

筑波大学博士（言語学）学位請求論文

現代日本語の文章における
漢字使用の実態に関する通時的・計量的研究

菅野 倫匡

2020 年度

目次

序章		1
1	本論文の目的	1
2	本論文の立場	2
2.1	現代日本語について	2
2.2	漢字使用の実態について	6
3	本論文の構成	7
第 1 章	先行研究の概観	9
1	小説を対象として「漢字含有率」を調査した研究	10
1.1	安本 (1963) の研究	11
1.2	宮島 (1988a) の研究	12
1.3	中村 (2003) の研究	14
2	統計的な手法の扱い方について疑問を呈した研究	14
2.1	綾 (2013) の研究	15
2.2	石井 (2013) の研究	15
3	別の視点を導入することの必要性を指摘した研究	17
3.1	杉山・森岡 (1969) の研究	17
3.2	野村 (1988b) の研究	19
3.3	佐竹 (1982, 1987a) の研究	20
4	小説以外を対象に「漢字含有率」を調査した研究	22
4.1	梶原 (1982) の研究	24
4.2	土屋 (2000a, b) の研究	25
5	現状の整理	26

第 2 章	用語の定義	31
1	「字」及び「文字」	31
2	「字種」	32
3	「漢字含有率」	33
第 3 章	小説の漢字含有率は減少し続けるのか	37
1	調査の方法	37
1.1	対象	37
1.2	手順	38
2	分析の手法	39
2.1	相関分析	39
2.2	Kruskal-Wallis 検定	39
2.3	蛇行箱型図	40
3	調査の結果	40
4	結果の分析	43
4.1	相関分析	43
4.2	Kruskal-Wallis 検定	46
4.3	蛇行箱型図	47
5	本章の結論	48
第 4 章	小説の漢字含有率はなぜ安定したのか	49
1	漢字含有率の安定を形作る要因を探る方法	49
2	調査の方法	51
2.1	対象	51
2.2	手順	51
3	分析の手法	54
4	調査の結果	55
5	結果の分析	58
5.1	語種別の延べ文字数の点から	58
5.2	語種別の延べ漢字数の点から	61
6	本章の結論	63

第 5 章	新聞の漢字含有率は如何なる傾向にあるのか	65
1	調査の方法	65
2	分析の手法	66
3	調査の結果	66
4	結果の分析	69
4.1	相関分析	69
4.2	漢字含有率の変遷	70
5	変遷を形作る要因についての検討	72
5.1	相関分析	76
5.2	漢字含有率の変遷	77
6	本章の結論	79
第 6 章	漢字含有率に見る現在の漢字使用の実態	80
1	小説と新聞とにおける漢字含有率の変遷	80
2	相違点から見る小説や新聞の文章における漢字使用の実態	81
3	共通点から見る現代日本語の文章における漢字使用の実態	83
4	本章の結論	83
終章		85
1	本論文の結論	85
2	本論文の意義	87
3	本論文の課題	89
参考文献		91
本論文の構成と既発表論文との関係		100

凡例

- 図，表，式，用例等の番号はそれぞれ通し番号とする。
- 注は脚注によって示す。また，注の番号も通し番号とする。
- 句読点は「。」と「，」とを用いる。ただし，引用に際しては原文に拠る。
- 引用に際して縦組みは横組みに改める。また，改行位置は改めることがある。
- 特に断らない限り，圏点（傍点）や下線は原文に拠る。
- 上つきのルビは原文のままであることを示す「〔ママ〕」を除いて原文に現れるルビを示すものである。また，下つきのルビは本論文において施したものである。
- 用例を挙げる際には著者名，作品名，掲載誌名・書籍名，発行年の4点あるいは紙名，発行年月日，朝刊・夕刊の別，面種の4点を併記する。
- 参考文献は原則として著者の姓と丸括弧で括った発行年（西暦）とを併記する方式によって示す。ただし，丸括弧内では発行年を括る丸括弧は省略する。また，書誌情報は一括して稿末の「参考文献」欄に掲げる。
- 参考文献のうち辞事典については辞事典名と版数とを併記する方式によって示すことがある。ただし，改版のないものは版数を省略する。

序章

1 本論文の目的

本論文は日本語の表記，特に現代日本語の文章に用いられる漢字の使われ方についてその一端を明らかにするものである。以下に本論文の目的を示す。

本論文の目的 現代日本語の文章における漢字使用の実態を通時的・計量的な観点から解明すること

この目的を達するために本論文では文章に占める漢字の割合（「漢字含有率」）に着眼する。その上で実際に「漢字含有率」を調査し，これに基づいて漢字使用の実態について現在に至るまでの変遷を明らかにすることとする。

そもそも「漢字含有率」とは佐竹（1987b:109）によれば，「文章表記の様相を手軽に示す指標」あるいは「表記法のありようの特徴を示す手軽で簡明な尺度」であり，各時点におけるそれを比較したり時間軸に沿って並べたりすることが可能であるという利点から「表記の変遷を示すのにも有効な尺度」として多くの研究において用いられてきた。この利点こそが本論文においても「漢字含有率」に着眼する所以である。

なお，「漢字含有率」は土屋（2000a）の指摘するように複数の要因によって変動するものであり，それぞれの要因による個々の影響について全てを反映した値であることから「漢字含有率」に基づいて漢字使用の実態を明らかにしたとしても「漢字含有率」に影響を与えた具体的な要因——どのような要因がどの程度の影響を及ぼしたことによって漢字使用の実態に如何なる変化が生じたのか——までをも読み取るためには別の手立てが必要になる。そのための手立てとして本論文では野村（1988b）や土屋（2000b）の示唆するところに従って語種別に表記を考察することとし，これを通して「漢字含有率」に影響を与える具体的な要因をも探り出さんとする。

また、「漢字含有率」は野村（1980, 1988b）や佐竹（1982）も指摘するように媒体やジャンルによって異なる可能性があることから単一の媒体やジャンルのみを対象として「漢字含有率」を調査すると他の媒体やジャンルの文章にも広く見られる一般的な傾向と当該の媒体やジャンルの文章に見られる特有の傾向とを峻別し得ないという問題が生じてくる。この問題を解決するための手立てとして本論文では複数の媒体やジャンルを対象として「漢字含有率」を調査することとし、その比較を通して当該の媒体やジャンルの文章にも他の媒体やジャンルの文章にも見られる共通点としての一般的な傾向と当該の媒体やジャンルの文章に見られるものの他の媒体やジャンルの文章には見られない相違点としての特有の傾向とを共に捉えんとする。

すなわち、本論文は複数の媒体やジャンルを対象として「漢字含有率」を調査し、これに基づいて漢字使用の実態について現在までの変遷を明らかにし、その変遷を形作る要因についても語種に基づいて検討し、これによって所期の目的を達せんとするものである。

2 本論文の立場

前節に示した本論文の目的については「現代日本語」と「漢字使用の実態」との2点を順に取り上げ、それに関わる本論文の立場に関して以下に説明を加える。

2.1 現代日本語について

まず、本論文の目的に掲げた「現代日本語」について述べる。この「現代日本語」は「現代日本語である」ということを意図したものというよりは「日本語以外の他言語でもなく、現代語よりも前の時期の日本語でもない」ということを意図したものである。このように消極的ながらも論題や目的に「現代日本語」を掲げてあるのは中国語や朝鮮語¹の文章においても上代（日本）語や中古（日本）語の文章においても「漢字使用の実態」を扱うことが可能だからであり、本論文において「漢字使用の実態」を明らかにせんとする以上、その「漢字使用の実態」については（少なくとも他言語や他の時期の日本語の文章におけるそれではないという）何らかの限定が必要になるからである。

¹「中国語」も「朝鮮語」も特定の時期や政体に拠らない総称として用いる。なお、これらの言語名は3つの専門事典（『日本語学研究事典』、『日本語大事典』、『日本語学大辞典』）における立項状況を踏まえたものである。また、「中国語」という言語名に関しては田野村（2018）、「朝鮮語」という言語名に関しては内山（2004）も参考にした。

一般に「現代（日本）語」と言うと明治以降の日本語を指す場合²（佐藤 1995; 鈴木 1988; 浅川 2013 など）と昭和初年以降の日本語を指す場合³（鈴木 1988; 杉戸 2018）と（第二次世界大）戦後の日本語を指す場合⁴（吉田 1989; 飛田 2007; 浅川 2013; 真田 2014; 小柳 2018; 杉戸 2018; 沖森 2018 など）とがあり、その定義については統一を見ていない現状にある⁵。また、その前提となる時代区分⁶についても議論の分かれるところである。

そもそも時代区分は「切れ目のないところに強いて切れ目を入れて分割する」（小柳 2018:16）ものであり、それ自体が困難な試みであると言える⁷。また、仮に時代区分を設けるにしても「いくつに分けるべきかが第一の問題で、いつを境界とすべきかについては、その後に考えられるべきであろう」という前田（1985:845）の指摘もあり、「いつを境界とすべきか」という点のみならず、全体について「いくつに分けるべきか」という点にも注意を払う必要がある⁸。

² 時期については「明治期以降」（鈴木 1988:104）、「明治以降」（佐藤 1995:14）、「明治元年（一八六八年）から」（浅川 2013:44）とあるのを一括して「明治以降」とした。

³ 時期については「昭和初年以降」（鈴木 1988:104）、「昭和初年から」（杉戸 2018:341）とあるのを一括して「昭和初年以降」とした。

⁴ 時期については「1946～」（吉田 1989:283）あるいは「第2次世界大戦の敗戦から」（吉田 1989:303）、「太平洋戦争終了後、すなわち昭和二〇年八月一五日の終戦から」（飛田 2007:517）、「昭和二一年（一九四六年）から」（浅川 2013:44）、「第二次世界大戦後から」（真田 2014:714）、「1945年-」（小柳 2018:16）、「1945年以降の戦後」（杉戸 2018:341）、「第二次世界大戦後（一九四五～）」（沖森 2018:9）とあるのを一括して「（第二次世界大）戦後」とした。

⁵ なお、浅川（2013）は日本語のうち特に「東京語」に限定するが、本論文では特定の方言に限定せず、専ら時期について考える。また、「現代日本語といえば、本来は方言も含まれるわけであるが、普通は共通語としての日本語を意味する。その共通語は、東京語を中核として成立したから、現代語の変遷は、東京語の変遷とほぼ重なり合うことになる。」という鈴木（1988:104）の指摘については踏まえつつ本論文では「現代日本語」に「本来は方言も含まれる」という立場を採る。

⁶ 小松（1999:13）は「時間的な流れを区分するなら時期区分という用語が自然である」とし、「時代区分」という用語を斥けるが、本論文では用語について異論のあることを示すに留める。

⁷ なお、小柳（2018:16）は時代区分を設ける理由として「それでも、歴史を記述するのに適度な（細かすぎず大まかすぎずの）区分があれば便利」であることを挙げ、時代区分は「本質的に仮説である」ことを指摘した。また、これに関連して佐藤（1995:13）は研究を成り立たせる上で「とどまるところを知らない時間の流れに、研究の分野、研究者の立場などに従って、適当に時間の区切りを設けて限定」することは必要であり、「無限定を相手にはできない」ことを指摘した。これが本論文において「何らかの限定が必要になる」と述べる所以である。

⁸ なお、全体について「いくつに分けるべきか」という点は重要であるが、本論文の論旨には関わらないことから措くこととする。また、前田（1985）は特に語彙史を念頭に置いたものであることを申し添える。

一方、「日本語の歴史の時代区分を考えるに当たっては、巨視的な立場からの区画論を否定するわけではないが、それに先立って、音韻・文法・文字・語彙・文体などの項目ごとに、その歴史的境界を考え、その上で、総合的な検討に進むことが望ましいと思われる」という築島（1988:59-60）の指摘がある⁹。これに照らせば、文字・表記史の時代区分やそれにおける「現代日本語」について考えることが本論文においては重要である。

築島（1988）は「音韻史」、「アクセント史」、「文法史」、「文字史・表記史」、「語彙史」の各項目について時代区分を論じ、「文字史・表記史の時代区分」においては「昭和21年は、国語表記史の上の一区分をなす年と見ることができよう」（築島 1988:63）と述べた。

また、概ね同様の立場を採り、「音韻」、「文法」、「文字表記」、「語彙」、「文章」に分けて時代区分を論じた大木（2018）は以下のように述べた¹⁰。

音韻・文法の大きな画期が室町～江戸にかけてであるのに対して、文字表記・語彙のきわめて大きな画期は明治以降であるということはいえるであろう（文字史の画期はひらがな・カタカナの成立、表記史は近代言語政策が実施された一九四六年という見方もあろう）。そして、音韻・文法にとって重要な室町～江戸にかけての時期は、文字表記・語彙にとっては画期とは無縁であり、逆に、文字表記・語彙にとって画期の明治以降は、音韻・文法にとって画期とは全くいえない。つまり、日本語全体に対して時代区分をおこなうのはほとんど不可能であるというべきではないか。（大木 2018:29）

⁹「項目ごとに、その歴史的境界を考え、その上で、総合的な検討に進む」（築島 1988:60）ということについては「語彙史の時代区分が明確になれば、音韻史の時代区分、文法史の時代区分などと考え合わせ総合することにより、国語史の時代区分が明らかになるはずである」という前田（1985:826）の指摘もあるが、「文字」や「表記」に言及があることから築島（1988）を引いた。なお、小松（1999:12）は「言語と密接な関わりをもつが、言語そのものではない」ことから「漢字の導入や仮名／片仮名の発達などは日本語史の圏外にある」とするが、これに対して乾（2001:28）は「文字・表記の史的研究」について「小松が指摘するように、言語および言語史の研究とは、本質的に一線を画すべきであるというのも、一面においては首肯できる。しかしながら、ことに日本語においては文字と言語の関係が不即不離である点を鑑みれば、ことはそう単純ではない。」として「文字史の時代区分」を論じており、本論文では「文字と言語の関係が不即不離である」（乾 2001:28）という点については留保するものの「文字論はけっして言語学の外に位置づけられるものではない」という乾（2014:2009）の指摘も踏まえ、文字・表記史の時代区分も「文字・表記の史的研究」（乾 2001:28）も「日本語史の圏外にある」（小松 1999:12）とは看做さない立場を採る。また、この点については「言語史の一部としての文字・表記史」と「言語史から独立した分野としての文字・表記史」との「いずれもが文字・表記史である」という矢田（2016:142-143）の指摘や「これからの言語学」は「文字や文字体系の特性」などについて「理解を深めていく必要がある」という斎藤（2018:294）の指摘も参考にした。

このように見ると文字・表記史、特に表記史における「現代日本語」は1946年からの日本語を指すと考えることも可能ではある。これを踏まえ、本論文では「現代日本語」について基本的には1946年から現在までの日本語——概ね戦後の日本語——を指すものとして解することとする。

ただし、「政治上の事件のように、何年何月に勃発したとか終結したとかはいえず、前代末と次代初とは線で区切られるものではなく、相互に重複している」という吉田(1989:283)の指摘も参考とし、1868年から1945年までの日本語——明治以降で第二次世界大戦の終結より前の時期の日本語——についても「現代日本語」ではないと積極的に看做す立場は採らず、その境界は緩やかなものであると考える。これに照らせば、「現代日本語」について広くは1868年から現在までの日本語——明治以降の日本語——を指すものとして用いているということにもなる。

更に「現代日本語」について考えるに当たっては以下に示す沖森(2018)の指摘も重要である。

現代日本語は一九四五年以降の日本語だけを対象にできるかと言えば、必ずしもそうではない。一九四五年以前に言語習得した人たちのことばを近代語とは言いがたい。また、今後一九四五年以降の生まれしかなくなったとしても、現代語という言葉の主たる意味は、一九四五年以降の日本語というよりも、〈いま現在実際に使われている日本語〉に傾いている。その意味で、日本語史における「現代」は時代区分として喚起される度合いが低い。(沖森 2018:9)

このように「現代日本語」は日本語史の時代区分としての「現代日本語」という側面と現在の日本語を指すことに重きを置いた「いま現在実際に使われている日本語」(沖森 2018:9)としての「現代日本語」という側面とを共に持つものであり、本論文においてはいずれの側面も含むものとして「現代日本語」を捉えることとする。すなわち、時代区分としての「現代日本語」のうち特に現在の日本語に焦点を当てるということである。

これに加えて一般的には共時態と看做される「現代日本語」と本論文の目的に掲げた「通時的な観点」との関係についても補足する。

¹⁰ 本論文では「総合的な検討」(築島 1988:60)あるいは「日本語全体の時代区分」(大木 2018:30)の可否や是非については措くこととする。

半澤 (2007:32) は「通時」が成り立つためには「共時」の設定が必要であるが、一つの体系が不変状態であるとする時間幅は、何らかの変化が認められた結果として相対的・部分的にしか設定できないものであって、「通時」の所与の前提にはなりえない」とし、「現実には、便宜的に、社会的・政治的な歴史区分にならって、たとえば、上代語・中古語あるいは奈良時代語・平安時代語などのように称し、それぞれを「共時」的な言語体系とかりにみなしているにすぎない」と指摘している。また、「現代日本語」についても「昭和初年から約 90 年、戦後約 70 年を経た中では、「現代の」とひとくくりにするのがためらわれるような様々な変容が見られる」という杉戸 (2018:341) の指摘がある。すなわち、「現代日本語」を「共時的な言語体系」(半澤 2007:32) と看做するのは便宜的なものであって実際には過去に用いられていた日本語から現に用いられている日本語へと徐々に移り変わってきた日本語の歴史の所産とも言えるものである。本論文はその移り変わりを捉えんとするものであり、このような観点から「現代日本語」を見るものであると言える。

なお、本論文は得られた結果から「現代日本語」の画期を定めんとするものではないことを申し添える。

2.2 漢字使用の実態について

「漢字使用の実態」という用語は斎賀 (1960)、中村 (2003)、土屋 (2000a) などの各研究や 2010 年の「改定常用漢字表」にも見られるものである。これは「漢字の使い方」と「漢字の使われ方」との両方を指せるが、本論文では「漢字の使われ方」と捉えることとする。このように捉えるのは本論文があくまで事実としての表記を見るものであって表記主体——書き手——の意図 (や表現としての効果 (ののようなもの)) には立ち入らないということである。

また、「漢字使用の実態」は総体としての漢字の使われ方も個々の漢字の使われ方も含むものであるが、本論文は前者を見る (ことに重きを置いた) ものである (=巨視的な立場)。そのためには当然ながら計量的な観点が必要となる。しかし、計量的な観点に立つことは必ずしも巨視的な立場を採ることにはならず、微視的な立場であっても計量的な観点は可能である (いずれにしても量に還元することによって見えてくる点と見えなくなる点とがある)。双方の立場は対立するものではなく、相互に補完するものである。

これに関連して水谷 (1977) は以下のように述べている。

そうした宿命の語彙論で語彙の全般に関わるような事柄を見ようとすると、何と
言っても手っ取り速いのは統計量の利用である。統計量は対象の集団的特性を簡
潔に記述し得るものとして案出された。(水谷 1977:46)

統計的扱いによって語のいわば個性が切り捨てられてしまうではないかという
心配は(しばしば非難の形で現れたが)別問題である。これは言語の統計的研究一
般に起こる問題で、今それについて詳論する場ではないから簡単に触れるにとどめ
るが、どんな方法にも(したがって非計量的方法にも)限界があること、すべての
《個性的な》事柄に価値があるとは限らないこと、計量も問題の立て方によっては
個性が扱えること、質的特性と量的特性とは全く異なるという単純な二元論があま
り有効でないことを、言っておこう。量的特性というのも、手をこまねいていて初
めから対象の側に具わっているといえるようなものではない。自然科学の場合も
事情は同じであった。量的特性は、それを認識する人間の側の発明に係る。

(水谷 1977:47)

すなわち、「漢字使用の実態」の場合にも総体としての「漢字使用の実態」を考える場合
には計量的な観点を採ることが有効であり、このような事情から本論文では計量的な観
点を採ることになるのである。

3 本論文の構成

続いて本論文の構成を以下に示す。

序章では本論文の目的を示し、この目的に関連して本論文の採る立場について述べた
上で本論文の構成を示した。

第1章では漢字使用の実態に関する計量的な研究のうち特に「漢字含有率」に関する
ものを取り上げ、それらを「小説を対象として「漢字含有率」を調査した研究」、「統計的
な手法の扱い方について疑問を呈した研究」、「別の視点を導入することの必要性を指摘
した研究」、「小説以外を対象に「漢字含有率」を調査した研究」の4つに分け、それぞ
れの研究の概要を第1節から第4節までに述べる。また、これらの研究を踏まえて第5節
に現状を整理する。

第2章では本論に入るための準備として本論文における用語の定義について述べる。

また、文章に占める漢字の割合を指す用語とその算出方法とを整理した上で本論文の採用する定義についても明示する。

第3章では小説を対象に各作品の「漢字含有率」を実際に調査し、その変遷を明らかにする。その上で特に「漢字含有率」が時間の変化に伴ってどのように変動するかという点に焦点を当て日本語の文章から漢字が無くなるという安本(1963)の予測について検証する。本章では調査の結果に基づいて「漢字含有率」が20世紀の中葉からは減少せず、安定するようになるということを統計的な手法を用いて提示する。これにより、安本(1963)の予測に改めて疑問を呈すると共に見解の相違する宮島(1988a)と中村(2003)との両研究のうち前者の予想が裏づけられたことを指摘する。

第4章では第3章で扱った作品を再び取り上げ、小説における「漢字含有率」が安定したことの要因について明らかにする。特に本章においては語種に着目した考察を通して20世紀の中葉までは見られた語種の構成比率や語種別に見た漢字使用の実態の変化が20世紀の末葉から見られなくなったことにより、「漢字含有率」が安定するに至ったことを主張する。また、20世紀の中葉まで見られた「漢字含有率」の減少は主として漢字表記の漢語の減少に起因している可能性を指摘する。更に漢字表記の和語の減少という従来の指摘が仮名表記の和語の増加に伴う和語全体の増加による相対的な減少として捉えられる可能性も併せて指摘する。

第5章では第3章と概ね同時期の新聞を対象に各記事の「漢字含有率」を調査し、その変遷を明らかにする。その上で新聞における「漢字含有率」が減少することを示した上でその減少が1996年4月1日の「数字の表記」の変更によるものであり、その影響を取り除くと「漢字含有率」が安定しているということを指摘する。更に語種の構成比率や語種別に見た漢字使用の実態にも変化が見られないことが安定の要因となっていることを主張する。なお、本章は第3章と第4章とにおいて用いた調査・分析の手法が新聞という他の媒体やジャンルを対象とした調査においても概ね適用し得ることを示すものでもある。

第6章では第3章や第4章において取り上げた小説における「漢字含有率」の変遷と第5章において取り上げた新聞における「漢字含有率」の変遷とを照らし合わせることで、相違点としての個々の媒体やジャンルにおける漢字使用の実態と共通点としての現代日本語の文章における漢字使用の実態とを共に明らかにする。

終章では第3章から第6章までの各章の結論を改めて整理した上で本論文の持つ意義と残された課題とについて述べる。

第 1 章

先行研究の概観

ここでは現代日本語の文章における漢字使用の実態について扱っている先行研究を概観する。ただし、そのような研究は数多く存在し、その全てを取り上げることは困難であることから本章においては計量的な研究、特に「文章表記の様相を手軽に示す指標」（佐竹 1987b:109）として広く用いられ、既に多くの蓄積がある「漢字含有率¹¹」に関する研究に限って取り上げることにする。

また、本章では「漢字含有率」に関する研究について「小説を対象として「漢字含有率」を調査した研究」、「統計的な手法の扱い方について疑問を呈した研究」、「別の視点を導入することの必要性を指摘した研究」、「小説以外を対象に「漢字含有率」を調査した研究」の 4 つに便宜的に分け、それぞれの概要について第 1 節から第 4 節までの各節において順に述べる。更に第 5 節において現状の整理を試みる。

なお、本章においては改稿や再録に当たるものは原則として取り上げないが、安本（1963）を改稿した安本（1966）、宮島（1988a）を再録した宮島（2019）、中村（2003）を再録した中村（2012）、石井（2013）を再録した石井（2019）、杉山・森岡（1969）を再録した杉山・森岡（1991）、野村（1988b）を再録した野村（2008）については初出のものと照らし合わせ、本章において取り上げる範囲において論旨に関わる改変が施されていないことを確認してある。また、この点については個々の研究を取り上げる際に必要に応じて改めて言及することがある。

¹¹ 文章に占める漢字の割合を指す用語やその算出方法については第 2 章において詳しく取り上げることとし、本章では文章に占める漢字の割合を一括して「漢字含有率」とする。

1 小説を対象として「漢字含有率」を調査した研究

初めに小説を対象として「漢字含有率」を調査した研究として本節では安本（1963）、宮島（1988a）、中村（2003）の各研究を取り上げる¹²。

なお、このように小説を対象としたものを取り上げるのは実際に「漢字含有率」を調査した研究として特に広く知られている安本（1963）の研究が小説を対象としているからであり、この研究に続く一連の研究を概観するに当たり、まずは小説を対象としたものに限って取り上げることとする¹³。また、これは媒体やジャンルによって「漢字含有率」が異なる可能性があるという野村（1980, 1988b）や佐竹（1982）の指摘を踏まえてのことでもある。

¹² これ以外にも小説を対象としたものとしては内田（1964）の研究や Tanaka(1998)の研究もあるが、論旨に大きくは関わらないものと考え、取り上げないこととする。なお、前者は1887年から1957年までの計46篇の小説を対象に各作品の「漢字含有率」を調査したものであり、安本（1963）の研究は挙げていないものの「漢字含有率」が次第に減少することを指摘したという点では安本（1963）に通ずるものである。また、後者は「漢字含有率」に関する研究についての概説的な内容を主とするものであり、概ね田中（1987, 1989）を改稿したものと見えるが、新たに1篇の小説を対象に「漢字含有率」を調査したものでもあり、その結果に基づいて安本（1963）の予測に対して疑問を呈したものである。これは調査の結果に基づいて安本（1963）の予測に疑問を呈するという点では宮島（1988a）と概ね同様の企てである。これに加えて水谷（1983）や野村（1988b）に言及が見られる賀来（丸山）直子氏の卒業論文は未公開であって入手困難であることから取り上げないが、水谷（1983:100-101）によれば、これは1945年から1975年までの小説を対象に各作品の「漢字含有率」を調査したものであり、野村（1988b）によれば、宮島（1988a）の調査と「同様の傾向をしめすもの」（野村 1988b:129）であるとのことである。

¹³ なお、菅野（2017:481）では安本（1963）の研究を「漢字の割合に関する研究の嚆矢」と位置づけたが、本章では「実際に「漢字含有率」を調査した研究として特に広く知られている」ものとして取り上げることとする。このように改めたのは文章に占める漢字の割合を実際に調査して示すということ自体には麻生（1955）の研究や菅野（2017）においても言及した中村（1958）の研究などの先例があり、必ずしも安本（1963）の研究が最初のものであるとは言えないことを踏まえたものである。また、安本（1963）の研究について「特に広く知られている」としたのは田中（1987）、宮島（1988a）、野村（1988b）、犬飼（2002）などの記述を踏まえたものである。ただし、管見の限りでは文章に占める漢字の割合を実際に調査することに主眼を置いたものとしては安本（1963）の研究が最初のものであると考えられること、この研究から派生した研究が数多く見受けられること、それらの研究の大半が安本（1963）よりも前の研究を直接的な先行研究として挙げていないことの3点に照らせば、安本（1963）の研究をそれ以降の一連の研究の嚆矢と位置づけることは可能であるものと考えられる。また、野元（1989:211）によれば、安本（1963）の研究は「実際の漢字使用の調査によって将来を予測したもの」としては「まず最初であろう」とのことである。

1.1 安本（1963）の研究

まず、安本（1963）の研究について述べる。安本（1963）は現在の漢字使用の実態とそこに至るまでの変遷とを解明することが必要であるとし、小説の文章における漢字使用の実態について通時的に明らかにするとして1900年から1954年までの計100篇の小説を対象に各作品の「漢字含有率¹⁴」を調査した。

その方法は最初に100人の作家を選び、次いで各作家の作品から代表作として1篇を採り、各作品から1,000字ずつ抜き出し、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¹⁵。なお、本文は『現代日本文学全集』（筑摩書房）に拠る。

安本（1963）は「漢字含有率」を作品の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、「漢字含有率」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「漢字含有率」とする単回帰分析を実施し、以下に示す式（1）を得た。

$$y = -1.244x + 2726.17 \quad (x : \text{作品の発表年}, y : \text{「漢字含有率」}) \quad (1)$$

式（1）に $y = 0$ を代入すると2191年には「漢字含有率」が0となる。このことから安本（1963）は「漢字含有率」が直線的に減少し、なおかつ、今後も同様の傾向が続くという仮定に基づけば、22世紀の末葉には日本語の文章から漢字が無くなるという予測が立つとした。

これに加えて安本（1963）は漢字が少なくなれば、漢字を減らそうとすることに対する抵抗が大きくなり、減少の傾き $\frac{dy}{dx}$ が次第に小さくなるという仮定に基づいた式（2）と反対に漢字が少なくなれば、漢字を減らそうとする力が増し、減少の傾きが徐々に大きくなるという仮定に基づいた式（3）とを示している。安本（1963）の示した3つの予測式を合わせて図1に示す。

¹⁴ なお、安本（1963）の用語としては「漢字の使用率」であるが、既に述べたように本章においては「漢字含有率」に一括し、詳細は第2章において検討する。

¹⁵ 調査の方法について安本（1963）は詳しく述べていないことから後述する宮島（1988a）に倣って安本（1965）によって補うこととする。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とのことである。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ている。選ばれた代表作から1,000字を抜き出すに当たっては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段を選び、各段の冒頭50字を抜き出したとのことである。

$$y = 360.17e^{-0.0038659x+7.34521} \quad (2)$$

$$y^2 = -1074.14x + 2181603.11 \quad (3)$$

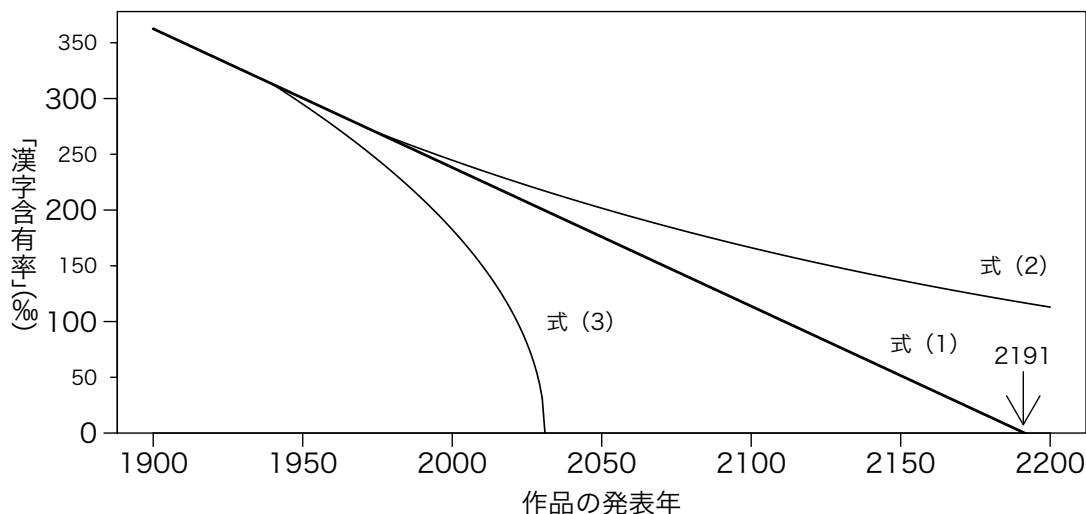


図1 安本（1963）の予測¹⁶

これらの3つの式に基づく予測について安本（1963）は式（2）と式（3）とを極端であるとし、「漢字含有率」が直線的に減少し、なおかつ、今後も同様の傾向が続くという仮定に基づいた式（1）が実際の「漢字含有率」に最も合致していると述べて22世紀の末葉には日本語の文章から漢字が無くなるという前掲の予測を最終的な結論とした¹⁷。

1.2 宮島（1988a）の研究

続いて宮島（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受け、この予測を検証するとして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）を対象に各作品の「漢字含有率¹⁸」を調査した¹⁹。

¹⁶ これは安本（1963）の図2、図3、図4と式（1）、式（2）、式（3）とに基づいて作図したものである。また、作図に当たっては『図説日本語—グラフで見ることばの姿—』（角川書店）を参考とした。

¹⁷ 最終的な結論について安本（1963）を改稿した安本（1966）においても同様であることは確認済みであることを改めて申し添える。

¹⁸ なお、宮島（1988a）の用語としても「漢字含有率」であるが、本章においては既に述べたように個々の研究における用語や算出方法については措いた上で一括して「漢字含有率」としており、詳細は第2章において検討する。

その方法は各作品から 1,000 字ずつ抜き出し、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²⁰。なお、本文は『文藝春秋』掲載のものに拠る²¹。

宮島（1988a）は安本（1963）に倣って「漢字含有率」を 5 年ごとに平均し、大局的には「漢字含有率」が減少してきたと結論した。また、「漢字含有率」と作品の受賞年との相関係数が -0.463 であり、「漢字含有率」と作者の誕生年との相関係数が -0.454 であることから漢字使用が漢字を習得する時期の習慣によって決定づけられるという仮説は成り立たないとした。更に宮島（1988a）は雑誌や新聞を対象とした他の調査（国立国語研究所 1987; 梶原 1982）を援用しつつ 20 世紀の中葉まで「漢字含有率」が減少し続けるという点ではいずれの調査も一致しているとして安本（1963）の調査結果が裏づけられたと述べた。

ただし、宮島（1988a）は芥川賞作品においては 1960 年を境に「漢字含有率」が安定していることを指摘した上で 30 年後、50 年後も安定しているかは不明であるとしつつも 5 年後、10 年後の「漢字含有率」は安定ないし微増するとの予想を示した。

これに加えて宮島（1988a）は各作品から抜き出した 1,000 字について和語の延べ文字数、漢語の延べ文字数、外来語の延べ文字数の 3 つに振り分け、語種の構成比率や語種別の「漢字含有率」も調査し、その結果から 20 世紀の中葉まで「漢字含有率」が減少した背景に漢語の減少と漢語・和語の「漢字含有率」の減少とがあることを指摘し、それらの減少が安定に転じたことによって各作品の「漢字含有率」が安定するに至ったという可能性を示唆した²²。

¹⁹ 宮島（1988a）は作品の選び方が結果を左右する可能性を指摘しており、調査を継続するためには作品を選定する基準（以下、作品の基準）が一定である方が望ましいとして芥川賞作品であることをその基準として採用した。

²⁰ 宮島（1988a）によれば、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 100 字ずつ抜き出したとのことである。なお、調査の再現や検証を容易にするために乱数表を用いず、文章の最初を外すために冒頭 2 行を飛ばしたとのことである。

²¹ 宮島（1988a）によると本文を『文藝春秋』掲載のものとするのは『芥川賞全集』（文藝春秋）のように受賞後に改めて編まれたものには表記の改変が見られ、表記を対象として調査するには適していないからである。なお、この点については宮島（1988a）に言及はないものの「古文獻でなくても、本文の問題に気をつけなくてはならない」という水谷（1983:26）の指摘と関連するものと見える。また、このような表記の改変について「漢字含有率」を用いつつ検討したものとして田原（1975）の研究やこれに基づいた田中（1987, 1989）の研究がある。

²² なお、宮島（1988a:56）の用語としては「和語表記字数」、「漢語表記字数」、「外来語表記字数」であるが、本節ではそれぞれ和語の延べ文字数、漢語の延べ文字数、外来語の延べ文字数に改めた。この方法については第 4 章において検討する。

1.3 中村（2003）の研究

更に中村（2003）の研究について述べる。中村（2003）は「漢字含有率」を調査したものであるとして中村（1958）の調査²³と安本（1965）の調査²⁴とを挙げ、両調査には最近の作品が含まれていないことを指摘した上で小説の文章における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篇の小説を対象に各作品の「漢字含有率²⁵」を調査した。

その方法は対象とした各作品から一定の間隔で計5ページを抜き出し、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²⁶。なお、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篇の小説から得られた「漢字含有率」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「漢字含有率」が月日を経るごとに減少してきたと結論した²⁷。

2 統計的な手法の扱い方について疑問を呈した研究

次に「漢字含有率」を調査した研究における統計的な手法の扱い方に疑問を呈した研究として本節では綾（2013）の研究と石井（2013）の研究とを取り上げる。

²³ これは「漢字含有率」を調査することに主眼を置いたものではないが、1898年から1955年までの計50篇の小説を対象に各作品から冒頭100字を抽出し、各作品の「漢字含有率」を算出したものである。

²⁴ これは宮島（1988a）が安本（1963）の調査の方法を補うために参照していることから明らかであるように「漢字含有率」に関する部分に限れば、安本（1963）と同様のものである。

²⁵ なお、中村（2003）の用語としては「漢字使用率」であるが、既に述べたように本章では「漢字含有率」に一括し、詳細は第2章において検討する。

²⁶ 中村（2003）は対象とする作品を「新しい作品に重点をおいて選定した」（中村2003:8）と述べるに留まっており、作品の基準は不明確である。なお、中村（2003）の調査の方法について菅野（2017:483）においては「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、」と書いたが、これはやや不正確であり、本章では「計5ページを抜き出し、」に改めた。

²⁷ 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）においても取り上げられていることから正しくは14である。なお、この点は中村（2003）を再録した中村（2012）においても改められていないことを申し添える。また、「日輪」の発表年について中村（1958）は1929年としているが、これは安本（1965）や中村（2003）にもある通り、1923年が正しい。

2.1 綾 (2013) の研究

まず、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とが問題であると指摘した。

前者の問題点は安本 (1963) が散布図を示しておらず、相関係数²⁸や決定係数²⁹なども記していないことであり、後者のそれは調査期間の凡そ5倍も外挿して予測していることである。また、綾 (2013) は散布図や決定係数が示されていたならば、多くの人が予測に疑問を持ったはずであるとし、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と結論づけた。

なお、綾 (2013) は宮島 (1988a) の手法については特に問題視しておらず、単に宮島 (1988a) の調査結果を用いて単回帰分析を実施し、その結果を示すことに留まっている。

2.2 石井 (2013) の研究

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³⁰」の手法を用いて安本 (1963) が「漢字含有率」の減少傾向を捕捉する過程について検討し、「漢字含有率」を5年という等間隔の区分に分け、各区分の代表値として平均値を採用し、代表値の変動のみを全体の傾向を代表するものと看做した安本 (1963) の手順に問題があることを指摘した。

石井 (2013) によれば、観測値を等間隔の区分に分けることの問題点は各区分における観測値の数の偏りが生じ、その数の少ない区分において代表値が不安定になるということである。また、各区分の代表値として平均値を採用することの問題点も平均値が外れ値の影響を受け易いことから代表値が不安定になるということである。更に代表値の変動のみを全体の傾向を代表するものと看做したことの問題点は様々なノイズを含む代表値の変動

²⁸ 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。

²⁹ 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。ただし、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おり、それによると「漢字含有率」が 0 となるのは 2187 年である。

³⁰ これは石井 (2013:129-130) によると「統計的仮説を検証するための「確認的データ解析」(推測統計学) に対して「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」ための手法であり、「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」ものである。なお、詳細については石井 (2013) も挙げている渡部・鈴木・山田・大塚 (1985) や「探索的データ解析」を切り拓いた Tukey (1977) に譲る。

が直ちに意味のある変動を示すとは限らないということである。

これらの問題を解決するために石井（2013）は各区分を構成する観測値の数を等しくし、各区分の代表値として中央値を採用した上で五数要約値³¹を示し、代表値以外の変動をも把握することが不可欠であるとした上で五数要約値に基づいて各区分の箱ひげ図³²を描き、これを平滑化³³して図示する「蛇行箱型図³⁴」（渡部・鈴木・山田・大塚 1985）という手法を提示した。また、この手法を用いて安本（1963）の調査結果と宮島（1988a）の調査結果とを検証した。その結果、小説の文章における「漢字含有率」は図2のように安本（1963）の調査結果が描き出すS字カーブに宮島（1988a）の調査結果が描き出すS字カーブが重なっており、全体としても大きなS字カーブを描きながら減少しているという結論が示された³⁵。

³¹ これは中央値、最大値、最小値、上ヒンジ、下ヒンジの5つのことである（石井 2013）。なお、上ヒンジとは中央値より大きな観測値の中央値のことであり、下ヒンジとは中央値より小さな観測値の中央値のことである。

³² これを指す用語として渡部・鈴木・山田・大塚（1985）は Tukey(1977) が「図式図 (schematic plot)」と「箱ヒゲ図 (box-and-whisker plot)」とを描き方によって区別しているということに触れた上で（特に前者を指すものの）両者を区別せずに指す「箱型図 (boxplot)」あるいは「平行箱型図 (parallel boxplots)」を採用しており、これに倣って石井（2013）も「箱型図」あるいは「蛇行箱型図」としている。しかし、いずれの描き方によるものも一般的には小林（2019）のように「箱ひげ図」とすることが多いことから本論文においては「箱ひげ図」に統一する。

³³ これを指す用語として渡部・鈴木・山田・大塚（1985）は「ならし (smoothing)」を採用しており、これに倣って石井（2013）も「ならし」としている。しかし、「smoothing」の訳語は一般的には柴田（2015）のように「平滑化」とすることが多いことから本論文では「平滑化」に統一する。

³⁴ この手法について Tukey(1977) は「wandering schematic plots」と記しており、これを踏まえて「探索的データ解析」の手法について解説した渡部・大塚・鈴木・山田（1984）は「蛇行図式図 (wandering schematic plot)」としているのに対し、渡部・鈴木・山田・大塚（1985）は「蛇行箱型図 (wandering boxplot)」としており、これに倣って石井（2013）も「蛇行箱型図」としているという点は確認済みである。ただし、「探索的データ解析」に関する「主要な文献の中では、蛇行図式図はほとんど扱われていない」という渡部・大塚・鈴木・山田（1984:75）の指摘もあり、この手法を指す一般的な用語やその訳語については判じ難い点もあることから本論文では便宜的に「蛇行箱型図」とする。また、この手法について菅野（2017）では「箱ひげ図の平滑化」と書いたが、「(箱ひげ図にも描かれる) 中央値や上下ヒンジを平滑化して図示したものであって変形した箱ひげ図に当たるもの」を「箱ひげ図を平滑化したもの」としたのはやや不正確であり、本論文では改めた。

³⁵ 石井（2013:129）は「言語変化の進行（普及・伝播）は「S字カーブ（成長曲線）」を描くことが多いものの「現実のデータからS字カーブを発見することは、必ずしも容易ではない」とした上で「蛇行箱型図」という手法を用いることによって、従来S字カーブとは考えられていなかった言語変化の中にS字カーブを描くものがあることを示すことに主眼を置いたものである。なお、石井（2013:150）によれば、「蛇行箱型図」を作図する手順のうち「隣接多角形」や「外側値や極外値」を書き加えることは「S字カーブの発見には直接かかわらないので、省略する」とのことである。

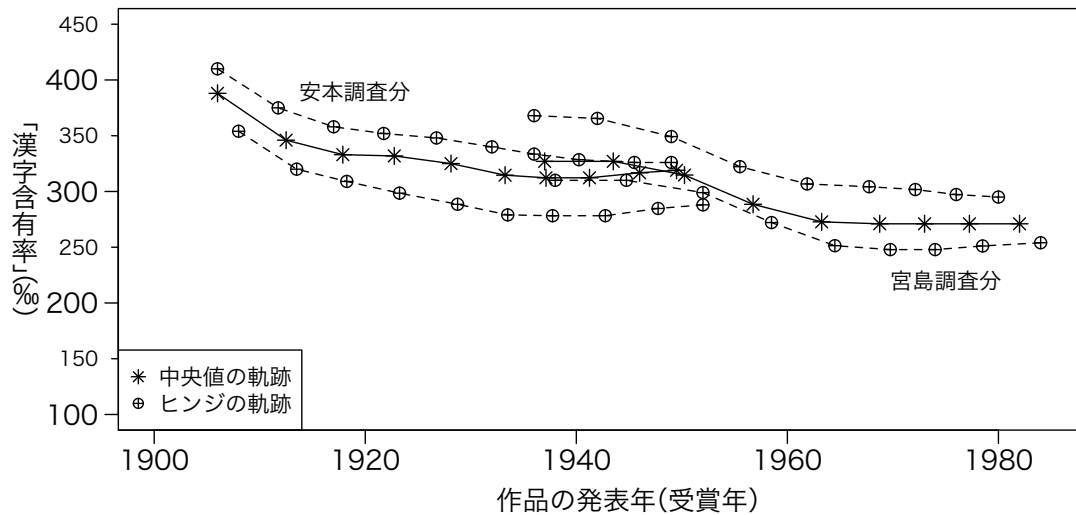


図2 石井（2013）による安本（1963）と宮島（1988a）との検証結果³⁶

3 別の視点を導入することの必要性を指摘した研究

続いて「漢字含有率」を調査した研究に対して別の視点を導入することの必要性を指摘した研究として本節では漢字使用の実態を捉えるために語種に着目することが必要であるとした杉山・森岡（1969）、野村（1988b）の両研究と表記の実態を捉えるために「カタカナ含有率³⁷」を考慮することが必要であるとした佐竹（1982, 1987a）の研究とを取り上げる³⁸。

3.1 杉山・森岡（1969）の研究

まず、杉山・森岡（1969）の研究について述べる。杉山・森岡（1969）は現在の漢字・漢語使用の実態とそこに至るまでの変遷とを質的・量的の双方の観点から明らかにするとして1879年から1968年までの新聞を対象に「漢字表記語率³⁹」を調査した⁴⁰。

³⁶ これは石井（2013）の図8を概ね再現したものである。

³⁷ 一括して「漢字含有率」としたのと同様に本章においては文章に占めるカタカナの割合を指す用語やその算出方法については立ち入らず、一括して「カタカナ含有率」とする。

³⁸ なお、前掲した宮島（1988a）については語種に着目することが必要であるとした研究としても位置づけ得る。

³⁹ 杉山・森岡（1969）の用語としては「漢字表記語率」あるいは「漢字表記語（の）使用率」であるが、本論文では文章に占める漢字表記の語の割合を指す用語として「漢字表記語率」を採用し、これに一括することとする。なお、この用語を採用するに当たっては専門事典を引き、『漢字百科大事典』（田中1996b）や『計量国語学事典』（高田2009）の記述も参考にした。また、本章では算出方法については措き、これに関連する議論は第4章に譲る。

その方法は『朝日新聞』の社会面を対象に約2年ごとに計50箇所から約1,000字ずつを抜き出し、それを「 β 単位」(国立国語研究所1962)を参考にした独自の規則に従って語に分割し、そこに占める漢字表記の語の割合を算出するというものである⁴¹。

調査に先立って杉山・森岡(1969)は安本(1963)の研究を挙げ、この研究について小説の文章における漢字使用の実態を明らかにしたという点において一定の意義を有することを認めつつも日本語の文章に見られる一般的な傾向を捉えたものとは言えないことを指摘した。また、漢字使用の実態を捉えるためには字を単位とするのではなく、語を単位とした調査が必要であるとし、「漢字表記語率」を調査することの意義を強調した。

その上で杉山・森岡(1969)は調査によって明らかにした「漢字表記語率」が年代が下るに連れて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、「漢字表記語率」の減少率が次第に大きくなるとも述べて t を年代、 y を「漢字表記語率」とする指数回帰によって式(4)を得た⁴²。

$$y = 100 - 6.27511e^{0.018904t} \quad (t: \text{年代}, y: \text{「漢字表記語率」}) \quad (4)$$

これに加えて杉山・森岡(1969)は「漢字表記語率」が減少する要因を明らかにする必要があるとして和語の割合、仮名表記の和語の割合、仮名表記の漢語の割合をそれぞれ被説明変数とする回帰式⁴³を計算し、それらの結果を図3のように示した。

図3から杉山・森岡(1969)は「漢字表記語率」が減少したことの要因として和語が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た。また、仮名表記の漢語は僅かに増加しているに過ぎず、「漢字表記語率」が減少したことには大きく関わらないとした。

なお、外来語について杉山・森岡(1969)は和語や漢語と同様に重要な位置を占めているとした上で100年間に約3.7%の増加に留まることから仮に外来語を図3に含めたとしても全体の傾向を歪めるものではないとした。

⁴⁰ 本節では主として「量的考察」に当たる部分を取り上げ、「質的考察」に当たる部分は取り上げないこととする。

⁴¹ 杉山・森岡(1969)によれば、賽子や乱数表を用いて対象とするページを決め、該当するページの各段に番号を振って抜き出す段を無作為に選び、選んだ段の各行にも新たに番号を振って抜き出す行を無作為に選び、選んだ行の冒頭約1,000字を抜き出したとのことである。なお、抜き出す約1,000字には助詞や助動詞、固有名詞を含めないとのことである。

⁴² この式(4)については「数学を知っている方にはすぐ分るが重大な誤植がある。引用に当たって訂正した。」という水谷(1977:84)の指摘に従って訂正した上で引用した。

⁴³ 本節では煩雑なることを避け、個々の式については取り上げないこととする。

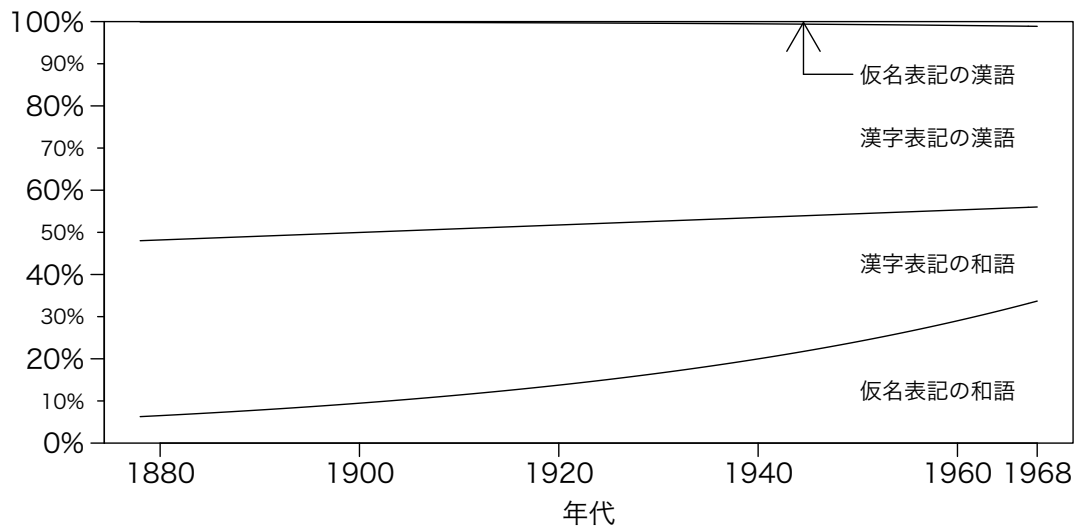


図3 杉山・森岡（1969）による「漢字表記語率」の変遷⁴⁴

3.2 野村（1988b）の研究

続いて野村（1988b）の研究について述べる。野村（1988b）は安本（1963）、宮島（1988a）、杉山・森岡（1969）の各研究に基づいて「漢字含有率」や「漢字表記語率」が今後も減少し続けることを予想した⁴⁵。

野村（1988b）は宮島（1988a）の調査結果に基づけば、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安本（1963）の予測は修正を要するとしつつも野村（1980）の示唆するところによれば、雑誌では外来語が増加しつつあるとし、小説とは異なる媒体やジャンルにおいて小説には見られない変化が起きている可能性を考慮する必要があることを指摘した。

また、野村（1988b）は安本（1963）の予測について「漢字含有率」が直線的に減少し、なおかつ、今後も同様の傾向が続くという仮定に加えて語種（和語・漢語・外来語）の構成比率に変化がないという暗黙の仮定にも基づいているとし、漢字使用の実態の解明

⁴⁴ これは杉山・森岡（1969）の第70図を概ね再現したものである。

⁴⁵ 本節では主として本書の第3章の第6節「漢字はなくなるか」を取り上げ、それ以外の節は取り上げないこととする。また、本書の一部について疑問を呈したものとして飛田（1995）の研究があるが、本論文では論旨に関わらないことから取り上げないこととする。なお、飛田（1995）の研究を取り上げるに当たっては杉山・森岡（1969）の研究について「漢字含有率」を調査したものと誤認した海保（1983）の研究を引き写したものと見えるという点と野村（1988b:235）が「漢字表記語率」と「漢字含有率」との両者が一致するものではないことに注意を促しているという点とに留意する必要がある。

や予測には別の視点——すなわち、語種の構成比率という点——を加えて考察することが必要であることを指摘し、その手掛かりとして杉山・森岡（1969）の研究を挙げた。

その上で野村（1988b）は和語の増加よりも今後は外来語の増加によって漢語が減少し続けるとし、杉山・森岡（1969）の研究と種々の調査結果⁴⁶とを参考にして示した5つの仮定⁴⁷に基づいて「漢字表記語率」が図4の太線のように減少し、これに伴って「漢字含有率」が今後も減少し続けるとの予想を示した。

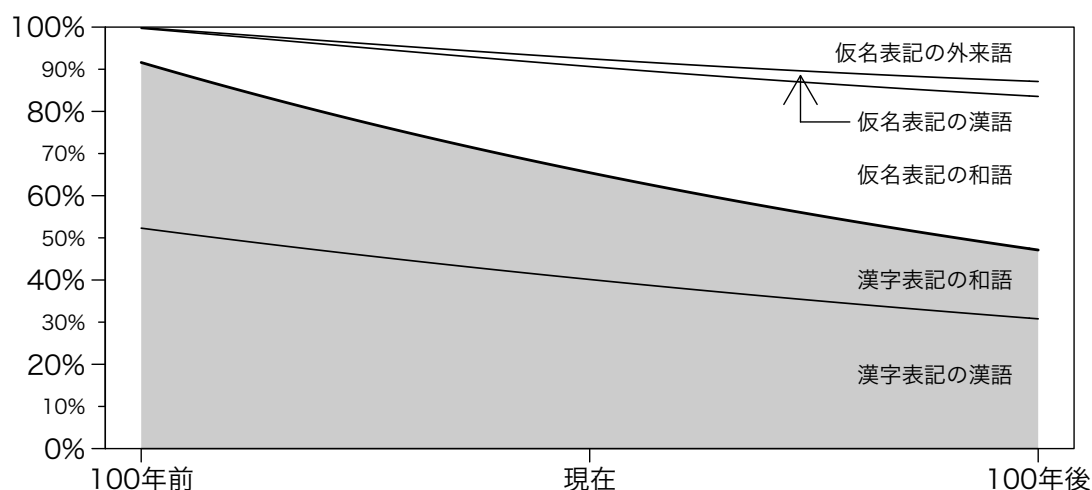


図4 野村（1988b）による「漢字表記語率」の変遷に関する予想⁴⁸

3.3 佐竹（1982, 1987a）の研究

更に佐竹（1982）の研究について述べる。佐竹（1982）は対象とする媒体やジャンルが異なることによって生ずる「漢字含有率」の差やそこに見られる法則性を明らかにするとして1979年の雑誌について月刊誌を中心とする計55誌を対象に「漢字含有率⁴⁹」を調査した。

⁴⁶ 野村（1988b:233）よれば、杉山・森岡（1969）の研究「をはじめ、これまでに引用した、各種調査のデータを利用している」とのことである。

⁴⁷ この仮定は単に図4における各曲線の通る座標を述べたものであり、別に掲げずとも図4によって明らかであることから取り上げないこととする。

⁴⁸ これは野村（1988b）の図26を概ね再現したものである。

⁴⁹ なお、佐竹（1982）の用語としても「漢字含有率」であるが、本章においては既に述べたように個々の研究における用語や算出方法については措いた上で一括して「漢字含有率」としており、詳細は第2章において検討する。

その方法は各誌の各記事を対象に計 196 箇所から最大で見開き 2 ページを抜き出し、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⁵⁰。

調査に先立って佐竹（1982）は「漢字含有率」について以下のような問題を提起した。

漢字含有率は確かに文章表記の様相を知るのに便利な指標ではある。しかし、指標として漢字含有率だけでよいかどうかとなると疑問が生じる。漢字含有率は、漢字の量と漢字以外の文字または表記記号の量との相対的な比率によって定まる値である。それゆえ漢字以外の文字または表記記号の量についての考慮がなくてはならないはずである。（佐竹 1982:329）

また、特に「カタカナ含有率⁵¹」が増加しつつある可能性を挙げ、これを考慮することも「漢字含有率」を調査するに当たって必要であることを指摘し、実際に「漢字含有率」に加えて「ひらがな含有率⁵²」や「カタカナ含有率」なども算出した。

なお、この指摘に関連して佐竹（1982）の調査結果を改めて分析した佐竹（1987a）は更に踏み込んで以下のように述べた⁵³。

かつての文章においては、漢字含有率を知れば、その文章表記のありようをある程度把握することができた。しかし、最近の文章では片仮名の比率は無視できないものが出現してきており、漢字含有率だけでは表記の姿をとらえることができなくなってきた。表記のありようを認識するには、漢字・平仮名とともに片仮名の比率をも知らねばならないのである。（佐竹 1987a:19）

⁵⁰ 佐竹（1982）によれば、1979年7月～8月発行の63誌を対象とし、小説、評論・論文、実用・解説、ルポ・報告、インタビュー・座談会、随筆・エッセー、読者投稿の7項目に該当する全ての記事を選定した上で各項目について約5万字になるように1つの記事から最大で見開き2ページ（1,000字～2,000字程度）を抽出することとし、最終的に計55誌の計196箇所から35万～36万字程度を抽出したとのことである。

⁵¹ なお、佐竹（1982）の用語としては「カタカナの比率」あるいは「カタカナ比率」であるが、本章においては既に述べたように一括して「カタカナ含有率」とする。

⁵² なお、佐竹（1982）の用語としては「ひらがなの比率」あるいは「ひらがな比率」であるが、「漢字含有率」や「カタカナ含有率」と同様に本章では文章に占めるひらがなの割合を指す用語やその算出方法については立ち入らず、一括して「ひらがな含有率」とする。

⁵³ 語句の異同は認められるものの佐竹（1987b）にも佐竹（1982）の調査結果を示した上で概ね同様の言及があることを申し添える。なお、佐竹（1987a）の研究は佐竹（1982）の調査結果に基づいて「漢字含有率」と「ひらがな含有率」と「カタカナ含有率」との間の相関を分析したものである。また、佐竹（1987a）は対象とした1979年の雑誌について計54誌しか掲げていないが、計196箇所から約36万字を抽出したとあることから総合誌の『世界』を掲げ忘れたものと見える。

その上で佐竹（1982）は「総合誌・純文学系・実用科学誌」では「カタカナ含有率」が一定で「漢字含有率」が減少すると「ひらがな含有率」が増加するという関係が見られるのに対し、「大衆文学系・女性誌・男性誌」では「漢字含有率」が減少すると「ひらがな含有率」が増加せずに「カタカナ含有率」が増加するという関係が見られるとし、これを一般化して以下のような「文章表記における字種比率に関する仮説」（佐竹 1982:345）を提示した。

- a. 漢字含有率の比較的高い文章、たとえば評論や論文などでは、漢字とひらがなとの間に補完性がある。
- b. 漢字含有率の比較的低い文章、たとえば実用記事や解説文などでは、漢字とカタカナとの間に補完性がある。
- c. 文章において、漢字が減少して仮名が増加する現象が生じる場合、まずひらがなが増加するが、ある程度の限界があり、その限界を越えるとカタカナが増加する⁵⁴。

（佐竹 1982:345）

なお、佐竹（1982:345）によれば、「補完性」とは「一方の字種比率が多くなると他方が少なくなる関係」を指すものである⁵⁵。また、これに関連して佐竹（1987a:23）は相関係数に基づいて「マイナスの値が大きいほど補完関係が強い」ことを主張した。

4 小説以外を対象に「漢字含有率」を調査した研究

最後に小説以外を対象に「漢字含有率」を調査した研究として本節では新聞を対象に「漢字含有率」を調査した梶原（1982）、土屋（2000a, b）の両研究を取り上げる。

そもそも小説以外のものを対象として「漢字含有率」を調査した研究は数多く存在するが、その多くは雑誌を対象としたものと新聞を対象としたものとの2つに大別し得るものと見える⁵⁶。

⁵⁴ 佐竹（1982:345-346）によれば、「これは模式的に説明しただけで、実際に一つの文章に対して漢字表記を仮名書き化する場合のことを述べたものではないし、漢字含有率が将来減少していくことを予言しているものでもない」とのことである。

⁵⁵ なお、佐竹（1982）の用語としての「字種」は「漢字の異なり種ではなく、文字体系の種類をさす」（佐竹 1982:328）ものである。この用語については第2章において検討する。

⁵⁶ 勿論、それ以外にも教科書を対象としたもの（例えば、吉田・福田・檜岡 1995 など）や作文を対象としたもの（例えば、島村 1987 など）があるが、論旨に大きくは関わらないことから本論文では取り上げないこととする。

そのうち雑誌を対象としたものとしては土屋（1967）、野村（1980）、佐竹（1982, 1987a）、石井（2010）の各研究がある⁵⁷。ただし、野村（1980）の研究や前掲した佐竹（1982, 1987a）の研究はいずれも共時的な観点に基づいて1979年の雑誌を対象に「漢字含有率」を横断的に調査したものであり、これを本論文の目的に照らして一定の期間に亘って継続することは困難である⁵⁸。また、土屋（1967）の研究や石井（2010）の研究は通時的な観点に基づき、特定の雑誌を対象として縦断的に調査したものであるが、土屋（1967）の研究は既に廃刊になった雑誌を対象としたものであり、石井（2010）の研究は国立国語研究所の語彙調査（国立国語研究所1987）を引き継いだ石井・入江（2015）の調査結果を利用したものであり、そのまま調査を継続するに当たってはいずれも難があると言える⁵⁹。

一方、新聞を対象としたものとしては既に述べた通り、梶原（1982）、土屋（2000a, b）の両研究がある⁶⁰。また、両者はいずれも通時的な観点に基づいて『毎日新聞』とその前身に当たる『東京日日新聞』とを対象として「漢字含有率」を縦断的に調査したものであり、そのまま調査を継続することも可能である。これに加えて小説は最も書き手の裁量が大きいと考えられる媒体やジャンルであるのに対し、新聞は最も書き手の裁量が小さい

⁵⁷ これ以外にも雑誌を対象としたものとしては水谷（1960）の研究や寿岳（1963）の研究がある。なお、前者は共時的な観点に基づき、国立国語研究所の語彙調査の一環として1956年の雑誌を対象に「漢字含有率」を調査したものである。また、後者は「漢字含有率」を調査することに主眼を置いたものではないが、1962年の週刊誌を対象に「漢字含有率」も調査したものである。

⁵⁸ なお、野村（1980）の研究は媒体やジャンルが異なることによって生ずる「漢字含有率」の差や算出方法が異なることによって生ずる「漢字含有率」の差を明らかにするとして1979年の雑誌について週刊誌を主とする計27誌を対象に「漢字含有率」を調査したものである。この研究については一部を第2章において取り上げる。

⁵⁹ 実際に土屋（1967）の調査を引き継いだ土屋（2000a）はそのまま調査を継続することに難があるとし、後述するように対象を改めていることを申し添える。なお、土屋（1967）の研究は1895年から1938年までの総合誌の『太陽』を対象として「漢字含有率」を調査し、「漢字含有率」が減少することを指摘したものである。また、石井（2010）の研究は「漢字含有率」を調査することに主眼を置いたものではないが、1906年から2006年までの総合誌の『中央公論』を対象として「漢字含有率」も調査し、「漢字含有率」が1946年を境に大きく減少することを指摘したものである。

⁶⁰ これ以外にも新聞を対象としたものとしては海保（1980）の研究や小林（2002）の研究がある。なお、前者は時期が不明であるが、後者は2002年の新聞を対象に「漢字含有率」を調査したものである。また、「漢字含有率」を調査することに主眼を置いたものではないが、1966年の新聞を対象とした斎藤（1968）の研究や2003年の新聞を対象とした矢澤（2012）の研究もある。

と考えられる媒体やジャンルであり、雑誌は両者の間に位置づけられる媒体やジャンルであるということに照らせば、小説とは対蹠的な関係にある媒体やジャンルとしての新聞を対象に「漢字含有率」を調査した研究を取り上げることは第1節において取り上げた小説を対象に「漢字含有率」を調査した研究と好対照を成すものであり、比較するに当たって適したものと言える。なお、これは両極に位置する2つの媒体やジャンルを網羅することになることから現代日本語の文章に広く見られる一般的な傾向をも捉えんとする本論文の目的にも適うものである。

このように雑誌を対象とした研究と新聞を対象とした研究とを比べると調査を継続する上でも小説を対象とした研究と比較する上でも後者の益するところが大きいことから本節では小説以外のものを対象に「漢字含有率」を調査した研究として特に新聞を対象とした梶原（1982）、土屋（2000a, b）の両研究を取り上げるに至ったのである⁶¹。

4.1 梶原（1982）の研究

まず、梶原（1982）の研究について述べる。梶原（1982）は1877年から1947年までの新聞を対象に「漢字含有率⁶²」を調査した。

その方法は『東京日日新聞』（『大阪毎日新聞』と共に1943年に題号を『毎日新聞』に統一）を対象に10年ごとに1日～3日分の全紙面を抜き出し、そこ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⁶³。

梶原（1982）は大局的に見ると70年間の「漢字含有率」は次第に減少してきたと結論した。また、その主たる要因として漢字制限と口語化⁶⁴とを挙げた。

⁶¹ なお、土屋（2000a, b）の研究は第3節において取り上げた語種に着目することが必要であるとした研究としても位置づけ得る。

⁶² なお、梶原（1982）の用語としても「漢字含有率」であるが、本章においては既に述べたように個々の研究における用語や算出方法については措いた上で一括して「漢字含有率」としており、詳細は第2章において検討する。

⁶³ 梶原（1982）によれば、1877年においては西南戦争（同年1月30日～同年9月24日）を避けて無作為に11月10日を選び、10年ごとに11月10日の紙面から抜き出すことにしたとのことである。また、それぞれの年について抜き出した紙面の延べ文字数を等しくするために1877年は3日分、1887年は2日分、それ以外の年は1日分を抜き出すとして1877年は11月10日・12日・13日、1887年は11月10日・11日、それ以外の年は11月10日の全紙面を抜き出したとのことである。

⁶⁴ 梶原（1982）によれば、1927年には社説も含めて概ね全て口語化し、1937年には完全に口語化したとのことである。

4.2 土屋（2000a, b）の研究

続いて土屋（2000a）の研究について述べる。土屋（2000a）は現在の漢字使用の実態とそこに至るまでの変遷とを明らかにした研究として安本（1963）、杉山・森岡（1969）の両研究を挙げた上で両者と概ね同様の方法を試みた土屋（1967）の研究に触れ、これをそのまま引き継ぐには難があるとしつつも改めて同様の方法を試みるとして1872年から1962年までの『東京日日新聞』（『大阪毎日新聞』と共に1943年に題号を『毎日新聞』に統一）を対象に「漢字含有率⁶⁵」と「漢字表記語率⁶⁶」とを調査した。

その方法は『東京日日新聞』ないしは『毎日新聞』の無署名の記事を対象に15年～23年ごとに計6箇所から約500文節ずつを抜き出し、そこに占める漢字の割合を算出すると共に文節を基本としつつも独自の規則に従って語に分割し、そこに占める漢字表記の語の割合も算出するというものである⁶⁷。

土屋（2000a）は新聞における「漢字含有率」が小説や雑誌におけるそれよりも相対的に高いとした上で90年間の「漢字含有率」が次第に減少してきたことを指摘し、同じく『東京日日新聞』を対象に「漢字含有率」を調査した梶原（1982）の調査結果と概ね同様の結果が得られたと述べた。

また、「漢字含有率」が減少する要因となるものとして①仮名表記の語の増加、②送り仮名の増加、③仮名表記の外来語の増加、④国語施策、⑤口語化を挙げ、「漢字含有率」が複数の要因によって変動することを指摘した。

その上で漢字使用実態を明らかにするためには「漢字表記語率」に着眼する必要があるとして調査結果を図5のように示した。

⁶⁵ なお、土屋（2000a）の用語としても「漢字含有率」であるが、本章においては既に述べたように個々の研究における用語や算出方法については措いた上で一括して「漢字含有率」としており、詳細は第2章において検討する。また、土屋（2000a）の調査を引き継いだ土屋（2000b）の用語としても「漢字含有率」であることを申し添える。

⁶⁶ 土屋（2000a）は文章に占める漢字表記の語の割合を指す用語について明確には言及していないが、本論文では既に述べたように個々の研究における用語や算出方法は措いた上で一括して「漢字表記語率」とする。なお、土屋（2000a）の調査を引き継いだ土屋（2000b）は「漢字表記語率」としていることを申し添える。

⁶⁷ 土屋（2000a）によれば、1872年、1895年、1917年、1932年、1947年、1962年の記事を選んだとのことであるが、記事を選んだ基準については不明確である。また、記事から約500文節を抜き出す方法は不明であるが、土屋（1967）と同様であるならば、各記事の冒頭から抜き出したものと見える。なお、1872年は479文節、それ以外の年は500文節をそれぞれ抜き出し、固有名詞は含めたとのことである。

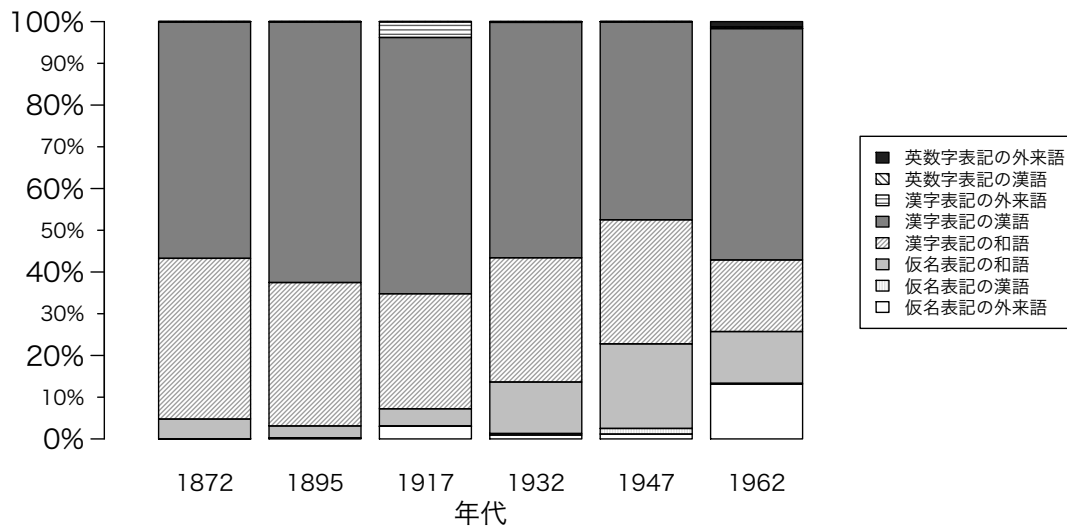


図5 土屋（2000a）による「漢字表記語率」の変遷⁶⁸

図5から土屋（2000a）は「漢字表記語率」が減少していることを指摘し、その要因として口語化と仮名表記の和語の増加とを挙げた。また、和語の増加は認められないとして杉山・森岡（1969）の指摘は裏づけられなかったと述べた⁶⁹。

これに加えて土屋（2000b）の研究について述べる。土屋（2000b）は土屋（2000a）の調査を継続するとして1975年、1990年、2000年の『毎日新聞』の一面を対象に計3箇所から500文節ずつを抜き出し、概ね同様の方法によって調査した。

土屋（2000b）は土屋（2000a）の調査結果と合わせた計128年間の「漢字含有率」が減少し続けることを指摘する一方、仮名表記の和語が1975年を境に減少に転じたことを報告した。また、その要因としてワープロによる影響を挙げた⁷⁰。

5 現状の整理

ここでは「漢字含有率」に関する研究の現状について整理する。

⁶⁸ これは土屋（2000a）の図3を概ね再現したものである。

⁶⁹ これに関連するものとして近藤（2019a）や近藤（2019b）の研究がある。このうち前者は特に口語化と和語の増加との関係について考える上では重要な研究であり、土屋（2000a）を挙げてはいないものの杉山・森岡（1969）研究の再録に当たる杉山・森岡（1991）にも言及がある。

⁷⁰ なお、土屋（2000b:304）は「ワープロが表記に影響を及ぼすことを、計量的に確かめたのは、おそらくこれが初めてであろう」と記しているが、そのような研究としては鶴岡（1991）の研究や金・樺島・村上（1993）の研究が既にあることを申し添える。

まず、一連の研究の嚆矢とも目せられる安本（1963）の研究が小説を対象として「漢字含有率」を調査し、「漢字含有率」が次第に減少することを報告した。また、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かと予測した。

これについて本節においては「この傾向が今後も続くか」という点と「小説における「漢字含有率」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」という点とを取り上げ、それぞれの問題について以下に検討を試みる。

初めに前者の問題について述べる。まず、第1節において取り上げた小説を対象として「漢字含有率」を調査した研究を整理すると小説における「漢字含有率」が20世紀の初葉から中葉に掛けて減少したという点ではいずれの調査も一致している。しかし、その傾向がその後も続くかという点では宮島（1988a）の予想と中村（2003）の結果とが対立している。

また、各研究における統計的な手法の扱い方については綾（2013）や石井（2013）が安本（1963）のそれに対して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）のそれにも該当するものである。これらの指摘を考慮すれば、20世紀後半の小説における「漢字含有率」の変遷についての見解の相違は対象の選び方や「漢字含有率」の算出方法が異なることに加えて統計的な手法の扱い方に問題があったことにも起因していると言えるのではないかと。

更に小説においては21世紀に入ってから新たな調査が管見の限りでは存在せず、現在に至るまで小説における「漢字含有率」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にある。すなわち、これまで述べてきた問題を克服し、その後の変遷を明らかにする調査が早急に求められている⁷¹。

これを踏まえ、第3章においては統計的な手法の扱い方に留意しつつ実際に小説を対象として「漢字含有率」を調査し、小説の文章における漢字使用の実態について現在に至るまでの変遷を明らかにせんとする。また、調査に当たっては佐竹（1982, 1987a）の指摘を踏まえ、漢字のみならず、漢字以外の文字体系をも考慮することとする。

⁷¹ 犬飼（2002）は綾（2013）や石井（2013）の指摘するような問題には触れていないものの宮島（1988a）以降の「漢字含有率」がどのような変遷を辿っているのかを明らかにすることは「調査・研究に値する課題」（犬飼 2002:140）であると述べている。

なお、「漢字含有率」を調査するに当たってはルビの有無を考慮することが必要であるという田中（1987, 1989）の指摘もあるが、これは「総ルビ印刷の場合には、漢字含有率は、漢字のウェイトを測る尺度にはなりえない（このことは、漢字表記語率でも同様である）」（田中 1987:37）とした上でルビのことを考慮せずには「明治・大正期の表記における、漢字のウェイトを測ることは、不可能である」（田中 1987:37）ことを指摘したものであり、本論文の論旨に大きくは関わらないものと見える⁷²。

次に後者の問題について述べる。まず、安本（1963）は小説における「漢字含有率」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「漢字含有率」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⁷³。

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。

すなわち、「小説における「漢字含有率」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」という安本（1963）の予測は偽であり、小説の文章における傾向を日本語の文章にまで一般化することには極めて慎重になる必要がある。

これらの2つの問題に続いて立ち現れる問題として本節においては「漢字含有率」が変動する要因は何か」という点についても以下に取り上げる。

20世紀の中葉に至るまで「漢字含有率」が減少し続けてきたことの要因について改めて整理すると宮島（1988a）は漢語の減少、漢字表記の和語の減少、漢字表記の漢語の減少の3点、杉山・森岡（1969）は和語の増加、漢字表記の和語の減少の2点、土屋（2000a）は仮名表記の和語の増加の1点をそれぞれ挙げている⁷⁴。更に外来語が増加するという野村（1988b）の予想や仮名表記の和語が減少しているという土屋（2000b）の報告も

⁷² 勿論、「漢字の増減を論じるさいには、振り仮名の有無との関係に、十分に注意する必要がある」という田中（1989:291）の指摘は重要である。また、第4章においては一部のルビを考慮しており、ルビを全く考慮していないわけではないことを申し添える。

⁷³ これは安本（1963）の式（1）のように漢字の割合が今後も直線的な減少し続けるという仮定の上では真であるが、本論文はこのような仮定に依拠するものではない。また、これが仮に真であったとしてもそのことは本論文の議論に何ら影響を与えない。なお、安本（1963）の式（3）はある非負の定数に上から漸近する場合に該当する。

⁷⁴ 杉山・森岡（1969）は漢語の減少や仮名表記の和語の増加にも言及しており、土屋（2000a）は仮名表記の漢語の増加や外来語の増加にも言及しているが、本章においては主たる要因と看做し得るものに限って取り上げてある。

踏まえる必要がある。すなわち、「漢字含有率」が変動することの要因として挙げられてきたものを集約すると語種に着目したものが多くことが読み取れる。

勿論、これ以外にも梶原（1982）が挙げる漢字制限、梶原（1982）と土屋（2000a）とが挙げる口語化、土屋（2000b）が挙げるワープロによる影響などの語種に着目したもので以外についても要因として考えることは可能である。しかし、このうちの口語化は文語体を念頭に置いたものであって殆どが口語体の文章を対象とする本論文において主要な要因とは看做し難いものと考えられる。また、20世紀の末葉から現在に至るまでの情報機器の発達・普及による書字環境の変化や国語施策の転換に鑑みれば、ワープロによる影響や漢字制限に着目することは重要であるが、このような言語外的な要因については言語内的な要因とは別に検討する必要があるものと考えられる。なお、ワープロの使用前後では「漢字含有率」に大きな変化が見られないという金・樺島・村上（1993）の報告もあり、これに従うならば、まずは言語内的な要因に力点を置くことが重要であると言える。

これを踏まえ、第4章では多くの研究において「漢字含有率」が変動することの主要な要因と看做されてきた語種に関わるものに着目し、これに基づいて第3章において明らかにする小説の文章における現在の漢字使用の実態とそこに至るまでの変遷とを形作る要因について考察することとする。

これに加えて最後の問題として本節では「他の媒体やジャンルにおいては如何なる傾向にあるのか」という点について以下に取り上げる。

既に述べた通り、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も考慮する必要があるという野村（1988b）の指摘がある。この指摘を踏まえれば、広く現代日本語の文章における漢字使用の実態を明らかにせんとする本論文の目的を達するためには小説以外の媒体やジャンルを取り上げることが不可欠である。

また、小説は最も書き手の裁量が大きいと考えられる媒体やジャンルであるのに対し、新聞は最も書き手の裁量が小さいと考えられる媒体やジャンルであり、雑誌は両者の間に位置づけられる媒体やジャンルであるということに照らせば、小説と対蹠的な関係にある新聞を取り上げることは両極に位置する2つの媒体やジャンルについて網羅することを可能にするという点において現代日本語の文章に広く見られる一般的な傾向をも捉えんとする本論文の目的に適うものであることも既に述べた通りである。

更に雑誌においては21世紀に入ってから新たな調査として石井（2010）の研究がある

のに対して新聞においては小説と同様に 21 世紀に入ってから新たな調査が管見の限りでは存在せず、現在に至るまで新聞における「漢字含有率」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にある⁷⁵。すなわち、これまでに述べてきた問題を克服し、その後の変遷を明らかにする調査が早急に求められているのは小説においても新聞においても同様であると言えるのである。

これを踏まえ、第 5 章においては新聞を対象に「漢字含有率」を調査し、新聞の文章における漢字使用の実態について現在に至るまでの変遷を明らかにせんとする。また、調査に当たっては小説を対象として調査する際に考慮した種々の点に加えて新聞に特有の事情にも留意しつつ第 3 章や第 4 章において整備する方法論の適用を試みることにする。

⁷⁵ なお、既に述べたように石井（2010）の研究は国立国語研究所の語彙調査（国立国語研究所 1987）を引き継いだ石井・入江（2015）の調査結果を利用したものであり、「漢字含有率」を調査することに主眼を置いたものではないことから「漢字含有率」を調査することに主眼を置いた他の研究や本論文の調査結果と比較する上では調査や分析の方法に課題を残すものであると考えている。勿論、国立国語研究所の語彙調査（国立国語研究所 1987）を継続すること自体は極めて意義深いことであり、それ自体を否定する意図は全くないが、調査や分析の方法は研究の目的に応じて変わるものであり、「漢字含有率」を調査することに主眼を置いて考えると再考の余地があるという（だけの）ことである。

第2章

用語の定義

ここでは本論に入るための準備として本論文における用語の定義について示す。初めに「字」, 「文字」, 「字種」を取り上げ, それらの定義を示し, 次に第1章において一括して「漢字含有率」とした文章に占める漢字の割合についても検討し, 先行研究における用語とその算出方法とを整理した上で本論文において採用する定義について述べることとする。

1 「字」及び「文字」

まず, 「文字」を定義した樺島(1977)の研究を取り上げる。樺島(1977)は「言語の構造に対応して記号列を作る記号を表記記号と呼び, 附属文字の中で自立文字と形が同じもの, 同じでないものを区別すると, 記号は次のように分類することができる」(樺島1977:31)として以下の分類を提示した。

- I 表記記号(言葉に対応して記号列を作るもの)
 - 1 表記要素列の構成要素となる。
 - 1・1 表記要素の構成要素となる。
 - 1・1・1 単独で表記要素になる(自立文字) ……①
 - 1・1・2 単独では表記要素にならない(附属文字)
 - 1・1・2・1 自立文字と同じ形を持つ。 ……②
 - 1・1・2・2 自立文字と異なる形を持つ。 ……③
 - 1・2 表記要素(列)の反復を示す。 ……④
 - 2 表記要素列の構成要素とならない。 ……⑤

II 表記記号ではない記号（交通標識、地図の記号など）

（樺島 1977:31-32）

ここから樺島（1977:32）は「①—③を文字とするのが合理的」であるとし、「文字」を以下のように定義した。

文字とは、言葉に対応して、表記要素の構成要素となる記号をいう。

（樺島 1977:32）

なお、樺島（1977）によれば、「一（長音符）」は③に該当し、「々（同の字点）」や「ヽ（一の字点）」は④に該当し、括弧や句読点は⑤に該当するとのことである。

これを踏まえ、本論文では原則として樺島（1977:32）の定義に従って文字という用語を用いることとする。ただし、便宜的に①から⑤までの表記記号を指して字と数えることがある。

2 「字種」

続いて「字種」について述べる。これを取り上げるのは佐藤（2013）の述べるところによれば、「字種」という術語に樺島（1977）の言うところの「文字体系」を指す場合と「文字素」を指す場合とがあるとのことであり、両者の混同を避けるためである⁷⁶。

なお、「文字体系」について樺島（1977:26）は「発生的、歴史的に一つのまとまりを形作る文字の集合」と定義しており、その例として「漢字、ひらがな、カタカナ、ラテン文字、ハングル」を挙げている。また、「文字素」について樺島（1977:33-34）は「文字が書かれたり印刷されたりしたときに示す幾何学的な形を字形とよぶ」とし、「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を文字素とよぶ」と定義している⁷⁷。

一方、「漢字含有率」を調査した研究を見ると「字種」に言及しているのは佐竹（1982, 1987a）、宮島（1988a）、土屋（2000a）の各研究である。

⁷⁶ 菊地（2020）は樺島（1975）の「文字素」について「「字種」に該当するものか、「字体」に該当するものかが明確でない」（菊地 2020:462）と述べており、この点は樺島（1977）の「文字素」についても明確ではない可能性があるが、論旨には関わらないことから措くこととする。

⁷⁷ 本論文では如何にして同じと判定するかという問題については立ち入らないこととする。

特に佐竹（1982）は「字種（漢字の異なり種ではなく、文字体系の種類をさす）」（佐竹 1982:327-328）と述べており、明確に樺島（1977）の「文字体系」を指すものと解せる。

また、宮島（1988a）は各作品から抜き出された 1,000 字に占める「字種ごとの数」を示しており、この「字種ごとの数」はひらがな、カタカナ、ローマ字、漢字という 4 つの項目に分けられていることから宮島（1988a）の「字種」が樺島（1977）の「文字体系」を指すものと読み取れる。更に土屋（2000a）は「字種・語種の判定」（土屋 2000a:197）と述べており、これも「文字体系」を指すものと読み取れる。

すなわち、「漢字含有率」を調査した研究においては「字種」が樺島（1977）の「文字体系」を指す語として用いられていることから本論文もこれに倣うこととした⁷⁸。

3 「漢字含有率」

更に先行研究における文章⁷⁹に占める漢字の割合を指す用語とその算出方法とを整理すると共に本論文において採用する定義について述べる。

まず、文章に占める漢字の割合を指す用語としては実際に調査した研究を中心に見ると「漢字の使用率」とする研究（安本 1963; 内田 1964）や「漢字使用率」とする研究（金・樺島・村上 1993; 中村 2003）があるのに対して「漢字含有率」とする研究（水谷 1960; 土屋 1967, 2000a, b; 田原 1975; 野村 1980; 梶原 1982; 佐竹 1982, 1987a; 田中 1987, 1989; 宮島 1988a; 鶴岡 1991; 石井 2010; 羽山 2017）があり、研究によって用語が異なっている現状にある⁸⁰。しかし、専門事典に当たると文章に占める漢字の割合を指して「漢字（の）使用率」とするものは見られない一方で「漢字含有率」とするものは数多く見られることから後者については術語として認める立場のあることが読み取れる⁸¹。これに従って本論文では文章に占める漢字の割合を指す用語として「漢字含有率」を採用する。

⁷⁸ なお、文字・表記論に関する基本的な術語について解説した水谷（1987:6）は「本来的に別の文字体系を成すと言える、漢字・仮名・ローマ字・アラビア数字などの別を、字種の違いという」と述べているのに対し、文字・表記論に関する「共通概念や術語の整備が急務である」とした犬飼（1988:83）は「漢字と仮名とローマ字の別をどう呼べば良いか。漢語と和語と外来語の別を語種と呼ぶのにならって字種と呼ぶことはできない。」と述べているが、本節においては異論のあることを示すに留める。

⁷⁹ 本論文における「文章」とは樺島（1977）が述べている「言葉に対応した「記号列」に相当する。

⁸⁰ 水谷（1983）、佐竹（1987b）、野村（1988b）はいずれも安本（1963）について「漢字含有率」を調査した研究として位置づけており、文章に占める漢字の割合を「漢字含有率」とするものは数多く見られる。

次に算出方法について整理する。「漢字含有率」の算出方法については算出式の分子を延べ漢字数とする点ではいずれも共通しているが、小説を対象に調査した研究を例として見ると分母は延べ表記記号数とする安本（1963）の研究、延べ文字数とする宮島（1988a）の研究、延べ仮名数と延べ漢字数との和とする中村（2003）の研究があり、研究によって異なっている現状にある⁸²。また、この点は専門事典でも同様である⁸³。

なお、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⁸⁴を $N_T(G)$ 、異なり字数⁸⁵を $V_T(G)$ とそれぞれ表現することとし⁸⁶、小説を対象として調査した研究を中心に算出方法の整理を試みる。

まず、安本（1963）の「漢字の使用率」の算出方法を式（5）に示す⁸⁷。なお、延べ句点数と延べ読点数とをそれぞれ仮に $N(\{.\})$, $N(\{, \})$ と置いた⁸⁸。更に安本（1965）は

⁸¹ 『漢字百科大事典』によると「漢字含有率」（田中 1996a:133）は文章に占める漢字の割合を指すのに対して「漢字使用率」は「文書データの表記に現れた漢字の総数（延べ語数）に対する、個々の漢字の使用度数（頻度）の比率」（田中 1996b:141）を指すものであり、そもそも両者を別の術語とする立場もある。

⁸² ここでは「漢字含有率」の算出方法を分類した野村（1980）を参考にしつつ分母について延べ表記記号数とする式、延べ文字数とする式、延べ仮名数と延べ漢字数との和とする式として整理した。

⁸³ 「漢字含有率」を立項する『漢字百科大事典』には「文書データの表記に用いられている総文字数に対する、漢字数の割合」（田中 1996a:133）との記述があり、『計量国語学事典』には「文字全体のなかで漢字が占める割合（漢字含有率）」（横山 2009:57）との記述がある。また、『計量国語学事典』には「文書全体の表記に用いられている総文字数に対する漢字数の割合」（高田 2009:50）との記述もある。更に索引に「漢字含有率」を掲げる『日本語学大辞典』には「一定の文章中に用いられる文字（場合によって表記記号も加える）の総数に対する漢字数の比率」（野村 2018:173）との記述と「使われた漢字・ひらがな・カタカナの総数に対して漢字が何パーセントを占めているかを計算した値を「漢字含有率」（梶原 2018:548）とするとの記述とがある。このように分母については延べ文字数と解し得る記述が多いものの研究の目的に応じて異なる算出式も用いられている。なお、『日本語学研究事典』や『日本語大事典』の索引には「漢字含有率」が掲げられていないことを申し添える。

⁸⁴ 野村（1988a:340）によれば、「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。このうちの「漢字」を本論文では「文字」に読み替える。また、如何にして同一の文字と判定するかという問題には立ち入らない。

⁸⁵ 野村（1988a:341）によれば、「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論文においては「漢字」を「文字」に読み替える。

⁸⁶ 文章 T を逐一明示すると煩雑になることから以降は省略する。

⁸⁷ 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。

漢字以外の字種の内訳を示しておらず、これを本論文では $N(H \cup K \cup R \cup A)$ とした⁸⁹。

$$\text{漢字の使用率} = \frac{N(C)}{N(H \cup K \cup R \cup A) + N(C) + N(\{。 \}) + N(\{、 \})} \times 1000 \quad (5)$$

続いて宮島 (1988a) の「漢字含有率」のそれを式 (6) 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、これを本節では延べ英数字数と呼び、 $N(R \cup A)$ と表現した。

$$\text{漢字含有率} = \frac{N(C)}{N(H) + N(K) + N(C) + N(R \cup A)} \times 1000 \quad (6)$$

更に中村 (2003) の「漢字使用率」のそれを式 (7) に示す。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、本節では延べ仮名数を $N(H \cup K)$ とした。

$$\text{漢字使用率} = \frac{N(C)}{N(H \cup K) + N(C)} \times 1000 \quad (7)$$

なお、土屋 (2000a) は漢字の割合について 2 つの算出方法を示しており、それぞれを「含有率 1」、「含有率 2」と呼んでいる。この「含有率 1」は分母が全ての表記記号の延べ字数⁹⁰となっており、概ね安本 (1963) の「漢字の使用率」に相当するものと見える⁹¹。また、土屋 (2000a) の「含有率 2」は宮島 (1988a) の「漢字含有率」に相当するものと見える⁹²。

⁸⁸ 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ことから延べ字数 $N_T(G)$ のように延べ句点数と延べ読点数とを表現することは不可能であるが、この問題について本論文では一旦留保し、仮に $N(\{。 \})$ と $N(\{、 \})$ とそれぞれ置いた。

⁸⁹ 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ものの含めないとも述べておらず、本節では暫定的に含められていると考えた。また、句読点以外の記号についての扱いも不明であり、それらも本来は分母に含まれていた可能性がある。

⁹⁰ 全ての表記記号について延べ字数を定義することは野村 (1988a:340) の「延べ字数」の定義における「漢字」を「文字」ではなく、本論文における「字」、すなわち、樺島 (1977:31-32) の①から⑤までの全ての表記記号として読み替えれば、可能になるものの文章 T における各文字体系 G の延べ字数 $N_T(G)$ とする表現についても修正の必要が生じてくることから本論文では改善の必要性を指摘することに留める。なお、これは先行研究を整理する際に起こる問題であって本論文の議論には直接関わらないものである。

⁹¹ 実際には安本 (1963) の「漢字の使用率」の分母は不明であるが、仮に全ての表記記号の延べ字数ということであったとすれば、土屋 (2000a) の「含有率 1」と安本 (1963) の「漢字の使用率」とは同一ということになる。

⁹² 土屋 (2000a) は「含有率 2」について「記号類を除いた文字の中における漢字の占める比率」(土屋 2000a:198) としており、これを踏まえたものである。ただし、「漢字含有率の計算に際しては、漢字と仮名のみを数えることも考慮した」とも述べていることから中村 (2003) の「漢字使用率」と同様である可能性もある。また、土屋 (2000b) は「含有率 2」について「句読点・英数字等を除いて漢字と仮名のみを取り出した中での含有率である」(土屋 2000b:304) と述べており、これらの異同については不明である。

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約 3% とする報告もあり⁹³、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論文の位置づけをも左右するものである。

本論文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を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は本論文の対象に含まれない。これらの点から本論文では文章に占める漢字の割合の算出方法として宮島（1988a）の「漢字含有率」を採用し、これを漢字含有率とする⁹⁴。

なお、基本的には安本（1963）や宮島（1988a）に倣って漢字含有率を千分率によって示しているが、百分率か千分率かは割合の示し方の差であって算出方法には含まれないとの立場を採ることを申し添える。

⁹³ 野村（1980）は 1979 年の週刊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、この結論を得ている。なお、どの程度変動するかは改めて検討する必要がある。

⁹⁴ なお、「漢字含有率」という用語については田中（1996a:133）が「調査対象とする文書データに用いられている文字の種類（平仮名・片仮名・ローマ字・漢字）の比率の中から、漢字の比率を取り上げたものである」と述べており、これは本論文において採用した宮島（1988a）の「漢字含有率」と概ね一致する。また、「漢字使用率」という用語については田中（1996b:141）が「文書データの表記に現れた漢字の総数（延べ語数）に対する、個々の漢字の使用度数（頻度）の比率」と述べており、中村（2003）の「漢字使用率」と全く異なるという点に留意する必要がある。

第3章

小説の漢字含有率は減少し続けるのか

ここでは小説を対象に文章に占める漢字の割合を実際に調査し、漢字使用の実態の現在に至るまでの変遷を明らかにする。また、漢字の割合の変動については先行研究の指摘を踏まえつつ統計的な分析に基づいて考察する。

1 調査の方法

初めに調査の方法を対象と手順とに分けて以下に述べる。

1.1 対象

まず、本論文では作品の基準を一定にするという前掲の宮島（1988a）の方針¹⁹を踏襲し、芥川賞作品における漢字含有率を調査の対象とする。また、既に宮島（1988a）が1935年度から1985年度までの各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⁹⁵。なお、対象とする本文は前掲の宮島（1988a）の指摘²¹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、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

⁹⁵ 宮島（1988a）が「年度」を単位としたのに対し、本論文は「年」を単位としたことから米谷ふみ子「過越しの祭」が双方の調査に含まれている。このことを利用して本論文ではこの作品における字種ごとの延べ字数が宮島（1988a）と本論文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。

である⁹⁶。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。

これについて更に説明を加えるために図 6 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 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 行目の先頭から 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のような文字以外のものを取り除いてある。なお、抜き出す 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。

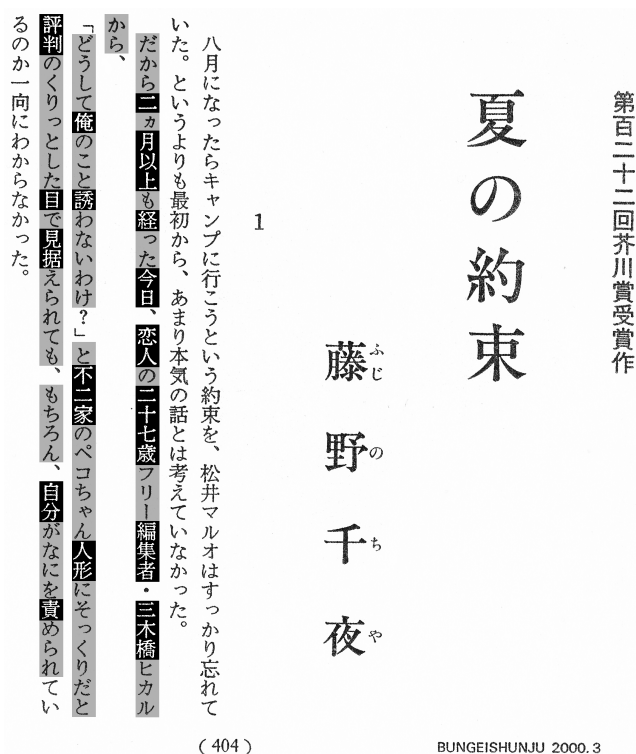


図 6 藤野千夜「夏の約束」の冒頭 1 ページ⁹⁷ (一部)

1.2 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) を使用

⁹⁶ 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。

⁹⁷ これは 2000 年 3 月発行の『文藝春秋』78 巻 4 号 (三月特別号) 404 ページから一部分を引用したものである。また、説明のために網掛けを施した。

し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照らし合わせ、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。

2 分析の手法

続いて本章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、それぞれの字種の変遷を散布図によって図示し、調査の全体を概観する。また、相関分析を実施することにより、漢字とそれ以外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について Kruskal-Wallis 検定を用いて分析する。更に箱ひげ図を示し、これを平滑化することにより、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.3.3.2) を用いることとする。また、手法の概要については以下に述べる。

2.1 相関分析

初めに相関分析について簡単に述べる。まず、相関係数とは「数量データと数量データの間“直線的”な関連が見受けられるかどうかを明らかにする指標」(高橋 2004:120)であり、 -1 から $+1$ までを取り、2つの変数が強く関連していると ± 1 に近づき、反対に関連していないと 0 に近づくというものである。

また、これに対する検定は竹内 (2014:122) によれば、「相関係数が無相関（つまりゼロ）である確率を計算し、それが一定の水準よりも低ければ、得られた相関係数が「無相関（ゼロ）とはいえない」と主張するために」用いられる。

2.2 Kruskal-Wallis 検定

次に本論文が採用する 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な検定である (Kruskal & Wallis 1952)。すなわち、この手法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利点を有している（田中・垂水 1999）。

このような手法を採用する背景には本論文において扱う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ということがある。また、正規分布を仮定する手法を用い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、本論文では広く用いられる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 蛇行箱型図

続いて作図の手法について述べる。まず、箱ひげ図とは渡部・鈴木・山田・大塚（1985）によれば、観測値の分布を図によって要約する手法の 1 つであり、箱の中心にある直線が中央値を示し、箱の幅に全観測値の約 50% が収まるように描かれるものである。

次に変形した箱ひげ図に当たる蛇行箱型図とは渡部・鈴木・山田・大塚（1985:81）によれば、「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を描くこと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、同一の時点に複数の観測値がある場合にも有効であるという利点を有している（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本論文の目的に適うものであり、この手法を本論文において採用する所以である⁹⁸。

3 調査の結果

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。

⁹⁸ 推測統計学（「確認的データ解析」）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は第 4 節第 3 項に記す。

表 1 芥川賞作品（1986年～1999年）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	$N(H)$	$N(K)$	$N(R \cup A)$	$N(C)$	$V(C)$
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	瀧澤美恵子「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「運転士」	626	118	0	256	159
1992 下	1960	1993	多和田葉子「犬婚入り」	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	笙野頼子「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「この人の闘」	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「プ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年～2015年）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	$N(H)$	$N(K)$	$N(R \cup A)$	$N(C)$	$V(C)$
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	吉村萬壺「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	糸山秋子「沖で待つ」	689	52	4	255	172
2006 上	1971	2006	伊藤たかみ「八月の路上に捨てる」	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「ボトスライムの舟」	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

4 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7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。

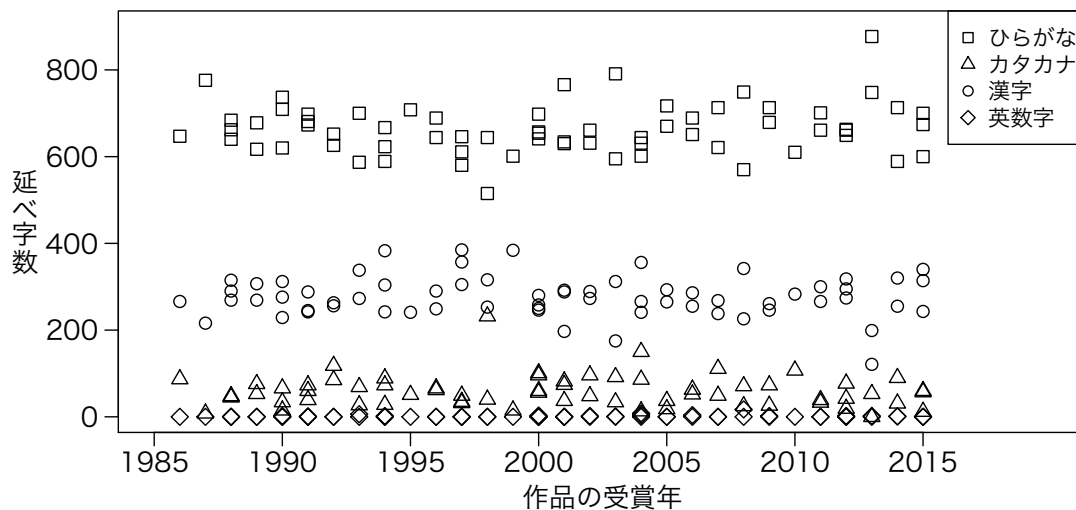


図7 字種ごとの延べ字数の変遷

これらの調査結果に対する分析として以下では初めに相関分析を実施して全体の動向を概観し、次に Kruskal-Wallis 検定を用いて漢字含有率の変動について分析する⁹⁹。更に蛇行箱型図によって漢字含有率の変遷を素描する。

4.1 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。

4.1.1 延べ漢字数とその他の字種の延べ字数との関係

まず、延べひらがな数と延べ漢字数との関係を図8に示し、延べカタカナ数と延べ漢字数との関係を図9に示す。

⁹⁹ 定義より、本章における1,000字に占める延べ漢字数の値は漢字含有率の値と一致する。

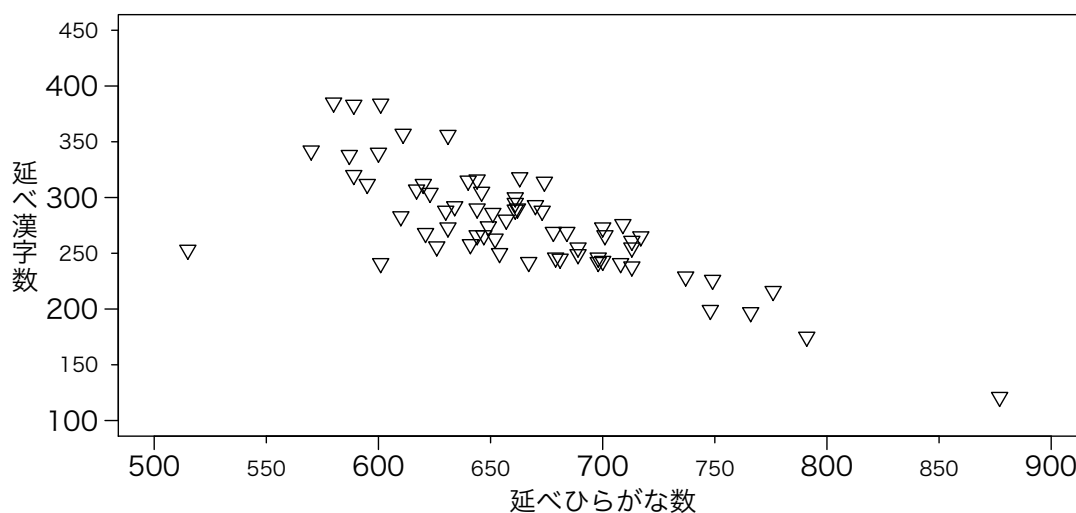


図8 延べひらがな数と延べ漢字数との関係

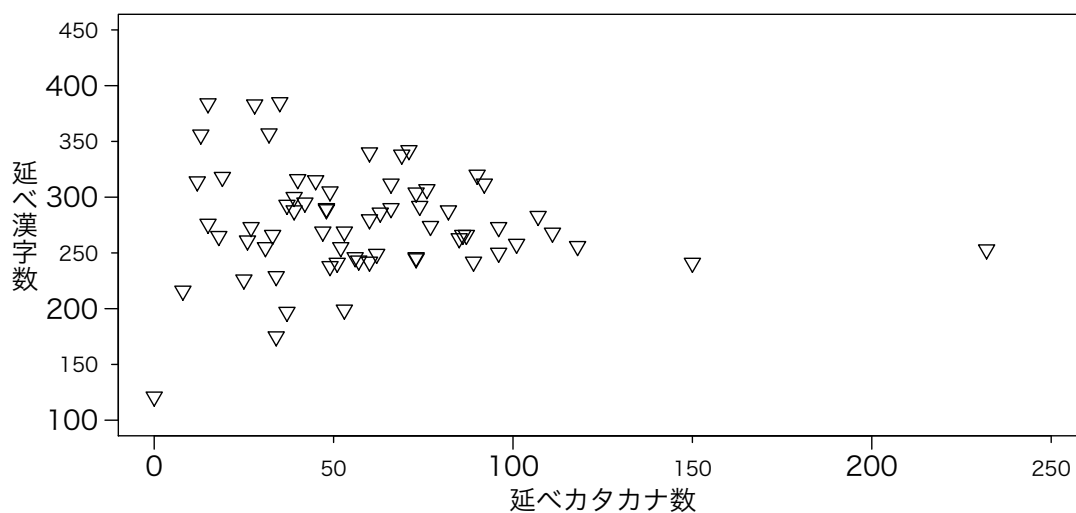


図9 延べカタカナ数と延べ漢字数との関係

続いて相関分析を実施した結果、図8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見られた ($r = -.773, p < .005$)。また、図9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なかった ($r = -.066, p = .528, n.s.$)。このことは純文学系の雑誌において「漢字含有率」と無関係に「カタカナ含有率」が一定であるのに対し、漢字とひらがなとの間には「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も負の相関が見られた ($r = -.581, p < .005$)。

4.1.2 受賞年や誕生年と延べ漢字数との関係

次に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。初めに作品の受賞年と延べ漢字数との関係を図 10 に示す。

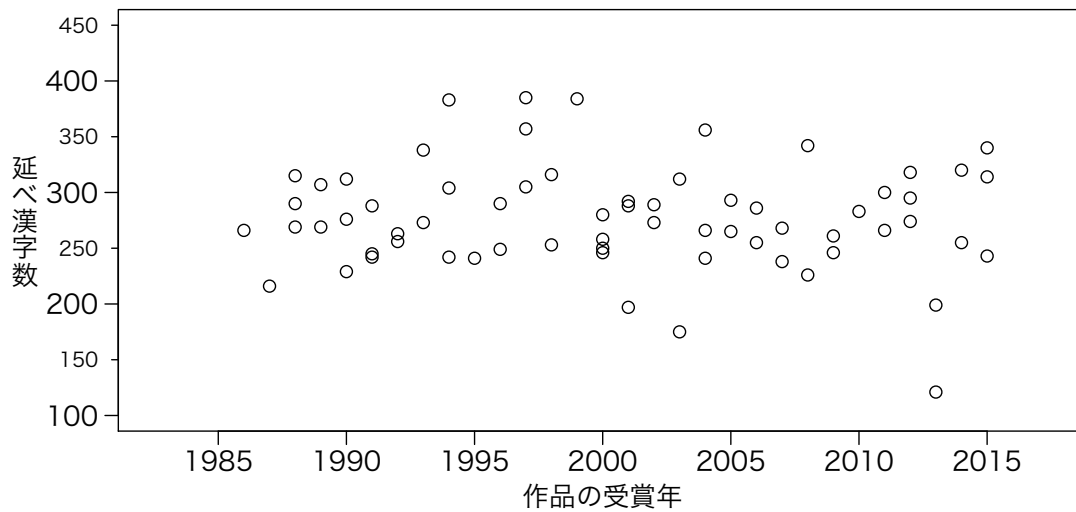


図 10 受賞年と延べ漢字数との関係

図 10 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ところ、両者の間に相関は認められなかった ($r = -.081, p = .516, n.s.$)。次に作者の誕生年と延べ漢字数との関係を図 11 に示す。

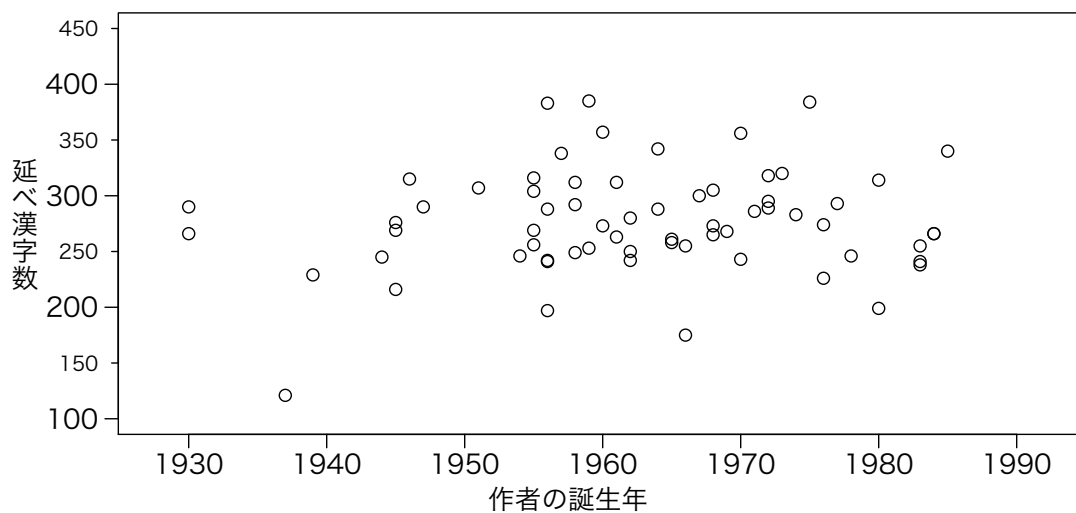


図 11 誕生年と延べ漢字数との関係

図 11 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、「漢字含有率」が次第に減少するという先行研究に見られたような傾向が本章において対象とした 1986 年から 2015 年までの芥川賞作品においても見られるか否かを相関分析によって分析したところ、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次第に減少するという結果は得られなかった¹⁰⁰。

4.2 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)、宮島 (1988a) の両研究において「漢字含有率」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) の指摘に従っ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。各区分の箱ひげ図を図 12 に示す。

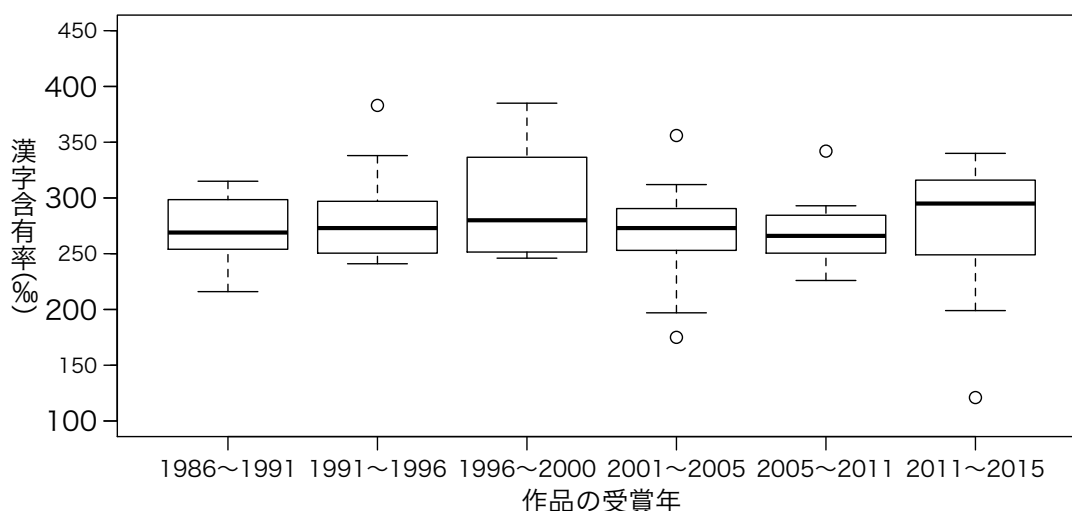


図 12 各区分の箱ひげ図

また、図 12 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得られた。

¹⁰⁰ ここでは作品の漢字含有率が作品の受賞年によって決定づけられるのか作者の誕生年によって決定づけられるのかという点については扱わない。

4.3 蛇行箱型図

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。

これを踏まえ、本章では検定の枠組みのみならず、これを補うものとして蛇行箱型図を用いて漢字含有率の変遷を素描することとし、図 13 に蛇行箱型図によって描いた漢字含有率の変遷を示す。

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、本章調査分は本章の調査結果を石井（2013）に倣って作図したものである¹⁰¹。また、先行調査分と本章調査分とは別々に作図し、両図の繋ぎ目は平滑化せずに破線によって繋いでいる。このように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。

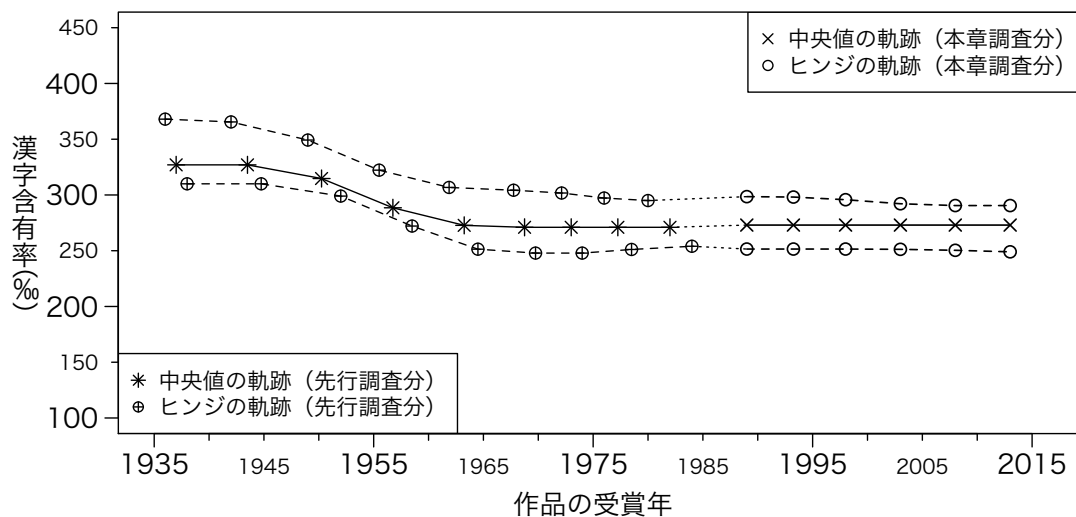


図 13 漢字含有率の変遷

図 13 を見ると中央値の軌跡も上下ヒンジの軌跡も横ばいである。このことは本章の調査において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唆している。

¹⁰¹ 米谷ふみ子「過越しの祭」は先行調査分に含めず、本章調査分に含めた。

すなわち、この結果から 1986 年以降について漢字含有率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。

5 本章の結論

第 3 章では 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、本章において示されたような小説における「漢字含有率」の安定傾向または増加傾向¹⁰²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するのも真であることから小説における「漢字含有率」にこのような非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らないという結論を導くことが可能となる。

¹⁰² これは「減少傾向」の否定——すなわち、「安定傾向または増加傾向」——である。

第4章

小説の漢字含有率はなぜ安定したのか

第3章においては1986年から2015年までの芥川賞作品を対象に各作品の漢字含有率を調査し、この30年間は漢字含有率が安定しつつあるという傾向について報告した。

しかし、なぜ安定するに至ったのかという点は依然として明らかになっていない現状にある。これを踏まえ、本章では語種に着目して漢字含有率の安定する傾向を形作る要因について分析を試みる。

1 漢字含有率の安定を形作る要因を探る方法

初めに安定を形作る要因を探る方法について述べる。本章では宮島（1988a）の方法に従って各作品から抽出した1,000字について語種別に延べ平仮名数、延べ片仮名数、延べ漢字数、延べ英数字数を算出し、その変遷を後述する手法を用いて明らかにすることによって漢字含有率が安定した要因を探ることとする。

この方法について宮島（1988a）は「漢字表記語率」と関連するものであるとした上で「一般に、語種の増減をみるには単語単位ですべきだが、ここでは、漢字含有率についての話なので、便宜上、文字単位でいく」（宮島 1988a:56）と述べて便宜的なものとした。

しかし、本論文では語種と表記との関係を見るという目的に限れば、この方法を用いることは「漢字表記語率」を用いることに比して利点があるものとする。その利点とは如何にして語を認定するかという点と如何なるものを（「漢字表記語率」の分子となる）漢字表記の語と看做すかという点とを回避し得るということである。

「漢字表記語率」を調査した研究を見ると如何にして語を認定するかという点について杉山・森岡（1969）は「 β 単位」（国立国語研究所 1962）を参考に独自の規則によって語を認定しているのに対し、土屋（2000a, b）は文節を用いている。また、混種語については杉山・森岡（1969）が 0.5 語とするのに対し、土屋（2000a）は和語、漢語、外来語のいずれかになるまで小さく切り分けるとして統一を見ていない現状にある。当然、この点が異なれば、同じ文章を対象としていても結果が大きく変わり得ることから如何にして語を認定するかという点は「漢字表記語率」を扱う際には研究の根幹に関わる問題となる¹⁰³。

同様に如何なるものを漢字表記の語と看做すかという点については杉山・森岡（1969）は明確に述べていないが、水谷（1977:71）によれば、「二字以上で書いてある時その中の一字でも漢字で表わしていれば（訓であっても）漢字表記語と認めている」とあり、土屋（2000a）は挙例を見る限り、漢語は漢字のみのものを漢字表記の語と看做し、和語や外来語は漢字を含むものを漢字表記の語と看做しているものと見える。当然、漢字のみのものを漢字表記の語とすることに異論を差し挟む余地はなく、議論が分かれるのは漢字以外も含むものを扱う場合である¹⁰⁴。確かに土屋（2000a）のように語種に応じて漢字表記の語と看做す基準を変えることも一案ではあるが、混種語の扱いに難がある。また、漢字のみか否かという点によっても漢字を含むか否かという点によっても送り仮名を多く（あるいは少なく）送るような場合については捨象したものとなることにも留意する必要がある。しかし、既に述べた通り、これらの 2 点は「漢字表記語率」を用いる場合に問題となるのであって宮島（1988a）の方法を用いる場合には問題とはならない¹⁰⁵。また、この方法は文字を単位とすることから送り仮名を多く（あるいは少なく）送る場合も反映したものとするという利点も有しており、これも宮島（1988a）の方法を採用した理由である。

¹⁰³ 土屋（2000a）も述べているように長い語を認めると混種語が多くなることから如何にして語を認定するかという点は混種語の扱いと密接に関わるものである。また、混種語を和語、漢語、外来語のいずれかになるまで小さく切り分けることには「調査単位の均質性」（土屋 2000a:196）に問題があり、これは杉山・森岡（1969）のように一律に 0.5 語として和語、漢語、外来語に加算する場合も同様である。なお、水谷（1983:54）は「語の認め方のゆらぎ」が結果を歪める可能性を指摘しており、如何にして語を認定するかという点を考える上で「調査単位の均質性」は重要である。

¹⁰⁴ 本論文では既に「漢字表記語」を定義している田島（1998）を踏まえ、その語の表記について漢字のみのものを「漢字表記語」とし、漢字を含む語を「漢字書き語」として呼び分け、両者を区別せずに呼ぶ際は「漢字表記の語」のように「の」を入れてある。これは「仮名」の場合も同様である。

¹⁰⁵ これは語種と表記との関係を見る場合に限ってのことである。当然ながら品詞と表記との関係を見るような場合には品詞の認定は語の認定に基づくものであることから如何にして語を認定するかという点に立ち入らずに論を進めることは困難である。

2 調査の方法

次に調査の方法についてを第3章と同様に対象と手順とに分けて以下に述べる。

2.1 対象

まず、対象については第3章と同様に1986年から2015年までの計66篇の芥川賞作品を取り上げ、各作品からそれぞれ抽出した1,000字とする。ただし、本章においては宮島(1988a)との連続性を確認するために1979年から1985年までの計10篇の芥川賞作品から改めて1,000字ずつ抽出し、これも調査の対象に加えることとする。

また、各作品から1,000字ずつ抽出する際は各作品について『文藝春秋』掲載の当該の作品の冒頭から可能な限り長い間隔、かつ、等間隔に10ページを選び、該当するページの3行目の先頭からそれぞれ100字を抽出するという宮島(1988a)、第3章の方法を採る¹⁰⁶。なお、この方法に従って章や節を示す語や本文前にある引用などは除外する。

2.2 手順

次に調査の手順については第3章と同様に各作品から1,000字ずつ抽出した上で宮島(1988a)に倣って語種別に延べ平仮名数、延べ片仮名数、延べ漢字数、延べ英数字数に振り分けるという流れである¹⁰⁷。

また、語種を認定する際は形態素解析器 MeCab (Ver.0.996) と解析用辞書「現代書き言葉 UniDic」(Ver.2.2.0) とを用いることとする。小椋・小磯・富士池・宮内・小西・原(2011:67)によれば、この解析用辞書の語種は「原則として『新潮現代国語辞典』第2版(新潮社)による」ものであり、「『新潮現代国語辞典』第2版の見出しにない語は、

¹⁰⁶ 10ページを選ぶための式を既に第3章に示している。ただし、この式を用いると広告のみのページにページ番号がある場合(柳美里「家族シネマ」430ページのみ)を除外し得ないことから本章ではページ番号の有無に拘わらず、広告のみのページを一律に除外して10ページを選んだ。

¹⁰⁷ 例えば、3字から成る「段取り」という語は混種語であるが、本章では漢語の延べ漢字数、和語の延べ漢字数、和語の延べ平仮名数の各1字とする。このように語としては混種語であっても個々の文字は和語、漢語、外来語のいずれかに振り分けられるのである。なお、少数ながら「描く」の「描」などのように1字のうちに複数の語種を含むものは仮名表記にした際の文字数に応じて按分し、漢語の延べ漢字数、和語の延べ漢字数の各0.5字のようにする。(ルビは脚注109に従う)

『日本国語大辞典』第2版（小学館）を主たる資料として」認定したとある¹⁰⁸。なお、この解析用辞書には語種として和語、漢語、外来語、混種語、固有名、記号が立てられており、対象とする文字を含む語がこれらのうちの混種語、固有名、記号に該当する場合や解析用辞書に登録のない場合は主として『新潮現代国語辞典』第2版、『新潮国語辞典—現代語・古語—』第2版、『日本国語大辞典』第2版（以下、3辞典）を参照して和語、漢語、外来語のいずれかに振り分けることとする。

更に宮島（1988a:56）は語種を認定する際に「^{エキゾチック}異国的」のように漢字にヨーロッパ語のふりがなのあるものは、ふりがなを無視して、漢字の方をよむという点と「現代中国音のものは外来語にいれ」という点とを明示しており、この2点は解析用辞書の認定とは異なることから宮島（1988a）に合わせて修正する¹⁰⁹。また、この解析用辞書は位取りを示す語のない2桁以上の数字に位取りを示す語を補う「NumTrans」と呼ばれる変換処理を事前に施してから用いることがあるが、このような処理は施さない方針を採る¹¹⁰。

これに加えて語種の認定に際しては「スイメン」とも「みなも」とも読める「水面」のように読み方によって語種が変わるものがあることを踏まえ、以下の手順によって読み方を認定することとする¹¹¹。

¹⁰⁸ 西尾（2002:94）によれば、「多数の語について一応語種を決定し、それに基づいて何か調査する必要に面したときは、ひらがなとかたかなの使い分けによってすべての見出し語の語種を明示している『新潮国語辞典』に依拠するやり方が多くとられている」とのことである。

¹⁰⁹ 後者についてはルビも参考しつつ「高粱」^{ジヤンチエン}、「張倩」^{シヤ シキョウ}などは外来語とし、「謝志強」などは漢語とする。また、朝鮮語の「由熙」^{ユ ヒ}なども外来語とする。なお、本章では上ルビで原文にあるルビを示し、下ルビの平仮名、片仮名、ローマ字でそれぞれ和語、漢語、外来語と認定することを示す。

¹¹⁰ 例えば、「27歳」^{ニナナサイ}、「二七歳」^{ニシチサイ}、「二十七歳」^{ニジュウナナサイ}に「NumTrans」という処理を施すといずれも位取りを示す「十」の入った漢数字の「二十七歳」となるが、本章では変換処理によって本文を改めることはせず、前二者は位取りを示す「十」を落として1桁ずつ「27歳」^{ニナナサイ}、「二七歳」^{ニシチサイ}と読み、後一者はそのまま「二十七歳」^{ニジュウナナサイ}と読むこととする。また、「27日」^{ニナナサイ}は「27日」^{ニシチ}とするように同じ「27」^{ニジュウナナサイ}であっても後続する語に応じて読み方の変わるものについては考慮してある。なお、最高位が1であるものや0を含むものは「120人」^{イチニヒト}、「一〇人」^{イチヒト}、「百二十人」^{ヒヤクニジュウニヒト}のようにする。なお、この点については石井（2002）も参考にした。

¹¹¹ 田中（2020）は『吾妻鏡』や川端康成『雪国』の例を挙げつつ読み方によって語種が変わるものがあることは和語の表記に漢字を用いる以上、避けられない問題であるとし、これによって語種の構成比率も算出し得ないことを指摘している。また、読み方が定められないことは山崎（2013）の挙げる「語彙調査に関する未解決の問題」の1つであり、コーパスを構築する際にも問題となる点であるが、本論文では試みに読み方を認定する手順を提示し、便宜的ながらも解決を図ることとする。なお、この点については尾山（2020）も参考にした。

- (1) a. 本文の該当する箇所あるいは初出の箇所にルビのある場合にはルビに従って読み方を認定する。ただし、洋語のルビ（「恋愛」など）や仮名以外を含むルビ（「ラバイ」や「d o p e」など）は除く¹¹²。
- b. 読み方に迷うものについて『芥川賞全集』や各作品の単行本・文庫本の該当する箇所あるいは初出の箇所にルビのある場合にはそのルビに従って読み方を認定する。ただし、洋語のルビや仮名以外を含むルビは除く¹¹³。
- c. 人名は(2)によって読み方を認定し、それ以外のものは内省によって読み方を認定する¹¹⁴。ただし、実在する地名などの固有名詞は調べた上で認定する。
- (2) a. 本文に姓ないしは名の読み方を示す箇所が別にあれば、それに従って読み方を認定する（「タカコ」と呼ばれる「貴子」など）。
- b. 本文に姓ないしは名的一部分を含む呼び名が別にあれば、そこから推し量って読み方を認定する（「シュンちゃん」と呼ばれる「俊治」など）。
- c. 姓は『苗字 8 万よみかた辞典』（以下、『苗字 8 万』）を引き、名は『名前 10 万よみかた辞典』（以下、『名前 10 万』）を引き、1 通りの読み方しか無ければ、それに従って読み方を認定し、複数通りの読み方があれば、改めて姓は『人名よみかた辞典—姓の部—』新訂第 3 版（以下、『姓の部』）を引き、名は『人名よみかた辞典—名の部—』新訂第 3 版（以下、『名の部』）を引き、最も多くの実例が挙げられている読み方に従って読み方を認定する¹¹⁵。
- また、認定した読み方に基づいて語種を認定する手順を整理すると以下の通りである。
- (3) a. 解析用辞書「現代書き言葉 UniDic」(Ver.2.2.0)によって和語、漢語、外来語と認定し得る場合にはそれに従って語種を認定する。
- b. それ以外の場合には 3 辞典によって語種を認定する。ただし、人名として項

¹¹² いずれも岡田（2003）の「異語ルビ」に相当するものである。

¹¹³ 本章では読み方によって語種が変わるものに限って『芥川賞全集』などに当たることとする。

¹¹⁴ 内省によって認定するものは主に筆者の内省によって認定してあるが、特に読み方に迷うものは 3 名の大学院生の内省によって認定している。なお、この 3 名はいずれも東京都（島嶼部を除く）で言語形成期を過ごした 1990 年代生まれの日本語母語話者である。

¹¹⁵ 最も多くの実例が挙げられている読み方が 1 つに定まらなければ、そのうちのいずれかを内省によって採ることとする。なお、「読み方」を「表記」に読み替えれば、(3d)でも同様である。

目が無くても地名として項目があれば、準用することとする。

- c. 姓ないしは名が明らかであって当該の姓ないしは名の一部分を含む呼び名のうち外来語と迷うものは外来語とする¹¹⁶。
- d. 仮名書き語としてのみ現れる人名の仮名の部分は読み方に基づいて姓は『苗字 8 万』を引き、名は『名前 10 万』を引き、1 通りの表記しか無ければ、それに従って語種を認定し、複数通りの表記があれば、その表記に基づいて姓は『姓の部』を引き、名は『名の部』を引き、最も多くの実例が挙げられている表記に従って語種を認定する¹¹⁷。
- e. 明確に外来語と認定し得るものを除けば、字音で読めるものは漢語とし、それ以外は和語とする¹¹⁸。

3 分析の手法

本章では語種別の延べ文字数や延べ漢字数の変遷を描き出し、その変遷と第 3 章において明らかにした各作品の漢字含有率の変遷とを照らし合わせながら分析を試みる。

また、それぞれの変遷については第 3 章と同様に蛇行箱型図（渡部・鈴木・山田・大塚 1985）という手法によって描き出すこととする。この手法は渡部・鈴木・山田・大塚（1985:81）によれば、「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。

¹¹⁶ 例えば、金原ひとみ「蛇にピアス」に現れる「アマ」という人物は自身の呼び名の由来について (i) のように語っているが、読み進めると (ii) のように「雨田」という姓であることが明らかになる。このような場合には「アマ」を外来語とする。

(i) 「ねえアマ、あんたの名前何ていうの？ Ama(deus) 天野？ スアマ？」

「何だよスアマって。俺のアマはね、アマデウスのアマなの。」

(ii) 「雨田和則さん、ですね」

警察が用紙に目を通しながら言った。私は初めてアマの名前を知った。アマデウスじゃ、ないじゃん。（金原ひとみ「蛇にピアス」『文藝春秋』第 82 巻 4 号、2004 年）

¹¹⁷ 例えば、多和田葉子「犬婿入り」に現れる「みつこ」について『名前 10 万』を引くと 33 通りの表記があり、それぞれの表記について『名の部』を引くと最も多くの実例が挙げられている表記は「美都子」（7 例）である。これに従って本章では「みつこ」のように語種を認定する。

¹¹⁸ なお、地名については『角川日本地名大辞典』、『日本歴史地名大系』、『大日本地名辞書』増補版を補助的に参照することがある。

4 調査の結果

1979年から2015年までの芥川賞作品から抽出した1,000字の内訳を表3、表4、表5に示す¹¹⁹。なお、第3章と同様に「々」は漢字とし、「ー（長音符）」やこれと同等と看做せる「～」は前接する文字の繰り返しとして処理してある。

表3 芥川賞作品（1979年～1985年）から抜き出した1,000字の内訳

受賞年度	受賞年	受賞作品	和語				漢語				外来語			
			ひ	カ	漢	英数	ひ	カ	漢	英数	ひ	カ	漢	英数
1979 上	1979	重兼芳子「やまあいの煙」	692	0	175	0	2	0	119	0	0	11	1	0
1979 上	1979	青野聰「愚者の夜」	674	6	92.5	0	10	0	151.5	0	0	64	2	0
1979 下	1980	森禮子「モッキングバードのいる町」	612	0	143	0	21	0	119	0	0	101	0	4
1980 上		該当作なし												
1980 下	1981	尾辻克彦「父が消えた」	702	7	127	0	26	0	129	0	0	9	0	0
1981 上	1981	吉行理恵「小さな貴婦人」	603	11	193	0	5	0	105	0	0	79	2	2
1981 下		該当作なし												
1982 上		該当作なし												
1982 下	1983	唐十郎「佐川君からの手紙」	704	14	138	0	10	0	102	0	5	21	1	5
1982 下	1983	加藤幸子「夢の壁」	653	4	158	0	17	2	109	0	0	21	36	0
1983 上		該当作なし												
1983 下	1984	笠原淳「柰二の世界」	664	19	116	0	32	0	138	0	0	31	0	0
1983 下	1984	高樹のぶ子「光抱く友よ」	701	2	161.7	0	29	0	102.3	0	0	4	0	0
1984 上		該当作なし												
1984 下	1985	木崎さと子「青桐」	660	0	188	0	20	0	106	0	0	26	0	0
1985 上		該当作なし												

¹¹⁹ 表中では延べ平仮名数、延べ片仮名数、延べ漢字数、延べ英数字数（延べローマ字数と延べ算用数字数との和）をそれぞれ「ひ」、「カ」、「漢」、「英数」と略記した。なお、本章でも漢字含有率の分母を延べ文字数としたことから文字を対象として1,000字を抽出していることから文字以外の表記記号（以下、記号）は抽出せず、結果として平仮名、片仮名、漢字、ローマ字、算用数字の5つの文字体系のいずれかに属する文字を抽出したが、別の箇所に見えるキリル文字やハングルも文字である以上、抽出し得るとの立場を採る。また、本章では『現代日本語書き言葉均衡コーパス』におけるサンプリングの対象について述べた丸山・柏野（2014）に基づいて「非日本語」あるいは「非現代語」に該当する場合には抽出する箇所から除くことを検討し、「非日本語」は除き、「非現代語」は除かないという暫定的な方針を立てていたが、結果として除いた箇所は無かったことを申し添える。このうち「非日本語」に関する議論については山崎（2013）や伊藤（2001）も参考にした。

表 4 芥川賞作品（1986年～2000年）から抜き出した1,000字の内訳

受賞年度	受賞年	受賞作品	和語				漢語				外来語			
			ひ	カ	漢	英数	ひ	カ	漢	英数	ひ	カ	漢	英数
1985 下	1986	米谷ふみ子「超越しの祭」	639	0	130	0	10	0	136	0	0	85	0	0
1986 上		該当作なし												
1986 下		該当作なし												
1987 上	1987	村田喜代子「鍋の中」	739	0	120	0	39	0	96	0	0	6	0	0
1987 下	1988	池澤夏樹「スティル・ライフ」	653	0	138	0	31	0	131	0	0	47	0	0
1987 下	1988	三浦清宏「長男の出家」	630	8	158	0	32	2	129	0	0	38	3	0
1988 上	1988	新井満「尋ね人の時間」	624	9	192.5	0	16	1	122.5	0	0	35	0	0
1988 下	1989	南木佳士「ダイヤモンドダスト」	602	4	179	0	16	0	128	0	0	71	0	0
1988 下	1989	李良枝「由熙」	665	4	147	0	13	5	116	0	0	44	6	0
1989 上		該当作なし												
1989 下	1990	大岡玲「表層生活」	589	0	120	0	32	0	192	0	0	65	0	2
1989 下	1990	瀧澤美恵子「ネコババのいる町で」	719	12	139	0	18	4	90	0	0	18	0	0
1990 上	1990	辻原登「村の名前」	694	0	165	0	16	0	100	0	0	14	11	0
1990 下	1991	小川洋子「妊娠カレンダー」	682	0	143	0	16	0	99	0	0	60	0	0
1991 上	1991	辺見庸「自動起床装置」	651	31	119	0	33	2	124	0	0	39	0	1
1991 上	1991	荻野アンナ「背負い水」	663	19	184	0	11	7	101	0	0	13	2	0
1991 下	1992	松村栄子「至高聖所」	640	3	131.5	0	12	1	130.5	0	0	82	0	0
1992 上	1992	藤原智美「運転士」	599	20	147	0	27	12	109	0	0	86	0	0
1992 下	1993	多和田葉子「犬婿入り」	682	8	176	0	18	0	95	0	0	21	0	0
1993 上	1993	吉目木晴彦「寂寥郊野」	573	0	163	0	14	0	175	0	0	69	0	6
1993 下	1994	奥泉光「石の来歴」	577	3	188	0	12	0	194	0	0	25	1	0
1994 上	1994	室井光広「おどるでく」	609	27	107	0	14	3	197	0	0	43	0	0
1994 上	1994	笠野頼子「タイムスリップ・コンビナート」	625	0	147	0	35	0	95	0	10	86	0	2
1994 下		該当作なし												
1995 上	1995	保坂和志「この人の闘」	684	16	134	0	24	2	107	0	0	33	0	0
1995 下	1996	又吉栄喜「豚の報い」	624	16	189	0	20	2	101	0	0	48	0	0
1996 上	1996	川上弘美「蛇を踏む」	674	38	159	0	15	7	90	0	0	17	0	0
1996 下	1997	辻仁成「海峡の光」	576	0	182	0	5	2	202	0	0	33	0	0
1996 下	1997	柳美里「家族シネマ」	633	0	173.5	0	13	0	128.5	0	0	50	2	0
1997 上	1997	目取真俊「水滴」	599	13	226	0	12	4	131	0	0	15	0	0
1997 下		該当作なし												
1998 上	1998	藤沢周「ヴェニスアイレス午前零時」	507	38	152	0	8	8	101	0	0	186	0	0
1998 上	1998	花村萬月「ゲルマニウムの夜」	628	0	141.5	0	16	0	174.5	0	0	40	0	0
1998 下	1999	平野啓一郎「日蝕」	597	0	250	0	4	0	132	0	0	15	2	0
1999 上		該当作なし												
1999 下	2000	玄月「蔭の棲みか」	632	16	152	0	10	0	106	0	0	84	0	0
1999 下	2000	藤野千夜「夏の約束」	645	31	145	0	11	3	101	0	0	62	2	0

表5 芥川賞作品（2000年～2015年）から抜き出した1,000字の内訳

受賞年度	受賞年	受賞作品	和語				漢語				外来語			
			ひ	カ	漢	英数	ひ	カ	漢	英数	ひ	カ	漢	英数
2000上	2000	町田康「きれぎれ」	650	7	151	0	10	0	125	3	0	50	4	0
2000上	2000	松浦寿輝「花腐し」	667	14	155	0	32	0	89	0	0	41	2	0
2000下	2001	青来有一「聖水」	589	6	177	0	23	0	116	0	22	67	0	0
2000下	2001	堀江敏幸「熊の敷石」	621	0	128	0	9	0	161	0	0	81	0	0
2001上	2001	玄侑宗久「中陰の花」	745	6	110	0	21	0	85	0	0	31	2	0
2001下	2002	長嶋有「猛スピードで母は」	626	15	152	0	35	0	137	0	0	33	0	2
2002上	2002	吉田修一「パーク・ライフ」	613	1	171.5	0	19	0	99.5	0	0	94	2	0
2002下	2003	大道珠貴「しょっぱいドライブ」	779	9	102	0	12	2	73	0	0	23	0	0
2003上	2003	吉村萬壺「ハリガネムシ」	581	45	208	0	14	6	102	0	0	41	2	1
2003下	2004	金原ひとみ「蛇にピアス」	594	31	168.5	3	9	3	72.5	5	0	114	0	0
2003下	2004	綿矢りさ「蹴りたい背中」	631	20	156	0	14	0	110	0	0	65	0	4
2004上	2004	モブ・ノリオ「介護入門」	625	3	204	0	6	0	152	0	0	10	0	0
2004下	2005	阿部和重「グランド・フィナーレ」	696	0	122	0	22	2	143	0	0	15	0	0
2005上	2005	中村文則「土の中の子供」	652	0	152	0	18	2	141	0	0	35	0	0
2005下	2006	絲山秋子「沖で待つ」	664	2	145	0	25	2	108	0	0	48	2	4
2006上	2006	伊藤たかみ「八月の路上に捨てる」	641	0	147	0	10	5	139	0	0	58	0	0
2006下	2007	青山七恵「ひとり日和」	684	2	161	0	29	2	77	0	0	45	0	0
2007上	2007	諏訪哲史「アサッテの人」	610	7	138	0	11	4	128	0	0	100	2	0
2007下	2008	川上未映子「乳と卵」	720	7	109	0	29	0	116	0	0	18	1	0
2008上	2008	楊逸「時が滲む朝」	563	7	155	0	8	0	176	0	0	64	10	17
2008下	2009	津村記久子「ボトスライムの舟」	655	21	123	0	19	0	125	1	4	51	0	1
2009上	2009	磯崎憲一郎「終の住処」	687	0	136	0	26	0	125	0	0	26	0	0
2009下		該当作なし												
2010上	2010	赤染晶子「乙女の密告」	585	2	133	0	23	1	150	0	2	104	0	0
2010下	2011	朝吹真理子「きことわ」	693	3	178	0	8	1	88	0	0	29	0	0
2010下	2011	西村賢太「苦役列車」	653	4	142	0	8	4	156	0	0	31	2	0
2011上		該当作なし												
2011下	2012	円城塔「道化師の蝶」	659	0	116	0	2	0	179	0	0	42	0	2
2011下	2012	田中慎弥「共喰い」	656	0	221	0	7	0	97	0	0	19	0	0
2012上	2012	鹿島田真希「冥土めぐり」	645	0	143	0	4	0	131	0	0	77	0	0
2012下	2013	黒田夏子「abさんご」	784	0	69	0	87	0	52	0	6	0	0	2
2013上	2013	藤野可織「爪と目」	722	0	120	0	26	0	79	0	0	53	0	0
2013下	2014	小山田浩子「穴」	655	4	172	0	12	0	106	0	0	51	0	0
2014上	2014	柴崎友香「春の庭」	580	0	151	0	10	0	169	0	0	90	0	0
2014下	2015	小野正嗣「九年前の祈り」	655	6	137	0	46	4	101	0	0	46	5	0
2015上	2015	又吉直樹「火花」	661	0	164	0	13	0	150	0	0	12	0	0
2015上	2015	羽田圭介「スクラップ・アンド・ビルド」	578	0	139	0	22	0	200	0	0	61	0	0

5 結果の分析

以下では各作品の漢字含有率が安定しつつある要因について語種別の延べ文字数の点と語種別の延べ漢字数との点との2点から分析を試みる。なお、本章において先行調査分とは1935年から1985年までの芥川賞作品における1,000字の内訳（宮島1988b）を用いて作図した部分であり、本章調査分とは1979年から2015年までの芥川賞作品における1,000字の内訳（表3、表4、表5）を用いて作図した部分である。

これに基づいて改めて漢字含有率の変遷を描き出したものが図14である。これと照らし合わせながら以下では論を進めることとする。

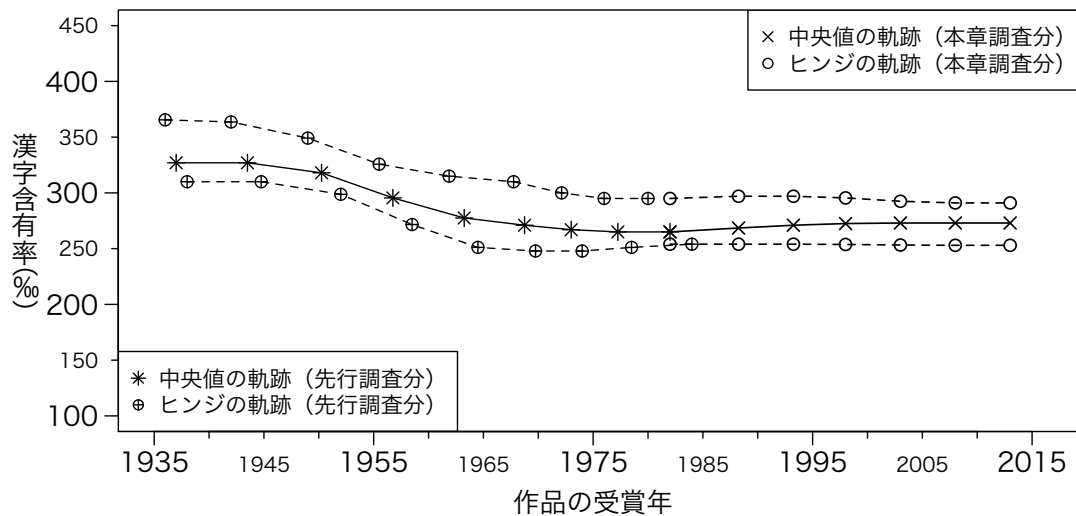


図14 漢字含有率の変遷

5.1 語種別の延べ文字数の点から

初めに図15に和語の延べ文字数の変遷を示す。

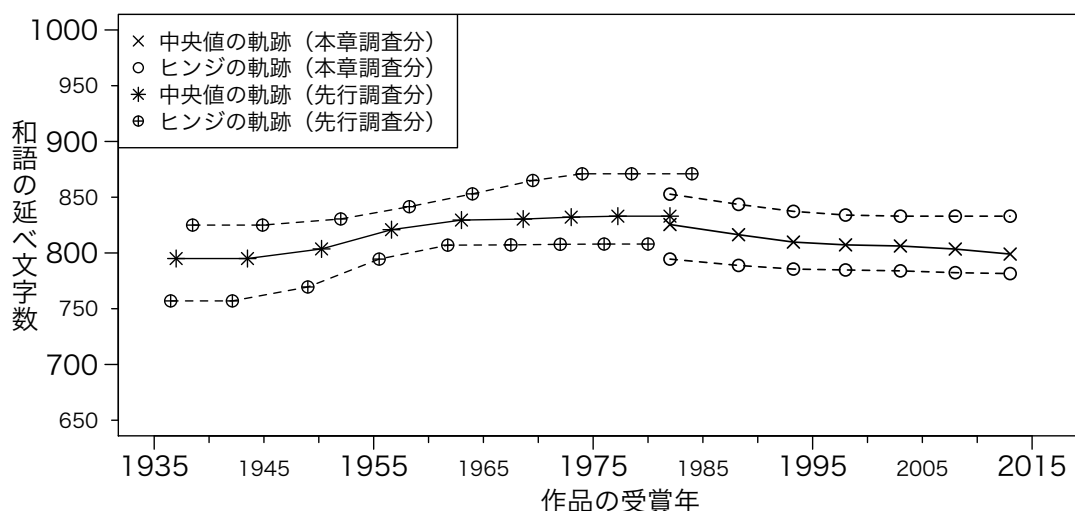


図 15 和語の延べ文字数の変遷

図 15 を見ると先行調査分では和語の延べ文字数が増加していることが読み取れる¹²⁰。これは既に明らかにした漢字含有率の変遷（図 14）とも対応しており、20 世紀の中葉まで見られた漢字含有率の減少と和語の増加との関連を示唆するものである。当然、杉山・森岡（1969）の指摘した和語の増加とも関わるものと見える。また、繋ぎ目に当たる部分は先行調査分よりも低くなっているが、この点を加味しても本章調査分では和語の延べ文字数が微減しているものと見受けられる¹²¹。

次に図 16 に漢語の延べ文字数の変遷を示す。図 16 を見ると先行調査分においては漢語の延べ文字数が減少していることが読み取れる。これは漢字含有率の変遷とも対応しており、宮島（1988a）の指摘するように 20 世紀の中葉まで見られた漢字含有率の減少と漢語の減少との関連を示唆するものである。また、繋ぎ目に当たる部分は先行調査分よりも高くなっており、この点を加味すると本章調査分では漢語の延べ文字数は微増ないしは横ばいであるものと見受けられる。

¹²⁰ 渡部・鈴木・山田・大塚（1985:93）によると中央値の軌跡は「X の変化に伴う Y の主たる変化のようす」を示し、上下のヒンジの軌跡に「囲まれた部分は分布の中央部おおよそ 50% のデータの動き」を示すものである。また、増減については「代表値の変動だけでなく、たとえばデータの半数の変動を把握する」ことが必要であると述べている石井（2013: 136）を参考に判断した。

¹²¹ 重複する作品については各作品の延べ平仮名数、延べ片仮名数、延べ漢字数、延べ英数字数が 1 字を除いて宮島（1988a）と完全に一致しており、抽出した箇所は同一であると考えられるが、語種別に見た延べ平仮名数、延べ片仮名数、延べ漢字数、延べ英数字数は完全には一致していない場合がある。

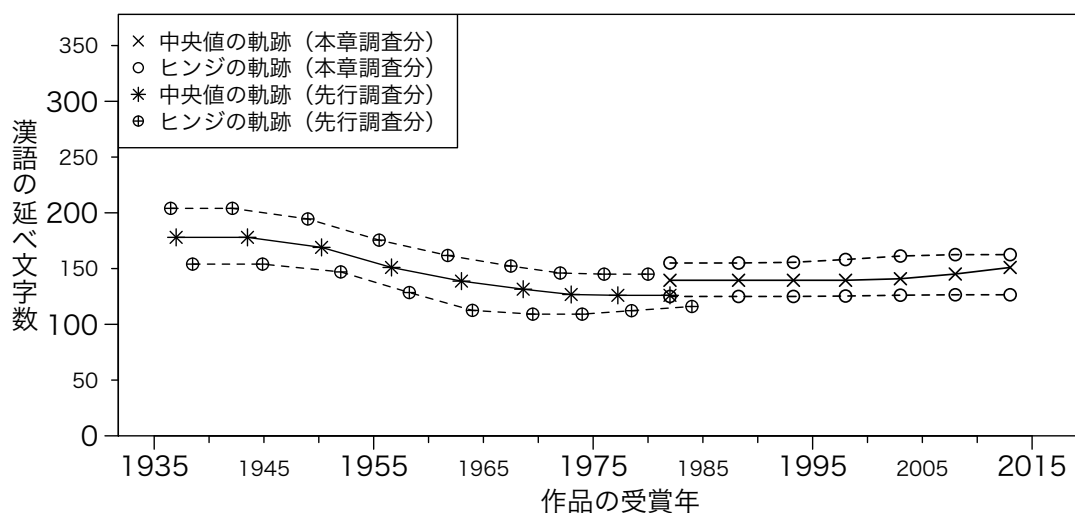


図 16 漢語の延べ文字数の変遷

続いて図 17 に外来語の延べ文字数の変遷を示す。

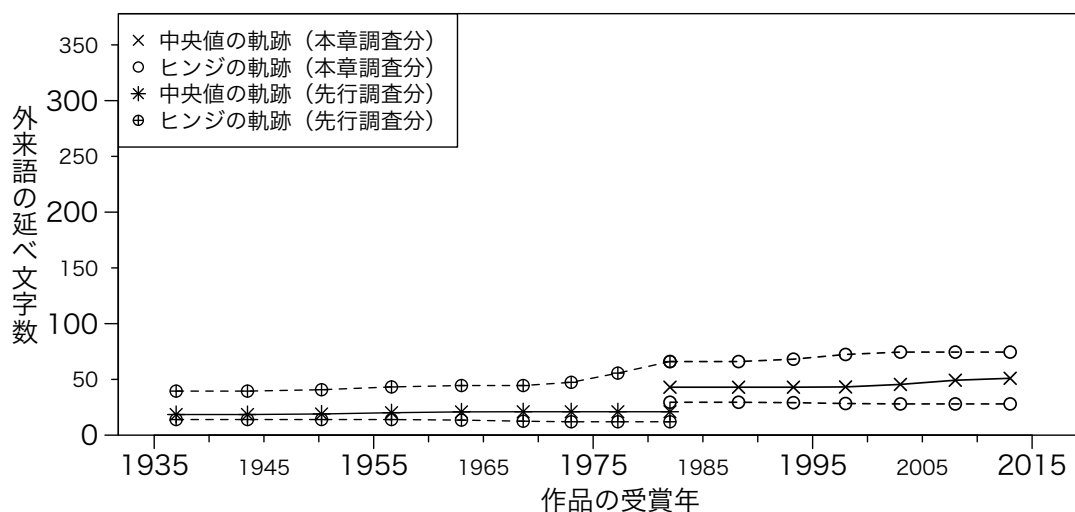


図 17 外来語の延べ文字数の変遷

図 17 を見ると先行調査分では外来語の延べ文字数が微増ないしは横ばいであることが読み取れる。これについては漢字含有率の変遷と対応しているとも言い難く、「外来語はやや増加しているものの、数がすくないので、全体の漢字含有率にあたえる影響は大きくない」という宮島 (1988a:56) の指摘の通りであると考えられる。また、繋ぎ目に当たる部分は先行調査分よりも高くなっており、この点を加味すれば、本章調査分でも外来語の

延べ文字数は微増ないしは横ばいであるものと見受けられる。なお、仮に微増しているとしても外来語が増加することによって漢字含有率が減少するという野村（1988b）の予想を裏づけるものとは看做し難い結果である。

すなわち、語種別の延べ文字数に基づいて語種の構成比率の変遷を見た結果を整理すると先行調査分では和語が増加するに連れて漢語が減少し、外来語は微増ないしは横ばいであることが改めて示された。また、本章調査分では和語が微減する一方で漢語や外来語は微増すると消極的には読み取り得るものの 20 世紀の中葉までに見られたような顕著な傾向であるとは言い難く、いずれも概ね横ばいであるとも看做し得る結果である。

5.2 語種別の延べ漢字数の点から

初めに図 18 に和語の延べ漢字数の変遷を示す。

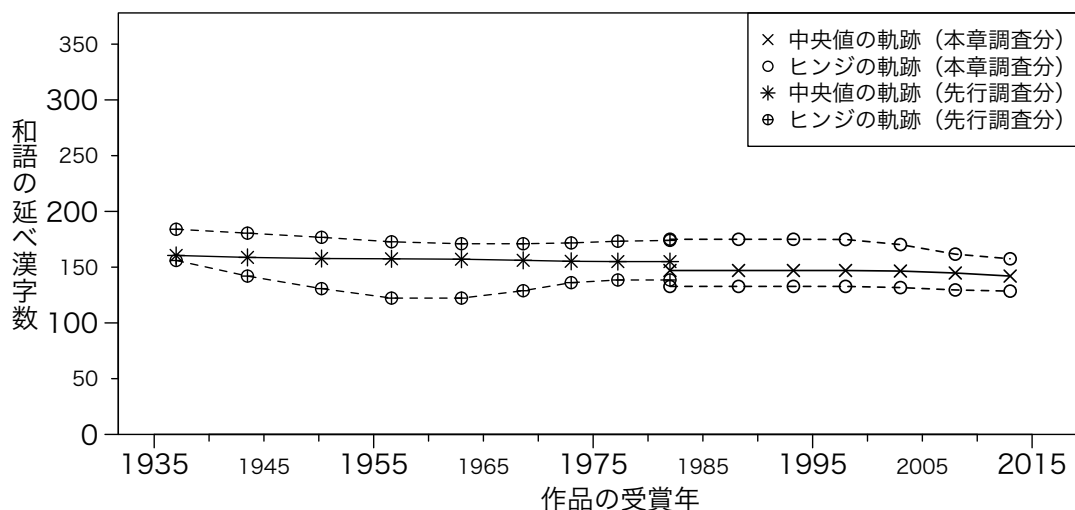


図 18 和語の延べ漢字数の変遷

図 18 を見ると先行調査分においては和語の延べ漢字数が微減ないしは横ばいであることが読み取れる。また、繋ぎ目に当たる部分は先行調査分よりも低くなっており、この点を加味すると本章調査分でも微減ないしは横ばいであるものと見受けられる。なお、先行調査分において和語の延べ漢字数が横ばいであるとする漢字表記の和語が減少しているとした杉山・森岡（1969）や宮島（1988a）の指摘と整合していないようにも見えるが、図 15 に見たように和語の延べ文字数が増加していることに照らせば、和語の延べ文字数に占める和語の延べ漢字数の割合（和語の漢字含有率）は減少しているということになる

(図 19)。

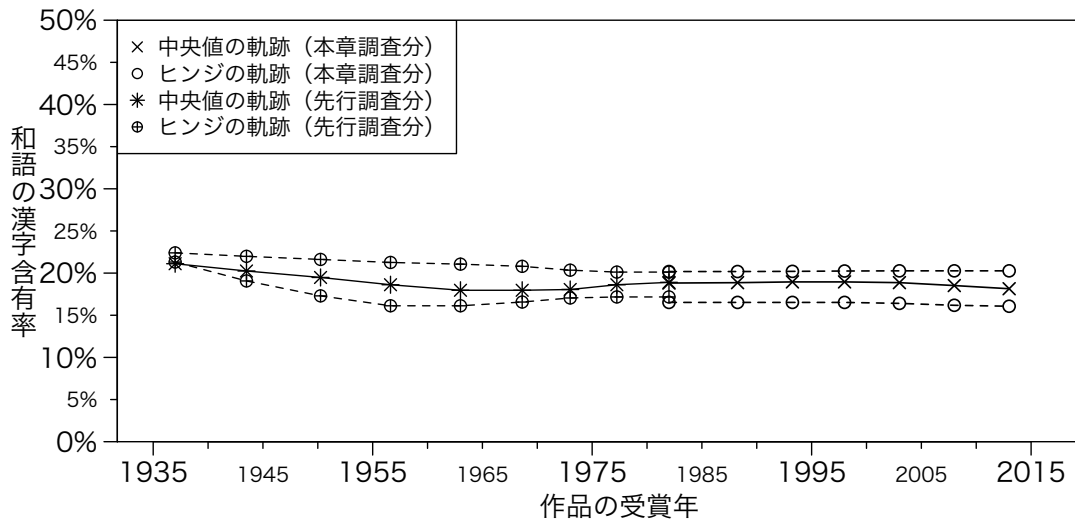


図 19 和語の漢字含有率の変遷

すなわち、図 15 と図 18 とを踏まえると先行調査分においては漢字表記の和語が減少したというよりも仮名表記の和語の増加によって和語全体が増加し、それによって漢字表記の和語が相対的に減少したという可能性が新たに導かれるのである。

次に図 20 に漢語の延べ漢字数の変遷を示す。

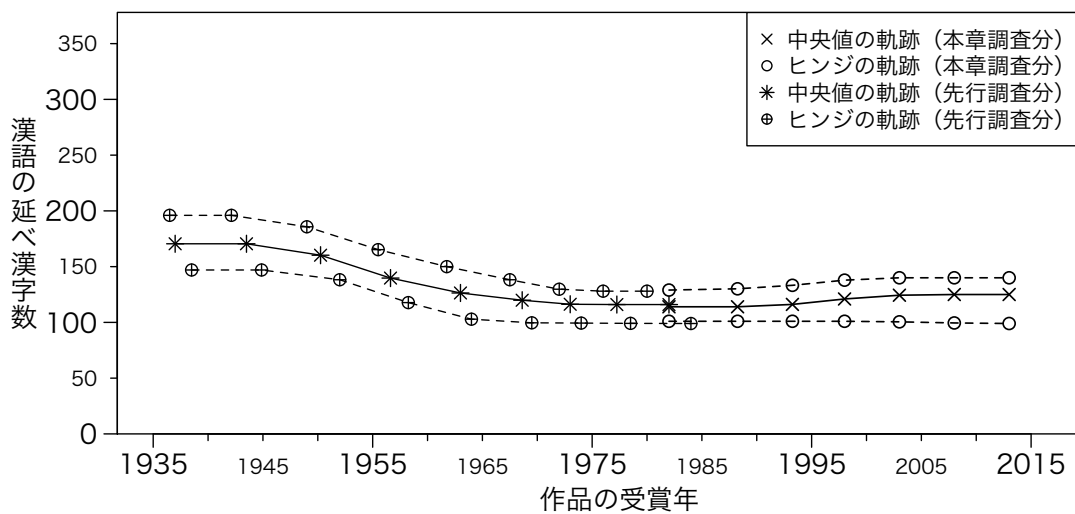


図 20 漢語の延べ漢字数の変遷

図 20 を見ると先行調査分では漢語の延べ漢字数が減少していることが読み取れる。こ

れは漢字含有率の変遷とも軌を一にしており、宮島（1988a）の指摘しているように 20 世紀の中葉まで見られた漢字含有率の減少と漢字表記の漢語の減少との関連を示唆するものである。また、既に見たように和語の延べ漢字数（漢字表記の和語）が横ばいであり、漢字含有率の変遷と対応していると言い難いこと、そもそも外来語の延べ漢字数（漢字表記の外来語）は極めて少ないことの 2 点に照らせば、20 世紀の中葉まで見られた漢字含有率の減少は主として漢語の延べ漢字数（漢字表記の漢語）が減少したことによって生じているとも言えるのではないか。

一方、繋ぎ目に当たる部分は概ね一致しており、本章調査分では微増ないしは横ばいであるものと見受けられる。これが仮に微増しているとする漢字含有率には大きな変動が見られない以上、他の要因（和語の延べ漢字数が微減していることなど）と相殺して影響を及ぼしていないという可能性が考えられる。しかし、20 世紀の中葉まで見られたような顕著な傾向であるとは言い難く、いずれも概ね横ばいであるとも看做し得る結果である。

6 本章の結論

第 4 章では 1979 年から 2015 年までの芥川賞作品を対象に各作品の語種別の延べ文字数や延べ漢字数を算出し、これを用いて各作品の漢字含有率が安定しつつある要因について分析を試みた。

初めに語種別の延べ文字数を用いて語種の構成比率の変遷を明らかにした結果、和語が微減して漢語や外来語が微増するとも読み取れるものの 20 世紀の中葉まで見られたような顕著な傾向であるとは言えず、概ね横ばいであると看做し得ることを指摘した。次に語種別の延べ漢字数の変遷を明らかにした結果、漢字表記の和語（和語の延べ漢字数）が微減して漢字表記の漢語（漢語の延べ漢字数）が微増するとも読み取れるものの 20 世紀の中葉まで見られたような顕著な傾向であるとは言い難く、概ね横ばいであると看做し得ることを指摘した。すなわち、語種の構成比率や語種別に見た漢字使用の実態が大きくは変化していないことによって漢字含有率が安定しつつあるものと考えられる。

また、宮島（1988a, b）の数表に基づいて 1935 年から 1985 年までの芥川賞作品を対象に同様の方法によって改めて検証を試みた。その結果、20 世紀の中葉まで漢字含有率（各作品の延べ漢字数）が減少したのは主として漢字表記の漢語（漢語の延べ漢字数）が

減少したことに起因している可能性が示された。更に漢字表記の和語（和語の延べ漢字数）は仮名表記の和語（和語の延べ仮名数）の増加に伴う和語全体（和語の延べ文字数）の増加によって相対的に減少したと捉えられるという可能性も示された。

第 5 章

新聞の漢字含有率は如何なる傾向にあるのか

第 3 章においては 1986 年から 2015 年までの芥川賞作品を対象に漢字含有率を調査し、この 30 年間は漢字含有率が安定しつつある傾向について報告した。また、続く第 4 章では語種に着目した検討により、その安定の要因として語種の構成比率や語種別に見た漢字使用の実態に大きな変化が見られなくなったことが挙げられることを指摘した。

ただし、野村（1988b）の示唆するところによれば、小説以外の他の媒体やジャンルにおける漢字含有率は小説におけるそれとは異なる傾向を示す可能性がある。この点を踏まえ、本章では小説とは異なる媒体やジャンルとして新聞を取り上げ、新聞における漢字含有率の変遷を明らかにする。

1 調査の方法

本章では第 3 章において対象とした芥川賞作品と概ね同時期に当たる 1985 年から 2015 年までの『毎日新聞』を対象に 5 年ごとに 12 記事ずつ計 84 記事を抽出し、各記事の漢字含有率を調査することとする。それぞれの年の 12 記事はプログラミング言語 Python (Ver.2.7.10) を使用し、特定の曜日に偏らないように各月から 1 日ずつ無作為に抽出した計 12 日のそれぞれの日の朝刊の一面トップ記事（題字の左隣にある記事）である。

なお、各記事の本文はそれぞれの年の『毎日新聞縮刷版』（毎日新聞社）に拠ることとし、見出しや図表、語句等の説明、解説、「【山田太郎】」や「(○面に関連記事)」のような

部分は除外する。また、第3章や第4章と同様に「々」や「ー（長音符）」は前接する文字の繰り返しとし、合字や組文字、縦中横、丸つき数字はそれを構成する文字に戻してから漢字含有率を算出する。

2 分析の手法

本章でも漢字含有率の変遷を素描する方法として探索的データ解析の手法の1つである蛇行箱型図（渡部・鈴木・山田・大塚 1985）を採用する。

これは石井（2013）によれば、単に箱ひげ図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、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り、漢字含有率の変遷を素描することに適した手法とのことである。

なお、これまでの新聞における「漢字含有率」を調査した研究では1つの時点から1つの観測値を採っているのに対し、本章では同一時点から複数の観測値を採っており、これに適した手法を採用したということになる。

3 調査の結果

各記事から抽出した表記記号の内訳を表6、表7、表8に示す。

表6 毎日新聞（1985年）から抜き出した表記記号の内訳

発行年月日	対象となる記事の見出し	和語				漢語				外来語				記号
		ひ	カ	漢	英数	ひ	カ	漢	英数	ひ	カ	漢	英数	
1985/01/18 (金)	自民三役は留任	314	0	64	7	4	1	324	26	0	151	0	2	80
1985/02/01 (金)	原発で「石橋見解」も	324	2	64	3	7	0	491	62	0	31	0	18	81
1985/03/25 (月)	衆院審議、初日で中断 <small>予算委</small>	423	0	63	0	2	0	424	14	0	28	0	0	84
1985/04/17 (水)	横浜で短銃銀行強盗	326	0	63	11	2	1	287	39	0	62	0	3	48
1985/05/28 (火)	都市交 早朝スト突入	469	0	78	5	0	2	406	91	0	238	0	12	120
1985/06/16 (日)	「5億円はP3C工作費」	465	0	79	5	3	0	433	9	0	0	0	0	91
1985/07/06 (土)	宇宙へ「本格、挑戦	379	0	35	3	0	4	308	35	0	55	0	27	61
1985/08/13 (火)	臨時国会は11月初旬召集	363	0	156	6	4	1	273	32	0	32	0	0	67
1985/09/16 (月)	524人乗せ墜落	291	0	55	3	4	0	313	27	0	57	0	0	66
1985/10/10 (木)	大統領権限を強化	366	0	63	10	12	0	336	45	0	58	4	1	47
1985/11/10 (日)	NY連銀が円買い介入	349	0	65	10	0	0	415	47	0	4	0	0	73
1985/12/25 (水)	ソ連国民に「共存」訴え	299	0	65	11	0	0	329	47	0	0	0	0	58

表7 毎日新聞（1990年～2000年）から抜き出した表記記号の内訳

発行年月日	対象となる記事の見出し	和語				漢語				外来語				記号
		ひ	カ	漢	英数	ひ	カ	漢	英数	ひ	カ	漢	英数	
1990/01/05 (金)	「追加公認含め過半数」	293	0	76	6	2	1	311	23	0	44	0	8	60
1990/02/17 (土)	東大医科研 倫理委「脳死肝移植」を初承認	500	5	135	8	7	0	250	31	0	150	0	4	111
1990/03/08 (木)	国民負担率 40% 突破	508	2	161	0	8	0	350	14	0	3	0	0	98
1990/04/10 (火)	総選挙に 200 億円	308	0	36	1	2	0	338	0	0	19	0	30	60
1990/05/09 (水)	下水道使い都市通信網	538	0	89	1	6	1	439	20	0	34	14	0	89
1990/06/15 (金)	ソ連・インドネシア 首脳会談「北方領土」に言及	314	0	75	2	0	1	310	21	0	5	0	0	68
1990/07/16 (月)	空中給油機導入へ	269	1	98	3	0	0	281	7	0	3	0	0	57
1990/08/12 (日)	脳死心臓移植を正式承認	493	0	123	3	4	0	404	8	0	6	0	2	105
1990/09/24 (月)	IMF 融資を弾力運用	377	0	68	4	0	1	460	53	0	56	0	5	95
1990/10/10 (水)	自衛隊「併任」で決着	525	2	173	17	14	2	528	39	0	36	0	8	136
1990/11/01 (木)	協力法案、廃案やむなし	627	0	111	7	4	0	593	52	0	61	0	0	110
1990/12/18 (火)	共和国の 連邦離脱 国民投票結果を容認	290	0	56	4	1	0	387	72	0	12	0	3	69
1995/01/22 (日)	届かぬ救援物資	287	0	78	10	16	0	207	83	0	17	0	0	134
1995/02/10 (金)	東京共和・ 安全信組 10 億円超の預金 24 件	330	4	42	1	0	6	371	49	0	48	0	10	68
1995/03/21 (火)	不審な男、つぶさに目撃	279	0	49	1	2	0	398	8	0	4	12	3	65
1995/04/16 (日)	行動部隊支援の「諜報省」	236	0	48	16	0	0	289	51	0	121	0	0	68
1995/05/27 (土)	3 人に死刑執行	210	0	25	5	0	1	286	37	0	56	0	0	34
1995/06/15 (木)	「殺せば助けてやる」	327	3	57	2	0	1	326	8	0	22	0	16	64
1995/07/10 (月)	特別手配の 2 人逮捕	536	2	135	10	4	0	616	33	0	21	0	0	99
1995/08/16 (水)	円急落 96 円台	522	0	79	2	4	0	510	18	0	51	0	3	99
1995/09/21 (木)	住専処理 新たな妥協案	375	0	79	5	5	0	445	24	0	0	0	0	67
1995/10/28 (土)	住専 設立母体 銀行損失を一括償却	242	0	80	8	3	0	253	17	0	7	0	0	64
1995/11/28 (火)	山口元労相が「裏保証」	192	0	48	1	0	0	297	6	0	20	0	0	60
1995/12/04 (月)	全元大統領を逮捕	258	0	40	14	5	0	325	39	0	0	0	0	85
2000/01/21 (金)	職員会議に法的根拠	339	0	30	0	4	0	374	8	0	21	0	0	69
2000/02/19 (土)	都「03 年度に義務化」案	206	0	24	1	2	0	191	31	0	45	0	15	37
2000/03/07 (火)	「持ち回り」で処分決裁	351	3	70	2	5	1	308	13	0	1	0	0	67
2000/04/01 (土)	有珠山 噴火口 5 カ所	476	0	106	7	5	2	539	74	0	61	39	2	104
2000/05/31 (水)	「利上げ遅れが要因」	318	0	65	1	0	1	348	27	0	49	0	0	74
2000/06/13 (火)	衆院選 きょう告示	445	0	97	2	2	0	473	24	0	17	0	0	114
2000/07/24 (月)	安理の改革「不可欠」	297	2	58	0	0	1	470	17	0	80	0	25	85
2000/08/27 (日)	泥酔者死なせ虚偽報告	289	0	38	0	0	0	248	5	0	23	0	0	40
2000/09/03 (日)	主要 5 行、債権放棄へ	299	3	101	2	0	0	250	22	0	21	0	0	59
2000/10/05 (木)	150 億円申告漏れ	327	0	64	8	10	0	321	61	0	78	0	0	57
2000/11/17 (金)	多数派工作、自民で激化	342	0	101	7	2	0	481	23	0	12	0	0	83
2000/12/11 (月)	チーノ、どうしてる？	389	71	77	4	7	0	234	33	0	106	0	0	107

表 8 毎日新聞（2005 年～2015 年）から抜き出した表記記号の内訳

発行年月日	対象となる記事の見出し	和語				漢語				外来語				記号
		ひ	カ	漢	英数	ひ	カ	漢	英数	ひ	カ	漢	英数	
2005/01/28 (金)	苦手な漢字 時代を反映	284	0	70	0	6	2	218	0	0	37	0	0	69
2005/02/05 (土)	英滞在 1 ヶ月で感染か	312	0	74	0	3	0	490	0	0	93	0	0	83
2005/03/18 (金)	歴史問題「謝罪、反省を」	439	4	135	0	8	0	226	0	0	88	0	0	64
2005/04/04 (月)	ローマ法王死去	355	0	97	0	0	0	415	0	0	19	0	3	71
2005/05/29 (日)	テロ特措法 2 年延長へ	252	3	87	0	4	1	420	0	0	3	0	0	70
2005/06/14 (火)	OB 会で「配分表」	371	0	144	0	0	0	311	0	0	35	0	0	87
2005/07/06 (水)	郵政法案 5 票差可決	263	0	119	0	0	3	323	0	0	55	0	2	58
2005/08/23 (火)	「郵政」だけでは改革なし	199	0	51	0	0	3	249	0	0	78	0	0	36
2005/09/22 (木)	改革総仕上げ強調	203	0	41	0	0	0	312	0	0	12	0	0	44
2005/10/24 (月)	自民・川口氏が大勝	254	0	56	0	0	0	319	0	0	16	0	0	54
2005/11/17 (木)	重電 5 社 本格捜査へ	169	3	51	0	0	3	339	0	0	23	3	0	43
2005/12/14 (水)	「女系天皇」容認 7 割	202	0	44	0	2	0	385	0	0	32	3	0	48
2010/01/12 (火)	防災拠点に穴	189	0	37	0	0	0	192	0	0	12	0	0	31
2010/02/20 (土)	高橋「誇り」の銅	670	0	120	0	6	0	768	0	0	94	0	3	116
2010/03/24 (水)	生方氏解任 撤回	390	0	74	0	2	1	543	0	0	40	0	0	138
2010/04/07 (水)	米、非保有国に核不使用	1011	10	207	0	8	1	1030	0	0	44	0	1	250
2010/05/21 (金)	「瀬戸際」過激化	474	0	72	0	6	1	465	0	0	167	0	16	76
2010/06/27 (日)	「懲戒免」緩和の動き	396	0	48	0	2	0	307	0	0	136	0	0	57
2010/07/04 (日)	放駒親方を代行指名へ	365	0	43	0	0	0	394	7	0	26	0	10	70
2010/08/03 (火)	首相 平身低頭	407	0	51	0	10	1	505	0	0	52	2	0	92
2010/09/27 (月)	樽床氏に迂回献金か	324	0	41	0	2	1	325	0	0	42	0	26	65
2010/10/11 (月)	前部長 覚悟の否認	576	0	96	0	6	0	686	7	0	11	0	0	119
2010/11/11 (木)	複数管区で閲覧可能	394	0	81	0	6	0	402	0	0	0	0	0	82
2010/12/17 (金)	個人増税 6200 億円	286	2	35	0	2	0	397	0	0	32	0	0	51
2015/01/12 (月)	反テロ百数十万人行進	592	2	132	0	8	0	645	0	0	38	0	0	113
2015/02/17 (火)	砂糖業界 100 万円献金	499	0	91	0	4	0	460	2	0	17	0	18	104
2015/03/23 (月)	公明「環境権」の除外検討	484	7	182	0	4	0	461	0	0	26	0	0	82
2015/04/03 (金)	応急手当てに保険	254	0	64	1	2	0	406	3	0	16	2	0	65
2015/05/31 (日)	商業権転売価格 自在に	511	0	141	0	2	0	490	20	0	87	0	26	90
2015/06/10 (水)	政府「行使は限定的」	385	0	56	0	4	0	440	3	0	81	0	6	70
2015/07/31 (金)	米、車関税 20 年で撤廃	498	4	96	0	2	1	536	0	0	6	0	0	64
2015/08/22 (土)	少女殺害 45 歳男逮捕	503	0	163	3	2	6	628	13	0	109	0	14	114
2015/09/23 (水)	五輪「復興工事に影響」8 割	393	4	44	1	0	0	508	5	3	55	0	0	92
2015/10/29 (木)	データ不正 釧路でも	526	0	128	0	4	3	463	0	0	132	0	7	91
2015/11/26 (木)	昨年衆院選「違憲状態」	249	0	28	0	4	0	207	0	0	65	0	3	40
2015/12/22 (火)	保育無料を拡大	238	0	64	0	2	0	281	0	0	18	0	0	41

4 結果の分析

ここでは漢字と他の文字体系との関係について相関分析に基づいて報告すると共に新聞における漢字含有率の変遷を素描することとする。

4.1 相関分析

まず、漢字含有率と同様にひらがな含有率とカタカナ含有率とを算出した上でひらがな含有率と漢字含有率との関係を図 21 に示し、カタカナ含有率と漢字含有率との関係を図 22 に示す。

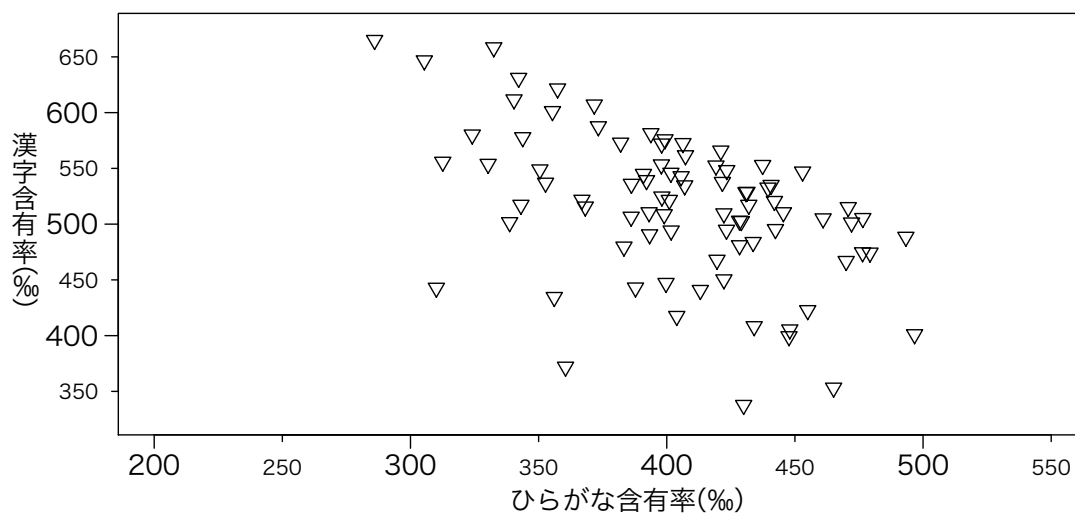


図 21 ひらがな含有率と漢字含有率との関係

これらの関係について相関分析を実施した結果、ひらがな含有率と漢字含有率との間に負の相関が認められ ($r = -.500, p < .001$)、カタカナ含有率と漢字含有率との間にも負の相関が認められた ($r = -.633, p < .001$)。また、図は省略するが、ひらがな含有率とカタカナ含有率との間には相関が認められなかった ($r = -.185, p = .091, n.s.$)。

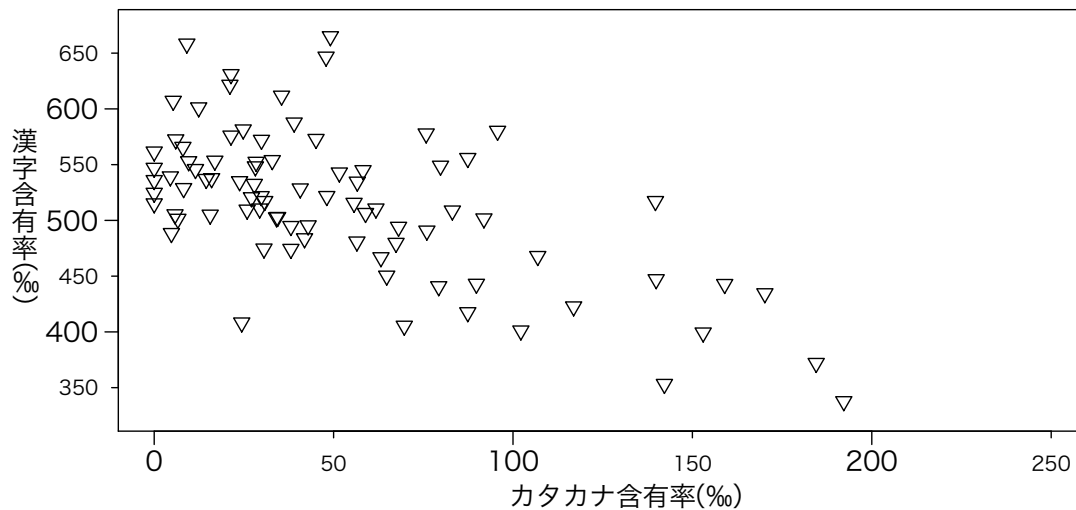


図 22 カタカナ含有率と漢字含有率との関係

これに関連して雑誌における「漢字含有率」を調査した佐竹（1982:345）は「漢字含有率の比較的高い文章」では漢字含有率とひらがな含有率の間には負の相関が見られ、漢字含有率とカタカナ含有率の間には相関が見られないという「文章表記における字種比率に関する仮説」を示しているが、新聞における漢字含有率は雑誌におけるそれよりも高いにも拘わらず、カタカナ含有率の間にも負の相関が見られており、この仮説とは異なる結果を示している。

4.2 漢字含有率の変遷

次にそれぞれの年の漢字含有率について箱ひげ図を作図した上で中央値と上下ヒンジとを平滑化して図 23 に漢字含有率の変遷を示す。

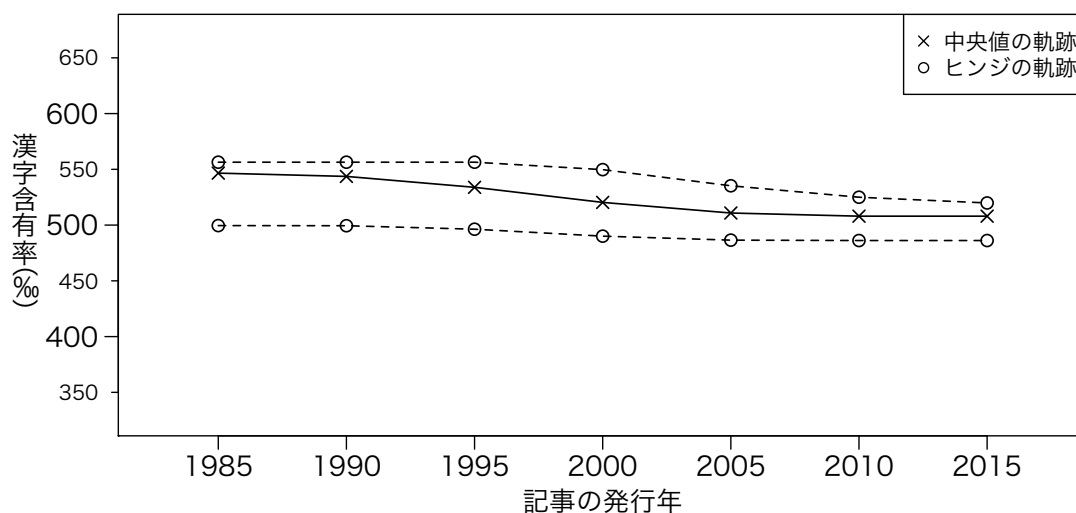


図 23 漢字含有率の変遷

図 23 を見ると漢字含有率は次第に減少する傾向にあることが読み取れる。しかし、平滑化する前の箱ひげ図 (図 24) に照らせば、これは 1995 年と 2000 年との間に生じた差を反映したものであり、この差は 1996 年 4 月 1 日に「数字の表記」が漢数字から算用数字に改められたこと (図 26) によって生じている可能性がある。

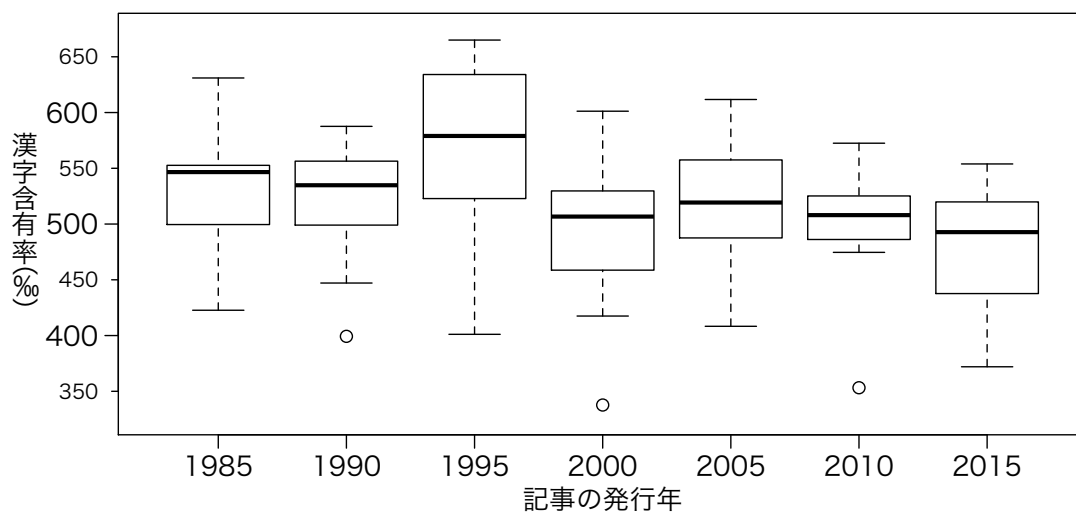


図 24 各区分の箱ひげ図

実際に紙面を見ると 1996 年 3 月 31 日の朝刊一面 (図 25) では漢数字が用いられているのに対し、翌日 (1996 年 4 月 1 日) の朝刊一面 (図 27) では算用数字が用いられて

いる。

契約が切れる四月一日以降、那覇防衛
この決定で、楚辺通信所(象のオリ)の

図 25 1996年3月31日
朝刊一面(一部)

「余録」も大きな活字に
漢数字が原則だった数字の表記をきょうから
洋数字に変えます。また、「余録」の活字が
大きくなりました。よりの読みやすくなった毎日
新聞にご期待下さい。
毎日新聞社

図 26 社告
(翌日朝刊一面)

在沖繩米軍の楚辺通信所(象のオリ)中
(の)の一部用地が3月31日で賃貸契約期限

図 27 翌日
朝刊一面(一部)

5 変遷を形作る要因についての検討

これを踏まえ、抽出した計 84 記事に対して形態素解析器 MeCab (Ver.0.996) と解析用辞書「現代書き言葉 UniDic」(Ver.2.2.0) とを用いて形態素解析を施し、「数字の表記」による影響や語種と表記との関係についても検討する。

なお、第 4 章と同様に解析の誤りは全て人手で修正し、語種については「固有名」や「記号」を認めず、この解析用辞書における「語種」を認定する基準となっている『新潮現代国語辞典』第 2 版等を用いて改めて和語、漢語、外来語、混種語のいずれかに認定することとする。

解析の結果に基づいて「名詞-数詞」(以下、数詞)に該当する語(短単位)の表記に含まれる表記記号を表 6, 表 7, 表 8 から除いたものを毎日新聞から抜き出した表記記号(数詞以外)の内訳として表 9, 表 10, 表 11 に示す。

表 9 毎日新聞（1985 年）から抜き出した表記記号（数詞以外）の内訳

発行年月日	対象となる記事の見出し	和語				漢語				外来語				記号
		ひ	カ	漢	英数	ひ	カ	漢	英数	ひ	カ	漢	英数	
1985/01/18 (金)	自民三役は留任	592	2	130	0	8	0	615	0	0	38	0	0	80
1985/02/01 (金)	原発で「石橋見解」も	499	0	89	0	4	0	412	0	0	17	0	18	81
1985/03/25 (月)	衆院 <small>予算委</small> 審議、初日で中断	484	7	174	0	4	0	409	0	0	26	0	0	84
1985/04/17 (水)	横浜で短銃銀行強盗	254	0	60	0	2	0	381	0	0	16	2	0	48
1985/05/28 (火)	都市交 早朝スト突入	511	0	135	0	2	0	432	0	0	87	0	26	120
1985/06/16 (日)	「5 億円は P3C 工作費」	385	0	56	0	4	0	406	0	0	81	0	6	91
1985/07/06 (土)	宇宙へ “本格、挑戦	498	4	90	0	2	1	507	0	0	6	0	0	61
1985/08/13 (火)	臨時国会は 11 月初旬召集	503	0	157	0	2	6	561	0	0	109	0	14	67
1985/09/16 (月)	524 人乗せ墜落	393	4	41	0	0	0	480	0	3	55	0	0	66
1985/10/10 (木)	大統領権限を強化	526	0	117	0	4	3	422	0	0	132	0	7	47
1985/11/10 (日)	NY 連銀が円買い介入	249	0	26	0	4	0	204	0	0	65	0	3	73
1985/12/25 (水)	ソ連国民に「共存」訴え	238	0	62	0	2	0	261	0	0	18	0	0	58

表 10 毎日新聞（1990年～2000年）から抜き出した表記記号（数詞以外）の内訳

発行年月日	対象となる記事の見出し	和語				漢語				外来語				記号
		ひ	カ	漢	英数	ひ	カ	漢	英数	ひ	カ	漢	英数	
1990/01/05 (金)	「追加公認含め過半数」	189	0	36	0	0	0	183	0	0	12	0	0	60
1990/02/17 (土)	東大医科研倫理委「脳死肝移植」を初承認	670	0	118	0	6	0	727	0	0	94	0	3	111
1990/03/08 (木)	国民負担率 40% 突破	390	0	65	0	2	1	474	0	0	38	0	0	98
1990/04/10 (火)	総選挙に 200 億円	1011	10	196	0	8	1	914	0	0	44	0	1	60
1990/05/09 (水)	下水道使い・都市通信網	474	0	67	0	6	1	427	0	0	167	0	16	89
1990/06/15 (金)	ソ連・インドネシア「北方領土」に言及 首脳会議	396	0	43	0	2	0	292	0	0	136	0	0	68
1990/07/16 (月)	空中給油機導入へ	365	0	43	0	0	0	389	0	0	26	0	10	57
1990/08/12 (日)	脳死心臓移植を正式承認	407	0	48	0	10	1	490	0	0	52	2	0	105
1990/09/24 (月)	IMF 融資を弾力運用	324	0	39	0	2	1	317	0	0	42	0	26	95
1990/10/10 (水)	自衛隊「併任」で決着	576	0	88	0	6	0	670	0	0	11	0	0	136
1990/11/01 (木)	協力法案、廃案やむなし	394	0	78	0	6	0	393	0	0	0	0	0	110
1990/12/18 (火)	共和国の 連邦離脱 国民投票結果を容認	286	2	32	0	2	0	385	0	0	32	0	0	69
1995/01/22 (日)	届かぬ救援物資	284	0	70	0	6	2	204	0	0	37	0	0	134
1995/02/10 (金)	東京共和・ 安全信組 10 億円超の預金 24 件	312	0	61	0	3	0	406	0	0	93	0	0	68
1995/03/21 (火)	不審な男、つぶさに目撃	439	4	129	0	8	0	196	0	0	88	0	0	65
1995/04/16 (日)	行動部隊支援の「諜報省」	355	0	95	0	0	0	394	0	0	19	0	3	68
1995/05/27 (土)	3 人に死刑執行	252	3	79	0	4	1	356	0	0	3	0	0	34
1995/06/15 (木)	「殺せば助けてやる」	371	0	140	0	0	0	284	0	0	35	0	0	64
1995/07/10 (月)	特別手配の 2 人逮捕	263	0	112	0	0	3	304	0	0	55	0	2	99
1995/08/16 (水)	円急落 96 円台	199	0	50	0	0	3	211	0	0	78	0	0	99
1995/09/21 (木)	住専処理 新たな妥協案	203	0	36	0	0	0	301	0	0	12	0	0	67
1995/10/28 (土)	住専 設立母体 銀行損失を一括償却	254	0	52	0	0	0	287	0	0	16	0	0	64
1995/11/28 (火)	山口元労相が「裏保証」	169	3	49	0	0	3	270	0	0	23	3	0	60
1995/12/04 (月)	全元大統領を逮捕	202	0	39	0	2	0	365	0	0	32	3	0	85
2000/01/21 (金)	職員会議に法的根拠	339	0	30	0	4	0	374	0	0	21	0	0	69
2000/02/19 (土)	都「03 年度に義務化」案	206	0	24	0	2	0	187	0	0	45	0	15	37
2000/03/07 (火)	「持ち回り」で処分決裁	351	3	70	0	5	1	306	0	0	1	0	0	67
2000/04/01 (土)	有珠山 噴火口 5 カ所	476	0	104	1	5	2	537	0	0	61	39	2	104
2000/05/31 (水)	「利上げ遅れが要因」	316	0	64	0	0	1	348	0	0	47	0	0	74
2000/06/13 (火)	衆院選 きょう告示	445	0	97	0	2	0	473	0	0	17	0	0	114
2000/07/24 (月)	安保理の改革「不可欠」	297	2	58	0	0	1	470	0	0	80	0	25	85
2000/08/27 (日)	泥酔者死なせ虚偽報告	289	0	38	0	0	0	245	0	0	23	0	0	40
2000/09/03 (日)	主要 5 行、債権放棄へ	299	3	101	0	0	0	246	0	0	21	0	0	59
2000/10/05 (木)	150 億円申告漏れ	327	0	64	0	10	0	310	0	0	78	0	0	57
2000/11/17 (金)	多数派工作、自民で激化	342	0	101	0	2	0	480	0	0	12	0	0	83
2000/12/11 (月)	チーノ、どうしてる？	389	71	77	2	7	0	233	0	0	106	0	0	107

表 11 毎日新聞（2005 年～2015 年）から抜き出した表記記号の内訳

発行年月日	対象となる記事の見出し	和語				漢語				外来語				記号
		ひ	カ	漢	英数	ひ	カ	漢	英数	ひ	カ	漢	英数	
2005/01/28 (金)	苦手な漢字 時代を反映	287	0	76	1	16	0	206	1	0	17	0	0	69
2005/02/05 (土)	英滞在 1 カ月で感染か	330	4	42	0	0	6	369	0	0	48	0	10	83
2005/03/18 (金)	歴史問題「謝罪、反省を」	279	0	46	0	2	0	397	0	0	4	12	3	64
2005/04/04 (月)	ローマ法王死去	236	0	48	2	0	0	286	0	0	121	0	0	71
2005/05/29 (日)	テロ特措法 2 年延長へ	210	0	25	2	0	1	284	0	0	56	0	0	70
2005/06/14 (火)	OB 会で「配分表」	327	3	56	0	0	1	326	0	0	22	0	16	87
2005/07/06 (水)	郵政法案 5 票差可決	536	2	135	0	4	0	616	0	0	19	0	0	58
2005/08/23 (火)	「郵政」だけでは改革なし	522	0	79	0	4	0	508	0	0	51	0	3	36
2005/09/22 (木)	改革総仕上げ強調	375	0	79	1	5	0	445	0	0	0	0	0	44
2005/10/24 (月)	自民・川口氏が大勝	242	0	80	0	3	0	253	0	0	7	0	0	54
2005/11/17 (木)	重電 5 社 本格捜査へ	192	0	48	0	0	0	297	0	0	20	0	0	43
2005/12/14 (水)	「女系天皇」容認 7 割	258	0	40	0	5	0	325	0	0	0	0	0	48
2010/01/12 (火)	防災拠点に穴	293	0	76	0	2	1	309	0	0	44	0	8	31
2010/02/20 (土)	高橋「誇り」の銅	500	5	135	0	7	0	249	0	0	150	0	4	116
2010/03/24 (水)	生方氏解任 撤回	508	2	161	0	8	0	349	0	0	3	0	0	138
2010/04/07 (水)	米、非保有国に核不使用	308	0	36	0	2	0	338	0	0	19	0	30	250
2010/05/21 (金)	「瀬戸際」過激化	538	0	89	0	6	1	438	0	0	34	14	0	76
2010/06/27 (日)	「懲戒免」緩和の動き	314	0	75	1	0	1	310	0	0	5	0	0	57
2010/07/04 (日)	放駒親方を代行指名へ	269	1	98	0	0	0	281	0	0	3	0	0	70
2010/08/03 (火)	首相 平身低頭	493	0	123	0	4	0	404	0	0	6	0	2	92
2010/09/27 (月)	樽床氏に迂回献金か	377	0	68	0	0	1	457	0	0	56	0	5	65
2010/10/11 (月)	前部長 覚悟の否認	525	2	173	4	14	2	527	0	0	36	0	8	119
2010/11/11 (木)	複数管区で閲覧可能	627	0	110	1	4	0	591	14	0	61	0	0	82
2010/12/17 (金)	個人増税 6200 億円	290	0	56	0	1	0	371	0	0	12	0	3	51
2015/01/12 (月)	反テロ百数十万人行進	314	0	64	0	4	1	321	0	0	151	0	2	113
2015/02/17 (火)	砂糖業界 100 万円献金	324	2	63	0	7	0	482	0	0	31	0	18	104
2015/03/23 (月)	公明「環境権」の除外検討	423	0	62	0	2	0	424	0	0	28	0	0	82
2015/04/03 (金)	応急手当てに保険	326	0	63	0	2	1	284	0	0	62	0	3	65
2015/05/31 (日)	商業権転売価格 自在に	469	0	78	0	0	2	382	0	0	238	0	12	90
2015/06/10 (水)	政府「行使は限定的」	465	0	79	0	3	0	433	0	0	0	0	0	70
2015/07/31 (金)	米、車関税 20 年で撤廃	379	0	35	0	0	4	302	0	0	55	0	27	64
2015/08/22 (土)	少女殺害 45 歳男逮捕	363	0	155	3	4	1	273	0	0	32	0	0	114
2015/09/23 (水)	五輪「復興工事に影響」8 割	291	0	54	0	4	0	313	0	0	57	0	0	92
2015/10/29 (木)	データ不正 釧路でも	366	0	63	0	12	0	336	0	0	58	4	1	91
2015/11/26 (木)	昨年衆院選「違憲状態」	349	0	64	4	0	0	415	0	0	4	0	0	40
2015/12/22 (火)	保育無料を拡大	299	0	65	10	0	0	325	0	0	0	0	0	41

5.1 相関分析

漢字含有率の変遷と「数字の表記」による影響との関係について検討することに先立ち、漢字と他の文字体系との関係について改めて検討する。

まず、図 21 に見たように漢字含有率とひらがな含有率の間には負の相関が見られていたが、「数字の表記」による影響を取り除き、数詞に該当する語の表記に含まれる漢字を除いた延べ漢字数を分子とする漢字含有率と数詞に該当する語の表記に含まれるひらがなを除いた延べひらがな数を分子とするひらがな含有率との関係について改めて相関分析を実施した結果、両者の間には相関が認められなくなった ($r = -.190, p = .084, n.s.$)。

一方、図 22 に見た漢字含有率とカタカナ含有率との間に負の相関が認められたことについては「数字の表記」による影響を取り除いても大きくは変化せず、数詞に該当する語の表記に含まれる漢字を除いた延べ漢字数を分子とする漢字含有率と数詞に該当する語の表記に含まれるカタカナを除いた延べカタカナ数を分子とするカタカナ含有率との関係について相関分析を実施した結果、両者の間には負の相関が認められた ($r = -.674, p < .001$)。

すなわち、図 21 に見たように漢字含有率とひらがな含有率との間に負の相関が認められたのは「数字の表記」による影響を受けて漢字含有率が減少したことを反映したものであり、そのような影響を除いても漢字含有率とカタカナ含有率との間には負の相関が認められていることの方が新聞を特徴づけるものとして重要であると考えられる。これは「漢字含有率の比較的高い文章」である新聞において漢字含有率とカタカナ含有率との間に負の相関が認められたという点において「漢字含有率の比較的高い文章」では漢字含有率とひらがな含有率との間に負の相関が見られ、「漢字含有率の比較的低い文章」では漢字含有率とカタカナ含有率との間に負の相関が見られるという佐竹 (1982:345) の「文章表記における字種比率に関する仮説」には反する結果である。

なお、佐竹 (1982) の仮説は複数の媒体やジャンルを対象としているものの基本的には同時期のものを対象として導き出されたものであり、本章において対象としている新聞のように時間的な幅を持ち、なおかつ、その中で変動するような場合を想定していない可能性がある。この点を加味した結果、新聞においては漢字含有率とひらがな含有率の間には負の相関が見られるのではなく、漢字含有率とカタカナ含有率との間に負の相関が見られるということを明らかにし得たのである。

5.2 漢字含有率の変遷

数詞に該当する語の表記に含まれる漢字を除いた延べ漢字数を分子とする漢字含有率の変遷を図 28 に示す¹²²。

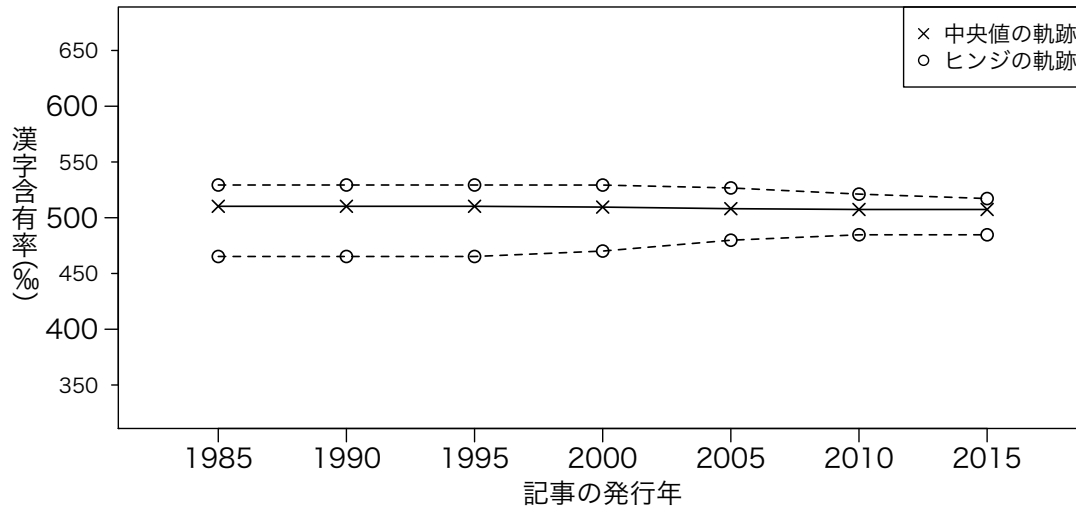


図 28 漢字含有率（分子から数詞を除く）の変遷

図 28 を見ると漢字含有率は概ね横ばいである。このことは「数字の表記」による影響を除けば、漢字含有率が安定していることを示唆している。

また、図 29 に和語の漢字含有率を示し、図 30 に漢語の漢字含有率を示すと漢語の漢字含有率の方が和語の漢字含有率に比べて「数字の表記」による影響を大きく受けていることが読み取れる。これは延べ文字数で見れば、数詞の表記に用いられている全 3298 字のうち和語は 390 字 (11.8%) であるのに対し、漢語は 2902 字 (88.0%) であるように数詞の多くが漢語であることに起因しているものと見える。

¹²² 数詞に該当する語の表記に含まれる漢字を除いた延べ漢字数を分子とすることは野村 (1980) が「漢字含有率」の算出方法を整理する際に漢数字を除いた延べ漢字数を分子とする「漢字含有率」を示していることに着想を得たものである。なお、分子のみから除いているのは数詞に該当する語の表記に含まれる漢字あるいは漢数字を分母となる延べ文字数からも除くことに比して分母が同一であることから比較することが容易であり、除いたことによる影響を適切に捉えられるという利点があるものと考えられる。また、単に漢数字を除くと「三位一体」の「三」や「一」も「純一郎」の「一」も除くことになるが、野村 (1980) は数詞に該当する場合に限って漢数字を除いており、これと同様に「数字の表記」による影響を捉えるためには数詞に該当する語に含まれる漢字のみを除くことが望ましいものと考えられる。

一方、和語の漢字含有率は「数字の表記」による影響が相対的に小さいことに加え、そもそも増加しているとは看做しがたく、これは土屋（2000b）の報告とは異なる結果である。

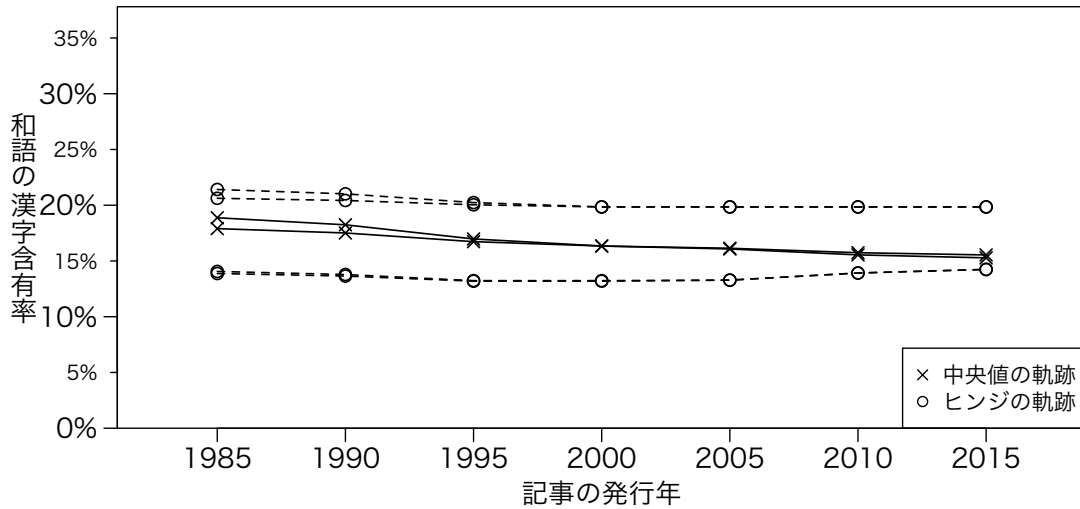


図 29 和語の漢字含有率の変遷（数詞の含むか除くかの差）

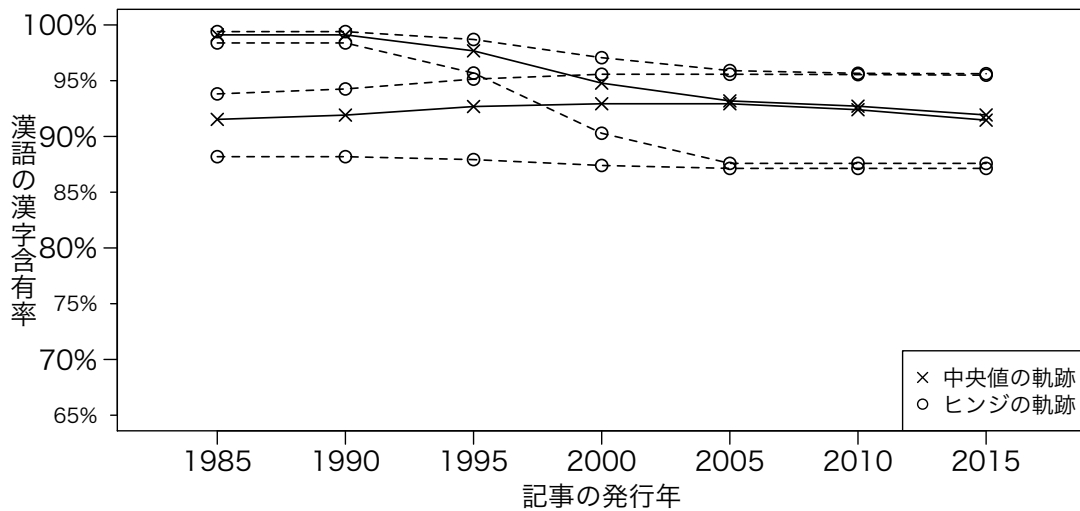


図 30 漢語の漢字含有率の変遷（数詞の含むか除くかの差）

なお、外来語が増加するという野村（1988b）の予想を検証するために概ね第 4 章と同様に 1,000 字当たりの外来語の延べ文字数の変遷についても図 31 に示す¹²³。

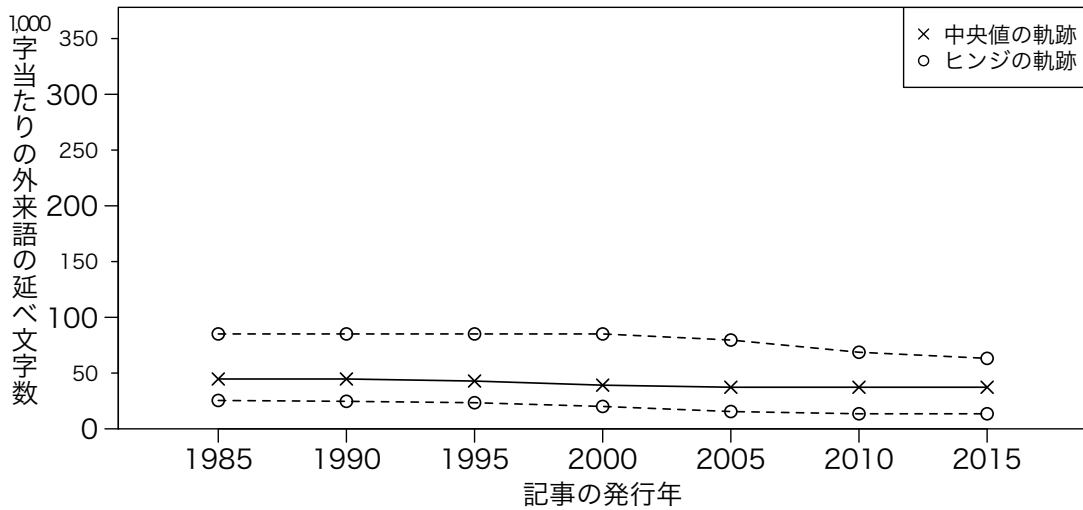


図 31 1,000 字当たりの外来語の延べ文字数の変遷

図 31 を見ると 1,000 字当たりの延べ文字数は概ね横ばいであり、外来語が増加するという野村（1988b）は裏づけられなかった。

6 本章の結論

第 5 章では 1985 年から 2015 年に掛けての新聞における漢字含有率は一見すると減少しているが、これは 1996 年 4 月 1 日に「数字の表記」が漢数字から算用数字に改められたことによるものであり、この影響を取り除けば、安定しているということを明らかにした。

また、宮島（1988a）に倣って和語や漢語の漢字含有率を算出し、その変遷を素描すると「数字の表記」による影響は認められるもののそれを除けば、大きな変化は見られず、このことが前掲した漢字含有率が安定していることの要因であると考えられる。

¹²³ 本章において実施した新聞を対象とした調査では抽出した記事の延べ文字数が記事によって異なっていることから第 4 章と同様に比較するために 1,000 字当たりの延べ文字数（外来語含有率とも呼び得る）に換算して示す。

第6章

漢字含有率に見る現在の漢字使用の実態

ここでは小説を対象とした第3章と第4章との調査結果と新聞を対象とした第5章の調査結果とを踏まえ、漢字含有率を通して見た漢字使用の実態について整理する。

1 小説と新聞とにおける漢字含有率の変遷

初めに小説における漢字含有率の変遷と新聞における漢字含有率の変遷とをそれぞれ図32と図33とに改めて掲げる。

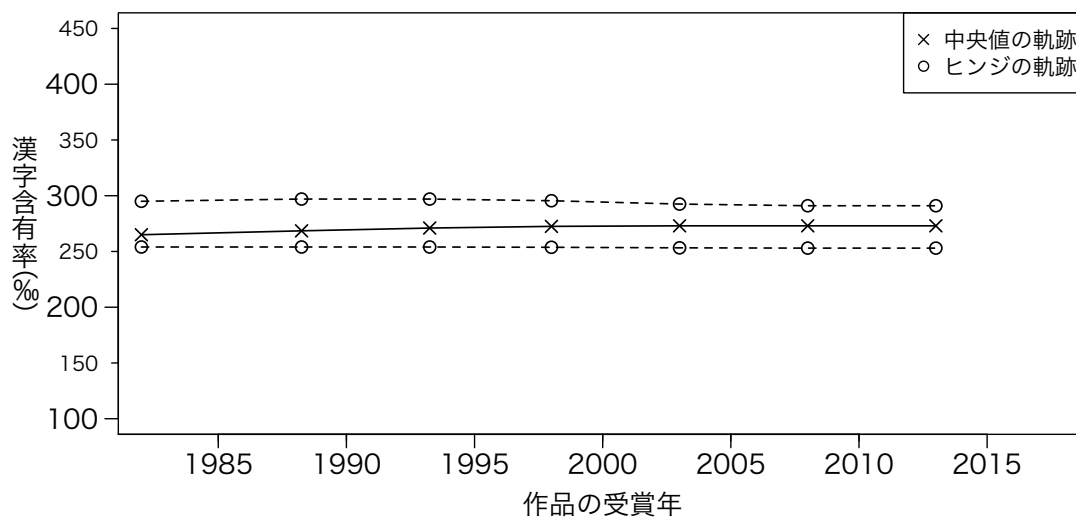


図32 小説における漢字含有率の変遷¹²⁴

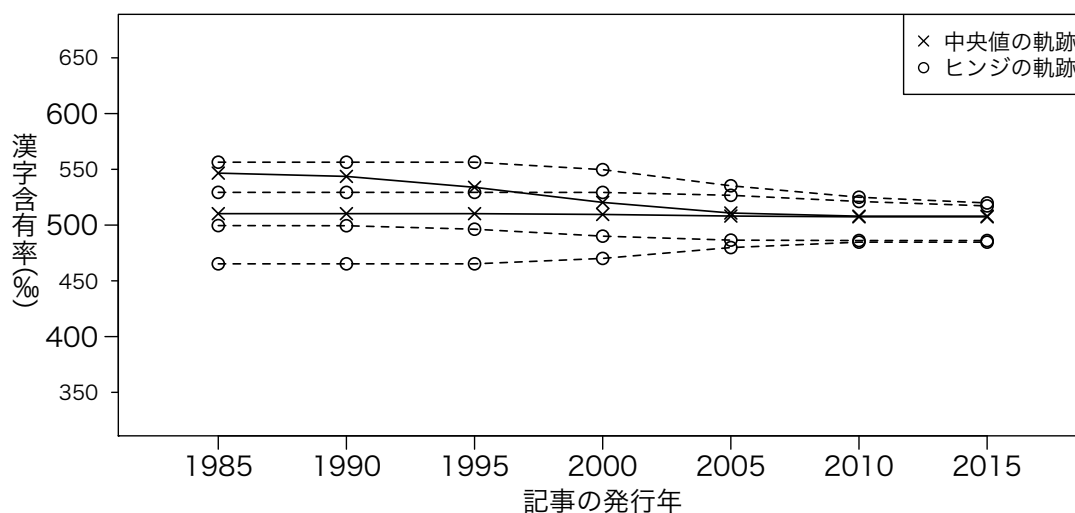


図 33 新聞における漢字含有率の変遷¹²⁵ (数詞の含むか除くかの差)

2 相違点から見る小説や新聞の文章における漢字使用の実態

まず、小説における漢字含有率の変遷と新聞における漢字含有率の変遷とを比較すると前者は 30 年に亘って漢字含有率が安定しているのに対し、後者は漢字含有率が減少しているという点が相違点として挙げられる。すなわち、両者の比較においては漢字含有率が安定しているということは小説に特有の傾向であり、漢字含有率が減少しているということは新聞に特有の傾向ということになる。

なお、既に述べた通り、新聞において漢字含有率が減少したのは 1996 年 4 月 1 日の「数字の表記」による影響であり、実態としては図 33 に示したように緩やかな減少したのではなく、同日を境に急激に減少したものである。このような言語外的（あるいは人為的）な要因の影響を受け得るという点は小説においては見受けられなかった点であり、小説の方が書き手の裁量が相対的に大きく、反対に新聞の方が書き手の裁量が相対的に小さいということを反映したものと見える。

¹²⁴ これは第 4 章に示した図 14 における本章調査分を抜き出したものである。

¹²⁵ これは第 5 章に示した図 23 と図 28 とを合わせたものである。

また、佐竹（1982）の「補完性¹²⁶」という用語に従って整理すれば、小説では漢字とひらがなの間に「補完性」があり、漢字とカタカナの間には「補完性」がないのに対し、新聞においては漢字とひらがなの間にも漢字とカタカナの間にも「補完性」があるという結果が得られた。これは一見すると漢字とひらがなの間に「補完性」があるという点において小説と新聞とが共通しているようにも見えるが、既に述べた通り、「数字の表記」による影響を除くと新聞においては漢字とひらがなの間の「補完性」が見られなくなることが明らかになった。すなわち、「数字の表記」による影響を除くことにより、小説はひらがなの間に「補完性」があって新聞はカタカナとの間に「補完性」があるという相違点先鋭化するのである。このように「数字の表記」による影響があるか否かという点は小説と新聞との両者の相違点としては大きいものであると言える。

更に小説における漢字含有率が相対的に低く、新聞における漢字含有率が相対的に高いという点も両者の相違点として挙げられる。これは媒体やジャンルによって漢字含有率が異なるという野村（1980, 1988b）や佐竹（1982）の指摘を裏づけるものである。また、第4章において確認したように漢字含有率と漢語の大部分を占める漢字表記の漢語とは密接に関わっていることから両者における漢字含有率の差は漢語の構成比率の差を反映したものであると考えられる¹²⁷。漢語の構成比率は樺島（1963）によれば、名詞の構成比率と密接に関わるものであり、名詞の構成比率はいわゆる「樺島の法則」（樺島 1955）として知られているようにジャンルと関わるものである。このように考えても小説における漢字含有率と新聞における漢字含有率との差は媒体やジャンルの差を反映したものであると考えられるのである。

¹²⁶ 佐竹（1982:345）によれば、「一方の字種比率が多くなると他方が少なくなる関係」のことであり、佐竹（1987a）は相関係数について「マイナスの値が大きいほど補完関係が強い」（佐竹 1987a:23）と述べている。例えば、漢字含有率（1,000字当たりの延べ漢字数）とひらがな含有率（1,000字当たりの延べひらがな数）との間に負の相関が見られれば、「漢字とひらがなの間に補完性がある」（佐竹 1982:345）ということになる。

¹²⁷ 佐竹（1987b）も「漢字含有率」と漢語の構成比率との間に正の相関があることを示唆している。

3 共通点から見る現代日本語の文章における漢字使用の実態

確かに新聞においては漢字含有率が減少していると言えるが、既に述べた通り、「数字の表記」による影響を除けば、安定しているということが明らかになった。また、小説においては漢字含有率が安定していることから「数字の表記」を変更するという明らかに言語外的な要因による影響を除けば、小説においても新聞においても漢字含有率は安定しているということになる。

更に外来語が増加するという野村（1988b）の予想は裏づけられず、小説においても新聞においても語種の構成比率や語種別の漢字使用の実態には少なくとも30年に亘って大きな変化が生じていないということが明らかになった。

そもそも媒体やジャンルによって漢字含有率が異なっており、個々の媒体やジャンルによって特有の事情があるというのは勿論であるが、小説は書き手の裁量が最も大きいと考えられる媒体やジャンルであり、新聞は書き手の裁量が最も小さいと考えられる媒体やジャンルであることから両者において漢字含有率が安定しているというのは広く現代日本語の文章において一般的に見られる傾向であると言える可能性がある。すなわち、漢字使用の実態は個々の媒体やジャンルによって異なっているものの少なくとも30年に亘って大きくは変化していないという仮説が導き出せるのである。

4 本章の結論

第6章では小説における漢字含有率の変遷と新聞における漢字含有率の変遷との比較を通して以下の点を明らかにした。まず、新聞においては「数字の表記」による影響が見られる一方、小説においては変化が認められず、これが書き手の裁量の大きさを反映したものと考えられることを指摘した。次に新聞においては漢字含有率が相対的に高く、漢字含有率とカタカナ含有率との間に「補完性」があるのに対し、小説においては漢字含有率が相対的に低く、漢字含有率とひらがな含有率との間に「補完性」があるのは個々の媒体やジャンルの違いを反映したものと考えられることを指摘した。最後に個々の媒体やジャンルによって特有の事情はあるもののそれを除けば、新聞においても小説においても漢字

含有率が安定していることから現代日本語の文章における漢字使用の実態は個々の媒体やジャンルによって異なっているものの少なくとも 30 年に亘って大きくは変化していないという仮説が導き出せることを主張した。

終章

ここでは第 1 節において本論文の結論を示し、第 2 節において本論文の意義について述べることにする。また、第 3 節では残された課題について触れることにする。

1 本論文の結論

本論文では「現代日本語の文章における漢字使用の実態を通時的・計量的な観点から解明すること」を目的とし、この目的を達するために「漢字含有率」に着眼した。また、複数の媒体やジャンルを対象として実際に「漢字含有率」を調査し、これに基づいて漢字使用の実態について現在までの変遷を明らかにし、その変遷を形作る要因についても語種に基づいて検討することを通して所期の目的を達せんとして研究を進めてきた。以下では第 3 章から第 6 章までの内容を再掲しながら本論文の結論を示すことにする。

初めに第 3 章では実際に小説を対象として「漢字含有率」を調査し、その変遷について明らかにした。その結果、1986 年から 2015 年に至るまで「漢字含有率」は大きく変動しているとは言えず、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安本（1963）の予測の前提となっている「漢字含有率」が減少し続けるという傾向は裏づけられないという結論が得られた。また、「漢字含有率」が安定しつつあるという可能性が改めて示された。これは「漢字含有率」が安定ないしは微増するという宮島（1988a）の予想を裏づける結果である。

次に第 4 章では第 3 章において明らかにした「漢字含有率」が安定する傾向を形作る要因を探るために先行研究に倣って語種に着目し、宮島（1988a）の方法を発展的に踏襲しつつ第 3 章と同様に小説を対象として語種の構成比率や語種別の漢字使用の実態について明らかにした。その結果、20 世紀の末葉から「漢字含有率」が安定しているのは 20 世紀の中葉まで見られたような語種の構成比率や語種別に見た漢字使用の実態の変化が見

られなくなったからであるという結論が得られた。これは外来語が増加するという野村(1988b)に反する結果であるとも言える。また、宮島(1988a, b)の数表を用いて同様に分析した結果、20世紀の中葉まで「漢字含有率」が減少していたのは主として漢字表記の漢語の減少に起因している可能性が示された。更に漢字表記の和語は仮名表記の和語の増加に伴う和語全体の増加によって相対的に減少していると捉えられることも示された。これは杉山・森岡(1969)や宮島(1988a)の調査結果を再解釈し、内実を明らかにしたものであると言える。

続いて第5章では第3章において対象とした小説と概ね同時期の新聞を対象として「漢字含有率」を調査し、その変遷について明らかにした。その結果、1985年から2015年に至るまでの「漢字含有率」は1996年4月1日に「数字の表記」が改められたことによって減少しているが、それによる影響を除くと30年間の「漢字含有率」が安定しているという結論が得られた。また、このように安定していることの要因について探るために第4章に倣って語種別の漢字使用の実態について明らかにした結果、「数字の表記」による影響を除けば、語種別に見た漢字使用の実態が大きくは変化しておらず、このことが安定の要因となっている可能性が示された。

最後に第6章では小説における「漢字含有率」の変遷と新聞における「漢字含有率」の変遷とを比較し、相違点に基づいて個々の媒体やジャンルにおける漢字含有率の変遷を明らかにすると共に共通点に基づいて広く現代日本語の文章における漢字使用の実態を明らかにすることを試みた。まず、小説においては変化が見られず、「漢字含有率」が安定しているのに対して新聞においては「数字の表記」による影響によって「漢字含有率」が減少していることは書き手の裁量の大きさを反映したものと考えられることを指摘した。次に小説においては「漢字含有率」が相対的に低く、新聞においては「漢字含有率」が相対的に高いことは個々の媒体やジャンルの違いを反映したものと考えられることを指摘した。最後に「数字の表記」のような言語外的な要因による影響を除けば、小説においても新聞においても「漢字含有率」は安定しており、小説が最も書き手の裁量の大きいと考えられる媒体やジャンルであって新聞が最も書き手の裁量が小さいと考えられる媒体やジャンルであることも踏まえば、確かに個々の媒体やジャンルに特有の事情はあるものの現代日本語の文章における漢字使用の実態は少なくとも30年に亘って大きくは変化していないという仮説が導き出せることを主張した。

2 本論文の意義

本論文は「漢字含有率」を調査することを通して現代日本語の文章における漢字使用の実態を通時的・計量的な観点から明らかにしたものである。その具体的な成果については第1節において述べた通りであるが、その意義について本節では中心的なものから周辺的なものに至るまで説明を加えることとする。

まず、本論文は「漢字含有率」の現在に至るまでの変遷を明らかにしたものであり、小説においては2000年までの小説を対象として「漢字含有率」を調査した中村(2003)に続くものとして新聞においては2000年までの新聞を対象として「漢字含有率」を調査した土屋(2000b)に続くものとして21世紀に入ってからの漢字含有率の変遷を明らかにした点は第一義的な意義であると言える。また、安本(1963)の予測や野村(1988b)の予想には反するものの宮島(1988a)の予想を裏づける結果が得られたことを報告した点や佐竹(1982)の仮説について検証した点も「漢字含有率」の研究史の進展に直接的に貢献するものであると言える。

これに関連して高田(2009:51)は「漢字含有率」について「文字使用の実態を把握するための基本的な尺度」であるとした上で「含有率の推移などの要因を考察するためには、土屋(1967)のように、ある語が漢字表記か仮名表記かといった観点で、漢字表記語率や語種別の語表記率を見ていく必要がある」ことを指摘すると共に「パソコンの普及によって、文字や表記の計量的研究のための環境が以前に比べて整備されたように思われるが、近年の成果が乏しい分野である」と述べている。このうちの「含有率の推移などの要因を考察するためには、土屋(1967)のように、ある語が漢字表記か仮名表記かといった観点で、漢字表記語率や語種別の語表記率を見ていく必要がある」(高田 2009:51)という点については第4章において述べたように土屋(1967)やそれを受け継いだ土屋(2000a, b)の方法には難があることから語種と表記との関係を見ることに限れば、宮島(1988a)の方法を発展的に踏襲することが望ましいことを本論文においては主張した。すなわち、本論文は「ある語が漢字表記か仮名表記かといった観点で、漢字表記語率や語種別の語表記率を見ていく必要がある」という高田(2009:51)の指摘について宮島(1988a)の方法を発展的に踏襲しつつ更に精緻化することにより、「含有率の推移などの要因を考察するため」(高田 2009:51)の方法論として整備することを通して批判的に継承したもので

あり、「近年の成果が乏しい分野」となりつつあった「漢字含有率」の研究史に新たな成果を報告するものとして位置づけられるのである。

更に関連して山崎（2010）は日本語を対象とした計量的な研究について概観した上で「漢字含有率」を日本語研究における「代表的な指標」（山崎2010:273）の1つとして挙げつつ以下のように述べた。

計量的研究の醍醐味として、統計指標の作成がある。日本語研究では、代表的な指標として漢字含有率、カバー率（累積使用率）、語彙の類似度などがある。これ以外にも名前がないだけで臨時的に作られた指標もあるだろう。有益な統計指標ができれば、学界に裨益するところが大きい。発想の柔軟な若い研究者には意欲的に取り組んでいただきたい課題である。（山崎 2010:273）

当然、ある統計指標を「有益な統計指標」（山崎 2010:273）とするためには単に指標を作るということのみならず、当該の指標を用いた研究の方法論を整備することも不可欠であることに照らせば、新たな統計指標を作成したものではないものの「代表的な指標」（山崎 2010:273）である「漢字含有率」を取り上げ、それに基づく研究の方法論について未整備の部分の拡充した本論文も「統計指標の作成」（山崎 2010:273）と同様に「学界に裨益するところが大きい」（山崎 2010:273）ものと言えるのではないか。

なお、「漢字含有率」は現代語の表記のみならず、古典語の表記を見るためにも用いられており（麻生 1955; 前田 1971; 彦坂 1987; 矢野 1987; 斎藤 2011; 太刀岡 2014 など）、本論文において方法論を整備したことはそれらの研究にも資する可能性がある。ただし、太刀岡（2014）に対しては「漢字使用率が諸本比較に有効とする論は、親本をそっくりそのまま写す書写態度が前提になっているので再考した方がいいのではないか」という中川（2016:80）の指摘もあることから古典語の表記を扱う際には現代語とは異なる事情のあることに留意する必要がある¹²⁸。

また、本論文のように巨視的な立場から漢字使用の実態を捉えた研究は微視的な立場から漢字使用の実態を捉えた研究と対立するものではなく、全体の傾向を明らかにすることによって既に明らかになっている個別の実態やその変化を全体の流れの中に位置づけることを可能にするものである。すなわち、既に多くの蓄積がある個別の表記の研究に対して

¹²⁸ 中川（2016）の言う「漢字使用率」は本論文における「漢字含有率」を指すものである。また、中川（2016）の指摘は斎藤（2011）に対しても言えるものと考えられる。

も総体としての表記という視座を提供することにより、その中に個別の表記の研究が位置づけられるようになることから表記の研究史においても本論文の果たす意義は大きいものであると言える。

更に本論文によって現在の漢字使用の実態が明らかになったことは定期的な改定が決められた「常用漢字表」などの国語施策（特に漢字政策）の評価や見直しにも資するものであり、科学的な根拠に基づく政策評価のための下地を提供するものであるとも言える。

これに加えて読み方が定められないという「語彙調査に関する未解決の問題」（山崎 2013:28）の1つについて暫定的ながらも客観的で再現可能な手順を定めた点はコーパス言語学的に意義深いものであると言える。この問題は田中（2020）も指摘しているように現代語に限らないものであり、コーパスを構築するためには解決が求められているものである。また、「表記論は、ある語をどう書いているか、という議論であると同時に、どう読むか（読まれるか）という議論でもある」という尾山（2020:78）の指摘にもある通り、本論文において中心的に扱ってきた表記の研究と密接に関わる問題であると言える。

最後に石井（2018）は「探索的コーパス言語学」を提唱し、その一環として「探索的データ解析」の手法が日本語研究に有効であることを主張しており、本論文はその実証を担うものとしても位置づけられる。この方法論の有効性を実証的に提示したことは「漢字含有率」のみならず、広くは言語データから何らかの割合を産出し、その変遷を扱う多くの研究に示唆を与えるものである。

3 本論文の課題

最後に残された課題について以下に述べる。

まず、石井（2013:149）も指摘しているように「探索的データ解析」は仮説を得るための手法であり、これによって「仮説を得たとしても、それは確認的データ解析（推測統計学）にもとづく別の調査によって」裏づけることが必要である。また、そもそも安定していることや安定した要因となっていることは積極的に結論づけるのは困難であるということも留意する必要がある。

また、本論文では総体としての漢字使用の実態を捉えることに主眼を置いたために微視的な立場から漢字使用の実態を捉えることについては検討が及んでおらず、この点は課題の1つとして挙げられる。

更に国語施策による影響や 20 世紀の末葉から現在に至るまでの情報機器の発達・普及による書字環境の変化などの言語外的な要因については「漢字含有率」に影響を及ぼすことが考えられるが、十分には扱えていないことも課題である。ただし、この点については定性的に言及することは可能であっても定量的に立証するためには種々の条件を統制する必要がある。例えば、実際に「漢字含有率」を調査した研究のうち漢字制限による影響に着目した羽山（2017）の研究が漢字制限の実施前後を比較する方法、ワープロによる影響に着目した鶴岡（1991）の研究や金・樺島・村上（1993）の研究がワープロの使用前後を比較する方法をそれぞれ採っているように個別の要因に限定して考察することも一案ではあるが、条件の統制が困難であることや対象が限られることは難点として挙げられる。

これに加えて語を単位とした分析や品詞に着目した分析では様相が異なる可能性もあるが、これらの分析には第 4 章において提起した問題について解決する必要性が生じてくるのである。これも課題の 1 つである。

最後に本論文では最も書き手の裁量が大きいと考えられる小説と最も書き手の裁量が小さいと考えられる新聞とを比較する方法を採ったが、その間に位置すると考えられる雑誌などについては検討が及んでいない点も課題として挙げられる。また、紙媒体に限らなければ、電子媒体などを今後は取り上げる必要もあるものと考えられる。

参考文献

- 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, American Statistical Association.
- Tanaka, A. (1998) The number of Kanji in modern Japanese writing, *Journal of quantitative linguistics*, 5(1-2), 100-104, International Quantitative Linguistics Association.
- Tukey, J. W. (1977) *Exploratory data analysis*, Addison-Wesley.
- 浅川哲也 (2013) 「日本語史における「現代語」とは何か—日本語史の時代区分についての一提言—」『都大論究』(50), 39-49, 東京都立大学国語国文学会.
- 麻生磯次 (1955) 「漢字漢語の活用」麻生磯次『江戸文學と中國文學』再版, 370-387, 三省堂.
- 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, FIT (電子情報通信学会・情報処理学会) 運営委員会.
- 石井久雄 (2002) 「日本語の文字・表記における現代と歴史との断絶」『日本語の文字・表記—研究会報告論集—』, 7-22, 国立国語研究所.
- 石井久雄 (2010) 「漢字の出現年代」『同大語彙研究』(13), 15-31, 同志社大学大学院日本語学研究会.
- 石井久雄・入江さやか (2015) 「『中央公論』101 年間語彙表」『同志社日本語研究—別刊—』(2), 1-410, 同志社大学大学院日本語学研究会.
- 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.

- 石井正彦（2018）「探索的コーパス言語学のための覚書」『現代日本語研究』10, 81-98,
大阪大学大学院文学研究科日本語学講座現代日本語学研究室.
- 石井正彦（2019）「蛇行箱型図による S 字カーブの発見」石井正彦『探索的コーパス言語学—データ主導の日本語研究・試論—』, 304-331, 大阪大学出版会. [石井（2013）を再録したもの]
- 伊藤雅光（2001）「ポップス系流行歌の語彙調査における外来語と外国語の判定基準」『計量国語学』23（2）, 110-130, 計量国語学会.
- 乾善彦（2001）「語彙史の時代区分・文字史の時代区分」国語語彙史研究会（編）『国語語彙史の研究 20』, 23-36, 和泉書院.
- 乾善彦（2014）「文字論」佐藤武義・前田富祺（編）『日本語大事典（下）』, 2009-2010, 朝倉書店.
- 犬飼隆（1988）「文字・表記（史的研究）」『国語学』（153）, 78-83, 国語学会.
- 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.
- 内田善次郎（1964）「明治以後の文学作品にあらわれた文字使用の変遷—漢字使用状況について—」『国語教育』（12）, 8-15, 高知大学国語教育学会.
- 内山政春（2004）「言語名称「朝鮮語」および「韓国語」の言語学的考察」『異文化』5, 73-107, 法政大学国際文化学部.
- 大木一夫（2018）「日本語史をふたつにわけること」『日本語学』37（13）, 22-31, 明治書院.
- 岡田寿彦（2003）「ルビと漢字」前田富祺・野村雅昭（編）『朝倉漢字講座 3 現代の漢字』, 159-178, 朝倉書店.
- 沖森卓也（2018）「ことばの時代区分とは何か」『日本語学』37（13）, 2-10, 明治書院.
- 小椋秀樹・小磯花絵・富士池優美・宮内佐夜香・小西光・原裕（2011）「『現代日本語書き言葉均衡コーパス』形態論情報規程集」第 4 版, 国立国語研究所.
- 尾山慎（2020）「文字・表記（現代・理論）」『日本語の研究』16（2）, 77-83, 日本語学会.
- 海保博之（1980）「漢字の含有率」『日本心理学会第 44 回大会発表論文集』, 369, 日本心理学会.
- 海保博之（1983）「漢字の昨今と将来」海保博之・野村幸正『漢字情報処理の心理学』, 124-130, 教育出版.
- 梶原滉太郎（1982）「新聞の漢字含有率の変遷—明治・大正・昭和を通じて—」『研究報

- 告集 3 (国立国語研究所報告 71)], 209-227, 秀英出版.
- 梶原滉太郎 (2018) 「新聞」日本語学会 (編) 『日本語学大辞典』, 547-548, 東京堂出版.
- 「角川日本地名大辞典」編纂委員会 (編) (1978~1990) 『角川日本地名大辞典』, 角川書店.
- 樺島忠夫 (1955) 「類別した品詞の比率に見られる規則性」『国語国文』 24 (6), 55-57, 京都大學國文學會.
- 樺島忠夫 (1963) 「漢語をめぐって」『計量国語学』 (27), 14-19, 計量国語学会.
- 樺島忠夫 (1975) 「文字体系の構造」『計量国語学』 (75), 13-22, 計量国語学会.
- 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.
- 菅野倫匡 (2017) 「漢字は無くなるのか—再び「漢字の将来」を問い直す—」『計量国語学』 30 (8), 481-498, 計量国語学会.
- 菊地恵太 (2020) 「漢字字体史研究における字体・字形・書体及び字種の捉え方」『国語学研究』 (59), 459-472, 東北大学大学院文学研究科「国語学研究」刊行会.
- 金明哲・樺島忠夫・村上征勝 (1993) 「手書きとワープロによる文章の計量分析」『計量国語学』 19 (3), 133-145, 計量国語学会.
- 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.
- 国立国語研究所 (1962) 『現代雑誌九十種の用語用字—総記および語彙表— (国立国語研究所報告 21)』 第 1 分冊, 秀英出版.
- 国立国語研究所 (1987) 『雑誌用語の変遷 (国立国語研究所報告 89)』, 秀英出版.
- 小林一仁 (2002) 「新聞「文章」に見る読みやすさの研究—一文の長さ・漢字含有率—」『桜美林国際学論集 Magis』 (7), 1-13, 桜美林大学大学院国際学研究科.
- 小林雄一郎 (2019) 『ことばのデータサイエンス』, 朝倉書店.
- 小松英雄 (1999) 「プロローグ」小松英雄『日本語はなぜ変化するか—母語としての日本語の歴史—』, 1-17, 笠間書院.
- 小柳智一 (2018) 「言語の歴史、言語変化、その記述」小柳智一『文法変化の研究』, 1-19, くろしお出版.
- 近藤明日子 (2019a) 「語種率・品詞率からみる近代文語文の通時的变化」『日本語学論集』 (15), 97(64)-83(78), 東京大学大学院人文社会系研究科国語研究室.
- 近藤明日子 (2019b) 「語種率・品詞率から見る明治・大正期の口語体実用文」『近代語研

- 究 21』, 151-169, 近代語学会.
- 斎賀秀夫 (1960) 「漢字使用の実態」『国立国語研究所年報 11』, 8-29, 国立国語研究所.
- 斎藤秀紀 (1968) 「漢字仮名混り文のエントロピー」『計量国語学』(43), 39-45, 計量国語学会.
- 斎藤達哉 (2011) 「仮名写本における「改行」と「文字使用」—正徹奥書本源氏物語の事例から—」『専修大学人文科学研究所月報』253, 11-29, 専修大学人文科学研究所.
- 斎藤伸治 (2018) 「言語と文字」今井隆・斎藤伸治 (編)『21 世紀の言語学—言語研究の新たな飛躍へ—』, 237-299, ひつじ書房.
- 佐竹秀雄 (1982) 「各種文章の字種比率」『研究報告集 3 (国立国語研究所報告 71)』, 327-346, 秀英出版.
- 佐竹秀雄 (1987a) 「漢字・平仮名・片仮名の比率の関係」水谷静夫教授還暦記念会 (編)『計量国語学と日本語処理—理論と応用—』, 19-27, 秋山書店.
- 佐竹秀雄 (1987b) 「語と表記とのかかわり」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, アクセント史資料研究会.
- 佐藤武義 (1995) 「概説」佐藤武義『概説日本語の歴史』, 1-15, 朝倉書店.
- 真田信治 (2014) 「現代の日本語」佐藤武義・前田富祺 (編)『日本語大事典 (上)』, 714-715, 朝倉書店.
- 柴田里程 (2015) 『データ分析とデータサイエンス』, 近代科学社.
- 島村直己 (1987) 「児童の漢字使用—課題作文の漢字含有率から—」『研究報告集 8 (国立国語研究所報告 90)』, 77-94, 秀英出版.
- 寿岳章子 (1963) 「週刊誌グラビヤの言葉」『計量国語学』(26), 1-6, 計量国語学会.
- 杉戸清樹 (2018) 「現代の日本語」日本語学会 (編)『日本語学大辞典』, 340-343, 東京堂出版.
- 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.
- 杉山昌子・森岡健二 (1991) 「明治・大正・昭和の漢字・漢語の変遷」森岡健二 (編)『改訂近代語の成立—語彙編—』, 357-388, 明治書院. [杉山・森岡 (1969) を再録したもの]
- 鈴木英夫 (1988) 「明治・大正・昭和」金田一春彦・林大・柴田武 (編)『日本語百科大事

- 典』, 104-112, 大修館書店.
- 高田智和 (2009) 「文字の使用量」 計量国語学会 (編) 『計量国語学事典』, 49-51, 朝倉書店.
- 高橋信 (2004) 『マンガでわかる統計学』, オーム社.
- 竹内理 (2014) 「相関分析入門 (1) —関係を探るには—」 竹内理・水本篤 (編) 『外国語教育研究ハンドブック』 改訂版, 121-131, 松柏社.
- 田島優 (1998) 『近代漢字表記語の研究』, 和泉書院.
- 太刀岡勇氣 (2014) 「中古日記文学の計量国語学的分析と異本間の関係性の客観分析—『和泉式部日記』と『更級日記』を題材に—」 『計量国語学』 29 (6), 187-210, 計量国語学会.
- 田中章夫 (1987) 「近代の表記における漢字依存度の変遷」 水谷静夫教授還暦記念会 (編) 『計量国語学と日本語処理—理論と応用—』, 29-47, 秋山書店.
- 田中章夫 (1989) 「漢字依存度の推移」 佐藤喜代治 (編) 『漢字講座 11 漢字と国語問題』, 277-300, 明治書院.
- 田中章夫 (1996a) 「漢字含有率」 佐藤喜代治・遠藤好英・加藤正信・佐藤武義・飛田良文・前田富祺・村上雅孝 (編) 『漢字百科大事典』, 133, 明治書院.
- 田中章夫 (1996b) 「漢字使用率」 佐藤喜代治・遠藤好英・加藤正信・佐藤武義・飛田良文・前田富祺・村上雅孝 (編) 『漢字百科大事典』, 140-141, 明治書院.
- 田中草大 (2020) 「『吾妻鏡』の語彙」 安部清哉 (編) 『シリーズ〈日本語の語彙〉 3 中世の語彙—武士と和漢混淆の時代—』, 26-37, 朝倉書店.
- 田中豊・垂水共之 (1999) 『Windows 版統計解析ハンドブック—ノンパラメトリック法—』, 共立出版.
- 田野村忠温 (2018) 「中国語を表す言語名の諸相—その多様性、淘汰と変質、用法差—」 『待兼山論叢—文化動態論篇—』 (52), 67-102, 大阪大学大学院文学研究科.
- 田原圭子 (1975) 「原文と現代表記版との比較—森鷗外「高瀬舟」の表記調査—」 『言語生活』 (285), 49-55, 筑摩書房.
- 築島裕 (1988) 「日本語史の時代区分」 金田一春彦・林大・柴田武 (編) 『日本語百科大事典』, 59-65, 大修館書店.
- 土屋信一 (1967) 「雑誌「太陽」の用字の変遷」 『言語生活』 (193), 34-43, 筑摩書房.
- 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, 計量国語学会.
- 齋岡昭夫（1991）「ワープロの使用前と使用後—使用文字の計量的研究—」『上越教育大学国語研究』（5）, 1-12, 上越教育大学国語教育学会.
- 中川正美（2016）「文章・文体（史的研究）」『日本語の研究』12（3）, 79-84, 日本語学会.
- 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.
- 中村明（2012）「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』普及版, 1-32, 朝倉書店. [中村（2003）を再録したもの]
- 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, 日本霊長類学会.
- 西尾寅弥（2002）「語種」北原保雄（監）・斎藤倫明（編）『朝倉日本語講座 4 語彙・意味』, 77-109, 朝倉書店.
- 日外アソシエーツ編集部（編）（1998）『苗字 8 万よみかた辞典』, 日外アソシエーツ.
- 日外アソシエーツ編集部（編）（2002）『名前 10 万よみかた辞典』, 日外アソシエーツ.
- 日外アソシエーツ（編）（2004a）『人名よみかた辞典—姓の部—』新訂第 3 版, 日外アソシエーツ.
- 日外アソシエーツ（編）（2004b）『人名よみかた辞典—名の部—』新訂第 3 版, 日外アソシエーツ.
- 日本国語大辞典第二版編集委員会・小学館国語辞典編集部（編）（2000～2002）『日本国語大辞典』第 2 版, 小学館.
- 野村雅昭（1980）「週刊誌の漢字含有率」『計量国語学』12（5）, 215-225, 計量国語学会.
- 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.
- 野村雅昭（2008）「漢字はなくなるか」野村雅昭『漢字の未来』, 195-208, 三元社. [野

- 村 (1988b) を再録したもの]
- 野村雅昭 (2018) 「日本の漢字・Ⅱ」 日本語学会 (編) 『日本語学大辞典』, 172-174, 東京堂出版.
- 野元菊雄 (1989) 「未来社会と漢字」 佐藤喜代治 (編) 『漢字講座 11 漢字と国語問題』, 210-228, 明治書院.
- 林大 (監)・宮島達夫・野村雅昭・江川清・中野洋・真田信治・佐竹秀雄 (編) (1982) 『図説日本語—グラフで見ることばの姿—』, 角川書店.
- 羽山慎亮 (2017) 「近代日本の新聞における漢字制限が記事の文章に与えた影響—『東京日日新聞』『読売報知』の社説の計量調査による考察—」 『日本文化論叢』 17, 23-44, 大韓日本文化學會.
- 半澤幹一 (2007) 「日本語研究法—通時—」 飛田良文・遠藤好英・加藤正信・佐藤武義・蜂矢清人・前田富祺 (編) 『日本語学研究事典』, 32, 明治書院.
- 彦坂佳宣 (1987) 「洒落本の漢字」 佐藤喜代治 (編) 『漢字講座 7 近世の漢字とことば』, 160-179, 明治書院.
- 飛田良文 (1995) 「現代日本の漢字問題」 近代語研究会 (編) 『日本近代語研究 2』, 1-19, ひつじ書房.
- 飛田良文 (2007) 「現代語」 飛田良文・遠藤好英・加藤正信・佐藤武義・蜂矢清人・前田富祺 (編) 『日本語学研究事典』, 517-519, 明治書院.
- 平凡社・平凡社地方資料センター (編) (1979~2004) 『日本歴史地名大系』, 平凡社.
- 前田富祺 (1971) 「仮名文における文字使用について—変体仮名と漢字使用の実態—」 『東北大学教養部紀要』 (14), 99-134, 東北大学教養部.
- 前田富祺 (1985) 「語彙史の時代区分」 前田富祺 『国語語彙史研究』, 826-846, 明治書院.
- 丸山岳彦・柏野和佳子 (2014) 「サンプリング」 前川喜久雄 (監)・山崎誠 (編) 『講座日本語コーパス 2 書き言葉コーパス—設計と構築—』, 22-44, 朝倉書店.
- 水谷静夫 (1960) 「漢字含有率の推定」 『国立国語研究所年報 11』, 29-31, 国立国語研究所.
- 水谷静夫 (1977) 「語彙の量的構造」 大野晋・柴田武 (編) 『岩波講座日本語 9 語彙と意味』, 43-86, 岩波書店.
- 水谷静夫 (1983) 『朝倉日本語新講座 2 語彙』, 朝倉書店.
- 水谷静夫 (1987) 「序説」 水谷静夫 (編) 『朝倉日本語新講座 1 文字・表記と語構成』, 1

-19, 朝倉書店.

宮島達夫 (1988a) 「漢字の将来」その後『言語生活』(436), 50-58, 筑摩書房.

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, 計量国語学会.

宮島達夫 (2019) 「漢字の将来」その後 宮島達夫『言語史の計量的研究』, 444-456, 笠間書院. [宮島 (1988a) を再録したもの]

矢澤真人 (2012) 「新聞から日本語を考えよう—漢字と仮名—」森山卓郎 (編)『日本語・国語の話題 ネター実はずりたかっただ日本語のあれこれ—』, 38-41, ひつじ書房.

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, 筑摩書房.

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.

安本美典 (1966) 「漢字の将来—漢字の余命は、あと二百三十年か—」安本美典『文章心理学の新領域』, 139-174, 誠信書房. [安本 (1963) を改稿したもの]

矢田勉 (2016) 「言語史叙述と文字・表記史叙述—その共通点と相違点—」大木一夫・多門靖容 (編)『日本語史叙述の方法』, 123-144, ひつじ書房.

矢野準 (1987) 「人情本の漢字」佐藤喜代治 (編)『漢字講座 7 近世の漢字とことば』, 199-218, 明治書院.

山崎誠 (2010) 「日本語コーパス研究と計量手法」石川慎一郎・前田忠彦・山崎誠 (編)『言語研究のための統計入門』, 269-273, くろしお出版.

山崎誠 (2013) 「語彙調査の系譜とコーパス」前川喜久雄 (監・編)『講座日本語コーパス 1 コーパス入門』, 134-158, 朝倉書店.

山田俊雄・小林芳規・築島裕・白藤礼幸 (1995)『新潮国語辞典—現代語・古語—』第2版, 新潮社.

山田俊雄・白藤礼幸・築島裕・奥田勲 (2000)『新潮現代国語辞典』第2版, 新潮社.

横山詔一 (2009) 「漢字の使用量」計量国語学会 (編)『計量国語学事典』, 56-58, 朝倉書店.

吉田金彦 (1989) 「日本語史」加藤彰彦・佐治圭三・森田良行 (編)『日本語概説』, 281-306, 桜楓社.

吉田武尚・福田誠・樫岡健史 (1995) 「教科書表現の変遷—中学校技術教科書—」『奈良教育大学教育研究所紀要』(31), 39-50, 奈良教育大学教育研究所.

吉田東伍（1969～1971）『大日本地名辭書』増補版，富山房.

渡部洋・大塚雄作・鈴木規夫・山田文康（1984）「行動科学データ解析のための探索的方法」『行動計量学』12（1），59-80，日本行動計量学会.

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.

本論文の構成と既発表論文との関係

序章

書き下ろし

第1章 先行研究の概観

書き下ろし

第2章 用語の定義

書き下ろし

第3章 小説の漢字含有率は減少し続けるのか

菅野倫匡（2017）「漢字は無くなるのか—再び「漢字の将来」を問い直す—」『計量国語学』30（8），481-498，計量国語学会.

第4章 小説の漢字含有率はなぜ安定したのか

菅野倫匡（2021 予定）「語種の観点から見る漢字含有率の安定要因—芥川賞作品の分析を通して—」『計量国語学』32（8），計量国語学会.

第5章 新聞の漢字含有率は如何なる傾向にあるのか

菅野倫匡（2019）「新聞における漢字含有率」『計量国語学会第六十三回大会予稿集』，36-41，計量国語学会.

第6章 漢字含有率に見る現在の漢字使用の実態

書き下ろし

終章

書き下ろし