

医学書テキストに現れる文末表現の特徴：単語 N-gramを用いた分析

| | |
|-----|---|
| 著者 | 本多 由美子, 丸山 岳彦, 三枝 令子 |
| 雑誌名 | 言語資源活用ワークショップ発表論文集 |
| 巻 | 5 |
| ページ | 73-84 |
| 発行年 | 2020 |
| URL | http://doi.org/10.15084/00003147 |

医学書テキストに現れる文末表現の特徴

—単語 N-gram を用いた分析—

本多 由美子 (一橋大学) †
丸山 岳彦 (専修大学・国立国語研究所)
三枝 令子 (専修大学)

Features of Sentence Final Expressions Appeared in Medical Texts: Analysis by Word N-gram

Yumiko Honda (Hitotsubashi University)
Takehiko Maruyama (Senshu University / NINJAL)
Reiko Saegusa (Senshu University)

要旨

本発表では、一般的な文章とは異なる文体的特徴を持つと考えられる「医学書」を対象とした言語学的分析の一例として、医学書2冊のテキスト(延べ約282万語)に現れる文末表現の特徴を単語 N-gram を用いて分析した。比較対象として、『現代日本語書き言葉均衡コーパス』の「特定目的・ベストセラー」内の文学作品のデータを用いた。分析の結果、医学書の文末表現には文学作品と比べて①定型表現が多く用いられること、②動詞の使用傾向が強いこと、③受身形の使用に特徴があることが明らかになった。また、医学書の文末で使用頻度の高い動詞「ある」「する」「いる」では、「ある」は「形状詞+である」、「する」は「漢語名詞+する」、「いる」は「受身形+ている」という形が、文学作品に比べて顕著に多く出現することが分かった。これらの特徴は、医学書が疾患や症状を解説し、その処置や対応方法を述べる内容のテキストであることによると考えられる。

1. はじめに

発表者らは、日本の医師国家試験を目指す外国人学習者への効果的な支援を目的として、医学教材の開発を目指している。そのためには医学書に現れるテキストの言語学的分析が必要不可欠だが、医学用語の語彙の分析は着手されているものの(山崎・相良(2014), 東条他(2018), 三枝他(2019, 2020)), その構文的特点はいまだ明らかではない。本発表は医学書の文末表現に着目し、文末部分の単語 N-gram を用いてその特徴を明らかにしようとするものである。分析の際、比較対象として、一般的な文章として文学作品を用いる。

2. 先行研究と問題の所在

N-gram を用いて文末表現を分析した事例として、丸山(2012a, 2012b)が挙げられる。丸山(2012a)では、文末表現のバリエーションの豊かさがそのテキストの表現力に関する特性を表すという前提のもと、BCCWJに含まれる12種類のレジスターから2万文ずつを収集し、文末から逆方向に取得した1グラムから5グラムまでの範囲について、その異な

† nihonda[AT]hotmail.com

り率やバリエーションの豊富さを分析している。丸山（2012b）では、13種類の文末表現について、最大8グラムまでの範囲で比較・検討を行っている。丸山（2012a）が示したBCCWJの各レジスターにおいて頻出する文末表現を、参考のため、表1に示す。

表1 BCCWJの各レジスターで頻出する文末表現
(丸山2012a:594, 表4より抜粋)

| 出版書籍 | 図書館書籍 | 雑誌 | 新聞 | 白書 | 教科書 |
|------------|------------|------------|------------|--------------|------------|
| のである 1.85% | のである 2.11% | ています 1.27% | している 2.32% | っている 9.91% | ましょう 4.08% |
| ています 1.73% | なかった 1.62% | している 1.00% | っている 1.43% | している 9.08% | てみよう 2.54% |
| っている 1.41% | っている 1.48% | っている 0.94% | れている 1.04% | なっている 6.28% | ています 2.09% |
| している 1.40% | している 1.18% | れている 0.90% | なかった 0.92% | れている 5.41% | れている 2.04% |
| なかった 1.34% | であった 1.18% | なかった 0.65% | になった 0.74% | となっている 5.09% | っている 1.57% |
| れている 1.21% | ています 1.17% | のである 0.65% | していた 0.67% | されている 2.73% | している 1.50% |

| 広報紙 | ベストセラー | Yahoo!知恵袋 | Yahoo!ブログ | 法律 | 国会会議録 |
|-------------|------------|---------------|------------|------------------|---------------|
| ください 9.07% | なかった 2.21% | しょうか? 5.46% | ています 1.72% | ならない 21.55% | ございます 14.08% |
| ています 8.46% | のである 1.94% | でしょうか? 5.44% | いました 1.48% | なければならぬ 18.79% | でございます 12.35% |
| しています 3.59% | であった 1.58% | のでしょうか? 2.79% | りました 1.21% | ができる 13.79% | おります 10.98% |
| てください 2.76% | っていた 1.43% | 思います 2.77% | きました 1.15% | ことができる 13.78% | ております 10.84% |
| あります 2.66% | っている 1.23% | ています 2.77% | しました 1.11% | ることができる 11.95% | 思います 8.47% |
| しました 2.49% | していた 0.80% | ください 2.75% | 思います 0.82% | しなければならない 10.62% | あります 7.76% |

以上の分析を医学書のデータに適用すると、どのような結果が得られるだろうか。これまで様々な日本語コーパスが開発・公開されてきているが、医学書に特化したコーパスは存在せず、特にその文法的な特徴は明らかではない。本発表ではこの点を明らかにするため、医学書のテキストをコーパス化して、文末表現を収集し、その形態的特徴を探ることとする。なお、医学書の特徴をより明確にするため、本発表では参照用コーパスとして、BCCWJの「特定目的・ベストセラー」に収録された文学作品を利用する。

3. データ

3.1 「医学書」データ

本発表では、医学書のデータとして、長年版を重ね、現場の医師によく利用されている以下の2冊を選んだ。

- 『今日の治療指針 2018年版』医学書院（以下、「治療」とする）
- 『今日の診断指針 第7版（2015）』医学書院（以下、「診断」とする）

この2冊から本文テキストを抽出し、このうち句点で終わる行のみを対象に形態素解析を行った¹。総語数は2,819,117語（記号・補助記号・空白を除く）であった。さらにこの結果

¹ 形態素解析には MeCab の ver.0.996（辞書は UniDic-cwj-2.3.0）を使用した。

をもとに、短単位の N-gram データを作成した。文末を特定するために、品詞情報が「補助記号-句点」のデータを抽出し、前文脈を目視で確認して文末を認定した。その結果、89,289 文が分析対象となった。

本発表では、文末の記号を除いた位置から左向きに短単位の書字形出現形を順に取得し、N-gram データとした。文末の単語 1~4-gram が本発表の調査対象である。図 1 は単語 1~4-gram の例である。以下、このデータを「医学書」データ、あるいは単に「医学書」と呼ぶ。なお、ある語の 1 語前に接する語を「前接語」と呼ぶ。例えば、図 1 の (2) では、「た」の前接語は「い」である。

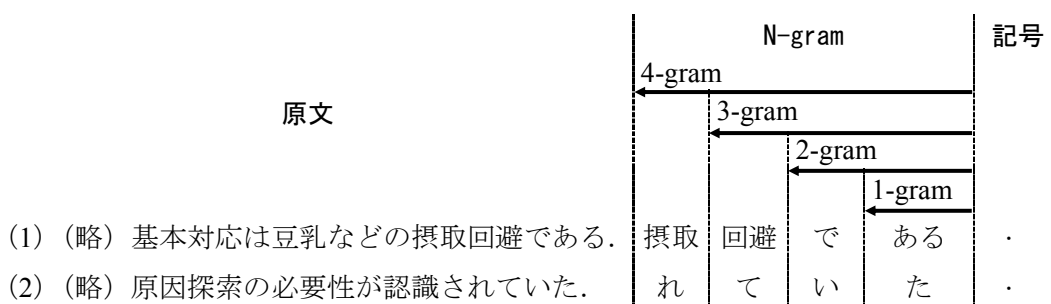


図 1 本発表における単語 N-gram の取り方

3. 2 「文学作品」データ

「医学書」データと比較する対象として、BCCWJ の「特定目的・ベストセラー」の中から NDC が「9 文学」のデータを用いた²。総語数は 2,257,398 語である（記号・補助記号・空白を除く）。検索アプリケーション「中納言」を用い、「文末から 1 語」が「補助記号-句点」および「文末から 1 語」がかぎ括弧であるものを抽出した³。医学書と同様に、文末の記号を除いた位置から左向きに単語 1~4-gram までを分析対象データとした⁴。以下、このデータを「文学作品」データ、あるいは単に「文学作品」と呼ぶ。

4. 結果

4. 1 文末表現の異なり率

図 2 は、「医学書」と「文学作品」に現れた文末表現について、単語 N-gram ごとに異なり数を求め、全体の文数で割った値を示したものである。全体的に「文学作品」よりも「医学書」のほうが文末に用いられている表現の異なりが少ない、すなわち定型的な表現が用いられることが多い傾向にあることがわかる。「文学作品」の文末表現のほうが異なりが多いことは、丸山 (2012a) の BCCWJ の書籍には多様な文末表現が含まれているという指摘と重なるものである。

² 2019 年 12 月~2020 年 1 月に検索したデータを用いた。「特定目的・ベストセラー」は、1976~2005 年の各年にベストセラーとなった書籍からサンプリングされたデータである（山崎編 2014）。本発表の分析対象となった 236 作品の作者の中で、生年が最も古いのは横溝正史（1902 年生まれ）であった。

³ 文末が「補助記号-句点」を抽出する際の検索式は、「キー: 品詞="補助記号-句点" ON 1 WORDS FROM 文末 AND 前方共起: 語彙素 LIKE "%" ON 1 WORDS FROM キー」である。

⁴ 文末が記号の連続である場合は、連続する記号を削除した位置から N-gram を取得した。例えば、文末が句点-括弧閉で終わる「私そろそろ帰るね。」のような発話の文の場合は、句点および括弧閉を削除し、“ね”から N-gram を取得した。

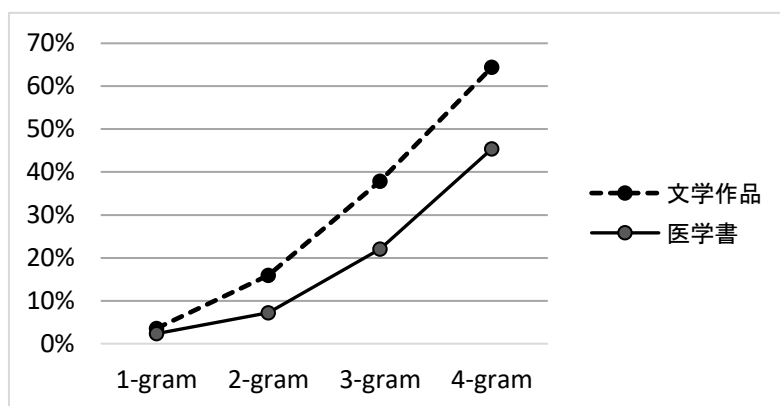


図2 「医学書」と「文学作品」における文末表現の異なり率

4. 2 1-gramの特徴

① 品詞

表2と表3は、「医学書」と「文学作品」の1-gramにどのような品詞の語が多く出現するかを示したものである。表2と表3から「医学書」と「文学作品」は文末の品詞の構成が大きく異なることがわかる。「医学書」と「文学作品」はいずれも上位2位までの品詞で全体の約7割を占めるが、「医学書」の上位2位は「動詞-非自立可能」と「動詞-一般」,「文学作品」の上位2位は「助動詞」と「助詞-終助詞」であり、全く異なっている。

表2 「医学書」の1-gram (品詞)

| 順位 | 品詞 | 頻度 | % |
|----|--------------|--------|--------|
| 1 | 動詞-非自立可能 | 47,685 | 53.4% |
| 2 | 動詞-一般 | 14,498 | 16.2% |
| 3 | 助動詞 | 10,887 | 12.2% |
| 4 | 形容詞-一般 | 5,010 | 5.6% |
| 5 | 名詞-普通名詞-サ変可能 | 2,491 | 2.8% |
| 6 | その他 | 8,718 | 9.8% |
| | 計 | 89,289 | 100.0% |

表3 「文学作品」の1-gram (品詞)

| 品詞 | 頻度 | % |
|-----------|---------|--------|
| 助動詞 | 77,030 | 55.8% |
| 助詞-終助詞 | 20,390 | 14.8% |
| 動詞-非自立可能 | 16,154 | 11.7% |
| 動詞-一般 | 5,028 | 3.6% |
| 形容詞-非自立可能 | 3,201 | 2.3% |
| その他 | 16,200 | 11.7% |
| 計 | 138,003 | 100.0% |

② 1-gramの高頻度語 (10語)

表4と表5は、「医学書」と「文学作品」における文末表現(1-gram)の頻度順上位10語(書字形出現形)である。網掛けの語は、「医学書」と「文学作品」のどちらにも見られる語である。上位10語のうち、4語が「医学書」と「文学作品」の両者に共通する語であるが、構成比は大きく異なる。例えば、「医学書」で1位の「ある」は、「文学作品」では3位であるが、その構成比は「医学書」では23.4%,「文学作品」では4.0%である。「文学作品」の1位の「た」は、「医学書」では10位であるが、構成比は、「文学作品」では36.4%,「医学書」では2.0%である。

「医学書」では、文末が「ル形」であるものが目立つ。「文学作品」で「た(助動詞)」が

1位であることとは対照的である⁵。「医学書」では症状の典型例を解説する文、診断や治療のポイントを述べる文が多いことによると思われる。

「医学書」にのみ見られる語には、「行う（動詞一般）」、「多い（形容詞一般）」という実質語や「れる」「られる」という受身の助動詞があり、特徴的である。他方、「文学作品」にのみ見られる語では、助動詞「だ」「です」「ます」や終助詞が目立つ。

表4 「医学書」の1-gram

| 順位 | 語 | 品詞 | 頻度 | % |
|----|-----|----------|--------|--------|
| 1 | ある | 動詞-非自立可能 | 20,936 | 23.4% |
| 2 | する | 動詞-非自立可能 | 16,542 | 18.5% |
| 3 | いる | 動詞-非自立可能 | 4,485 | 5.0% |
| 4 | 行う | 動詞-一般 | 3,910 | 4.4% |
| 5 | れる | 助動詞 | 3,712 | 4.2% |
| 6 | なる | 動詞-非自立可能 | 3,322 | 3.7% |
| 7 | 多い | 形容詞-一般 | 2,961 | 3.3% |
| 8 | られる | 助動詞 | 2,765 | 3.1% |
| 9 | ない | 助動詞 | 2,036 | 2.3% |
| 10 | た | 助動詞 | 1,795 | 2.0% |
| | その他 | | 26,825 | 30.0% |
| | 計 | | 89,289 | 100.0% |

表5 「文学作品」の1-gram

| 語 | 品詞 | 頻度 | % |
|----|----------|---------|--------|
| た | 助動詞 | 50,241 | 36.4% |
| だ | 助動詞 | 8,907 | 6.5% |
| ある | 動詞-非自立可能 | 5,536 | 4.0% |
| か | 助詞-終助詞 | 4,996 | 3.6% |
| よ | 助詞-終助詞 | 4,631 | 3.4% |
| いる | 動詞-非自立可能 | 4,444 | 3.2% |
| です | 助動詞 | 3,879 | 2.8% |
| ね | 助詞-終助詞 | 3,450 | 2.5% |
| ない | 助動詞 | 3,420 | 2.5% |
| ます | 助動詞 | 2,890 | 2.1% |
| | その他 | 45,609 | 33.0% |
| | 計 | 138,003 | 100.0% |

③ 文末の助動詞上位6語

表6と表7では文末1-gramの助動詞のみを示した。「医学書」において、助動詞の割合は12.2%と高くないが、「文学作品」では55.8%と半数以上を占める（表2、表3参照）。医学書では、受身形の「れる」「られる」の使用が最も多く、この2語で助動詞の約6割を占める。(1)(2)は受身形の例である。「医学書」では、症状の説明や、治療法に関する一般的な判断を述べる際にこれらの表現がよく用いられている。

- (1) 慢性期では人格変化，学業成績やIQ低下がみられる。(診断)
- (2) 本治療は，就寝時の歯ぎしりを抑制する作用があるとされる。(治療)

一方、「文学作品」では、「た」が最も多く、助動詞内の割合では2位以下と50ポイント以上の差があった。小説ではすでに起こった出来事を叙述することが多いからであろう。

⁵ 「文学作品」の2位の「だ（助動詞）」の前接語を見ると、「の（のだ）」「ん（んだ）」のように「助詞-準体助詞」が3,581語（「だ」の40.2%）で最も高く、次いで「名詞-普通名詞」が2,334語（同26.2%）であった。「読んだ」のように、前接語が「動詞-一般」で過去を表すものは1,367語（同15.3%）であった。

表6 「医学書」の1-gram (助動詞)

| 順位 | 語 | 頻度 | 助動詞内での割合 | 文末全体に占める割合 |
|----|-----|--------|----------|------------|
| 1 | れる | 3,712 | 34.1% | 4.2% |
| 2 | られる | 2,765 | 25.4% | 3.1% |
| 3 | ない | 2,036 | 18.7% | 2.3% |
| 4 | た | 1,795 | 16.5% | 2.0% |
| 5 | せる | 305 | 2.8% | 0.3% |
| 6 | たい | 222 | 2.0% | 0.2% |
| | その他 | 52 | 0.5% | 0.1% |
| | 計 | 10,887 | 100.0% | 12.2% |

表7 「文学作品」の1-gram (助動詞)

| 語 | 頻度 | 助動詞内での割合 | 文末全体に占める割合 |
|-----|--------|----------|------------|
| た | 50,241 | 65.2% | 36.4% |
| だ | 8,907 | 11.6% | 6.5% |
| です | 3,879 | 5.0% | 2.8% |
| ない | 3,420 | 4.4% | 2.5% |
| ます | 2,890 | 3.8% | 2.1% |
| だろう | 1,335 | 1.7% | 1.0% |
| その他 | 6,358 | 8.3% | 4.6% |
| 計 | 77,030 | 100.0% | 55.8% |

以下では、「医学書」の文末表現を詳しく見るために、表4の「医学書」で使用頻度の高い上位3語、「ある」「する」「いる」について考察する。

4. 3 文末に「ある」を取る表現の特徴

① 文末に「ある」を取る2-gramの表現

まず、表4で示した「医学書」の1-gramで最も頻度の高い「ある」について述べる。表8と表9は1-gramが「ある」の文について、2-gramの表現をまとめたものである。

「医学書」では、1位が「である」で62.4%と半数以上を占め、2位の「がある」を合わせると「ある」の約9割を占める。また、「である」は文末全体の14.6%を占め、「医学書」の文末において頻出する表現であることがわかる。「文学作品」でも1位の「である」と2位の「がある」を合わせると87.4%で9割に近く、「医学書」と同様の傾向を示している。しかし、文末全体では、両者を合わせても3.5%に過ぎない。

表8 「医学書」における
文末が「ある」の2-gram

| 順位 | 2-gram | 頻度 | 「ある」に占める割合 | 文末全体に占める割合 |
|----|--------|--------|------------|------------|
| 1 | である | 13,070 | 62.4% | 14.6% |
| 2 | がある | 5,731 | 27.4% | 6.4% |
| 3 | もある | 1,839 | 8.8% | 2.1% |
| 4 | にある | 136 | 0.6% | 0.2% |
| 5 | つつある | 102 | 0.5% | 0.1% |
| | その他 | 58 | 0.3% | 0.1% |
| | 計 | 20,936 | 100.0% | 23.4% |

表9 「文学作品」における
文末が「ある」の2-gram

| 2-gram | 頻度 | 「ある」に占める割合 | 文末全体に占める割合 |
|--------|-------|------------|------------|
| である | 4,217 | 76.2% | 3.1% |
| がある | 619 | 11.2% | 0.4% |
| もある | 290 | 5.2% | 0.2% |
| てある | 129 | 2.3% | 0.1% |
| その他 | 281 | 5.1% | 0.2% |
| 計 | 5,536 | 100.0% | 4.0% |

② 文末に「である」を取る 3-gram の表現

表 10 と表 11 は「である」の前接語に注目し、3-gram の表現を頻度順に示したものである。表 10 を見ると、「医学書」において「である」の前接語には「必要」「重要」「有用」「可能」「有効」「必須」など、形状詞（ナ形容詞）が目立つ。このことから、「医学書」の文末には、「形状詞＋である」が多く用いられることがわかる。「医学書」では、症状や疾患に対する処置や対応方法について述べる際にこれらの表現が用いられると思われる。また、「必要である」と「必須である」や、「有用である」と「有効である」などの類義の表現が見られる（例文（3）～（7））。これらの表現がどのように使い分けられているかについて、「医学書」の具体的な文脈に即して分析することは、今後の課題とする。

- (3) 放射線治療などの後療法をいかに組み合わせるかが重要である。（診断）
- (4) 本剤の投与には文書による同意が必要である。（治療）
- (5) 肉眼的血尿を主訴として来院した症例は、膀胱尿道鏡検査が必須である。（診断）
- (6) 後期新生児期（>生後7日）の敗血症スクリーニングには有用である。（診断）
- (7) 殺成虫治療としてピブラマイシンの長期投与（前出）が有効である。（治療）

表 10 「医学書」における
文末が「である」の 3-gram

| 順位 | 3-gram | 頻度 | 「である」に占める割合 |
|----|--------|--------|-------------|
| 1 | 必要である | 1,567 | 12.0% |
| 2 | 重要である | 1,385 | 10.6% |
| 3 | 有用である | 817 | 6.3% |
| 4 | 可能である | 575 | 4.4% |
| 5 | べきである | 520 | 4.0% |
| 6 | 疾患である | 310 | 2.4% |
| 7 | 的である | 304 | 2.3% |
| 8 | 有効である | 255 | 2.0% |
| 9 | 必須である | 237 | 1.8% |
| 10 | ことである | 219 | 1.7% |
| | その他 | 6,881 | 52.6% |
| | 計 | 13,070 | 100.0% |

表 11 「文学作品」における
文末が「である」の 3-gram

| 3-gram | 頻度 | 「である」に占める割合 |
|--------|-------|-------------|
| のである | 1,369 | 32.5% |
| ことである | 184 | 4.4% |
| からである | 182 | 4.3% |
| ようである | 166 | 3.9% |
| ものである | 148 | 3.5% |
| はずである | 69 | 1.6% |
| だけである | 68 | 1.6% |
| そうである | 45 | 1.1% |
| その他 | 1,980 | 47.0% |
| 計 | 4,211 | 100.0% |

一方、「文学作品」を見ると、表 11 では「のである」が「である」の約 3 分の 1 を占めることがわかる。その他にも、「ことである」「ものである」「はずである」などの形式名詞や「からである」という接続助詞が前接語の上位を占める。これは、「医学書」での上位が「必

要」「重要」「有用」などの実質語であるのとは大きく異なる⁶。

4. 4 文末に「する」を取る表現の特徴

次に、表4で示した「医学書」の1-gramで頻度順2位の「する」について述べる。表12と表13は1-gramが「する」の文について、「する」の前接語の品詞をまとめたものである。「医学書」では「する」は文末全体の18.5%を占めるが(表4参照)、「文学作品」では文末全体の0.7%(頻度899)であり、「医学書」とは約18ポイントの差があった。

表12から、「医学書」において「する」の前接語は「名詞-普通名詞-サ変可能」が90.5%、文末全体に占める割合でも16.8%を占めることがわかる。一方、「文学作品」でも「する」の前接語は「名詞-普通名詞-サ変可能」の割合が最も高いが、「する」に占める割合は43.5%、文末全体に占める割合は0.3%である。このことから、「文学作品」と比べ、「医学書」における2-gramの文末表現では「名詞-普通名詞-サ変可能+する」という組み合わせが非常に高い頻度で使われることがわかる。

なお、「文学作品」の文末が「名詞-普通名詞-サ変可能+します」であるものについても述べておく。「文学作品」において、「します」の頻度は216で、文末に占める割合は0.16%であった。そのうち、「名詞-普通名詞-サ変可能+します」は頻度61で、「します」に占める割合は28.5%、文末全体に占める割合は0.04%で0.1%に満たず、「名詞-普通名詞-サ変+する」よりも少なかった。

表12 「医学書」における
文末「する」の前接語の品詞

| 順位 | 品詞 | 「する」に 占める割合 | 文末全体に 占める割合 |
|----|------------------|----------------|----------------|
| 1 | 名詞-普通名詞 -サ変可能 | 90.5% | 16.8% |
| 2 | 助詞-格助詞 | 5.9% | 1.1% |
| 3 | 名詞-普通名詞-一般 | 1.6% | 0.3% |
| | その他 | 2.0% | 0.4% |
| | 計 | 100.0% | 18.5% |

表13 「文学作品」における
文末「する」の前接語の品詞

| 品詞 | 「する」に 占める割合 | 文末全体に 占める割合 |
|------------------|----------------|----------------|
| 名詞-普通名詞 -サ変可能 | 43.5% | 0.3% |
| 助詞-格助詞 | 36.3% | 0.2% |
| その他 | 20.2% | 0.13% |
| 計 | 100.0% | 0.7% |

表14-1、表14-2に「医学書」における「名詞-普通名詞-サ変可能+する」の内訳を示す。表14-1は「名詞-普通名詞-サ変可能」の語種をまとめたものである。漢語の割合が異なりで96.8%、延べで98.6%であり、「する」に前接する「名詞-普通名詞-サ変可能」は、ほぼ漢語であることがわかる。表14-2は「名詞-普通名詞-サ変可能」の高頻度10語である。思考や判断に関する語(考慮、確認、診断、注意、検討、評価)や、治療を含む患者に対する行為に関する語(併用、説明、投与、指導)が多く見られる。

⁶ 参考までに「です」の前接語を調べたところ、最も多かったのが「ん」、2位が「の」、3位が「こと」であった。文末「です」において、文末が「んです」または「のです」の割合が43.6%と4割強を占め、「である」と同様の傾向を示した。

このように、「医学書」における文末表現では、「漢語名詞+する」の使用が顕著に多いことが明らかになった⁷。

表 14-1 「名詞-普通名詞-サ変可能」の語種

| 語種 | 異なり (%) | | 延べ (%) | |
|------|---------|--------|--------|--------|
| 漢語 | 626 | 96.8% | 14,757 | 98.6% |
| 漢語以外 | 21 | 3.2% | 211 | 1.4% |
| 計 | 647 | 100.0% | 14,968 | 100.0% |

表 14-2 「名詞-普通名詞-サ変可能」の高頻度語 (10 語)

| 順位 | 語 | 頻度 | % |
|----|----|-----|------|
| 1 | 考慮 | 890 | 5.9% |
| 2 | 確認 | 740 | 4.9% |
| 3 | 併用 | 637 | 4.3% |
| 4 | 診断 | 558 | 3.7% |
| 5 | 注意 | 535 | 3.6% |
| 6 | 説明 | 485 | 3.2% |
| 7 | 投与 | 469 | 3.1% |
| 8 | 指導 | 398 | 2.7% |
| 9 | 検討 | 353 | 2.4% |
| 10 | 評価 | 272 | 1.8% |

4. 5 文末に「いる」を取る表現の特徴

ここでは、表 4 で示した「医学書」の 1-gram で頻度順 3 位の「いる」について述べる。表 15 と表 16 は 1-gram が「いる」の文について、2-gram の表現をまとめたものである。「医学書」では、1-gram が「いる」の文のうち、「ている」の割合が最も高く、98.5%である。一方、「文学作品」でも「ている」は 89.9%であり、最も高い。「医学書」と「文学作品」は、似た傾向を示している。

表 15 「医学書」における
文末が「いる」の 2-gram

| 順位 | 2-gram | 頻度 | 「いる」に 占める割合 | 文末全体に 占める割合 |
|----|--------|-------|----------------|----------------|
| 1 | ている | 4,416 | 98.5% | 4.9% |
| 2 | でいる | 50 | 1.1% | 0.1% |
| | その他 | 19 | 0.4% | 0.0% |
| | 計 | 4,485 | 100.0% | 5.0% |

表 16 「文学作品」における
文末が「いる」の 2-gram

| 2-gram | 頻度 | 「いる」に 占める割合 | 文末全体に 占める割合 |
|--------|-------|----------------|----------------|
| ている | 3,996 | 89.9% | 2.9% |
| でいる | 237 | 5.3% | 0.2% |
| その他 | 211 | 4.7% | 0.2% |
| 計 | 4,444 | 100.0% | 3.2% |

表 17 と表 18 は、2-gram が「ている」の文について 3-gram の表現をまとめたものである。「医学書」では、「れている」と「られている」という受身形の使用が特徴的で、両者を合

⁷ ちなみに、「名詞-普通名詞-サ変可能」において、漢語以外の語で出現頻度が高かった語は、「チェック (頻度 123)」、「モニター (頻度 20)」、「フォロー (頻度 17)」であった。

わせると、「ている」のうちの約63%を占める。一方で、「文学作品」でも上位に受身形は見られるが、「れている」と「られている」を合わせても約11%であり、「医学書」とは50ポイント以上の差が見られる。

「医学書」で「受身形+ている」の割合が高いことは、1-gramでも受身形の使用割合が高いこととも重なる(表4, 表6参照)。表19は4-gramの高頻度の表現である。「されている」「考えられている」「知られている」「行われている」「示されている」が上位5表現である。「医学書」では、様々な研究結果や知見を引用する際にこれらの表現が用いられていると思われる。

- (8) 足白癬が5人に1人、爪白癬が10人に1人の割合であるとされている。(治療)
- (9) 飛沫感染の可能性も否定できないが空気感染はしないと考えられている。(診断)

表17 「医学書」における
文末が「ている」の3-gram

| 順位 | 3-gram | 頻度 | 「ている」に占める割合 |
|----|--------|-------|-------------|
| 1 | れている | 2,291 | 51.9% |
| 2 | している | 787 | 17.8% |
| 3 | られている | 504 | 11.4% |
| 4 | なっている | 294 | 6.7% |
| 5 | きている | 79 | 1.8% |
| | その他 | 461 | 10.4% |
| | 計 | 4,416 | 100.0% |

表18 「文学作品」における
文末が「ている」の3-gram

| 3-gram | 頻度 | 「ている」に占める割合 |
|--------|-------|-------------|
| している | 606 | 15.2% |
| れている | 348 | 8.7% |
| なっている | 184 | 4.6% |
| られている | 82 | 2.1% |
| 見ている | 55 | 1.4% |
| その他 | 2,721 | 68.1% |
| 計 | 3,996 | 100.0% |

表19 「医学書」における「受身形+ている」の4-gram (上位5表現)

| 順位 | 4-gram | 頻度 |
|----|---------|-------|
| 1 | されている | 1,688 |
| 2 | 考えられている | 213 |
| 3 | 知られている | 135 |
| 4 | 行われている | 112 |
| 5 | 示されている | 108 |

また、「その他」の割合が「医学書」「文学作品」では、それぞれ10.4%、68.1%で、「文学作品」のほうが60ポイント近く高い。これは、「医学書」の異なりが156であるのに対し、「文学作品」では異なりが1,106であることによる。上述したように、「文学作品」よりも「医学書」のほうが定型的な表現が用いられていることがわかる(図2参照)。

以上のことから、「文学作品」と比較すると、「医学書」の文末表現では、「受身形+てい

る」の使用が顕著に多いことが明らかになった⁸。

5. まとめ

5. 1 医学書の文末表現の特徴

本発表では、医学書の文末表現の特徴を探るため、文末表現の単語 N-gram を用い、文学作品との対照を通して分析を行った。その結果、次のことが明らかになった。

- 文学作品と比較すると、医学書の文末は、表現の異なりが少なく、定型的な表現が多く用いられる傾向がある。
- 文末には、動詞の「ル形」が多い。
- 1-gram では「ある」「する」「いる」の3語の出現頻度が高い。
- 最も高頻度の「ある」については、「形状詞+である」の形が多い。
- 「する」「いる」については、「漢語名詞+する」「受身形+ている」の形が文学作品に比べて顕著に多く出現する。

こうした特徴は、医学書に、症状の典型例を解説したり、診断や治療のポイントを述べたりする文が多いことによると考えられる。丸山 (2012a) が述べた「文末表現のバリエーション」が示す傾向は、当該のレジスターに含まれるテキストが担う機能と関係すると考えられる。そのテキストがどのような目的によって書かれたものかという点を考えることによって、文末表現のバリエーションの偏りや、そこに現れる文法カテゴリの傾向の違いを説明することができる」という指摘は、本発表における医学書の分析結果にも通用すると考えてよいだろう。

5. 2 日本語学習者のための教材作成へ向けて

日本の医師国家試験を目指す外国人学習者向けの教材は、専門語彙だけでなく、構文についても医学書の特徴を踏まえて作成することが重要であると考えられる。さらに、そこにはどのようなサ変動詞、形状詞等がどのような構文の中で高頻度に用いられているか、日常語の用法とどのように違うか等の情報を提示する必要があるだろう。

本発表では、調査対象を文末に限定したことや、文末以外の部分との共起関係を見ていないため、部分的な分析に留まっている。今後は、調査対象の範囲を広げて、医学書の構文的な特徴と具体的な表現の実態を明らかにしていきたい。

謝辞

本研究は JSPS 科研費 JP18H00679 の助成を受けたものです。

⁸ なお、「文学作品」の「ています」についても述べておく。「文学作品」のうち「います」は、文末全体に占める割合は 0.4% (頻度 486) であり、「文学作品」の文末における「いる」の割合 (3.2%, 表 5 参照) よりも低い。「います」のうち、「ています」が占める割合は 93.2% (頻度 453) である。また、文末が「ています」の 4-gram では「しています」が 17.7% (頻度 77) で割合が最も高く、「れています」が 2 位で 8.9% (頻度 39) であった。表 18 の「ている」の 3-gram と比較すると、「文学作品」において「ています」も「ている」も内訳は同様の傾向を示していることがわかった。

文献

- 三枝令子・丸山岳彦・庵功雄・松下達彦・石川和信・小林元・品川なぎさ・稲田朋晃・山元一晃・遠藤織枝（2019）「動詞に見る医学用語の特徴—BCCWJ との比較から見えること—」『専門日本語教育研究』21, pp.69-76.専門日本語教育学会
- 三枝令子・丸山岳彦・松下達彦・品川なぎさ・稲田朋晃・山元一晃・石川和信・小林元・遠藤織枝・庵功雄（2020）「医学用語の収集と分類」『日本語教育』176, pp.33-47.日本語教育学会
- 東条佳奈, 内山清子, 岡照晃, 小野正子, 相良かおる, 山崎誠（2018）「実践医療用語に現れる語構成要素の辞書構築にむけて」『計量国語学会第 62 回大会予稿集』pp.7-12.計量国語学会
- 丸山岳彦（2012a）『『現代日本語書き言葉均衡コーパス』を用いた文末表現のバリエーションの分析』『言語処理学会 第 18 回年次大会発表論文集』pp.591-594.言語処理学会
- 丸山岳彦（2012b）『『現代日本語書き言葉均衡コーパス』を用いた文末表現のバリエーションの分析(2)』『第 2 回コーパス日本語学ワークショップ予稿集』pp.207-214.国立国語研究所
- 山崎誠・相良かおる（2014）「医療経過記録における漢字連続複合語の計量的分析」『じんもんこん 2014 論文集』3, pp.221-226.情報処理学会

調査資料

- 医学書院編（2018）『今日の治療指針 2018 年版』医学書院
- 金澤一郎, 永井良三 総編集（2015）『今日の診断指針 第 7 版』医学書院
- 『現代日本語書き言葉均衡コーパス』（Ver. 1.1）