様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

令和　６ 年　 ６ 月 １７ 日現在

機関番号：１２１０２
研究種目：研究活動スタート支援
研究期間：2022～2023
課題番号：２２Ｋ１９９８５

研究課題名（和文）日本語の文章における漢字使用の実態を解明するための近現代小説コーパスの構築

研究課題名（英文）Building a Corpus of Modern Fiction to Reveal the Usage of Chinese Characters in Japanese Sentences

研究代表者
菅野　倫匡（KANNO, Michimasa）

筑波大学・人文社会系・助教

研究者番号：３０９６１６５１

交付決定額（研究期間全体）：（直接経費）　　2,200,000円

研究成果の概要（和文）：本研究は日本語の文章における漢字使用の実態を解明するための近現代小説コーパスとして芥川龍之介賞受賞作品（芥川賞作品）のコーパスを構築することを目的としたものである。その主要な成果は21世紀以降の新しい作品を中心としてコーパスを試作したことである。また，コーパス構築に当たっては語彙調査に関する未解決の問題として知られているように読み方の定まらない語が問題となることから既に提示した暫定的な読み方を定める方法について本研究では同時代の資料に幅広く適用し得る汎用的な手順として再整理した。更に本研究では芥川賞作品のコーパス構築の方法について古典語のコーパス構築の事例に適用し，その汎用性の高さを実証した。

研究成果の学術的意義や社会的意義
本研究は現代日本語の文章における漢字使用の実態やその変遷を解明することに資するコーパスが新聞や雑誌を対象としたものに限られる現状にあって従来の研究を継承しながら小説を対象とした新たなコーパスを実際に構築することを試みたものである。そのためにコーパス構築の方法を検討し，古典語のコーパスを実際に構築する事例に適用することにより，その具体的な手順の汎用性の高さを実証的に示した。これはコーパスという言語資源の構築に関する方法論の整備に寄与するものである。また，日本語の表記や語彙の研究に資する近現代小説コーパスを構築することは将来の改定を見込む「常用漢字表」などの国語施策を議論する一助となるものである。

研究成果の概要（英文）：This study aimed to build a corpus of modern fiction based on Akutagawa Prize-winning works to reveal the usage of Chinese characters in Japanese sentences. The main contribution of this study is building the prototype corpus, which consists mainly of works from the 21st century. During the preparatory phase of this study, a method of annotation for heteronyms in corpora was proposed, addressing what is often considered an unresolved issue in vocabulary surveys. This study perfected the method, enabling its application to a wide range of contemporary language materials. Furthermore, the study developed a method for building a corpus, which specifically focused on Akutagawa Prize-winning works. The applicability of this method was further confirmed by its successful application to a corpus of classic works.

研究分野：日本語学

キーワード：表記　漢字使用の実態　芥川賞作品　コーパス

１．研究開始当初の背景

近現代日本語の表記や語彙については漢字や漢語に強い関心が寄せられ，多くの研究の蓄積がある。特に計量的な観点に限れば，文章に占める漢字の割合が次第に減少するものの 20 世紀中頃に安定へと転じたことが一連の研究によって実証的に示されてきた。また，そのような安定については語の出自（和語・漢語・外来語）やその表記（平仮名・片仮名・漢字）の構成比率の推移に大きな変化が見られなくなったことに起因することが明らかになってきた。

一方，漢字使用の実態を解明するためには総体としての漢字の使われ方を調査する巨視的な研究に加え，個々の漢字の使われ方を調査する微視的な研究も必要である。そのためには本文を電子化する際に表記に与える影響を考慮しつつ表記や語彙の研究に資するコーパスを構築することが求められる。

また，そのような研究を可能にする通時コーパスは非公開のものを除くと存在せず，新聞などを対象としたコーパス構築が進められつつあることが知られていたが，小説を対象としたものは皆無である。しかし，漢字使用の実態は新聞におけるそれと小説におけるそれとの間に相違が見られることも明らかになっており，漢字使用の実態を解明するためには新聞に加えて小説を対象とした通時コーパスが不可欠な状況にある。

なお，研究開始当初は近現代日本語の通時コーパスが存在していないに等しい状況にあったことから 2023 年 3 月に新聞などを対象とした『昭和・平成書き言葉コーパス』の公開に伴って状況は変わりつつあるが，小説を対象とした新たな通時コーパスが求められる状況は変わっていないと言える。

２．研究の目的

本研究の目的は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するために日本語の語彙や表記の研究に資する近現代小説コーパスとして芥川龍之介賞受賞作品（芥川賞作品）のコーパスを構築することである。

３．研究の方法

コーパスを構築するに当たっては対象とする本文を電子化し，それを各語に切り分けて品詞や語の出自などの情報を認定し，その情報を用いて各語を検索し得るコーパスとして整備するという段階を踏むこととなる。

特に各語に切り分けて品詞などの情報を認定する際には国立国語研究所のコーパスに倣って「短単位」に準拠し，既存のコーパスとの互換性を担保するが，語彙調査の未解決の問題として知られているように読み方の定まらない語については暫定的な読み方を認定する客観的で再現可能な手順を考案することとする。

また，既存のコーパス構築が大規模な研究機関によって進められてきた現状を踏まえ，互換性を担保しつつ独自にコーパスを構築する方法論についても実際のコーパス構築を通して整備を試みる。

なお，本文を電子化する段階では光学文字認識（OCR）ソフトウェアの誤りを人手で修正する作業を実施し，各語に切り分けて品詞などの情報を認定する段階でも形態素解析の誤りを人手で修正する作業を実施することにより，解析精度の評価や向上にも取り組むこととする。両段階において人手で修正する作業を進めるに当たっては国立国語研究所のマニュアルなどを参考に具体的な手順を作業用のマニュアルとして定めた上で作業補助者と事前に共有し，全体として一貫した処理になるように留意する。

４．研究成果

本研究の成果は（1）～（3）の 3 点である。それぞれの点について以下に概要を述べる。

(1) 21 世紀以降の作品を中心とした芥川賞作品コーパスの試作

本研究の主要な成果は芥川賞作品のコーパス構築に取り組み，特に 21 世紀以降の新しい作品を中心としてコーパスを実際に試作したことである。全作品について各語に認定する品詞などの情報を人手で修正する作業を完遂することは研究期間が短いことから不可能であるが，一部の作品について作業を完遂し得たことは今後のコーパス構築の一助となるものである。

(2) 読み方の定まらない語に暫定的な読み方を定める方法の提示

コーパスを構築するに当たっては読み方の定まらない語が問題となることから本研究を実施するための準備の一環として暫定的な読み方を定める方法を考案した。しかし，この方法は小説に現れる固有名詞に重点を置いたものであり，普通名詞などを考慮すると更なる精緻化の余地が残されていた。これを踏まえ，この方法について本研究では同時代の資料に幅広く適用し得る汎用的な手順として再整理した。（菅野倫匡（2023）「近現代語のコーパスを構築する際の「同字異訓」の問題に関する覚え書き」『筑波日本語研究』第 27 号，2023 年 2 月）

(3) 独自にコーパスを構築する方法論の整備

既存のコーパスは国立国語研究所などの大規模な研究機関が開発・構築を牽引してきたものであり，コーパスの公開と共にコーパスを構築する技術の公開も進められてきたが，その技術を用いて実際にコーパスを構築することは十分に普及したものとは言えない状況にある。これを踏まえ，本研究では国立国語研究所の開発した解析用辞書を用いて既存のコーパスとの互換性を担保しつつ芥川賞作品のコーパスを構築する作業の手順の簡略化を図り，コーパスを独自に構築する方法を考案した。また，古典語のコーパスを構築する事例に適用し，その汎用性の高さを実証的に示した。(菊池そのみ・菅野倫匡（2023)「中古和文資料『夜の寝覚』のコーパス構築の試み」言語資源ワークショップ 2023，2023 年 8 月)

5．主な発表論文等

〔雑誌論文〕　計1件（うち査読付論文　1件／うち国際共著　0件／うちオープンアクセス　1件）

1．著者名 菅野倫匡	4．巻 27
2．論文標題 近現代語のコーパスを構築する際の「同字異訓」の問題に関する覚え書き	5．発行年 2023年
3．雑誌名 筑波日本語研究	6．最初と最後の頁 1-34
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 －

〔学会発表〕　計1件（うち招待講演　0件／うち国際学会　0件）

1．発表者名 菊池そのみ・菅野倫匡
2．発表標題 中古和文資料『夜の寝覚』のコーパス構築の試み
3．学会等名 言語資源ワークショップ2023
4．発表年 2023年

〔図書〕　計0件

〔産業財産権〕

〔その他〕

－

6．研究組織

	氏名 （ローマ字氏名） （研究者番号）	所属研究機関・部局・職 （機関番号）	備考

7．科研費を使用して開催した国際研究集会

〔国際研究集会〕　計0件

8．本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関