国立国語研究所学術情報リポジトリ

Possibility of corpus stylistics with scenario corpora

メタデータ	言語: jpn
	出版者:
	公開日: 2019-02-14
	キーワード (Ja):
	キーワード (En):
	作成者: 松下, 晶子, 丸山, 岳彦, MATSUSHITA, Shoko
	メールアドレス:
	所属:
URL	https://doi.org/10.15084/00001659

脚本テキストに基づくコーパス文体論の可能性
―テレビドラマ脚本に注目して―

松下 晶子（専修大学大学院文学研究科）*
丸山 岳彦（専修大学文学部 / 国立国語研究所）

Possibility of corpus stylistics with scenario corpora

Shoko Matsushita (Graduate School of Letters, Senshu University)
Takehiko Maruyama (Senshu University / NINJAL)

要旨

現在、「日本脚本アーカイブズ推進コンソーシアム」により、1950 年代以降のテレビドラマの脚本を収集し、それらを体系的に保存・アーカイブ化する活動が進められている。脚本は、「話されることを前提とした書き言葉」という点で特徴的な書き言葉であるが、これまでの言語研究の中で顧みられることは少なかった。収集した脚本をコーパス化して定量的に分析することにより、新たな言語学的利用の可能性が開かれると考えられる。そこで本発表では、脚本のテキスト化・コーパス化を試験的に実施した経緯を述べ、そのデータを使ってどのような言語研究が可能になるかについて論じる。故市川森一氏による、1970 年代から 2010 年代までの脚本、32 作品をテキスト化し、パイロットスタディを実施した。このような分析は、近現代における言語の短期的な変化の研究、ある作家の作品に関するコーパス文体論的研究などにつながると考えられる。

1. はじめに

一口に「書き言葉」と言っても、その内実は一様ではない。実際の書き言葉は、ある特定のレジスター（言語使用域）の中で実現するものであり、レジスターの違いに応じて、多種多様な様式・スタイルを持つ。すなわち、書き言葉には、レジスターごとに多様な変異が存在する。

2011 年に公開された『現代日本語書き言葉均衡コーパス』（BCCWJ）は、書籍・雑誌・新聞・白書・教科書・広報紙・Yahoo!知恵袋・Yahoo!ブログ・韻文・法律・国会会議録という 11 種類のメディアから抽出した書き言葉のサンプルを収録しており、公開以来、メディアの違いに応じた言語変種（バラエティ）の分析が盛んに進められている。一方、BCCWJ に収録されなかったタイプの書き言葉も、多く存在する。本稿で取り上げる「脚本」も、その一つである。脚本は、これまで言語研究の中で分析対象データとして扱われることの少なかったメディアであり、その言語的な特徴の分析は、遠藤ほか (2014) を除き、ほとんど進んでいない。「話されることを前提とした書き言葉」という点で、脚本はユニークな性格を持つ書き言葉であるが、では、「脚本コーパス」を作った場合、言語分析の観点から、どのように利用できるであろうか。

本稿では、脚本をコーパスとして利用することにより、どのような言語学的分析が可能かに

* arukumatsushita@gmail.com

ついて論じる。現在、「日本脚本アーカイブズ推進コンソーシアム」によって実施されているテレビドラマの脚本をアーカイブ化する活動を紹介し、その中で筆者らが進めている脚本のテキスト化・コーパス化の経緯について述べる。さらに、そのデータを用いてどのような言語研究が可能になるかについて、パイロットスタディの結果を交えながら論じる。そこから、近現代における言語の短期的な変化の研究や、ある脚本家の作品を対象とした「コーパス文体論」の研究などに展開する可能性について指摘する。

2. 「日本脚本アーカイブズ推進コンソーシアム」の活動

一般社団法人「日本脚本アーカイブズ推進コンソーシアム（NKAC）」(1) は、2012 年 6 月に設立されて以降、テレビ・ラジオ放送（特にテレビドラマ）の脚本を収集し、アーカイブとして体系的に保存する活動を継続している。テレビドラマの脚本は出版物ではないため、図書館で収集・保存されることは原則的になく、制作後は放送局や番組関係者によって「保管」されることが一般的である。このため、体系的な収集・管理ができず、古い時代のものは散逸して失われてしまう可能性が高い。その一方、特に 1980 年代以前のテレビ放送は録画テープを重ね録りして使用していたため、当時の映像・音声自体が放送局に残っていないことも多い。

過去のテレビドラマの詳細を知る手掛かりとして、脚本を収集してアーカイブ化すれば、日本のテレビ放送史における放送文化研究やドラマ研究の分析対象データとして役立てることができる。また、脚本として書かれた日本語が、テレビ放送の歴史の中でどのように変化してきたのかを探るための言語資料として考えれば、言語学的な研究にとっても極めて有用であろう。

NKAC の活動により、2017 年までに、1930 年代から 2010 年代までの脚本、約 8 万点が集められた。これらは現在、国立国会図書館や川崎市民ミュージアムなどに分置されている。また、国立国会図書館に納められた脚本約 27,000 点のうち、3,000 点を超える分がデジタル化され、全国の図書館に配信されている。さらに、収集された脚本はデータベースに登録され、「脚本データベース(2) 」として、ウェブ上で公開されている。ここでは、ドラマタイトル、作家、放送局、放送日、キャストなどの情報を検索することができる(3)。なお、このデータベース化事業は、文化庁委託事業「文化関係資料のアーカイブ構築に関する調査研究～放送番組の脚本・台本のアーカイブ構築に向けた調査研究～」などによる成果である。

また、市川森一、永六輔、藤本義一といった脚本家については、その脚本作品のリストや本文（の画像）、脚本家の足跡や関係者のインタビューなどが、特設サイト上にまとめられている(4) (5) (6)。

2017 年度からは、科研費基盤 (B)「「脚本クロニクル」サイト構築とその教育活用および国際発信」もスタートし、その研究・教育への利用の模索も開始された。現在は、脚本の収集・データベース化と、その研究利用という両面から、活動を継続している。

---

(1) 「日本脚本アーカイブズ推進コンソーシアム」https://www.nkac.jp/
(2) 「脚本・台本の総合一覧 脚本データベース」http://db.nkac.or.jp/
(3) ただし、脚本の本文にあたる部分の閲覧・検索はできない。
(4) 「市川森一の世界」http://ichikawa.nkac.or.jp/
(5) 「永六輔バーチャル記念館」http://eirokusuke.nkac.or.jp/
(6) 「藤本義一アーカイブ」http://fujimotogiichi.nkac.or.jp/

3. 「市川森一 脚本コーパス」の構築

現在、科研費基盤 (B)「「脚本クロニクル」サイト構築とその教育活用および国際発信」における活動の一環として、「脚本」をテキストデータ化し、脚本の言語学的利用という観点からその利用可能性を探っている。以下では、これまでにパイロット的に作成している「市川森一脚本コーパス」について述べる。

故市川森一氏（1941～2011）は、かつて日本放送作家協会理事を務め、後の NKAC の発足に尽力した脚本家である。1960 年代に『快獣ブースカ』や『ウルトラマン A』などの子ども番組で活動を開始し、1970 年代以降は『傷だらけの天使』『黄金の日日』『花の乱』『港町純情シネマ』『淋しいのはお前だけじゃない』『異人たちとの夏』『幽婚』など、数々のテレビドラマ作品を手掛け、多くの受賞作品を生み出した。2003 年には国会の総務委員会において脚本アーカイブズ活動の必要性を提言し、その活動が、2012 年の NAKC の設立につながっている。

NKAC では、「デジタル脚本アーカイブズのトライアル」として、市川森一氏によるドラマ脚本やその関連書類（手書き原稿、創作ノート）をデジタル化し、「市川森一の世界」としてウェブ上に公開している。この延長線上にある活動として、今回、市川森一による脚本をテキスト化し、「市川森一 脚本コーパス」を作成した（今のところ、一般公開の予定はない）。

今回、コーパス化したのは、以下の 3 冊に含まれる脚本のテキストである。

1. 『市川森一 センチメンタルドラマ集』（1983 年、映人社） 14 作品収録
2. 『市川森一 メランコリックドラマ集』（1986 年、映人社） 10 作品収録
3. 『市川森一 メメント・モリドラマ集』（2012 年、映人社） 8 作品収録

全ページを OCR で読み込み、人手で校正したうえで、テキストファイルを作成し、さらに Mecab + unidic-cwj-2.3.0 で形態素解析を実施した。また、セリフとト書きの部分を区別するタグを付与して、両者を区別できるようにした。

次に、3 冊に含まれる脚本（32 作品）を放送年の順に並べ、全体を以下のように区分した。それぞれの総語数（句読点含む）と、各期に含まれる作品タイトルを、以下に示す。

第 1 期： 1969 年～1979 年 放送 （13 作品、165,042 語）

「仮面の墓場」「祭りのあとに」「林で書いた詩」「冬の時刻表」「夢に吹く風」「みどりもふかき」「紙コップのコーヒー」「夢のながれ」「霧の日の童話」「幻のぶどう園」「バースデイ・カード」「ラスト・ダンス」「露玉の首飾り」

第 2 期： 1980 年～1989 年 放送 （11 作品、137,411 語）

「春のささやき」「チャップリン暗殺計画」「蝶の鼓」「いもうと」「夢の指環」「鬼の恋舟」「受胎の森」「星の旅人たち」「途中下車」「ただ一度の人生」「中国服の女」

第 3 期： 1998 年～2011 年 放送 （8 作品、122,329 語）

「幽婚」「ここではない何処か」「乳房」「風の盆から」「銀河鉄道に乗って」「月の光」「旅する夫婦」「蝶々さん」

1 作品に含まれる語数は、最小で 7,780 語、最大で 39,086 語、平均で 13,274 語であった。

4. 分析

以下では、パイロットスタディとして、「市川森一 脚本コーパス」を分析した結果について示す。分析の観点は、(1) 終助詞の分布、(2) 二人称代名詞の使用、という 2 点である。

4.1 終助詞の分布

はじめに、脚本コーパスの中に現れた終助詞の分布について分析を行なう。

まず、表 1 は、第 1 期～第 3 期の各期に現れた終助詞の総数と出現率をまとめたものである。この表からは、時代を追うごとに、終助詞の出現率が減っていることが読み取れる。

表 1　各期における終助詞の総数と出現率

	総語数	総終助詞数	終助詞率
第 1 期	165,042	4,014	2.43%
第 2 期	137,411	2,776	2.02%
第 3 期	122,329	2,142	1.75%

次に、図 1 は、各期の終助詞のうち上位 5 つを抽出し、終助詞の総数に占める割合をグラフにしたものである。

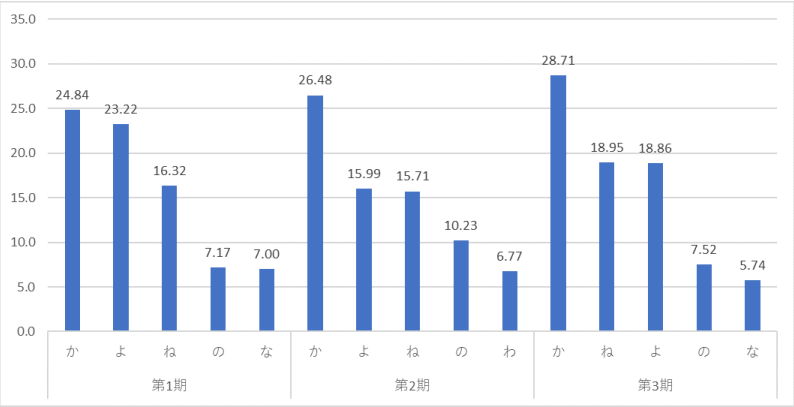

図 1　各期の終助詞（上位 5 位）の総終助詞数に占める比率

先にも述べた通り、時代を追うごとに終助詞の数は減少しているが、出現している終助詞の形式は、期によって異なる。どの期でも一番多いのは、「か」であるが、2 位と 3 位に注目すると、第 1 期では「よ」「ね」の登場回数は差が開いているが、第 2 期になるとその差が縮まり、第 3 期になるとわずかながら「ね」が「よ」を上回っている。

次に、終助詞の分布を「登場人物の性別」という観点から見てみよう。なお、今回の調査では、発話数が上位 5 位の中に入り、かつ 20 以上の発話をしている登場人物のセリフに限定して分析している。

図 2 は、女性登場人物が用いる終助詞の上位 5 つを抽出し、総終助詞数に対する比率をグラフにしたものである。ここではどの期でも登場する終助詞の種類は一緒だが、その順位が違う

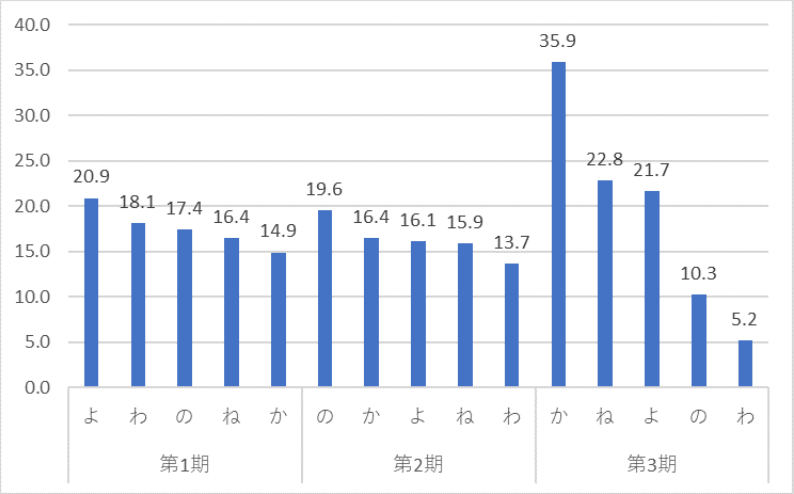

図 2　総終助詞数に対する各期の終助詞（上位 5 位）の比率（女性）

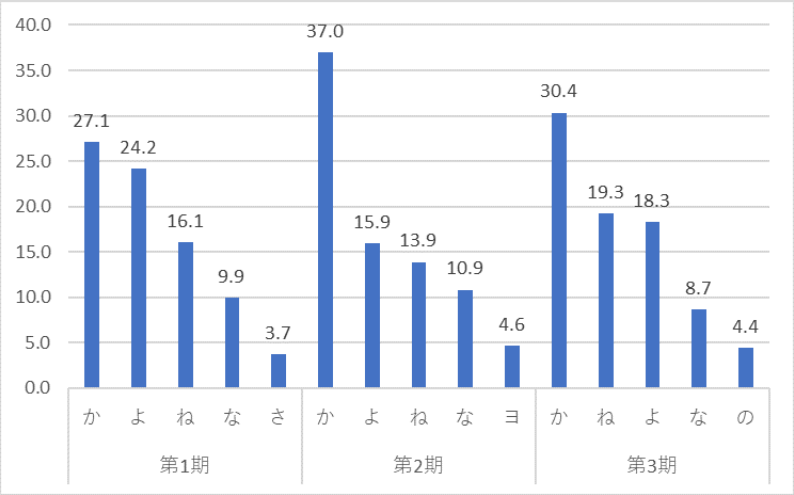

図 3　総終助詞数に対する各期の終助詞（上位 5 位）の比率（男性）

点が特徴的である。その中でも、特に女性らしい終助詞とされている「わ」は、第 2 期からその順位を下げ、上位 5 位に入ってはいるものの、その割合は大きく減少している。以下、各期の女性登場人物が用いる終助詞「わ」を含む文を挙げる。

(1)　「判ったわ。わざわざどうも（と席へ）」（第 1 期）

(2)　「懐しかったわ。ちっとも変わってなくて」（第 2 期）

(3)　「？……まだ、お義父さまにもご挨拶してないわ。」（第 3 期）

「女性らしさ」を表す終助詞として「わ」が用いられていることは、どの期でも変わっていない。しかし、割合が減少しているという事実から考えられることとして、女性らしさとして「わ」を用いる割合が減っていったというだけでなく、「女性」を象徴する役割語として「わ」を用いる割合も減っていった、という点が挙げられるのではないだろうか。この現象の要因は、当時の時代背景によるものなのか、脚本家自身考え方によるものなのか、現段階では分からない。今後、既存のコーパスと照らし合わせたり、脚本家自身の時代ごとの境遇を照合したりと

いった作業が必要である。

一方、図 3 は男性の登場人物が用いる終助詞の割合である。男性は、第 3 期で「ね」と「よ」が逆転したこと、どの期でも 5 位に異なる終助詞が挙げられている点が特徴的である。ここでは、第 1 期の男性の登場人物が用いた、終助詞「さ」を含む例を見ておこう。

(4) 「テニヲハが抜けてるンだ。ウエルニッケのいわゆる言語促迫。クルーベリン学派の術語では、支離滅裂というヤツさ」

(5) 「女を喰いもんにしていた二枚目の芋学生さ」

ここから、発話者の気取った発言に対し終助詞「さ」が用いられている印象を受ける。特に第 1 期には気取った発言をする性格の男性が多く登場する傾向があり、このために「さ」が上位に位置づけられたと考えられる。

なお、終助詞がカタカナで表記される場合が目についた。その比率を、表 2 に挙げる。

表 2　終助詞がカタカナで表記される割合

	第 1 期	第 2 期	第 3 期
ヨ	5.8% (57)	22.6% (130)	0.5% (3)
ネ	7.0% (49)	5.4% (25)	5.1% (22)
ワ	1.8% (4)	19.7% (46)	2.7% (2)
ナ	4.1% (12)	10.9% (20)	0.0% (0)

（カッコ内はひらがな終助詞とカタカナ終助詞の総数）

表 2 は、ひらがな終助詞とカタカナ終助詞の総数に対する、各期のカタカナ終助詞率をまとめたものである。ここでは、ひらがな終助詞の上位と重なっている「ヨ」「ネ」「ワ」「ナ」を抽出した。これを、1 万語あたりの調整頻度としてグラフ化すると、図 4 になる。

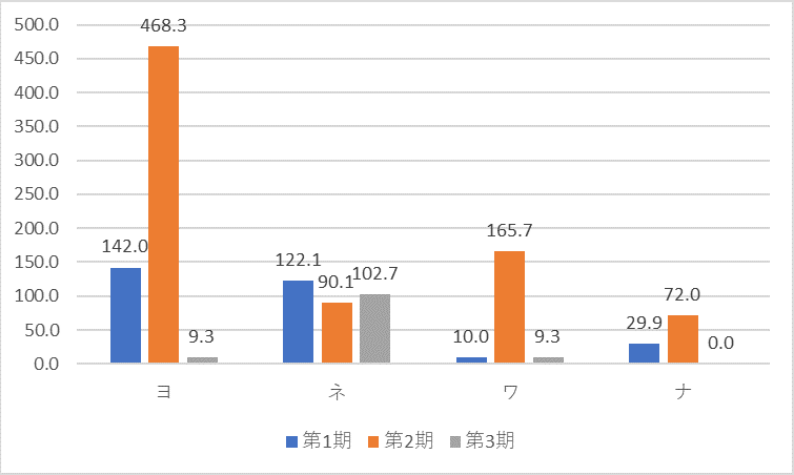

図 4　カタカナで表記された終助詞の頻度（1 万語あたり）

「ネ」「ナ」には目立った差はないが、「ヨ」「ワ」に関しては期によってかなりの差があることが、グラフから読み取れる。以下、カタカナで終助詞が用いられている例を挙げる。

(6) 「A さん、なに訊かれたのヨ」

(6) は女性の発話であるが、ひらがなの「よ」に比べて、上昇調のようなニュアンスが加わり、単なる「質問」に加えて、からかいや冷やかしのような意味合いが感じられる。

(7) 「全て、銭次第ですヨ」

(7) は男性の発話であるが、ひらがなの「よ」の場合に比べて、ここでは悪だくみや誘惑といったマイナスの言い含みを感じやすくなるように思われる。

このように、カタカナの終助詞を用いることで、ひらがなを用いる時とは印象が変わり、発話者の心情やニュアンスを豊富に表現する効果が感じられる。その他の種類の終助詞について、その表記の違いと用法・効果について分析することは、今後の課題の一つである。

4.2　二人称代名詞の使用

以下では、二人称代名詞の「あなた」「あんた」について分析してみよう。図 5 は、各期における「あなた」「あんた」の出現数を、1 万語あたりの調整頻度としてグラフ化したものである。

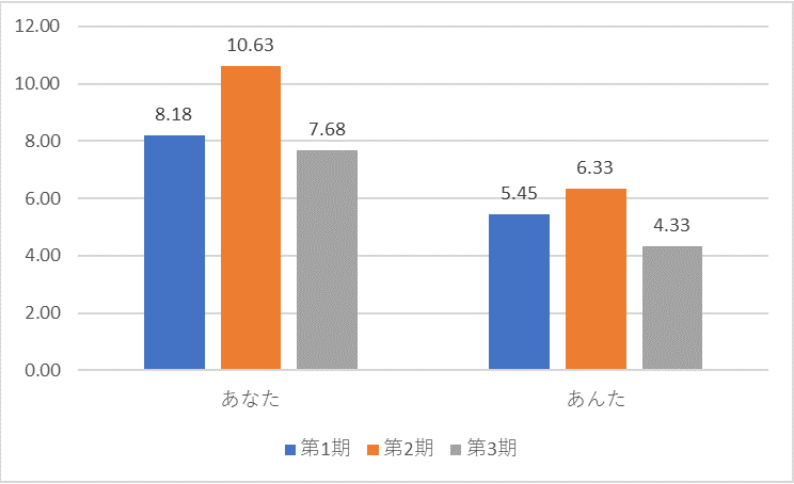

図 5　各期における「あなた」「あんた」の出現数（1 万語あたり）

第 1 期での「あなた」「あんた」の数が、それぞれが第 2 期で増えて、第 3 期で第 1 期を少し下回る数になっている。これには、何か要因があるのだろうか。

ここで、「あなた」「あんた」の出現数（粗頻度）を、登場人物の性別という観点から調査してみよう。ここでも、終助詞の場合と同様に、発話数の合計が各作品で上位 5 位までの人物で、かつ 20 以上の発話をしている人物に絞ったデータを使用する。

表 3　性別ごとの「あなた」「あんた」の出現数

	あなた		あんた	
	女性	男性	女性	男性
第 1 期	55	58	10	69
第 2 期	66	64	46	15
第 3 期	51	15	22	27

図 6 で特徴的なのは、それぞれ男女内でグラフが逆の形をしていることである。すなわち、

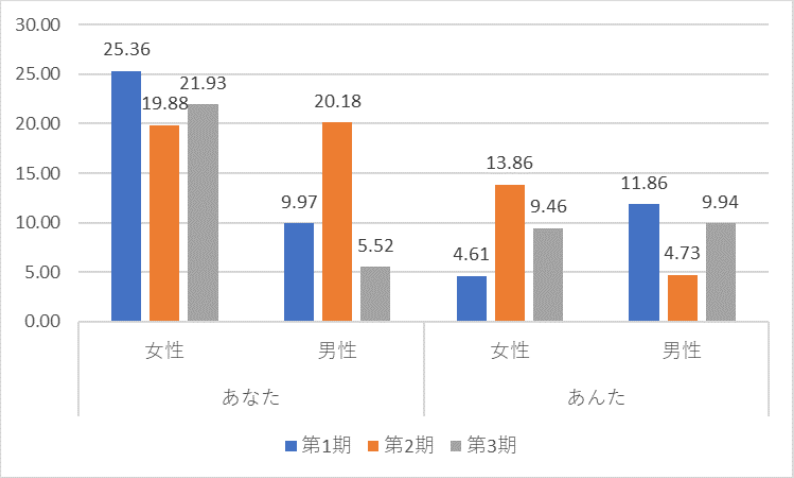

図 6　性別ごとの「あなた」「あんた」の出現数（1 万語あたり）

「あなた」と「あんた」の使用において、男女間で相補的な関係になっていることが分かる。この点については、今後、登場人物の男女それぞれがどのような相手に向かって「あなた」「あんた」を使っているのか、「あなた」「あんた」それぞれを使う登場人物がどのような性格なのかといった、具体的な場面分析を合わせて行なうことで、その要因を分析することができると考えられる。

5.　考察：「脚本」が言語研究に対して持つ可能性

ここまでの分析結果を受けて、脚本コーパスを言語研究に用いることの意義と可能性について、若干の考察を加えておく。脚本コーパスを言語研究に用いることの意義として、以下では 2 点に分けて整理しておこう。

1 点目は、「話すことを前提として書かれた言葉」が持つ言語的特徴を明らかにすることである。脚本で書かれたセリフの部分は、話すことを前提として書かれた書き言葉である。これは、実際の話し言葉とは大きく異なる（例えば、脚本には非流暢性がほとんど現れない）。では、同じ「書かれた話し言葉」である「小説の会話部分」に現れる言葉とは、同じだろうか、違うだろうか。

脚本には、例えば以下のように、「…」のみで構成されるセリフが多く見られる。

(8) 骨の折れる音が、場内に反響する。棒立ちのまま、息をのむ、犬尾、山口、ヨーコ。
静寂……。
犬尾「……」
ヨーコ「……」
山口「……」
ジュン「……」

これらは登場人物の沈黙を表すものであるが、これほど多くの「…」の連続は、小説にはまず出現しない。このような表現は、映像（沈黙する登場人物たち）を基盤とする脚本ならではのものであろう。今後は、BCCWJ の小説に現れる発話部分を取り出し、脚本のセリフと定量

的に比較・分析することで、脚本の言語的特徴を明らかにすることが課題の一つとなる。

2 点目は、個人の脚本家の作品群を集めることで、「コーパス文体論」の研究への道が拓かれるのではないか、という点である。1 人の文学作家の手による作品群を集めて、例えば「夏目漱石コーパス」や「太宰治コーパス」を作成し、両者の文体的な違いや時代ごとの特徴を定量的に比較・分析していく、という研究手法は、今後、大きな可能性を持つと考えられる。Stubbs(2014) はコンピュータを用いた文学テクストの量的分析の可能性について論じているが、これはすなわち、「コーパス文体論」の研究の方向性を示唆するものである。英語を対象としたコーパス言語学では、そのような試みがすでにいくつも発表されており（Studer 2012, Mahlberg 2015 など）、今後は日本語でもコーパス文体論的な研究が望まれる。

脚本を分析対象とする場合にも、同じことが言えると思われる。複数の脚本家による脚本を時代ごとに収集してコーパス化することにより、個人の作風の変化を探ったり、現代語の短期的な変化の影響を分析したり、異なる脚本家と文体を比較したりするような研究が可能になるだろう。今回パイロット的に作成した「市川森一 脚本コーパス」はそのような方向に向かう取り組みであり、「脚本コーパスに基づく文体論」につながるものと考えている。

6. おわりに

以上、本稿では、脚本のテキスト化・コーパス化を試験的に実施した経緯を述べ、そのデータを使ってどのような言語研究が可能になるかについて示した。「市川森一氏 脚本コーパス」を試作し、そのデータを用いたパイロットスタディとして、終助詞と二人称代名詞「あなた」「あんた」の分布について見た。

終助詞の調査では、性別や年代、表記といった観点からその分布を見た。特に、カタカナで表記される終助詞は、それが用いられている状況やそれを用いた登場人物の性格といった面から、カタカナ終助詞が持つ役割を分類できそうである。それらをひらがな終助詞と比較することにより、脚本の文体的な特徴を示す一例として考えることができると思われる。

また、二人称代名詞「あなた」「あんた」の調査では、第 2 期が特徴的な動きをしていることが判明した。これには、時代が影響しているのか、脚本家が置かれた当時の状況が影響しているのか、考えられる要因は多くある。今後、既存の書き言葉コーパス・話し言葉コーパスと照合しながら調査を進めることで、データに裏打ちされた脚本文体論の研究を進めることができるだろう。

脚本コーパスの量をさらに増やし、定量的な分析を実施する環境整備を行なうこともまた、今後の課題である。これにより、コーパスに基づく文体論研究の一例として、より精度を増した形で、脚本の言語学的な分析を進めたい。

謝　辞

本研究は、科研費基盤 (B)「「脚本クロニクル」サイト構築とその教育活用および国際発信」（17H02598）、科研費基盤 (B)「「昭和話し言葉コーパス」の構築による話し言葉の経年変化に関する実証的研究」（16H03426）、および国立国語研究所共同研究プロジェクト「大規模日常会話コーパスに基づく話し言葉の多角的研究」によるものである。

文　献

Mahlberg, Michaela (2015) *Corpus Stylistics and Dickens's Fiction.* Oxford: Routledge.

Stubbs, Michael（2014）Quantitative methods in literary linguistics. Peter Stockwell, P. and Whiteley, S. (eds.) *The Cambridge Handbook of Stylistics.* 46–62. Cambridge: Cambridge University Press.

Studer, Patrick (2012) *Historical Corpus Stylistics: Media, Technology and Change.* Continuum.

遠藤織枝・木村拓・桜井隆・鈴木智映子・早川治子・安田敏朗 (2004)『戦時中の話しことば ラジオドラマ台本から』ひつじ書房.