

ツリーバンクを用いたテキストジャンルの語彙的・ 文法的分析

メタデータ	言語: Japanese 出版者: 武蔵野大学文学部日本文学研究所 公開日: 2025-06-19 キーワード: 作成者: 三好, 伸芳 メールアドレス: 所属:
URL	https://mu.repo.nii.ac.jp/records/2000663

ツリーバンクを用いたテキスト ジャンルの語彙的・文法的分析

三好 伸芳

1. はじめに

「現代日本語書き言葉均衡コーパス」など、形態論情報を基本とする従来のコーパスでは、主語と述語の係り受け関係などを捉えることが困難であったが、統語解析情報付きコーパスである Kainoki Tree Bank (KTB) であれば、そのような調査が可能である。本研究では、KTB を用いて文学的文章と説明的文章における主述関係の分布の差異を調査し、各テキストジャンルの特性の一端を語彙的・文法的観点から明らかにする。具体的には、KTB の青空文庫と新聞記事における「普通名詞／固有名詞／代名詞」という3種類の名詞、およびそれらを主語とする「動詞述語／形容詞述語（イ形容詞とナ形容詞）／名詞述語」の用例を調査し、主述パターンの現れ方とテキストジャンルとの関係を検討していく。調査の結果、代名詞と形容詞述語の組み合わせは説明的文章に現れにくいこと、名詞述語文は文学的文章において指示代名詞と結びつく傾向にあることが分かった。これらの事実は、代名詞の語彙的性質や過去形の名詞述語文が持つ文法的性質によって説明することが可能である。特に後者の文法的性質については、ソ系指示代名詞と結びついた際に叙想的テンスとして解釈されやすくなるという過去形名詞述語文の特徴を量的な調査に基づいて示した点でも重要であると考えられる。

2. 前提

2.1. KTB の概要

KTB は必ずしも日本語研究において広く知られているコーパスであるとは言えない。そこで本節では、KTB が持つコーパスとしての特徴や検索方法について概観していく。KTB では、語や構成素などの文を構成する要素が樹形図（ツリー）状にアノテーションされている。以下に KTB のアノテーションの例を挙げる（以下、KTB の用例には例文 ID を付す）。

- (1) 陽は既に西に傾きかけている。 (Kainoki_224_aozora_Dazai_2_1940)

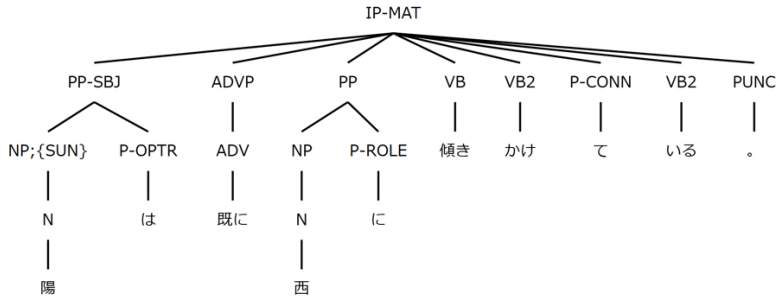


図1 KTBの統語構造

図1に示したように、KTBでは「N」、「P-OPTR」のような品詞レベルの節点が「NP」、「PP-SBJ」といった句レベルの節点や「IP-MAT」のような節レベルの節点とともに階層的にアノテーションされている。「傾き かけ」や「(て) いる」などの機能語的に用いられる動詞は、形態情報に基づくコーパスにおいて出現環境を問わず非自立的な動詞としてタグ付けされることになるが、KTBでは「VB2」のように文環境ごとに本動詞の用法から区別されている点も特徴的である（例えば、「壁にかける」などの本動詞としての「かける」の用法は「VB」となり、文環境によって異なるタグ付けがなされる）。

KTBでは、TGrep2と呼ばれる検索プログラムを用い、これらの語や節点の統語的關係性を指定することで検索を行う。例えば、「A < B」という検索式は、「Bというノードの直上にあるAというノード」が検索対象になる（逆向きの「A > B」は、「Bというノードの直下にあるAというノード」を表す）。また、「A \$ B」のような検索式は、「Bと統語的な姉妹関係にあるA」を検索対象とすることになる。これらの表現を括弧でまとめることでグループ化し、より複雑な統語環境を指定することも可能である。先述の「陽は既に西に傾きかけている」を例とすると、当該の文が検索対象となる検索式としては以下のようなものが挙げられる（図1と併せて参照されたい）。

- (2) a. /IP-MAT/ < /^VB\$/
= 直下に動詞述語を含む主節
- b. /^VB\$/ \$ /^VB2\$/
= 補助動詞と姉妹関係にある動詞述語
- c. (/PP-SBJ/ < (/P-OPTR/ < /は/))
= 「は」を含む主語助詞句
- d. /^NP\$/ > ((/PP/ < (/P-ROLE/ < /^に\$/)) > /IP-MAT/)

主節直下の「に」を伴う助詞句内に現れた名詞句

各ノードの表現を「/」（スラッシュ）で囲っているのは、これらの形式を正規表現として検索するためである。正規表現内部を「^」と「\$」で挟むと、完全一致指定の表現となる¹。これらはいくまで一例であり、別の検索式によっても調査することもできる。もちろん、上記の統語環境を含む文であれば、(1)の他にもさまざまな文が検索結果として表示される。

統語構造を指定して検索することにより、形態論情報に基づく従来のコーパスでは不可能な調査がKTBでは可能になる。例えば、以下の例文における主述の係り受け関係は、「現代日本語書き言葉均衡コーパス」などの従来のコーパスで検索することが難しい。

- (3) 老人は額を、元の窓の所へ立てかけて、席につくと、私にもその向う側へ坐る様に、手真似をしながら、私の顔を見つめて、こんなことを云った。
(Kainoki_132_aozora_Edogawa_1929)

述語「云う」の主語が「老人」であることは明らかであるが、このような線条的に離れた要素の関係を指定して目視に頼らず検索することは、従来のコーパスでは困難である。一方、KTBを用いて然るべき検索式で検索すれば、このような主述関係を指定して検索することが可能である。

以上、簡単にKTBの検索手法を紹介した。詳細はKTBの検索ガイド等も参照されたい。

2.2. テキストにおける品詞の分布

テキストにおいてどのような品詞がどのように分布しているかを明らかにすることは、日本語の語彙ないし文法の実態を基礎的なレベルで把握できるという意味で重要である。テキストジャンルと品詞の出現比率に一定の相関があることを指摘した先駆的な研究である樺島(1954, 1955)は、いずれのテキストジャンルにおいても名詞が最も多く用いられている一方で、日常会話、小説の会話文などにおいては相対的に動詞や形容詞の出現頻度が高く、哲学書、新聞記事などにおいては相対的に名詞の出現頻度が高いことを指摘している。これらのテキストは場面の依存度という点で異なった性質を持ち、日常会話などの場面や文脈への依存度が高いテキストにおいては、情報伝達において基本となる名詞の出現頻度が他のテキストに比べ相対的に少なくなるのに対し、新聞記事などの場面や文脈への依存度が低いテキストにおいては、より場面や文脈について説明を要するために名詞の出現頻度が高くなるのだと考えられる

(樺島 1954 : 17-19)。

このような品詞とテキストとの相関関係を明らかにすることは、日本語の実態を把握するという点でも有効であるだけでなく、それぞれのテキストが言語的にどのような性質を持っているのかを明らかにするうえでも欠かせないと言える。しかし、資料の問題もあり、テキストと統語構造の相関について論じた研究は管見の限り見られない²。形態レベルで「どのテキストにおいて、どの品詞が現れやすいのか」が問題にされるように、統語レベルにおいては、「どのテキストにおいて、どの品詞とどの品詞の組み合わせが現れやすいのか」という観点で、日本語の実態把握だけでなく、テキストジャンルの言語的性質の解明という点でも重要になると考えられる。そこで本研究では、テキストジャンルと統語構造にどのような対応関係が見られるのかという問題について、KTBを用いて調査することにしたい。

3. 調査方法

本研究では、三好 (2020) の調査に基づきテキストジャンルにおける統語構造の分布を調査する³。説明の都合上、本節の内容は三好 (2020) と一部重複がある。

3.1. 調査項目と資料

本研究では、KTBを用いて「どのテキストにおいて、どの品詞とどの品詞の組み合わせが現れやすいのか」を調査する。品詞の組み合わせといっても、副詞と動詞の関係や連体修飾成分と被修飾名詞句など、さまざまなものが存在するが、樺島 (1954, 1955) などで指摘されているように、文で表される事態の概念的な骨格が名詞によって形成されるという点を踏まえれば、項 (補語) の名詞句と述語の係り受け構造が最も基本的な統語的關係性であるように思われる。項となる名詞句には、「普通名詞／固有名詞／代名詞」などの意味的・語彙的分類が想定される。また、述語には「動詞述語／形容詞述語 (イ形容詞とナ形容詞)／名詞述語」などの区別が立てられる。これらを単純に組み合わせると「普通名詞／固有名詞／代名詞」を項とする「動詞述語／形容詞述語／名詞述語」を検索対象とすることになる。しかし、このような検索を行った場合の問題として、ヲ格やニ格を伴う目的語相当の要素は原則として動詞述語にしか現れないために、動詞述語の例を他の述語の例と比較しにくくなるという点が挙げられる⁴。そのため、いずれの述語でも現れうる主語名詞句のみを調査対象とする必要がある⁵。さらに、主題の「は」は従属節環境における分布に制限があるなど、統語的な性質は節環境によって異なり、より均質な統語環境を調査するためには、主節環境に限定した方が適切であると考えられる。このような前提に基づき、本研究では以下のような調査項目を立てる。

- | | | | | |
|---------------------------------------|---|----------------|---|---------------------|
| (4) 動詞述語
形容詞述語 (イ形容詞/ナ形容詞)
名詞述語 | } | を持つ主節に現れている主語の | { | 普通名詞
固有名詞
代名詞 |
|---------------------------------------|---|----------------|---|---------------------|

つまり、「普通名詞／固有名詞／代名詞」という3種類の名詞について、それぞれ「動詞述語／形容詞述語 (イ形容詞とナ形容詞) ／名詞述語」という述語との組み合わせで現れたもの (イ形容詞とナ形容詞を分けて考えれば3×4の12通り) を調査することになる。

調査資料としては、テキストジャンルと統語構造の相関関係を調べるため、KTBのうち「青空文庫」(文学的文章) と「新聞記事」(説明的文章) を検索対象とする。KTBにはいずれのテキストにも翻訳したものが含まれるが、それらは検索対象から外す。青空文庫は、エッセイを除いた14844 ツリー⁶、246011語を検索対象とする。新聞記事は、いずれも「河北新報」の新聞記事であり、5802 ツリー、86781語が検索対象である。

3.2. 検索式

2.1節で述べた通り、KTBではTGrep2という検索プログラムを用いて用例を収集する。以下、「動詞述語を持つ主節に現れている主語の普通名詞」に対応する検索式を例に説明する。

- (5) 「動詞述語を持つ主節に現れている主語の普通名詞」の検索式
`/^N$/ > (/^NP/ > (/PP-SBJ/ > (/IP-MAT/ < /^VB$/))`

上記の検索式のうち、末尾の「/IP-MAT/ < /^VB\$/」の部分は「動詞述語を持つ主節」を表しており、その左側で「主語の普通名詞」(より正確には「主語の助詞句内に現れた普通名詞」で、「PP-SBJ」が「主語の助詞句」に対応する) を指定している。このような指定により、「犯人は太郎が捕まえました／日本食が好きだ」のような目的語相当の要素を排除することができる。検索結果の一例を挙げておく。

- (6) 外は全く暗闇になっていた。 (Kainoki_53_aozora_Edogawa_1929)

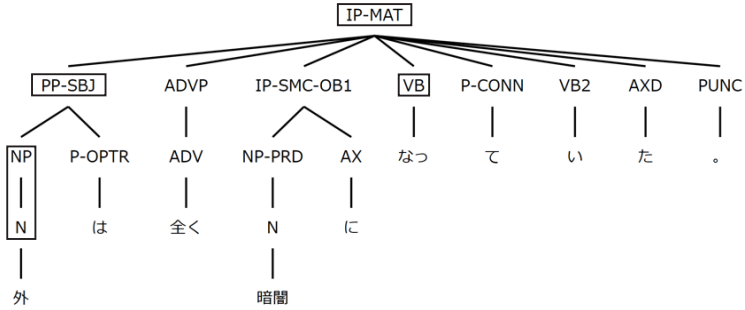


図2 主節動詞述語と主語名詞の例

(6)「外は全く暗闇になっていた」には、普通名詞主語の「外（は）」が現れており、さらに当該の要素は動詞述語を持つ主節の直下に現れている。この文のツリーは図2のようになっているが、検索式で指定されている節点およびその統語関係が(5)の検索式で意図した結果と合致していることが伺える。

ここまででは普通名詞かつ動詞述語を対象とする検索式を見てきたが、普通名詞や動詞以外が現れている統語環境も、上記の検索式の一部を入れ替えることで検索できる。例えば「^N\$」（普通名詞）を「^NPR\$」、「^PRO\$」に置き換えれば、それぞれ「固有名詞／代名詞」の用例を検索できる。同様に、末尾の「^VB\$」（動詞述語）を「^ADJIS\$」、「^ADJNS\$」、「^NP-PRD\$」に置き換えれば、それぞれ「イ形容詞述語／ナ形容詞述語／名詞述語」を検索することができる。

4. 調査結果

上述の検索式を用いて調査を行ったところ、全体で8541件の用例が収集できた。以下に主語名詞ごとの内訳と主節述語ごとの内訳をまとめる。括弧内の数値は、合計に対する割合を百分率で示したものである（以下、用例数を記載する場合にはこれに倣う）。

表1 主語名詞（青空文庫／新聞記事）の内訳

普通名詞	固有名詞	代名詞	合計
5211 (61.0%)	1870 (21.9%)	1460 (17.1%)	8541 (100.0%)

表2 主節述語（青空文庫／新聞記事）の内訳

動詞述語	イ形容詞述語	ナ形容詞述語	名詞述語	合計
6687 (79.2%)	509 (6.0%)	231 (2.7%)	1021 (12.1%)	8448 (100.0%)

当然のことではあるが、名詞の中で最も多く用いられているのは普通名詞であり、次いで固有名詞と代名詞が続いている。英語などの言語に比べれば日本語の代名詞は格変化などの文法的な機能に乏しいが、今回の調査資料の範囲ではそれなりの用例数があることが分かる。述語では、動詞が最も多く用いられており、次いで名詞述語、イ形容詞、ナ形容詞の順に続く。名詞は項などとしても用いられるため（表1の結果は、そのまま主節の主語として用いられた名詞の用例数でもある）、語彙的に見れば名詞が最も多く用いられているが、述語環境においては動詞のほうが優勢であると言える。一方で、イ形容詞述語とナ形容詞述語の合計よりも名詞述語のほうが多く、語彙量の違いが反映されていると考えられる。

テキストジャンルごとに見た各名詞と各述語の出現頻度は、以下の通りである。

表3 青空文庫における主語名詞と述語の分布

	動詞述語	イ形容詞述語	ナ形容詞述語	名詞述語	合計
普通名詞	2547 (76.7%)	301 (9.1%)	88 (2.6%)	386 (11.6%)	3322 (100.0%)
固有名詞	1248 (91.0%)	33 (2.4%)	26 (1.9%)	65 (4.7%)	1372 (100.0%)
代名詞	1144 (80.7%)	53 (3.7%)	44 (3.1%)	177 (12.5%)	1418 (100.0%)

表4 新聞記事における主語名詞と述語の分布

	動詞述語	イ形容詞述語	ナ形容詞述語	名詞述語	合計
普通名詞	1321 (69.9%)	146 (7.7%)	75 (4.0%)	347 (18.4%)	1889 (100.0%)
固有名詞	444 (89.2%)	4 (0.8%)	5 (1.0%)	45 (9.0%)	498 (100.0%)
代名詞	34 (81.0%)	0 (0.0%)	1 (2.4%)	7 (16.7%)	42 (100.0%)

表3と表4の結果を踏まえると、各タイプの名詞に見られる分布上の特徴は以下のようによまとめられる。

(7) a. 普通名詞

母数が多いため、各述語とともに用いられる割合は平均に近い値を示している。形容詞述語や名詞述語とも一定程度主述関係を持つ傾向にある。新聞記事における名詞述語文は相対的に見てやや多い（青空文庫：11.6%／新聞記事：18.4%）。ただし、普通名詞の中にもさまざまな種類の名詞が含まれている（形式名詞、相対名詞など）ため、分類次第で異なった見え方になる可能性がある⁷。

b. 固有名詞

青空文庫および新聞記事のいずれにおいても一定程度用例が見られる。ジャンルを問わず動詞述語と結びつきやすく、新聞記事においては形容詞述語との共起はあまりない。他の名詞と同じく、新聞記

事において名詞述語文が相対的に多い。

c. 代名詞

青空文庫においては固有名詞と同程度の用例が観察できるが、新聞記事においてはかなり少ない。全体としては動詞述語と用いられる傾向にあるが、新聞記事における名詞述語文の用例数は相対的に多い。

以下の節では、テキストジャンルと主語名詞、述語の分布の対応関係、特に主語名詞と形容詞述語、名詞述語の関係について分析していく。

4.1. 項の名詞と形容詞述語

まず、項となった名詞と形容詞述語の関係について議論していく。(7)で触れたように、新聞記事における固有名詞と代名詞については以下のような指摘ができる。

観察①: 代名詞は新聞記事において相対的に少なく、特に形容詞との共起例が少ない。

表3と表4で示した青空文庫と新聞記事における各名詞の出現頻度の合計と、それに対する各名詞の割合は以下のようにになっている。

表5 「青空文庫／新聞記事」における各名詞の頻度と割合

	青空文庫	新聞記事
普通名詞	3322 (54.4%)	1889 (77.8%)
固有名詞	1372 (22.4%)	498 (20.5%)
代名詞	1418 (23.2%)	42 (1.7%)
合計	6112 (100.0%)	2429 (100.0%)

表5に示したように、青空文庫では2割程度を占めている代名詞は、新聞記事においてそれほど用例が見られない。青空文庫と新聞記事の語数はそれぞれ246011語と86781語であり(およそ3:1)、母数の規模に照らして普通名詞がやや新聞記事において多いとはいえ、やはり代名詞の用例数は新聞記事において相対的に少ないと言えるだろう。

なぜこのような分布が生じるのであろうか。結論を先に述べれば、文学的文章においては物語の展開上繰り返し同一の人物が関係する動作や心理状態に言及する必要があるのに対し、説明的文章ではそのような方略をとらず、より簡潔で客観的な表現が選ばれるからであると考えられる。青空文庫において実際に用いられている代名詞の例を挙げる(以下、例文の下線は筆者によるもの)。

(9) 私はその犬を飼おうと思ったが、けれども、自分は軽はずみなことをしているのではないかという気もした。

けれどもまた考えてみるに、私の過去は軽はずみの連続のようなもので、もはやそのことでは私は自分自身を深く咎めだてする気にもなれないのである。

私はやはりいつもの伝でやることにした。

私は犬の顔を眺めながら、「私さえ保護者らしい気持を失わないならば、お互いがお互いを重荷に感ずるようなことはまずないだろう。」と思った。自信のあるような、ないような気持であった。

私はこれまで男の友達とは幾度か一緒に暮らしたことがあるが、いつも気まずい羽目になってしまったのである。

私はこの武蔵野市に移ってきてから、三年ほどになる。

私はある家の離れを借りて暮らしている。

(以上、aозora_Koyama_1955_1における13-20の例)

(9)の例では、代名詞「私」が8文中7回用いられている。文学的文章においては、物語を展開させるうえで登場人物の動向を辿っていくことになるため、「私は…気もした。…私は…なれないのである。私は…やることにした。私は…と思った。…」のように、必然的に同一の人物に繰り返し言及がなされる。特に一人称視点で物語が展開する場合、台詞と地の文を問わず人称代名詞が「私」などの一人称で固定され、代名詞を反復して使用することに繋がる。もちろん、(9)はあくまで数あるサンプルの1つを提示したに過ぎないが、今述べたことが文学的文章全体にわたる傾向として備わっていることは容易に想定される。文学的文章が持つこのような性質により、文学的文章では一人称を中心とする代名詞の例が多くなると考えられるのである⁸。

一方、新聞記事(説明的文章)では状況が異なる。

(10) a. 「彼らはあの寒い中、軽トラックの中で寝起きしていました。そこまでして支援して下さる方がいることが、本当に励みになりました」と当時を振り返ります。(27_news_KAHOKU_105)

b. それは福島とは異なる苦しみでもある。

(62_news_KAHOKU_102;K201401170A0T20XX00001)

説明的文章(新聞記事)において用いられる代名詞は、原則として(10a)のような取材対象者のコメントやインタビュー記事に現れる人称代名詞か、(10b)のようなわずかに用いられる文脈を指示する指示代名詞である。つまり、記事の執筆者が書

いた地の文には原則として指示代名詞しか現れず、一人称表現は用いられないということが指摘できる。このような事実は、新聞報道のような説明的文章において中立的・客観的な視点が求められるということを踏まえると、もっともなことであると言えるだろう。仮に一人称主語を明示して「{|私／筆者} は、…で…という事件があった }と考える／と聞いた}」のような表現を用いてしまうと、取材記者の視点や考えが前面に提示される形になってしまうため、事件等の報道として用いられるとは考えにくい。人称代名詞は使用者と聞き手がいる場を強く想起させる表現であり、そのような性質から、原則として報道などにおいては用いられないのである⁹。

「文学的文章は、同一の対象に反復的に言及する性質を持つ」という点は、青空文庫と新聞記事における固有名詞の分布からも伺うことができる。次の表は、青空文庫と新聞記事における主節動詞の主語となった固有名詞の異なり語数と延べ語数、およびその比率を示したものである。

表6 「青空文庫／新聞記事」における固有名詞の語数（動詞述語のみ）

	延べ語数	異なり語数	異なり語数／延べ語数
青空文庫	1372	131	0.095
新聞記事	498	371	0.745

「異なり語数／延べ語数」の値は、1に近いほどより多くの種類の語が使用されており、逆に0に近ければ少ない種類の語が繰り返し使用されていることを表す。青空文庫において、固有名詞の延べ語数は1372であるのに対し、異なり語数はわずか131である。異なり語数を延べ語数で割った値は0.095であり、これは1つの語が平均して10回ほど使用されていることを意味する。一方、新聞記事において、固有名詞は延べ語数が498であるのに対し異なり語数は371であり、1語につき平均して1.3回ほどしか使用されていない。以下に固有名詞の実際の用例も挙げておく。

- (11) a. すると六年生の一郎が来ました。

一郎はまるでおとなのようにゆっくり大またにやってきて、みんなを見て、「何した。」とききました。

みんなははじめてがやがや声をたててその教室の中の変な子を指さしました。

一郎はしばらくそっちを見ていましたが、やがて鞆をしっかりとかかえて、さっさと窓の下へ行きました。

みんなもすっかり元気になってついて行きました。

「だれだ、時間にならないに教室へはいつてるのは。」

一郎は窓へはいのぼって教室の中へ顔をつき出して言いました。

(以上、 aozora_Miyazawa_1934 における 31-37 の例)

- b. 復興の大目標に「減災都市」を掲げる多賀城市が、減災産業の集積に力を入れている。

東日本大震災の後、企業などの研究施設や工場の集積を図る減災リサーチパーク構想をいち早く発表。

既に「多賀城モデル」の防災用品が市場デビューを果たしている。多賀城市は、市域の3分の1が津波で浸水しながらも現地再建を決断した。

市震災復興推進局は「災害を完全に封じ込めるのは不可能。しなやかに受け流す減災都市を目指す。そのための協力を立地企業にもお願いしたい」と語る。

市の構想では、工業団地「復興パーク」に入居し、防災・減災に寄与する技術や製品の開発を進める企業に、年間120万円を上限に補助金を交付する。

既に八つの事業所や研究機関が「減災企業」の認定を受けた。

(以上、news_KAHOKU_82;K201402060A0T10XX00001 における 51-57 の例)

(11a) のように、固有名詞を用いた場合でも文学的文章では「一郎が来ました。一郎は…きました。…。一郎は…行きました。…」のように、同一人物への反復的な言及が見られる。一方、説明的文章ではそのような方略はとられず、(11b) のように一度「多賀城市」に言及したあとは、明確な場合は主語が省略されたり ((11b) の2文目)、「市」のような別の表現に置き換えられたりしている。これは、新聞記事が少ない紙幅でより多くの情報を伝えようとするという性質を持つのに加え、1つの記事の中では原則としてその記事のテーマが維持されるため、何についての言及かが明確であることに起因すると考えられる。以上のような事実からも、文学的文章は、同一の対象に反復的に言及する性質を持つという点が裏付けられるだろう¹⁰。

観察①で指摘したように、新聞記事における代名詞は全述語において用例数が少ないだけでなく、特に形容詞述語と共起する例がほとんど見られない。代名詞の使用が避けられるという新聞記事の文体的な傾向は、各テキストジャンルの述語に見られるこのような品詞の偏りとも密接に関わっているように思われる。以下に固有名詞と代名詞が形容詞の項となっている例を挙げる。

- (12) a. 手倉森ジャパンの一員として五輪代表入りを目指す藤村は意欲的だ。
(Kainoki_16_news_KAHOKU_117)
- b. 「生きているだけで私は幸せなんだ。

代名詞が形容詞とともに用いられている唯一の例である(12b)は、取材対象者のメッセージを引用したものであり、新聞記事の地の文ではない(つまり、代名詞が純粹な新聞文体の中で形容詞とともに用いられた例は、今回の調査では見出せなかったということになる)。固有名詞の例である(12a)にはそのような例外性はないが、固有名詞の場合も形容詞が共起した類例はそれほど多くない(青空文庫のイ形容詞+ナ形容詞:4.3%/新聞記事のイ形容詞+ナ形容詞:1.8%)。形容詞は本質的に相対的・連続的な尺度や他者から伺えない感情的・感覚的な心理を問題にする述語であり、例えばある人物が「意欲的か/意欲的でないか」というのを客観的に断定することは、一般的に困難である。当該の文のあとには、「試合に絡むチャンスがまた一つ増えた。即席のチームですけど次につながるプレー、自分のプレーを出していきたい。試合でしか感じられないものをつかみたい」と話す。」(Kainoki_17_news_KAHOKU_117)という「藤村」のコメントを紹介する文が続いているため、このような取材事実に基づいて「意欲的」という表現を使っているものと思われる。一方で、動詞述語においては、例えば「事件の犯人が{逮捕された/逮捕されていない}」のような事実認定は執筆者の立場によらず一定である。つまり、形容詞を用いて他者の感情や属性に根拠もなく言及することは新聞記事が持つ中立性や客観性といった特徴に抵触する可能性があるために避けられる傾向にあり、相対的に用例数が少なくなっているものと思われる。これは普通名詞や固有名詞にもある程度当てはまることであるが、特に代名詞の場合、代名詞自体の使用が新聞記事において抑制される傾向にあるので、形容詞述語とともに代名詞が用いられることはほとんどなくなるのである。

以上の分析をまとめると、代名詞は、文学的文章における場面展開の中で繰り返し言及されるために、当該のテキストジャンルにおける用例数が相対的に多く、ダイクシスによって表現主体の視点が前景化することを避けるという性質を持つ説明的文章(新聞記事)では、使用が抑制されるのだと考えられる。さらに、新聞記事においてはそもそも評価や感情を表す形容詞がテキストジャンルの特性と合致しないために、一般に避けられる傾向が見受けられる¹²。

4.2. 項の名詞と名詞述語

続いて、項となった名詞と名詞述語の関係に目を向ける。(7)でも言及したように、表3と表4からは次のような点が指摘できる。

観察②：いずれの名詞も、名詞述語と共起する例が青空文庫より新聞記事に相対的に多い。

形容詞述語は新聞記事に現れにくいので、単純にその部分を名詞述語がカバーしているとも考えられる。名詞述語文は、以下のように主語名詞句の所属や区分を述べることもあり、形容詞に比べて客観的な事実の報告となりやすい。

- (13) a. 父は久慈出身だった。
 (34_news_KAHOKU_303;K201401010W5XP4XX00001)
- b. 女川さんは元N T T関連会社のサラリーマン。
 (20_news_KAHOKU_1106;K201401010A0M106B00004)
- c. これこそが、自分が見逃していた点であり、防災力アップの鍵となるポイントである。
 (48_news_KAHOKU_34;K201404110A0T30XX00001)

例えば、(13ab) の述語名詞句「久慈出身」、「元N T T関連会社のサラリーマン」は、記事の関係者の基本的な社会的属性を読者と共有するうえで重要であるとともに、執筆者の主観が入り込む余地のない客観的な事実の報告である。もちろん、名詞述語が必ずこのようなものになるとは限らず、(13c) 「防災力アップの鍵となるポイント」のように、執筆者の評価を伴う名詞述語も少なくないが、(13ab) のような所属や区分を叙述する用法は原則として形容詞には見られないものであり、その点で新聞記事においても少なくない用例が見出されるものと思われる。

名詞述語文については、他にも注目すべき点がある。代名詞は新聞記事での用例が多くないため、代名詞の合計だけを母数として各述語との共起関係の比率を見ても、主述関係全体の中での位置づけが見えてこないという問題がある。そこで、すべての主述関係の用例を母数として、各主述パターンの割合を示すと、以下のようになる。

表7 青空文庫における全体を母数とした主語名詞と述語の比率

	動詞述語	イ形容詞述語	ナ形容詞述語	名詞述語	合計
普通名詞	2547 (41.7%)	301 (4.9%)	88 (1.4%)	386 (6.3%)	6112 (100.0%)
固有名詞	1248 (20.4%)	33 (0.5%)	26 (0.4%)	65 (1.1%)	
代名詞	1144 (18.7%)	53 (0.9%)	44 (0.7%)	177 (2.9%)	

表8 新聞記事における全体を母数とした主語名詞と述語の比率

	動詞述語	イ形容詞述語	ナ形容詞述語	名詞述語	合計
普通名詞	1321 (54.4%)	146 (6.0%)	75 (3.1%)	347 (14.3%)	2429 (100.0%)
固有名詞	444 (18.3%)	4 (0.2%)	5 (0.2%)	45 (1.9%)	
代名詞	34 (1.4%)	0 (0.0%)	1 (0.0%)	7 (1.7%)	

表7と表8は、それぞれのテキストジャンルの全体の用例数（表3と表4の「合計」

を足し合わせたもの)を母数として、改めて各主述パターンの比率を算出したものである。ここで目を引くのは、「代名詞+名詞述語」の組み合わせである。母数を各テキストジャンルの合計としても、名詞述語は新聞記事に多く現れる傾向が読み取れるが、興味深いことに、代名詞に関しては青空文庫のほうがより相対的に多く出現するという結果となっている(青空文庫:2.9%/新聞記事:1.7%)。これは、普通名詞や固有名詞と異なり、代名詞と名詞述語の組み合わせは全主述パターンとの関係において見ると、説明的文章よりもむしろ文学的文章において相対的に多く用いられていることを意味している。以下、代名詞と名詞述語の関係を中心に、このような分布が生じている要因を分析する。

青空文庫で見られる「代名詞+名詞述語」のパターンの例として、以下のようなものが挙げられる。

- (14) a. それは私も一面識のある森於菟彦さんだった。 (79_aozora_Hori_1969)
 b. 私は、よくよく不幸な男だ。 (257_aozora_Dazai_2_1940)
 c. 彼女は重症だった。 (121_aozora_Yamakawa_1962_1)

青空文庫における「代名詞+名詞述語」のパターンの特徴は、(14a)のような指示代名詞を主語としている例が相対的に多いという点である。無論、(14bc)のような人称代名詞の例も用いられるが、指示代名詞が主語のパターンに比べれば少ない傾向にある。

以下の表は、青空文庫における「代名詞+名詞述語」と「代名詞+動詞述語」という主述パターンについて、指示代名詞¹³とその他の代名詞が主語になった場合の用例数を比較したものである。

表9 青空文庫における指示代名詞の用例数

	指示代名詞主語	その他代名詞主語	合計
代名詞+名詞述語	122 (68.9%)	55 (31.1%)	177 (100.0%)
代名詞+動詞述語	61 (5.3%)	1083 (94.7%)	1144 (100.0%)

「代名詞+名詞述語」の場合には代名詞主語の半数以上を指示代名詞が占めているのに対し、「代名詞+動詞述語」の場合は、指示代名詞が1割にも満たない。このように、青空文庫に現れる代名詞主語は指示代名詞が相対的に多い。

では、このような分布はどのような要因によって生じているのだろうか。結論を先取りすると、指示代名詞を主語とする過去形の名詞述語文に、物語(文学的文章)を展開させるうえで有効な機能が備わっているからであると考えられる。この点について

て議論するために、述語のテンス形式について整理しておく。ここまでは詳細に取り上げなかったが、青空文庫で見られる述語として、新聞記事に比べて過去形の例が多いという特徴が挙げられる。以下に各主述パターンにおける過去形と非過去形の分布を示す¹⁴。

表 10 青空文庫における各主述パターンとテンス形式

	普通名詞			固有名詞			代名詞		
	過去	非過去	合計	過去	非過去	合計	過去	非過去	合計
動詞述語	1750 (68.7%)	797 (31.3%)	2547 (100.0%)	1073 (86.0%)	175 (14.0%)	1248 (100.0%)	906 (79.2%)	238 (20.8%)	1144 (100.0%)
イ形容詞述語	163 (54.2%)	138 (45.8%)	301 (100.0%)	25 (75.8%)	8 (24.2%)	33 (100.0%)	29 (54.7%)	24 (45.3%)	53 (100.0%)
ナ形容詞述語	51 (58.0%)	37 (42.0%)	88 (100.0%)	17 (65.4%)	9 (34.6%)	26 (100.0%)	18 (40.9%)	26 (59.1%)	44 (100.0%)
名詞述語	203 (52.6%)	183 (47.4%)	386 (100.0%)	45 (69.2%)	20 (30.8%)	65 (100.0%)	83 (46.9%)	94 (53.1%)	177 (100.0%)

表 11 新聞記事における各主述パターンとテンス形式

	普通名詞			固有名詞			代名詞		
	過去	非過去	合計	過去	非過去	合計	過去	非過去	合計
動詞述語	495 (37.5%)	826 (62.5%)	1321 (100.0%)	176 (39.6%)	268 (60.4%)	444 (100.0%)	0 (0.0%)	34 (100.0%)	34 (100.0%)
イ形容詞述語	28 (19.2%)	118 (80.8%)	146 (100.0%)	1 (25.0%)	3 (75.0%)	4 (100.0%)	0 (0.0%)	0 (0.0%)	0 (100.0%)
ナ形容詞述語	8 (10.7%)	67 (89.3%)	75 (100.0%)	1 (20.0%)	4 (80.0%)	5 (100.0%)	0 (0.0%)	1 (100.0%)	1 (100.0%)
名詞述語	31 (8.9%)	316 (91.1%)	347 (100.0%)	2 (4.4%)	43 (95.6%)	45 (100.0%)	1 (14.3%)	6 (85.7%)	7 (100.0%)

固有名詞と代名詞の新聞記事における用例はそれほど多くないため一部比較が難しいが、いずれの主述パターンにおいても青空文庫のほうが過去形が多く現れており、新聞記事のほうが過去形が多いという組み合わせは1つもない。このことから、青空文庫においては相対的に多くの過去形述語が用いられていると言える。

過去形述語との組み合わせにおいて、本節で問題にしている代名詞を主語とする名詞述語文は興味深い振る舞いを見せる。先ほど挙げた(14a)「それは私も一面識のある森於菟彦さんだった」は、叙想的テンス(ここではいわゆる発見の夕)の例であるが、他にも類例をいくつか挙げておく。

- (15) a. 嘉助はしばらく自分の目を疑って立ちどまっていましたが、やはりどうしても家らしかったので、こわごわもっと近寄って見ますと、それは冷たい大きな黒い岩でした。(483_aozora_Miyazawa_1934)
- b. 送ってくれた写真を見ると、メリー達のそばに、まぬけづらをした人間が写っていて、それはどうやら私のようであった。

c. 要するにそれは一種の甘い水薬であったのである。

(77_aozora_Terada_1929)

例えば、(15a)「冷たい大きな黒い岩でした」という述語は、近寄って見て確認した事実をもとになされた判断であり、単純な過去の意味ではなく叙想的テンス（発見のタ）と解釈するのが自然である。他の例も同様に、指示代名詞「それ」を主語とする過去形述語の例は、いずれも「確認してみると……だった」のような意味合いで解釈されるように思われる。このような事実は、内省判断によっても確認することができる。

(16) a. それは大きな施設だった。

b. そこは険しい山道だった。

ソ系指示詞を用いた(16)は、いずれも過去ではなく「確認してみると……だった」といった発見のタと解釈できる¹⁵。つまり、ソ系の指示代名詞と過去形の名詞述語文が結びつくと、叙想的テンスの解釈（発見の「タ」）に傾きやすいのである。ソ系の指示代名詞のこのような性質は、おそらく指示代名詞が持つ現場指示性が関わっている。指示詞によって現場指示ができるということは、指示詞の使用者は何らかの形で対象を知覚することができていることになる。そのような状況で名詞述語文の過去形を使用すれば、眼前に知覚していることを過去と関連づけることになり、必然的に叙想的テンスの解釈になるのである¹⁶。

「ソ系指示詞＋過去形の名詞述語」という組み合わせが持つ以上のような性質は、しばしば登場人物の視点から場面の転換や事態の把握を描写する文学的文章において、有効な表現手段になりうる。(15)の例はいずれも登場人物の視点による探索の結果、当該の対象が発見されたことが臨場感をもって描写されており、発見の「タ」が効果的に使用されている。すなわち、ソ系の指示代名詞を主語とする過去形の名詞述語文は、登場人物の視点から捉えた場面転換や事態把握を描写して物語を展開させる効果があり、青空文庫などの文学的文章において高頻度に現れるために、「代名詞＋名詞述語」という主述パターンの用例数の相対的な増加に繋がっているのだと考えられるのである。

5. おわりに

本稿では、KTBを用いてテキストジャンルと特定の統語構造との関連について調査・分析を行った。従来の分析においても形容詞が説明的文章において相対的に少な

くなることが指摘されていたが、本調査では述語環境を通じてその点を明らかにしたと言える。また、叙想的テンスについて、ソ系指示代名詞との関係において従来指摘の無かった特徴を示したほか、このような文法的概念を、量的な観点から文学的文章の文体と結びつけているという点でも重要であると考えられる。一方で、今回調査を行った範囲のデータに限っても、テンス表現と主述パターンの関係性など、十分に検討できたとは言えない論点も残されている。今後は名詞修飾表現など、従属的な環境における述語の使用と主節における使用なども比較し、より広範な調査と分析を試みたい。

付記

本稿は、日本語学会 2019 年度秋季大会（於：東北大学）でのブース発表がもとになっている。また、本研究は JSPS 科研費 23K12174 の助成を受けたものである。

- 1 「^」は前方一致、「\$」は後方一致を表し、これらを合わせると完全一致指定となる。完全一致指定をするのは、例えば「/VB/」のような部分一致表現を用いて「VB2」のような意図しないノードを検索対象としてしまうのを避けるためである（動詞とその補助動詞の用法を問わず全てを検索対象としたい場合には、「/VB/」で問題ない）。
- 2 Miyoshi (2021) では、テキストジャンルと非制限的連体修飾節の出現頻度の間に一定の相関があることを指摘している。
- 3 三好 (2020) では、KTB と関連するプロジェクトで開発された NINJAL Parsed Corpus of Modern Japanese (NPCMJ) を調査資料として用いており、検索式等に軽微な差異があるが、基本的な調査方法は同じである。
- 4 「彼は日本語に詳しい」などの二格をとる形容詞もあるが、全体としては多くないと考えられる。
- 5 KTB において「私には英語が分かりません（動詞述語）／私に妹はいません（形容詞述語）」のような二格が主語として扱われていることを踏まえると、形態的に見れば必ずしも均質とは言えない。一方で、このような例は述語全体に対してそれほど多くないと考えられるため、ここでは問題にしない。
- 6 「ツリー」はおおむね「文」という単位に相当するが、見出し語なども 1 ツリーとなるため、主述を備えた、いわゆる「文」の数とは必ずしも一致しない。
- 7 KTB では、いわゆる準体助詞の「の」が普通名詞（[N]）として扱われており、その点でも普通名詞の用例数は多くなっている可能性が高い。準体助詞の「の」がいずれの述語とも用いられることから、本稿では除外せず調査対象に含めたが、「財布を拾ったのは私です」のような分裂文の例が多い可能性などを踏まえると、「の」を除外することで多少異なった分布になる可能性はある。
- 8 これが三人称視点の物語となれば、主語は固有名詞などが用いられることになると考えられるため、その点で文学的文章に現れる人称表現は「一人称：代名詞／三人称：固有名詞」のようなある種の補完的な関係にある可能性もある。この点については、さらに実証的な調査を進めなければならないだろう。

- 9 もっとも、このような特徴は本研究で扱った説明的文章が新聞記事であったために観察されたものである。そのため、同じ説明的な文章であっても随筆や随想などの場合には、異なった特徴が現れる可能性もある。
- 10 これらの点を踏まえると、出現頻度のうえではそれほど違いの無いように見える青空文庫と新聞記事の固有名詞の分布は、かなり性質の異なるものであることが分かる。
- 11 用例の鉤括弧は原文ママ。
- 12 なお、樺島（1954：18）は新聞記事等で名詞の比率が大きくなるのは、名詞が「主要な概念項である」からであると述べているが、この主張は本稿の分析と矛盾するものではなく、むしろ補完的なものであると考える。新聞記事であっても、「心強い味方」のように述語が名詞修飾となる環境においては形容詞の用例数が増加する（普通名詞が被修飾名詞の場合、全体 2491 例に対し「動詞：1229（49.3%）／イ形容詞：310（12.4%）／ナ形容詞：533（21.4%）／名詞：419（16.8%）」であったことを踏まえると、「主要な概念項」であるかどうかという点から離れて、形容詞を用いた言い切りの叙述ないし判断を行うことが新聞記事において避けられていることが伺える。このことは、単に品詞の語彙数だけでなく統語環境もテキストの特性を明らかにするうえで重要であることを強く示唆するが、詳細は今後明らかにしたい。
- 13 具体的には、「これ・これら・ここ・こちら／それ・それら・そこ・そちら／あれ・あれら・あそこ・あちら」を対象としている（ただし、今回の調査では確認できなかったものもある）。「こいつ／そいつ／あいつ」については、人称代名詞とも捉えられるため対象から除外している。
- 14 過去形の用例数を調査する際には、次のような検索式を用いた（「普通名詞＋動詞述語」の例）。
 (i) $\text{/}^{\wedge}\text{N}\$ / > \text{/}^{\wedge}\text{NP} / > \text{/PP-SBJ} / > \text{((/IP-MAT} / < \text{/}^{\wedge}\text{VB}\$ / < \text{/}^{\wedge}\text{AXD}\$ / \text{))}$
- 15 コ系やア系では、少し違った傾向になる可能性がある。
 (i) a. ここは自分の家だった。
 b. あそこは自分の家だった。
 コ系はいわゆる想起のタ、ア系は過去の（通常の）タのように解釈できるが、詳細については今後の課題としたい。
- 16 叙想的テンスと過去の意味との関連については、金水（1998）、井上（2001）、定延（2004）、三好（2019）などを参照されたい。

【参考文献】

- 井上優（2001）「現代日本語の「タ」—主文末の「…タ」の意味について—」つくば言語文化フォーラム（編）『「た」の言語学』pp. 97-163. ひつじ書房。
- 樺島忠夫（1950）「類別した品詞の比率に見られる規則性」『国語国文』pp. 55-57.
- 樺島忠夫（1954）「現代文における品詞の比率とその増減の要因について」『国語学』pp. 15-20.
- 金水敏（1998）「いわゆる‘ムードの「タ」’について」東京大学国語研究室創設百周年記念 国語研究論集 編集委員会（編）『東京大学国語研究室創設百周年記念 国語研究論集』pp. 170-185, 汲古書院。
- 定延利之（2004）「ムードの「た」の過去性」『国際文化学研究 神戸大学国際文化学部紀要』21, pp. 1-68.

三好伸芳 (2019) 「叙想的テンスの意味と統語」 竹沢幸一他 (編) 『日本語統語論研究の広がり—記述と理論の往還—』 pp.107-126, くろしお出版.

三好伸芳 (2020) 「統語情報付きコーパスによる名詞句と述語の共起関係の分析」 『KLS Selected Papers』 2, pp. 215-226.

Miyoshi, Nobuyoshi (2021) Nonrestrictive Adnominal Clauses and Text Genre in Japanese. *NINJAL Research Papers* 21: pp. 163-178.

【調査資料】

Kainoki, Ed (2022) “The Kainoki Treebank – a parsed corpus of contemporary Japanese”
<https://kainoki.github.io> (アクセス日：2024/1/17) .