

『計量国語学』アーカイブ

ID	KK300801
種別	論文B
タイトル	漢字は無くなるのか —再び「漢字の将来」を問い直す—
Title	Will Chinese Characters Vanish from Japanese Sentences? Reconsidering ‘the Future of Kanji’ Once Again
著者	菅野 倫匡
Author	KANNO Michimasa
掲載号	30巻8号
発行日	2017年3月20日
開始ページ	481
終了ページ	498
著作権者	計量国語学会

論文B

漢字は無くなるのか

一再び「漢字の将来」を問い直す一

菅野 倫匡（筑波大学大学院生）

要旨

22 世紀の末葉に日本語の文章から漢字が無くなるという予測がある。この予測は文章に占める漢字の割合の通時的な調査に基づき、20 世紀の中葉に示されたものであるが、それから約半世紀を経た今、文章に占める漢字の割合が減り続けているかは疑問である。本稿ではこれまでの調査を概観すると共に実際に調査を実施して文章に占める漢字の割合を明らかにし、そこに見られる傾向について分析を試みた。その結果、20 世紀の末葉以降について漢字の割合が大きく変動しているとは言えず、漢字が無くなるという予測の前提となっている漢字の割合が減り続けるという傾向は裏づけられなかった。また、漢字の割合が安定しつつある可能性が改めて示された。

キーワード：漢字の将来、漢字使用の実態、漢字含有率、芥川賞作品

1. はじめに

安本（1963）は文章に占める漢字の割合についての通時的な調査に基づいて 22 世紀の末葉に日本語の文章から漢字が無くなるという予測を立てた。しかし、この予測から凡そ半世紀を経た今もなお、漢字が滅亡への途を歩み続けているかは疑問である。本稿はこの点について調査し、分析を試みるものである。まず、漢字の割合に関する研究の嚆矢たる安本（1963）の研究とそこから派生した研究とを概観し、現状の整理を試みる。

1.1 先行研究

初めに安本（1963）の研究について述べる。安本（1963）は小説文における漢字使用の実態を通時的に明らかにするとして 1900 年から 1954 年までの計 100 篇の小説について各作品の漢字の割合を調査した。その方法とは最初に作家を 100 人選び、次いで各作家の作品から代表作を 1 作品採り、各作品から 1,000 字ずつを抜き出し、それらの漢字の割合を算出するというものである¹。また、本文は『現代日本文学全集』（筑摩書房）に拠る。

¹ 調査の方法は安本（1963）が詳しく述べていないことから安本（1965）によって補った。それによれば、代表作の選び方は複数の文学辞典や作家の評論などにおいて最も共通して挙げられている作品を採るというものである。また、『現代日本文学全集』（筑摩書房）に収められていない作品や長さの短い作品は対象から外される。代表作が選ばれたら 1 つの作品を構成する全ての段から系統抽出法によって計 20 段が選ばれ、各段の冒頭 50 字が抜き出される。なお、安本（1963）は漢字の割合を「漢字の使用率」と呼んでおり、これについては第 2 章において検討する。

安本（1963）は漢字の割合を発表年の若い順に並べて 5 年ごとに平均すると山と谷とを繰り返しながら次第に減少すると述べた。また、漢字の割合が直線的に減少しているとも述べて x を作品の発表年、 y を漢字の使用率とする単回帰分析を実施し、次の式を得た。

$$y = -1.244x + 2726.17 \quad (x: \text{作品の発表年}, y: \text{漢字の使用率})$$

この回帰式に $y = 0$ を代入すれば、2191 年に文章に占める漢字の割合が 0 となり、漢字が滅亡するという前掲の結論が導かれる²。

続いて安本（1963）の研究から派生した研究を「調査に主眼を置いた研究」と「手法に疑問を呈した研究」とに大別し、それぞれの概要を以下に述べる³。更に安本（1963）の研究と深く関連する「語の表記に着眼した研究」についても併せてその概要を示す。

1.1.1 調査に主眼を置いた研究

初めに調査に主眼を置いた研究として宮島（1988a）と中村（2003）とを取り上げる。両者は安本（1963）以後の調査という点では共通しているが、文章に占める漢字の割合の減少傾向がその後も続いているかという点においては対蹠的な立場を取る。

まず、宮島（1988a）の研究について述べる。宮島（1988a）は安本（1963）の予測が示されてから四半世紀が経ったことを踏まえ、これを検討するとして 1935 年度から 1985 年度までの計 94 篇の芥川龍之介賞受賞作品（以下、芥川賞作品）の漢字の割合について調査した⁴。その方法は各作品から 1,000 字ずつを抜き出し、それらに占める漢字の割合をそれぞれ算出するというものである⁵。また、本文は『文藝春秋』掲載時のものに拠る⁶。

宮島（1988a）は安本（1963）に倣って漢字の割合を 5 年ごとに平均し、漢字が大局的には減ってきたと結論した。また、雑誌や新聞を対象とした漢字の割合についての調査を援用し、20 世紀前半に減少傾向を示すという点が共通していると述べて安本（1963）の調査結果は裏づけられたという結論を下した。ただし、1960 年以降は芥川賞作品の漢字の割合が減少していないとも述べて 5 年から 10 年後の漢字の割合は安定ないし微増すると予想している。また、漢字の割合と受賞年や誕生年との相関係数を算出して報告した⁷。

² 安本（1963）によれば、この式は漢字の使用率が直線的に減少しており、なおかつ、今後も同様の傾向が続くという仮定に基づくものである。なお、安本（1963）はこの式に加えて 2 つの予測式を示しているが、本稿では安本（1963）がこの式について実際の漢字の使用率に最も合致しているとして式の前提となる仮定を受け容れるとしたことからこの式のみを取り上げる。

³ 安本（1963）の研究から派生した研究は数多くあり、それらを網羅するのは困難であることからここでは安本（1963）と同様に小説を対象とした研究のみを取り上げる。また、これは小説と新聞などの小説以外のものとで傾向が異なる可能性を考慮してのことでもある。

⁴ 宮島（1988a）は作品の選び方によって結果が変わる可能性を指摘しており、調査を継続するためには作品の基準が一定である方が望ましいと述べて芥川賞作品をその基準とした。

⁵ 宮島（1988a）は各作品の最初のページから可能な限り、長い間隔、かつ、等間隔に 10 ページを決め、該当するページの 3 行目の先頭からそれぞれ 100 字を抜き出している。また、宮島（1988a）は漢字の割合を「漢字含有率」と呼んでおり、これについても第 2 章において扱う。

⁶ これは『芥川賞全集』（文藝春秋）のように発表後に改めて編まれたものを使用するとその表記が『文藝春秋』発表時のそれと著しく異なり、表記の調査として問題があることに因る（宮島 1988a）。

⁷ 受賞年との相関係数は -0.463 であり、誕生年とのそれは -0.454 である。ここから宮島（1988a）は漢字使用が漢字を習得する時期の習慣によって定まるとの仮説は成り立たないとした。

続いて中村（2003）の研究について述べる。中村（2003）は文章に占める漢字の割合の調査として中村（1958）の調査⁸と安本（1965）の調査⁹とを挙げ、いずれの調査も最近の作品が含まれていないことを指摘した上で漢字使用の実態を通覧するとして 1948 年から 2000 年までの計 34 篇の小説について漢字の割合を調査した。その方法は対象としたそれぞれの作品から一定の間隔で 5 ページずつを抜き出し、それらに占める漢字の割合を算出するというものである¹⁰。また、本文は各作品の単行本や文庫本、全集本などに拠る。

中村（2003）は 34 篇の小説に対する調査結果を前掲の 2 つの調査結果に加えて計 170 篇¹¹の小説における漢字の割合を列挙し、いくつかの区分を設けて各区分の平均値を示し、その結果から漢字の割合が月日を経るごとに減少してきたと結論した。

1.1.2 手法に疑問を呈した研究

次に手法に疑問を呈した研究として綾（2013）と石井（2013）とを取り上げる。両者は共に安本（1963）の研究と宮島（1988a）の研究とを取り上げ、特に安本（1963）の手法に対して疑問を呈するという点では共通しているが、問題とする点が異なっている。

初めに綾（2013）の研究について述べる。綾（2013）は安本（1963）の採用した予測式を用いる手法について結果の示し方とデータの扱い方が問題であると指摘した。前者の問題点は散布図が示されておらず、相関係数¹²や決定係数¹³なども記されていないことであり、後者のそれは調査期間の凡そ 5 倍も外挿していることである。綾（2013）によるとこのような統計的な問題を有する安本（1963）の予測は信頼し得ないものである。

続いて石井（2013）の研究について述べる。石井（2013）は探索的データ解析¹⁴の手法を用いて安本（1963）が漢字の割合の減少傾向を捕捉する過程を検討し、安本（1963）の手法が抱える問題点を指摘した。その問題点とは漢字の割合を 5 年という等間隔¹⁵の区分に分け、各区分の代表値に平均値¹⁶を採用し、各区分の代表値の変動を全体の傾向¹⁷を代表するものと看做したということである。

⁸ これは漢字の割合の検討を主としたものではないが、1898 年から 1955 年までの計 50 篇の小説について各作品の冒頭 100 字に占める漢字の割合を示している。

⁹ 安本（1965）を挙げているが、漢字の割合に関する部分は安本（1963）と同様である。

¹⁰ 中村（2003：8）は対象とする作品について「新しい作品に重点をおいて選定した」と述べるに留まっており、作品の基準は不明確である。また、中村（2003）は漢字の割合を「漢字使用率」と呼んでおり、これについても第 2 章において触れる。

¹¹ 中村（2003）は中村（1958）と安本（1965）とに重複する作品の数を 13 と記しているが、横光利一「日輪」は中村（1958）にも収められており、これを足すと 14 である。また、「日輪」の成立年が 1929 年となっているが、これは安本（1965）にもあるように 1923 年が正しい。

¹² これは宮島（1988a）も再計算の結果を示しており、いずれも -0.377 で一致している。

¹³ 綾（2013）が安本（1963）の調査結果から再計算した決定係数は $.114$ である。なお、再計算の結果によると回帰係数や切片の値が若干異なっている。

¹⁴ 石井（2013：130）によれば、これは「十分な数の無作為標本と正規分布する母集団とを前提とせず、どのような種類のデータでも解析の対象とする」もので「手元の限られたデータから何らかの仮説的な情報を得る」ことを可能にする手法である。なお、詳細については渡部・鈴木・山田・大塚（1985）に述べられている。

¹⁵ 観測値の数に偏りが生じ、その数が少ない区分では代表値が不安定となる（石井 2013）。

¹⁶ 平均値は外れ値の影響を受けやすい（石井 2013）。

¹⁷ 代表値の変動には様々なノイズの影響が含まれていることから代表値以外の変動についても把握することによって観測値に見られる傾向を確実に掴む必要がある（石井 2013）。

石井 (2013) はこれらの問題を解消するために各区分を構成する観測値の数を等しくし、代表値に平均値よりも頑健な中央値を採用した。また、五数要約値¹⁸を示して箱ひげ図を描き、更にこれを平滑化して図示している¹⁹。その結果、小説における漢字の割合は全体として大きな S 字カーブを描きながら減少するという結論が示された。

1.1.3 語の表記に着眼した研究

最後に語の表記に着眼した研究について杉山・森岡 (1969) と野村 (1988b) とを取り上げる。これらの研究は漢字の割合に関する直接的な研究ではないものの安本 (1963) の研究とそこから派生した研究とに深く関連するものである。

まず、杉山・森岡 (1969) の研究は安本 (1963) の研究について一定の意義を認めるとしながらも安本 (1963) のように字を単位とするのではなく、語を単位とする調査が必要であるとして 1879 年から 1968 年までの『朝日新聞』の社会面から抜き出した和語と漢語とについて漢字が用いられているか否かをそれぞれ調査したものである²⁰。その結果から杉山・森岡 (1969) は漢字表記の語が減少していることを指摘しており、その要因として和語が相対的に増加していることや漢字表記の和語が減少していることを挙げている。

続いて野村 (1988b) の研究について述べる。野村 (1988b) は日本語の文章から漢字が無くなるという安本 (1963) の予測について小説における漢字の割合が安定しつつあるとした宮島 (1988a) の研究に基づけば、修正する必要があると述べる一方、漢字使用の実態の解明や予測には小説を対象とする場合に見られない変化が小説以外のものを対象とした場合に現れる可能性をも検討する必要があると指摘している。また、これとは異なる観点として漢字の割合の変動する要因が杉山・森岡 (1969) の調査のように語種の構成比や語種ごとの表記を明らかにすることによって採り得ることを指摘し、この点についても留意する必要があると述べている。更に野村 (1988b) は外来語の増加などの 5 つの仮定を設けて漢字の割合が今後も減り続けるとの予想を示している²¹。

1.2 現状の整理

ここでは漢字の割合に関する研究の現状について整理する。まず、安本 (1963) の研究は小説の文章における漢字の割合の調査に減少傾向が見られることを報告し、この傾向が今後も続くならば、小説の文章から漢字が無くなるとして日本語の文章からも漢字が無くなる可能性を示唆したものであった。これについて本稿は「この傾向が今後も続くか」という点と「小説の文章における漢字の割合の減少傾向が今後も続くならば、日本語の文章から漢字が無くなるか」という点とを改めて問い直してみる必要があると考える。以下ではそれぞれの問題について現状の整理を試みる。

初めに前者の問題について述べる。これまでに見てきた小説の文章を対象とした調査の結果を整理すると 20 世紀の初葉から中葉に掛けて小説の文章に占める漢字の割合が減少したという点では一致している。しかし、その傾向がその後も続くかという点においては宮島 (1988a) の予想と中村 (2003) の結果とが対立している。

¹⁸ 中央値、最大値、最小値、上ヒンジ、下ヒンジの 5 つを指す (石井 2013)。

¹⁹ これは渡部・鈴木・山田・大塚 (1985) の「蛇行箱型図」という手法である。

²⁰ 紙幅の都合上、具体的な研究の方法について本稿では扱わない。

²¹ どのような仮定を設けるかという点は更なる検討を要するものである。

また、手法については綾（2013）や石井（2013）が安本（1963）のそれに対してそれぞれ疑問を投げ掛けており、特に石井（2013）の指摘は宮島（1988a）や中村（2003）の手法にも該当するものである。これらの指摘を考慮すれば、20 世紀後半の小説文における漢字の割合の変遷についての見解の相違は対象の選び方や漢字の割合の算出方法が異なることに加え、統計的な手法の扱い方に問題があったことにも起因するのではないか。

更に 2000 年以降の動向について報告する調査が管見の限りでは存在せず、現在に至るまで小説の文章における漢字の割合がどのような変遷を辿ってきたのかは明らかになっていないという現状がある。すなわち、これまでに述べてきた問題を克服し、その後の動向を明らかにする調査が早急に求められている²²。

次に後者の問題について述べる。まず、安本（1963）は小説の文章における漢字の割合の減少傾向が今後も続くならば、小説から漢字が無くなるとしたが、これは漢字の割合がある非負の定数に上から漸近するような場合が反例として与えられるから偽である²³。

また、安本（1963）は小説の文章から漢字が無くなるならば、日本語の文章からも漢字が無くなるとしているが、これは野村（1988b）が漢字使用の実態の解明や予測には小説において見られない変化が小説以外のものにおいて現れる可能性を考慮する必要があると指摘しているように日本語の文章が全て小説ではない限り、偽である。

このように小説の文章において見られる傾向が日本語の文章についても同様に見られるものか否かは慎重に吟味することが求められている²⁴。

1.3 研究の目的

本稿ではこのような現状を踏まえ、小説の文章に占める漢字の割合を通時的に調査することを目的とする。また、その変動については統計的な分析に基づいて考察する。

確かに土屋（2000a）が指摘するように漢字の割合は複数の要因によって変動し、その要因を特定することが困難であるという一面もある。しかし、表記の実態を反映する 1 つの指標としての漢字の割合は通時的に比較することが可能であり、そのことによって表記の変遷を描き出せるという特徴もある（佐竹 1987）。また、漢字の割合が変動する要因は野村（1988b）が指摘するように語を単位とする調査を実施し、語種の構成比や語種ごとの表記を明らかにすることによって探り得るものであり、本稿の調査はそのような調査の前段階と位置づけられるものである。それゆえ、本稿ではこの指標を用いることの得失の検討は一旦置き、表記の変遷を素描する試みの一環として特に漢字使用に着眼する。

2. 研究の方法

研究を進めるにあたり、その方法を以下に述べる。ここでは初めに用語の定義を明確にし、次いで本稿において実施する調査の概要を記し、最後に分析の手法を説明する。

²² 犬飼（2002）は統計的な問題には触れていないものの漢字の割合について宮島（1988a）以後の動向を明らかにすることは研究に値する課題であると述べている。

²³ これは漢字の割合が直線的な減少傾向を今後も示し続けるという安本（1963）の仮定に依拠するならば、真であるが、本稿はこの仮定に依拠するものではない。また、これがもし仮に真であったとしてもそのことは本稿の議論に影響するものではない。

²⁴ この点については本稿の第 4 章において改めて検討する。

2.1 用語の定義

まず、本稿における用語の定義を示す。ここでは最初に「字種」の定義を明らかにし、次に漢字の割合の算出方法を比較して本稿が採用する算出方法を明確に示す。

2.1.1 字種の定義

初めに「字種」について扱う。「字種」には樺島（1977）の言うところの「文字体系²⁵」を指す場合と「文字素²⁶」を指す場合とがある（佐藤 2013）。先行研究では宮島（1988a）のみが「字種」に言及しており、各作品から抜き出された 1,000 字について「字種ごとの数」が掲げられている。そこには内訳としてひらがな、カタカナ、ローマ字、漢字という 4 つの項目が立てられており、このことから宮島（1988a）の「字種」が前者を指すものと読み取れる。本稿ではこれを踏襲し、「字種」は「文字体系」を指すものとする。

2.1.2 漢字含有率の算出方法

次に先行研究で述べられている文章²⁷に占める漢字の割合の算出方法を整理し、本稿が採用する「漢字含有率」の定義を明確にする。なお、そのために以下ではひらがな、カタカナ、漢字、ローマ字、算用数字の各集合をそれぞれ H, K, C, R, A と置き、文章 T における各文字体系 G の延べ字数²⁸を $N_T(G)$ 、異なり字数²⁹を $V_T(G)$ とそれぞれ表現する³⁰。

まず、安本（1963）の「漢字の使用率」の算出方法を以下に示す³¹。なお、延べ句点数と延べ読点数とをそれぞれ暫定的に $N(\{. \})$ 、 $N(\{, \})$ とした³²。更に安本（1965）は漢字以外の字種の内訳を示しておらず、これを本稿では $N(H \cup K \cup R \cup A)$ と考えた³³。

$$\text{漢字の使用率} = \frac{N(C)}{N(H \cup K \cup R \cup A) + N(C) + N(\{. \}) + N(\{, \})} \times 1000$$

続いて宮島（1988a）の「漢字含有率」のそれを次に示す。なお、この研究ではローマ字に算用数字を含めており、これを本稿では延べ英数字数とし、 $N(R \cup A)$ と表現した。

²⁵ 樺島（1977：26）においては「発生的、歴史的に一つのまとまりを形作る文字の集合」であると述べられており、「漢字、ひらがな、カタカナ、ラテン文字、ハングルなど、それぞれを集合として考えると、これらは文字体系である。」

²⁶ 樺島（1977：34）によれば、「同じと判定される文字の集合について、字形の異なりを捨象して得られる文字観念」である。

²⁷ 本稿における「文章」とは樺島（1977）の「記号列」に相当する。

²⁸ 野村（1988a：340）によると「その調査に出現した漢字の総数。同一の漢字が反復して出現する場合には、それぞれ 1 回として数えた場合の数」である。本稿ではこのうちの「漢字」を「文字」に読み替える。また、何を以て同一の文字と看做するかという点は本稿では扱わない。

²⁹ 野村（1988a：341）によると「その調査に出現した漢字の種類を示す数。同一の漢字が反復して出現しても、それを 1 回として数えた場合の数」である。延べ字数の場合と同様に本稿においては「漢字」を「文字」に読み替える。

³⁰ 文章 T を逐一明示すると煩雑になるために以降では省略する。

³¹ 安本（1963）が詳細を述べていないことからここでも安本（1965）を手掛かりにした。

³² 樺島（1977）の定義の上では句読点は「文字」に該当するものではない。それゆえ、前掲の定義に従えば、延べ句点数と延べ読点数とを定義することは不可能である。本稿はこの問題を一旦留保し、仮に $N(\{. \})$ と $N(\{, \})$ とそれぞれ置いた。

³³ 安本（1965）はローマ字や算用数字を含めるとは述べていないが、含めないとも述べておらず、本稿では暫定的に含められていると考えた。また、句読点以外の記号の扱いも不明である。

$$\text{漢字含有率} = \frac{N(C)}{N(H) + N(K) + N(C) + N(R \cup A)} \times 1000$$

最後に中村（2003）の「漢字使用率」のそれである。この研究ではひらがなとカタカナとを合わせて仮名としており、本稿では延べ仮名数を $N(H \cup K)$ とした。

$$\text{漢字使用率} = \frac{N(C)}{N(H \cup K) + N(C)} \times 1000$$

これらの分子はいずれも延べ漢字数であるのに対し、分母は算出方法によって異なっている。このような分母の差が文章に占める漢字の割合に与える影響を凡そ 3% とする報告もあり³⁴、算出方法をどのように定義するかは調査の結果に影響する。また、それを定義することは本稿の位置づけをも左右するものである。

本稿は漢字使用の実態の変遷について検討するために文章に占める漢字の割合について分析するものであるから漢字とそれに置き換わる可能性を有する他の字種との関係は考慮する必要がある。また、漢字と直接的に喰い合う関係にあるとは考えにくい句読点や感嘆符や括弧や中黒のような文字以外のものは本稿の対象に含まれない。これらの点から本稿では文章に占める漢字の割合の算出方法として「漢字含有率」を採用する。

2.2 調査の概要

ここでは本稿において実施する調査の概要を述べる。本稿における調査は概ね宮島（1988a）の調査を踏襲するものである。これは宮島（1988a）が調査の結果に対する再現や検証を可能にするために対象とする作品を選定する基準や漢字含有率の算出方法を明確に述べていることに因る。以下に本稿において実施する調査の対象と手順とを述べる。

2.2.1 調査の対象

まず、調査の対象は作品の基準を一定にするという前掲の宮島（1988a）の方針を踏襲し、芥川賞作品における漢字含有率とする。また、既に宮島（1988a）が 1935 年度から 1985 年度までの作品を調査し終えていることから 1986 年から 2015 年までの計 66 篇の芥川賞作品について調査を実施する³⁵。なお、対象とする本文は前掲した宮島（1988a）の指摘に基づき、『文藝春秋』掲載時のものに拠る。

次に各作品からそれぞれ 1,000 字を抜き出す。これは初めに各作品から 10 ページずつを選び³⁶、次いで該当する各ページの 3 行目の先頭からそれぞれ 100 字を対象とするものである。なお、章や節を示す語や本文前にある引用などは行数から除外する。

³⁴ 野村（1980）は 1979 年の雑誌を対象とした調査において算出方法の差異による変動を検討し、この結論を得た。なお、どの程度変動するかは改めて検討する必要がある。

³⁵ 宮島（1988a）は「年度」を単位とし、本稿は「年」を単位としている。その結果、米谷ふみ子「過越しの祭」が双方の調査に含まれている。本稿においてはこの作品の字種ごとの延べ字数が宮島（1988a）と本稿とで一致していることを以て算出方法の再現性を確認した。

³⁶ 各作品の最初のページを p_f とし、最後のページを p_l とすると各作品から抜き出すページは $n = 1 \sim 10$ のとき、 $p_f + [(p_l - p_f)/9](n - 1)$ である。

これについて更に説明を加えるために図 1 に実際の『文藝春秋』に載せられている藤野千夜「夏の約束」から冒頭の 1 ページを例として示した。図中にある「1」のような部分が行数に含めない章や節を示す語である。また、これを除いて 3 行目に当たる箇所から 100 字に網掛けを施した。この図から明らかなように 100 字を抜き出す際には疑問符や鉤括弧などの文字以外のものを取り除いてある。なお、抜き出された 100 字のうち漢字に該当するのは黒い網掛けの部分である。

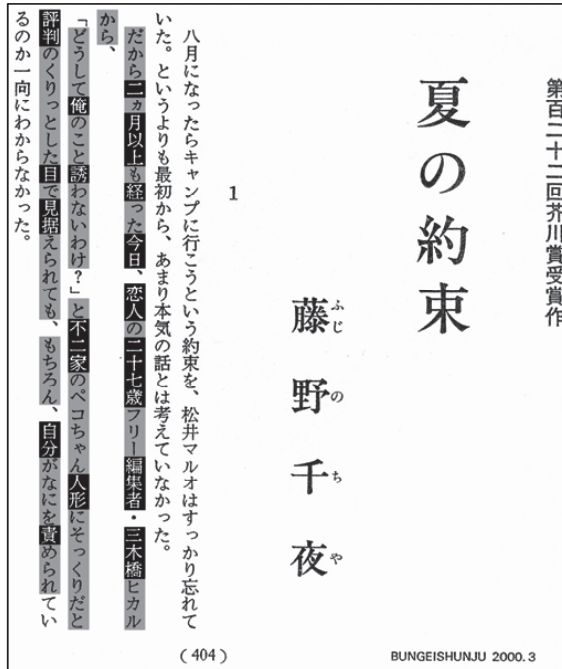


図 1：藤野千夜「夏の約束」の冒頭の 1 ページ³⁷（一部）

2.2.2 調査の手順

次に調査を実施する手順について簡単に述べる。初めに光学文字認識ソフトウェア読取革命 (Ver.15) を用いて各作品が載せられている『文藝春秋』の本文から振り仮名を除外した電子テキストを生成する。次いでプログラミング言語 Python (Ver.2.7.6) を用いて各作品のテキストから文字以外のものを取り除き、漢字含有率を算出する。

なお、漢字含有率の算出にあたり、テキストと本文とを校合し、読み取りの誤りがある場合には適宜、これを修正した。また、「々」は漢字とし、「一（長音符）」は前接する字種の繰り返しとして処理した。更に「～」が 1 例あり、これについても「一」と同様のものと考えて前接する字種の繰り返しとする処理を施した。また、対象とする字種は漢字含有率の分母と同様にひらがな、カタカナ、漢字、英数字である。

³⁷ これは 2000 年 3 月発行の『文藝春秋』78 巻 4 号（三月特別号）404 ページから引用したものである。また、網掛けは説明のために施したものであり、実際のページに示されているものではない。

2.3 分析の手法

本稿では調査によって明らかにした芥川賞作品の漢字含有率に対し、いくつかの分析を試みる。まず、各字種の変遷を散布図によって示し、調査の全体を概観する。また、相関分析を実施して漢字とその他の字種との関係や通時的な変化について分析する。

次に漢字含有率に着眼し、そこに見られる変動を Kruskal-Wallis 検定によって分析する。更に箱ひげ図やその平滑化によって漢字含有率の変遷を描き出す。なお、これらの検定や作図にはプログラミング言語 R (Ver.3.0.2) を使用し、手法の概要は以下に述べる。

2.3.1 Kruskal-Wallis 検定

初めに本稿が採用する Kruskal-Wallis 検定について述べる。これは独立な k 個の標本が異なる母集団分布から得られたものか否かを調べるためのノンパラメトリック検定である (Kruskal & Wallis 1952)。すなわち、この手法の特徴は正規分布のような特定の分布を仮定せずに用いることが可能であるという点にある (田中・垂水 1999)。

このような手法を採用する背景には本稿において用いる漢字含有率がどのような分布に従うか明らかになっていないことがある。また、正規分布を仮定する手法を採用するために変数変換を実施すると誤った結論が導かれる可能性もある (久保 2012)。これらの問題を可能な限り、克服するためには Kruskal-Wallis 検定の採用が求められる。

また、本稿では広く用いられる χ^2 近似による検定が保守的な傾向を示すという名取 (2014) の指摘に基づき、NSM3 という R のパッケージを用いて Kruskal-Wallis 検定を実施し、試行回数を 1,000,000 回とするモンテカルロ近似による検定を選択する。

2.3.2 箱ひげ図の平滑化

次に作図の手法について説明する。なお、ここでは箱ひげ図それ自体について扱うことは避け、前掲した箱ひげ図の平滑化について簡単に述べる。

まず、この手法は渡部・鈴木・山田・大塚 (1985: 81) によると「変数 X (横軸) の値の変化に従って、変数 Y (縦軸) の値がどのように変化していくかそのようすを図示したもの」である。また、これは単に箱ひげ図³⁸を描くよりも通時的な変化やそこに見られる傾向を読み取ることに適しており、同一時点に複数の観測値がある場合にも有効な手法であるという利点がある (石井 2013)。このことは漢字含有率の通時的な変化やその傾向を分析するという本稿の目的に適うものであり、本稿においても採用する理由となる³⁹。

3. 漢字含有率の変遷

ここでは初めに 1986 年から 2015 年までの芥川賞作品における漢字含有率の変遷を示し、次いで統計的な分析を実施し、そこに見られる傾向について明らかにする。

3.1 調査の結果

まず、各作品から抜き出した 1,000 字に占める字種ごとの延べ字数について表 1 と表 2 とに示す。また、漢字については延べ字数に加えて異なり字数も併記した。

³⁸ 渡部・鈴木・山田・大塚 (1985) は明示的に「平行箱型図」と呼び分けている。

³⁹ 推測統計学の手法である Kruskal-Wallis 検定と探索的データ解析の手法である蛇行箱型図とを併用する理由については第 3 章において述べる。

表 1: 芥川賞作品 (1986~2000) から抜き出した 1,000 字に占める字種ごとの延べ字数

受賞年度	誕生年	受賞年	受賞作品	N(H)	N(K)	N(R U A)	N(C)	V(C)
1985 下	1930	1986	米谷ふみ子「過越しの祭」	649	85	0	266	182
1986 上		1986	該当作なし					
1986 下		1987	該当作なし					
1987 上	1945	1987	村田喜代子「鍋の中」	778	6	0	216	142
1987 下	1945	1988	池澤夏樹「スティル・ライフ」	684	47	0	269	189
1987 下	1930	1988	三浦清宏「長男の出家」	662	48	0	290	185
1988 上	1946	1988	新井満「尋ね人の時間」	640	45	0	315	206
1988 下	1951	1989	南木佳士「ダイヤモンドダスト」	618	75	0	307	210
1988 下	1955	1989	李良枝「由熙」	678	53	0	269	153
1989 上		1989	該当作なし					
1989 下	1958	1990	大岡玲「表層生活」	620	66	2	312	205
1989 下	1939	1990	滝澤美恵子「ネコババのいる町で」	737	34	0	229	152
1990 上	1945	1990	辻原登「村の名前」	709	15	0	276	189
1990 下	1962	1991	小川洋子「妊娠カレンダー」	698	60	0	242	170
1991 上	1944	1991	辺見庸「自動起床装置」	682	72	1	245	154
1991 上	1956	1991	荻野アンナ「背負い水」	673	39	0	288	200
1991 下	1961	1992	松村栄子「至高聖所」	652	85	0	263	202
1992 上	1955	1992	藤原智美「運転土」	626	118	0	256	159
1992 下	1960	1993	多和田葉子「大婿入り」	700	27	0	273	195
1993 上	1957	1993	吉目木晴彦「寂寥郊野」	587	69	6	338	229
1993 下	1956	1994	奥泉光「石の来歴」	589	28	0	383	285
1994 上	1955	1994	室井光広「おどるでく」	623	73	0	304	203
1994 上	1956	1994	笠野頼子「タイムスリップ・コンビナート」	670	86	2	242	170
1994 下		1995	該当作なし					
1995 上	1956	1995	保坂和志「この人の関」	708	51	0	241	152
1995 下	1947	1996	又吉栄喜「豚の報い」	644	66	0	290	171
1996 上	1958	1996	川上弘美「蛇を踏みむ」	689	62	0	249	150
1996 下	1959	1997	辻仁成「海峡の光」	580	35	0	385	255
1996 下	1968	1997	柳美里「家族シネマ」	646	49	0	305	206
1997 上	1960	1997	目取真俊「水滴」	611	32	0	357	241
1997 下		1998	該当作なし					
1998 上	1959	1998	藤沢周「ブエノスアイレス午前零時」	515	232	0	253	177
1998 上	1955	1998	花村萬月「ゲルマニウムの夜」	644	40	0	316	228
1998 下	1975	1999	平野啓一郎「日蝕」	601	15	0	384	277
1999 上		1999	該当作なし					
1999 下	1965	2000	玄月「蔭の棲みか」	642	100	0	258	183
1999 下	1962	2000	藤野千夜「夏の約束」	654	96	0	250	177

表 2 : 芥川賞作品 (2000~2015) から抜き出した 1,000 字に占める字種ごとの延べ字数

受賞年度	誕生年	受賞年	受賞作品	N(H)	N(K)	N(R U A)	N(C)	V(C)
2000 上	1962	2000	町田康「きれぎれ」	660	57	3	280	203
2000 上	1954	2000	松浦寿輝「花腐し」	698	56	0	246	179
2000 下	1958	2001	青来有一「聖水」	634	74	0	292	219
2000 下	1964	2001	堀江敏幸「熊の敷石」	630	82	0	288	208
2001 上	1956	2001	玄侑宗久「中陰の花」	766	37	0	197	141
2001 下	1972	2002	長嶋有「猛スピードで母は」	661	48	2	289	184
2002 上	1968	2002	吉田修一「パーク・ライフ」	632	95	0	273	185
2002 下	1966	2003	大道珠貴「しょっぱいドライブ」	791	34	0	175	120
2003 上	1961	2003	吉村萬老「ハリガネムシ」	595	92	1	312	228
2003 下	1983	2004	金原ひとみ「蛇にピアス」	603	148	8	241	163
2003 下	1984	2004	綿矢りさ「蹴りたい背中」	645	85	4	266	192
2004 上	1970	2004	モブ・ノリオ「介護入門」	631	13	0	356	223
2004 下	1968	2005	阿部和重「グランド・フィナーレ」	718	17	0	265	197
2005 上	1977	2005	中村文則「土の中の子供」	670	37	0	293	204
2005 下	1966	2006	絲山秋子「沖で待つ」	689	52	4	255	172
2006 上	1971	2006	伊藤たかみ「八月の路上に捨てて」	651	63	0	286	187
2006 下	1983	2007	青山七恵「ひとり日和」	713	49	0	238	164
2007 上	1969	2007	諏訪哲史「アサッテの人」	621	111	0	268	187
2007 下	1976	2008	川上未映子「乳と卵」	749	25	0	226	145
2008 上	1964	2008	楊逸「時が滲む朝」	570	71	17	342	224
2008 下	1978	2009	津村記久子「ポトスライムの舟」	680	72	2	246	170
2009 上	1965	2009	磯崎憲一郎「終の住処」	713	26	0	261	186
2009 下		2010	該当作なし					
2010 上	1974	2010	赤染晶子「乙女の密告」	610	107	0	283	145
2010 下	1984	2011	朝吹真理子「きことわ」	701	33	0	266	152
2010 下	1967	2011	西村賢太「苦役列車」	661	39	0	300	223
2011 上		2011	該当作なし					
2011 下	1972	2012	円城塔「道化師の蝶」	661	42	2	295	216
2011 下	1972	2012	田中慎弥「共喰い」	663	19	0	318	214
2012 上	1976	2012	鹿島田真希「冥土めぐり」	649	77	0	274	176
2012 下	1937	2013	黒田夏子「ab さんご」	877	0	2	121	83
2013 上	1980	2013	藤野可織「爪と目」	748	53	0	199	128
2013 下	1983	2014	小山田浩子「穴」	713	31	1	255	170
2014 上	1973	2014	柴崎友香「春の庭」	589	90	1	320	207
2014 下	1970	2015	小野正嗣「九年前の祈り」	701	56	0	243	181
2015 上	1980	2015	又吉直樹「火花」	674	12	0	314	186
2015 上	1985	2015	羽田圭介「スクラップ・アンド・ビルド」	600	60	0	340	222

3.2 結果の分析

次に結果の分析を試みる。まず、調査結果を概観するために図2に字種ごとの延べ字数の変遷を示す。この図からはいずれの作品においても $N(H) > N(C) > N(K) > N(R \cup A)$ という関係が成立することが読み取れる。

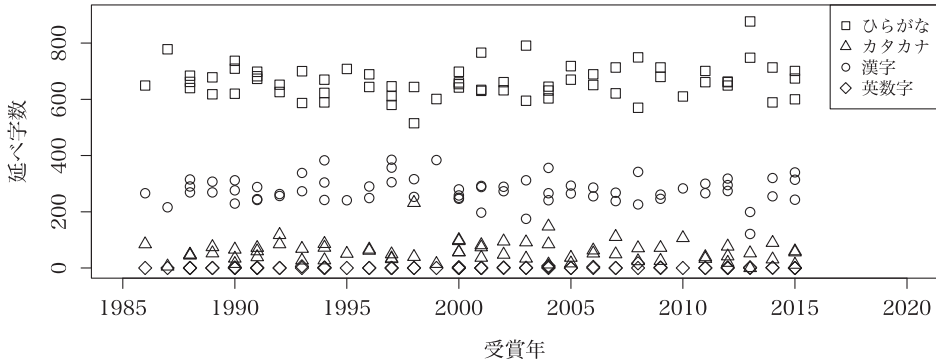


図2：字種ごとの延べ字数の変遷

以下ではこれらの調査結果に対する分析として初めに相関分析を実施して全体の動向を概観し、次いで Kruskal-Wallis 検定を用いて漢字含有率の変動について分析する⁴⁰。更に箱ひげ図の平滑化によって漢字含有率の変遷を素描する。

3.2.1 相関分析

ここでは散布図と相関分析とを用いて全体の動向を概観する。初めに延べ漢字数とその他の字種の延べ字数との関係を観察し、次いで作品の受賞年や作者の誕生年と延べ漢字数との関係について考える。

まず、延べひらがな数と延べ漢字数との関係を図3に示し、延べカタカナ数と延べ漢字数との関係を図4に示す。

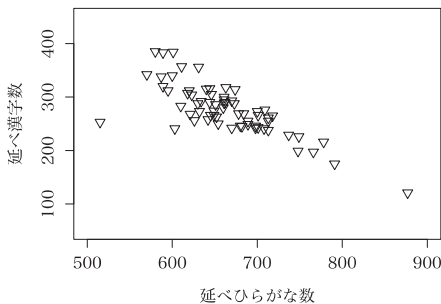


図3：延べひらがな数と延べ漢字数との関係

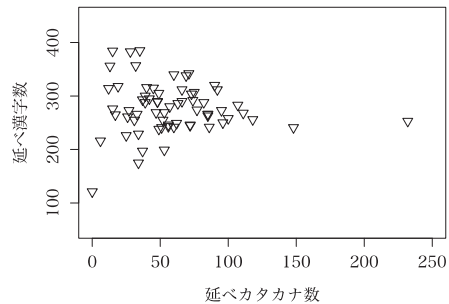


図4：延べカタカナ数と延べ漢字数との関係

⁴⁰ 本稿では1,000字に占める延べ漢字数の値は定義によって漢字含有率の値と一致する。

これらの関係についてそれぞれ相関分析を実施した結果、図 3 に示した延べひらがな数と延べ漢字数との間には強い負の相関が認められた ($r = -.773, p < .001$)。また、図 4 に示した延べカタカナ数と延べ漢字数との間には相関が見られず ($r = -.066, p = .528, n.s.$)、このことは純文学系の雑誌において文章に占める漢字の割合と無関係にカタカナの割合が一定で漢字とひらがなの間に「補完性」があるとする佐竹 (1982) の報告と一致するものである。

更に紙幅の都合上、図は省略するが、延べひらがな数と延べカタカナ数との間には負の相関が認められた ($r = -.581, p < .001$)。

続いて作品の受賞年や作者の誕生年と延べ漢字数との関係について検討する⁴¹。初めに作品の受賞年と延べ漢字数との関係を図 5 に示す。

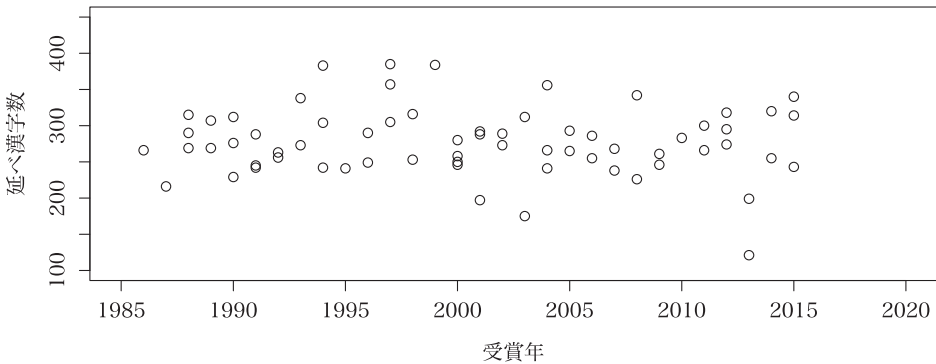


図 5 : 受賞年と延べ漢字数との関係

図 5 に示した受賞年と延べ漢字数との関係について相関分析を実施した結果、両者の間に相関は見られなかった ($r = -.081, p = .516, n.s.$)。

次に作者の誕生年と延べ漢字数との関係を図 6 に示す。

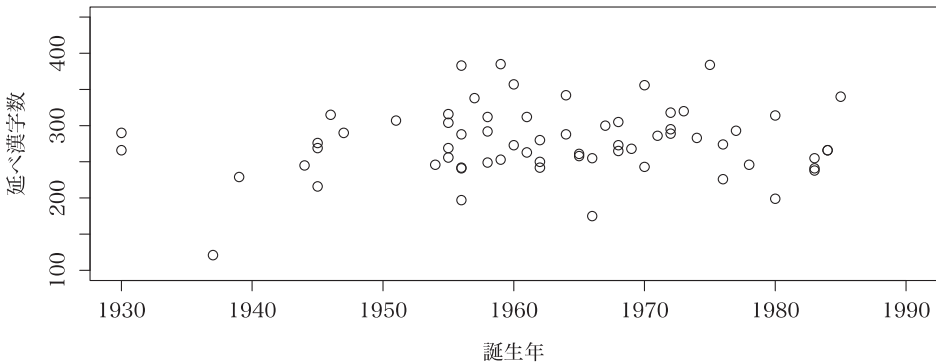


図 6 : 誕生年と延べ漢字数との関係

⁴¹ 作者の誕生年について本稿では『文藝春秋』に書かれているものを採用した。

図 6 に示した誕生年と延べ漢字数との関係についても相関分析を実施した結果、両者の間に相関は見られなかった ($r = .122, p = .330, n. s.$). すなわち、相関分析の結果からは作品の受賞年についても作者の誕生年についても漢字含有率が通時的な減少傾向を示しているという結果は得られなかった⁴².

3.2.2 Kruskal-Wallis 検定

更に漢字含有率の変動について検討を加える. 初めに全ての漢字含有率を 6 つの区分に分割する. これは安本 (1963) や宮島 (1988a) の研究において文章に占める漢字の割合が 5 年ごとに区切られていることに対応する. なお, 石井 (2013) の指摘に基づいて区分ごとの数が均等になるように分割した. すなわち, 表 1 と表 2 とを連結した表の上から順に 11 個ずつを各区分とするというものである. 各区分の箱ひげ図は図 7 に示した. 図に示した各区分について Kruskal-Wallis 検定を実施した結果, 有意差は認められなかった ($H = 2.019, p = .855, n. s.$). この結果から各区分の間に大きな差があるとは言えず, 漢字含有率が大きく変動しているとは言えないという結論が導かれた.

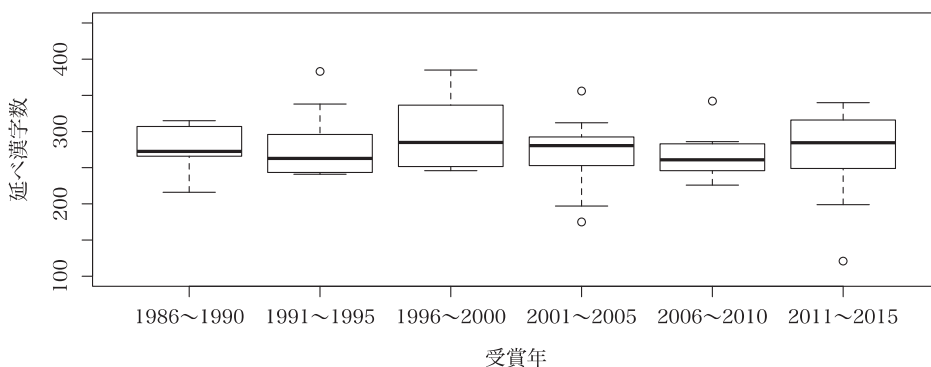


図 7: 各区分の箱ひげ図

3.2.3 箱ひげ図の平滑化

ここまでの分析において漢字含有率の減少傾向を裏づけるものは皆無である. しかし, 統計的な仮説検定が帰無仮説の棄却によってのみ結論を下せるという性質上, 漢字含有率の変動していないということを積極的に結論することは困難である. それゆえ, 本稿では検定のような枠組みのみならず, これを補うものとして箱ひげ図の平滑化を実施して漢字含有率の変遷を素描する.

図 8 に箱ひげ図の平滑化によって描いた漢字含有率の変遷を示した. なお, 先行調査分は宮島 (1988a) の調査結果に対し, 石井 (2013) が描画したものを概ね再現したものであり, 本稿調査分は本稿の調査結果から作図したものである⁴³. 先行調査分と本稿調査分とは別々に作図しており, 両図の繋ぎ目は平滑化せずに破線によって繋いだ. また, 両図が比較可能となるのは作品の基準と漢字含有率の算出方法とがいずれも等しいからである.

⁴² 本稿では作品の漢字含有率を決定づけるのが作品の受賞年か作者の誕生年かという点は扱わない.

⁴³ 米谷ふみ子「過越しの祭」は先行調査分に含めず, 本稿調査分に含めてある.

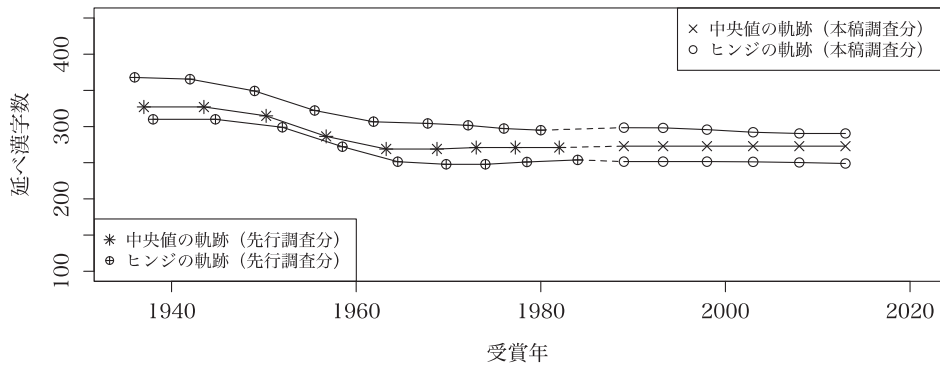


図 8 : 漢字含有率の変遷

この図では中央値の軌跡も上下ヒンジの軌跡も横ばいである。このことは本稿の調査が明らかにした各作品の漢字含有率の概ね半数が同様の傾向にあることを示すものである。

すなわち、この結果から 1986 年以降について文章に占める漢字の割合が安定する傾向にあるという仮説を導き出すことが可能となる。また、このことは消極的ではあるものの宮島 (1988a) の予想を概ね裏づけるものであると言える。

4. おわりに

最後に分析の結果から導き出された本稿の結論と今後の課題とを述べる。

4.1 本稿の結論

本稿では 1986 年から 2015 年までの芥川賞作品を対象に各作品から 1,000 字ずつを抜き出し、そこに占める漢字の割合を算出し、その変遷について分析を実施した。初めに相関分析を実施した結果、作品の受賞年や作者の誕生年と作品の漢字含有率との間にはいずれも相関が認められなかった。次に全ての漢字含有率を 6 つの区分に分け、Kruskal-Wallis 検定を実施した結果、有意差は認められなかった。更に各区分の中央値と上下ヒンジとを平滑化した結果、いずれの軌跡も横ばいになっており、このことは漢字含有率が概ね安定する傾向にある可能性を示唆している。

また、本稿において示された小説の文章における漢字の割合の安定傾向または増加傾向が今後も続くならば、小説から漢字が無くならないとするのは真であり、小説の文章から漢字が無くならないならば、日本語の文章からも漢字が無くならないというのも真であるから小説の文章における漢字の割合にこのような傾向が今後も続くならば、日本語の文章から漢字が無くならないという結論を導くことが可能となる。

4.2 今後の課題

ここでは本稿の課題について簡単に述べる。まず、本稿では漢字含有率の変動について安定する傾向にある可能性に言及したが、1986 年以降の漢字含有率が安定していると積極的に結論することは本稿において採用した手法の性質上、困難であり、このことは追試によって裏づけられることが望まれる。

次に具体的な漢字についての考察が必要である。本稿においては抜き出した字の少なから意図的に避けたが、伊藤（2002）によれば、全数調査を実施することによって初めて低頻度のものまで含めた質的な検討が可能となることから今後はそのような調査によって精緻に検討することが求められている。特に異なり漢字数に占める表外漢字の割合を算出し、比較するような検討は全数調査でなければ、不可能なものであり、このような検討は漢字政策の評価や見直しに繋がるものとして一定の意義がある。

更に本稿では字を単位として分析を実施したが、今後は語を単位とした分析を併用することが必要となる。これについて安本（1965）は語の決定の難しさを指摘すると共にどのような単位を用いても本質的な違いは無いと述べているが、「形態素解析」の発達によって語の決定は以前よりも容易になってきている。また、宮島（1988b）は動詞の補助的用法において漢字使用が減少していると述べており、土屋（2000b）は漢字表記の和語が復権してきていることを報告している。既に述べたように本稿は野村（1988b）の指摘にあるような語を単位とする調査の前段階と位置づけられるものであり、今後はこのような分析によって品詞や語種ごとに漢字使用の実態を観察することが必要となる。

また、本稿では小説を対象としたが、野村（1988b）の指摘するように漢字使用の実態の解明や予測には小説以外において異なる変化が見られる可能性があり、小説と小説以外とに見られる傾向がどのように異なるかを明らかにするような研究も求められている。

これらに加えて書字環境や書字方向についても漢字使用に影響を与える可能性があり、その他にも多くの課題があるものと考えられるが、本稿が提起した問題についての更なる研究の進展が俟たれるところである。

謝辞

本稿の執筆にあたり、筑波大学人文社会系教授の矢澤真人先生からご指導いただいた。また、筑波大学システム情報系助教の佐野幸恵先生にもご助言いただき、国立国語研究所言語変化研究領域教授の山崎誠先生からもご助言を賜った。加えて大塚貴史氏、佐藤知菜氏、垂石早紀氏、萩庭慎氏からも様々な形で助力を得た。記して感謝申し上げる。

なお、本稿の一部は筑波大学先導的研究者体験プログラム（種目 C）「日本人の漢字使用率の変遷について～芥川賞受賞作品に関する調査研究～」(研究期間：2014 年 8 月 1 日～2015 年 1 月 30 日、研究代表者：菅野倫匡) による助成を受けた。また、本稿は計量国語学会第五十九回大会（2015 年 9 月 26 日、神戸大学鶴甲第 1 キャンパス）で口頭発表した内容に加筆し、修正を施したものである。会場においては多くの有益なご助言を賜った。

最後に査読者の皆様、編集委員会の皆様に感謝申し上げ、筆を擱くことにする。

文献

- Kruskal, W. H., & Wallis, W. A. (1952) Use of ranks in one-criterion variance analysis, *Journal of the American statistical Association*, 47(260), 583-621.
- 綾皓二郎（2013）「線形回帰分析による「漢字の将来」の予測（1963）と 50 年後の漢字含有率の実際～分析方法と結果の再検討、および統計教育への教訓～」『情報科学技術フォーラム講演論文集』12(4):453-456.
- 石井正彦（2013）「探索的データ解析による言語変化研究—蛇行箱型図による S 字カーブ

- の発見—」相澤正夫（編）『現代日本語の動態研究』129-150. おうふう.
- 伊藤雅光（2002）『計量言語学入門』大修館書店.
- 犬飼隆（2002）『シリーズ〈日本語探究法〉5 文字・表記探究法』朝倉書店.
- 樺島忠夫（1977）「文字の体系と構造」大野晋・柴田武（編）『岩波講座日本語 8 文字』23-60. 岩波書店.
- 久保拓弥（2012）『データ解析のための統計モデリング入門—一般化線形モデル・階層ベイズモデル・MCMC—』岩波書店.
- 佐竹秀雄（1982）「各種文章の字種比率」『国立国語研究所研究報告集』3:327-346.
- 佐竹秀雄（1987）「語と表記のかかわり」水谷静夫（編）『朝倉日本語新講座 1 文字・表記と語構成』109-120. 朝倉書店.
- 佐藤栄作（2013）「字体とその示し方」『論集』9:1-16.
- 杉山昌子・森岡健二（1969）「明治・大正・昭和の漢字・漢語の変遷」森岡健二（編）『近代語の成立—明治期語彙編—』377-408. 明治書院.
- 田中豊・垂水共之（1999）『Windows 版統計解析ハンドブッケーノンパラメトリック法—』共立出版.
- 土屋信一（2000a）「明治・大正・昭和期の漢字使用—東京日日新聞を資料として—」前田富祺（編）『国語文字史の研究 5』193-212. 和泉書院.
- 土屋信一（2000b）「漢字使用の新しい傾向」『計量国語学』22(7):303-305.
- 中村明（1958）「コトバの美と力—句読点の心理学—」遠藤嘉基・波多野完治・小林英夫・興水実・宮城音弥・中島文雄・小保内虎夫（編）『コトバの科学 5 コトバの美学』1-36. 中山書店.
- 中村明（2003）「文学と漢字」前田富祺・野村雅昭（編）『朝倉漢字講座 3 現代の漢字』1-32. 朝倉書店.
- 名取真人（2014）「カイ二乗近似によるクラスカル・ウォリス検定と小標本」『霊長類研究』30(2):209-215.
- 野村雅昭（1980）「週刊誌の漢字含有率」『計量国語学』12(5):215-225.
- 野村雅昭（1988a）「日本語の漢字」金田一春彦・林大・柴田武（編）『日本語百科大事典』329-345. 大修館書店.
- 野村雅昭（1988b）「漢字はなくなるか」野村雅昭『漢字の未来』225-241. 筑摩書房.
- 宮島達夫（1988a）「漢字の将来」その後『言語生活』436:50-58.
- 宮島達夫（1988b）「芥川賞作品の漢字含有率」『計量国語学』16(6):271-279.
- 安本美典（1963）「漢字の将来—漢字の余命はあと二百三十年か—」『言語生活』137:46-54.
- 安本美典（1965）「文章の性格学」安本美典『文章心理学入門』89-247. 誠信書房.
- 渡部洋・鈴木規夫・山田文康・大塚雄作（1985）『探索的データ解析入門—データの構造を探る—』朝倉書店.

（2016 年 6 月 30 日受付, 2016 年 11 月 6 日再受付）

Paper B

Will Chinese Characters Vanish from Japanese Sentences?
Reconsidering ‘the Future of Kanji’ Once Again

KANNO Michimasa

(Graduate School of Humanities and Social Sciences, University of Tsukuba)

Abstract:

According to a prediction made in the mid-20th century, Chinese characters were expected to vanish from Japanese sentences by the end of the 22nd century. This prediction was based on research on the usage rate of Chinese characters in Japanese sentences. However, this is now doubtful because although more than half a century has passed, the use of Chinese characters in Japanese is still widespread. This study examines the rates of Chinese character usage in Japanese sentences in the past three decades. Through the examination, this study concludes that the hypothesis of decreasing rates, in accordance with the prediction, was not supported because a year-by-year analysis of the usage tendencies suggested that the rates have not substantially decreased. Moreover, the analysis suggested the possibility of the rates becoming stable.

Keywords: future of Kanji, Chinese character usage, usage rate of Chinese characters, Akutagawa prize-winning works