。149 作品で、どのような色彩表現がどの程度使用されているかを見るのである。色彩表現の使用率を見るために、「文字数」もカウントした。「文字数」のカウントは、色彩表現の抽出プログラムと同様、PASCAL でプログラムを作成した。ここから、「作品の文字数」と「色彩表現の用例数」は、比例しないということがわかった。作品が長編だから、色彩表現が増えるというわけではないのである。むしろ、作品のテーマと直結しているということがわかる。そのことは、色彩表現が見られない作品の内容を確認することでわかる。149 作品中、色彩表現の用例が見られないのは4作品であった。これらはいずれも、「心理描写中心」「時間の経過が実時間でない」というものであった。言い換えるならば、作品が心理描写に傾く、あるいは時間の経過が実時間でない場合は、色彩表現の使用が減少するのである。

次に、「機能別分類」を行った。多用されるのは「連体修飾用法」であった。これは、色彩表現は事物の説明に使用されることが多いという結果と合致する。反対に、最も用例数が少ないのは「述語的用法」であった。これは、「連体修飾用法」の用例数が多いことと、表裏一体である。

また「対象別分類」では、「自然物」「身体」「着物・装身具」の順で用例数が多く見られ、「心理・象徴」に属する用例数はごくわずかである。

「ジャンル別傾向」で見ると、芥川の自伝的内容である「保吉もの」での用例数が多く認められた。

さらに「各色の傾向」により、内容的な考察を行った。7色それぞれを考察することにより、芥川の使用する「色」の意味するものを次のように導き出した。

白…厳肅さ

赤…現実

青…追憶

黒…嫌悪

黄…死

紫…不安と憧憬

緑…自然と安心

最終章では、「結論」として本研究のまとめを行った。

文学作品の中に表れる「色彩表現」には、必ず作者の意図が伴っている。ただし、用例数が多いからと言って、それが特定の意味を含んだ色彩表現とは限らない。それは、かなりステレオタイプの使用で、一般的な描写であることのほうが多い。むしろ、効果を狙うために用例数を少なくしていることがある。しかし、それも全用例を対象として初めて言えることである。

論文審査の結果の要旨

文学作品の中の色彩表現に関する研究は既に多く存在するが、大半は一作品もしくは一部の作品の一部分のみを対象としたもので、作家の全作品を対象としたものはまだ存在しない。本論文は、芥川龍之介の小説ジャンルに属する全作品148編と『侏儒の言葉』の合計149編を収録したテキストデータベースを構築し、色彩表現の明確な抽出規準を設定し、コンピュータ処理を採用することにより、これら全作品を対象にすべての色彩表現を抽出し、芥川の色彩表現に関して種々の角度から考察を行なったものである。その主要な成果は次の通りである。

- (1) 全作品を対象とした色彩表現に関する研究が行なえるように、芥川龍之介の上述の149編の作品を収録したテキストデータベースを構築している。これにより、コンピュータ処理が可能になり、大量のデータを高速に処理することができる。
- (2) 色彩文字を含む表現を3つのタイプ、すなわち、色彩を感じることができ且つ他の機能に置き換えても同一の事物を指示するもの、他の機能に置き換えると別の事物を指示するもの、色彩を感じることのできないもの、に分類する抽出規準を提案している。これにより、文字に基づいた色彩表現の抽出、すなわち、コンピュータによる抽出を可能にしている。
- (3) 芥川龍之介の色彩表現の数量的傾向としては、作品の文字数と色彩表現の用例数は比例しない、すなわち、作品が心理的描写に傾く場合や作品中の時間経過が実時間でない場合は色彩表現の使用が減少すること、「連体修飾用法」が多いこと、「自然物」に対して用いられる用例数が一番多く、「心理・象徴」に属する用例数は少ないこと、芥川の自伝的内容である「保吉もの」での用例数が多いことを確認している。
- (4) 芥川龍之介の色彩表現の質的傾向としては、白は厳肅さ、赤は現実、青は追憶、黒は嫌悪、黄は死、紫は不安と憧憬、緑は自然と安堵を意味していることを明らかにしている。
- (5) さらに全体としては、「白」「黒」「赤」など用例数の多い色彩表現よりも、「黄」「紫」など用例数の少ない色彩表現に芥川龍之介の色彩感の特質が濃厚にあらわれていることを、他の多くの資料も援用して論証している。

ただし、本論文は、作家は「小説の展開において不必要な場面などは描かない」という前提に立って考察を行なっているので、ジョイス（James Joyce）風の「意識の流れ」の手法で書く作家の場合にはこの前提は当嵌らない。従って、そういうケースはどう扱うべきかについての考察が添えられていれば、より説得力をもつであろう。

しかし、全体としては、欧米語と違って分ち書きができず、又、ひらがな・カタカナ・漢字の混在する日本語の文字処理という困難を克服して芥川龍之介の149編の作品のテキストデータ化を遂行しただけではなくて、さらに、そこから抽出されたデータに基いてこの作家の色彩表現の諸特性について綿密な文学的考察を行ない、今後の芥川龍之介研究に新たな一石を投じた本論文の功績は大い。

以上により、本論文を博士（言語文化学）の学位請求論文として十分な価値を有するものと判定する。