

現代日本語の従属節に現れるモダリティ形式の分布

丸山 岳彦 (国立国語研究所 言語資源研究系)[†]

Distribution of Modal Expressions in Japanese Subordinate Clauses

Takehiko Maruyama (Dept. Corpus Studies, NINJAL)

1 はじめに

2011年に『現代日本語書き言葉均衡コーパス』(BCCWJ)が一般公開されたことにより、さまざまなレジスター(言語使用域、言語変種)における言語使用の実態を定量的に分析するための基盤が整えられ、どのような場面でどのような言語表現が使われるか(あるいは使われないか)を知ることができるようになった。例えば、言語の使用場面が変わることによって、複数のバリエーションを持つ文法形式の使用傾向に変化が見られるか否かを定量的に観察できるようになった。また、多様なレジスターを含むBCCWJを参照コーパスとして利用し、これまでに行なわれてきた定性的・定量的な文法記述の結果を検証したり、追試をしたりすることも可能になった。このような状況の出現は、母語話者の内省に基づく従来の記述的文法研究に対して、言語使用の実態に即した「現実的な文法研究」を大きく前進させる可能性をもたらしたとすることができる。

そこで本稿では、現代日本語における複文を構成する従属節のうち、特に連用節として機能する従属節がさまざまなレジスターの中でどのように出現しているかを定量的に分析することを目的とする。同時に、連用節を構成する述語句内にどのようなモダリティ形式が現れるかについても検討する。現代日本語文法の研究において、連用節とモダリティ形式の共起関係という問題はさまざまな角度から論じられてきたが、本稿ではBCCWJに基づいて、どのようなレジスターでどのような文法形式が観察されるかという点について、先行研究の検証・追試を交えながら論じていくことにしたい。

2 先行研究

日本語の述語句内部にどのようなモダリティ形式が現れ得るか、という問題については、国語学における助動詞の相互承接から見た構文論的研究(渡辺, 1953; 北原, 1970)や、三上章による活用形の研究(三上, 1953)などを背景として、現代における記述的文法研究の中に受け継がれている(仁田, 1991; 益岡, 1991, 2007)。また、連用節の述語句に現れ得るモダリティ形式の範囲を手がかりとして、連用節を3つのクラスに分類した南不二男の研究(南, 1974, 1993)により、文の階層的な成立段階がモデル化され、この枠組み(通称「南モデル」)によって多様な文法現象を説明できることが明らかにされた(田窪, 1987; 野田, 1995)。

一方、定量的な観点からは、南(1991)がさまざまな連用節の内部に取り得る要素(主題、補足語、修飾語、助動詞類)の出現数を調査し、数量的な分析を行なっている。ただし、調査資料の規模が少なかったためか、モダリティ形式を含む従属節(ガ、ケレドモ、カラ、ノデ、ナラ、タラなど)の用例数が十分に拾えているとは言い難く、調査では専ら、連用節内にある主題や補足語、修飾語が主要な分析対象となっている。

時代が下り、電子テキストが一般に普及した2000年代以降では、ナロック(2006)が、各種電子ブックや「CASTEL/J」などの電子テキストを大量に準備してブレンドし、そこから17万例以上におよぶ接続助詞の用例を検索・抽出して、連用節内部に現れる各モダリティ形式の分布を詳細に分析している。

[†] maruyama@ninjal.ac.jp

ナロック (2006) は、対比節 (ガ、ケレドモ)・理由節 (カラ・ノデ)・逆接節 (ノニ・ニモカカワラズ) という 3 グループ、合計 6 種類の連用節を対象として、そこに現れるモダリティ形式を調査した。調査したモダリティ形式は、「認識的モダリティ」「証拠的モダリティ」「根源的モダリティ」の 3 グループ、各 4 形式の、12 種類である。ナロックによる調査の結果をまとめ直したものを、表 1 に示す。なお、論文中で、出現数は 10000 語あたりの数に正規化されている。

表 1: ナロック (2006) による調査結果のまとめ (1)

| | | 対比節 | | 理由節 | | 逆接節 | |
|--------------|-----------|-----|-----|-----|----|-----|-------|
| | | ガ | ケレド | カラ | ノデ | ノニ | ニモ... |
| 認識的 モダリティ | ニチガイナイ | 36 | 67 | 111 | 20 | 39 | 69 |
| | ハズ | 37 | 60 | 83 | 8 | 375 | 46 |
| | ダロウ | 29 | 67 | 42 | 0 | 49 | 0 |
| | カモシレナイ | 96 | 130 | 51 | 18 | 15 | 0 |
| 証拠的 モダリティ | ソウ 1 (様相) | 14 | 11 | 37 | 22 | 12 | 0 |
| | ソウ 2 (伝聞) | 67 | 72 | 139 | 2 | 0 | 0 |
| | ラシイ | 38 | 79 | 45 | 26 | 6 | 9 |
| | ヨウ・ミタイ | 66 | 78 | 45 | 18 | 4 | 0 |
| 根源的 モダリティ | ナケレバナラナイ | 18 | 9 | 90 | 39 | 66 | 0 |
| | タイ | 145 | 37 | 57 | 18 | 24 | 2 |
| | テモイイ | 31 | 164 | 205 | 4 | 125 | 0 |
| | ベキ | 6 | 5 | 4 | 0 | 43 | 3 |

連用節の種類とモダリティ形式の種類を一つにまとめた上で、10000 語あたりの出現数を算出したナロックの調査結果を図式化すると、図 1 のようなグラフとなる。

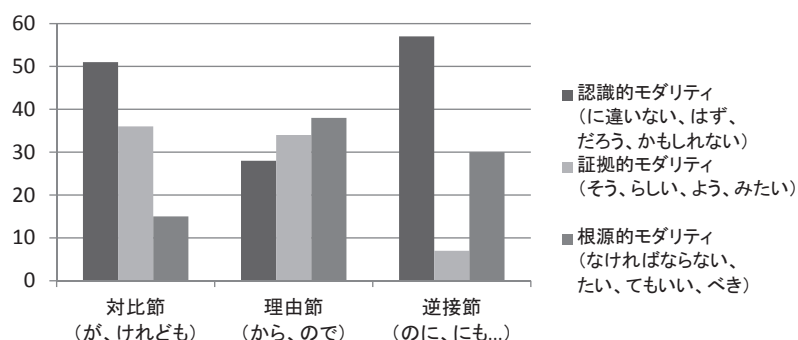


図 1: ナロック (2006) による調査結果のまとめ (2)

ここからナロックは、理由節では「根源 → 証拠 → 認識」の順で使用頻度が下がり、対比節では「認識 → 証拠 → 根源」の順で使用頻度が下がる、すなわち、理由節と対比節におけるモダリティ形式の使用傾向には逆の関係があることを指摘している。ナロック自身が述べている通り、使用したデータの集合が日本語のどの側面を代表しているのかが明確でないという問題点はあるものの、17 万例以上という大量の用例を集計し、各連用節が取るモダリティ形式の分布が明らかに異なることを実証したという点において、コーパス日本語学としての優れた実践例とすることができる。

3 本稿の目的

コーパスを利用して言語の使用実態・使用傾向を調査することの利点は、母語話者の内省では知り得ない言語事実を実証的に明らかにすることができる点にある。文法研究に即して言えば、どのような場面でどのような文法形式がどれだけ使われるか (または使われないか) を記述することは、現実の発話状況や使用場面をも取り込んだ、より「現実的な」文法記述への可能性を開くことになる。

これは、翻れば、「頻度情報付きの日本語記述文法書」が存在しないという現状の問題に結びつく。英語の記述文法書（レファレンスグラマー）としては、Quirk et al. (1972, 1985) や Biber et al. (1999)、Huddleston and Pullum (2002)、Carter and McCarthy (2006) など、コーパスに基づく大規模な記述文法書がすでに数多く公刊されている。一方、体系立てられた日本語記述文法書としては、益岡・田窪 (1992) や「日本語記述文法研究会」による7巻組の記述文法書（日本語記述文法研究会 (2010) など）があるものの、日本語コーパスで観察されたレジスターごとの頻度情報を正面から取り上げて掲載した日本語記述文法書は、まだない。

そこで本稿では、言語の使用場面によって文法形式の分布がどのように異なるかを記述することを目的として、連用節とモダリティ形式の分布について分析を行なう。BCCWJ を分析対象データとして、異なる性質の書き言葉の中でどのような連用節・モダリティ形式が現れるかという問題を考えた。分析の立脚点として、先に見たナロック (2006) を踏襲する。すなわち、ナロック (2006) における調査方法を BCCWJ における各レジスターに適用し、先行研究の検証・追試としながら、議論を進めることにする。

4 分析対象データ

4.1 『現代日本語書き言葉均衡コーパス』(BCCWJ)

ここでは、『現代日本語書き言葉均衡コーパス』(BCCWJ) の DVD 版に収録された長単位データを用いる。分析対象は、『BCCWJ-DVD 版』に収録された形態論情報付きテキストデータのうち、「韻文」を除く全体とした。韻文（俳句・短歌・詩）は、通常 of 書き言葉（いわゆる散文）とは明らかに異なる文体を持つため、連用節やモダリティを定量的に分析する上で、対象から除外することが適当であると判断した。タブ区切りの表形式になっている形態論情報付きテキストデータを Microsoft SQL Server に格納し、RDB に対してクエリを実行して用例の抽出と分類を行なった。

『BCCWJ-DVD 版』に収録されたデータ全体の語数を、レジスターごとに表 2 に示す。なお、本稿での「語数」とは、長単位として解析された単位数を指すものとする。

表 2: BCCWJ に収録された語数（長単位、概数）

| 出版サブコーパス | | 特定目的サブコーパス | |
|-----------|-----------|----------------|-----------|
| 書籍 (PB) | 2762.3 万語 | 白書 (OW) | 382.4 万語 |
| 雑誌 (PM) | 428.4 万語 | 教科書 (OT) | 92.5 万語 |
| 新聞 (PN) | 122.1 万語 | 広報紙 (OP) | 309.0 万語 |
| | | ベストセラー (OB) | 386.3 万語 |
| | | Yahoo!知恵袋 (OC) | 1030.0 万語 |
| | | Yahoo!ブログ (OY) | 1093.3 万語 |
| | | 法律 (OL) | 83.3 万語 |
| | | 国会会議録 (OM) | 449.8 万語 |
| 図書館サブコーパス | | | |
| 書籍 (LB) | 3027.4 万語 | | |

4.2 連用節の抽出手順

本稿で分析対象とする連用節は、ナロック (2006) での分析対象を一部入れ替えて、以下の 13 種類の形式とした。便宜的に、「並列節」「理由節」「条件節」という 3 つのグループに分類しておく。

並列節：ガ、ケレドモ、ケドモ、ケレド、ケド

理由節：カラ、ノデ

条件節：タラ、タラバ、ト、ナラ、ナラバ、レバ

ナロック (2006) で論じられていた「逆接節 (ノニ、ニモカカワラズ)」は、「条件節 (タラ、タラバ、ト、ナラ、ナラバ、レバ)」に入れ替えた。これは、条件節が日常の書き言葉の中に多く観察される代表的な連用節の一部であると考えられること、南モデルの B 類・C 類に分類される連用節であり、モダリティ形式との共起関係がしばしば問題とされること、大規模コーパスから逆接節としての「ノニ」を自動的に収集することが困難であること、などの理由による。

各形式の連用節について、RDB に格納された BCCWJ のデータ全体から用例を取り出した。その際、当該の語を中心として、前接する 10 語、後接する 10 語を結合し、前後 10 語ずつの KWIC 形式のデータを作成した。各語には、長単位として付与されている形態論情報のうち、書字形出現形 (OT)、語彙素 (LM)、品詞 (POS)、文節境界 (B) の情報を持たせた。KWIC データから一部を抜粋した例を、図 2 に示す。

| Sample_ID | order | CBL | B2 | OT02 | POS02 | B1 | OT01 | POS01 | OT00 | POS00 | OT_1 | POS_1 | B_2 | OT_2 | POS_2 |
|-----------|------------|-------|----------|------|-------|--------|------|-------|------|---------|------|--------|-----|------|----------|
| 1 | PB13_00019 | 24220 | keredomo | B | あっ | 動詞一般 | た | 助動詞 | けれど | 助詞-接続助詞 | も | 助詞-係助詞 | B | 書か | 動詞一般 |
| 2 | PB13_00045 | 14940 | keredomo | | ませ | 助動詞 | ん | 助動詞 | けれど | 助詞-接続助詞 | も | 助詞-係助詞 | | | 補助記号-読点 |
| 3 | PB13_00045 | 11590 | keredomo | B | ごさい | 動詞一般 | ます | 助動詞 | けれど | 助詞-接続助詞 | も | 助詞-係助詞 | | | 補助記号-読点 |
| 4 | PB12_00356 | 34480 | keredomo | | な | 助動詞 | んだ | 助動詞 | けれど | 助詞-接続助詞 | も | 助詞-係助詞 | | | 補助記号-読点 |
| 5 | PB12_00356 | 19560 | keredomo | B | あり | 動詞一般 | ます | 助動詞 | けれど | 助詞-接続助詞 | も | 助詞-係助詞 | | | 補助記号-読点 |
| 6 | PB12_00356 | 13230 | keredomo | | な | 助動詞 | んです | 助動詞 | けれど | 助詞-接続助詞 | も | 助詞-係助詞 | B | く | 補助記号-括弧閉 |
| 7 | PB13_00045 | 18490 | keredomo | | は | 助詞-係助詞 | わかる | 動詞一般 | けれど | 助詞-接続助詞 | も | 助詞-係助詞 | | | 補助記号-読点 |
| 8 | PB13_00045 | 19580 | keredomo | B | 思う | 動詞一般 | のです | 助動詞 | けれど | 助詞-接続助詞 | も | 助詞-係助詞 | | | 補助記号-読点 |
| 9 | PB13_00045 | 22570 | keredomo | B | 思い | 動詞一般 | ます | 助動詞 | けれど | 助詞-接続助詞 | も | 助詞-係助詞 | | | 補助記号-読点 |
| 10 | PB13_00045 | 23710 | keredomo | | ませ | 助動詞 | ん | 助動詞 | けれど | 助詞-接続助詞 | も | 助詞-係助詞 | | | 補助記号-読点 |

図 2: BCCWJ から作成した KWIC データの例 (一部抜粋)

ただし、この状態では、連用節以外の用例が KWIC データに含まれている。例えば、文節の冒頭に出現する「けれど」や「ですので」、「なら」、「そうすると」といった形式は、実質的には接続詞に相当するが、形態素解析の結果としては接続詞ではなく、「だ_助動詞 / けど_接続助詞」「です_助動詞 / ので_助動詞」「なら_助動詞」「そう_副詞 / する_動詞 / と_接続助詞」として解析されている(かつ、これは設計上、正しい解析結果である)。これらは、形式的には連用節の検索条件に適合するため、連用節の用例として抽出されてしまうことになる。そこで、以下に示すような表現が文節の始端に位置している場合は、接続詞に相当する用例と見なし、分析データから除外した。

そうすると、そうすれば、そしたら、なら、そうしたら、けれど、ですので、だとしたら、だが、だとすれば、だったら、だとすると、ですが、なので、だから、でしたら、すると、であれば、ですから、であるなら、けど

4.3 モダリティ形式の抽出手順

次に、連用節の述語句内に現れるモダリティ形式を抽出した手順を示す。ここでは、以下に示す 13 種類のモダリティ形式を検討する。これは、ナロック (2006) が「認識的モダリティ」「証拠的モダリティ」「根源的モダリティ」という 3 つのグループに分類して検討したモダリティ形式に対応する。

認識的モダリティ: だろう、かもしれない、はずだ、にちがいない

証拠的モダリティ: (し) そうだ、ようだ、みたいだ、らしい、(する) そうだ

根源的モダリティ: なければならない、たい、てもいい、べきだ

図 2 に示した KWIC データからこれらのモダリティ形式を抽出するために、検索用クエリを作成した。上記に挙げた各形式を基本形として、一部が漢字表記になる場合(「かもしれない」→「かも

知れない」など)、丁寧形になる場合(「だろう」→「でしょう」など)、過去形になる場合(「はずだ」→「はずだった」など)、否定形になる場合(「べきだ」→「べきではない」など)、交替形を取る場合(「なければならない」→「ないといけない」「なくちゃならん」など)をカバーするように検索式を作成した。

なお、形態素解析の結果によっては、同じモダリティ形式が異なる結果で解析されている場合がある。たとえば「かもしれません」というモダリティ形式は、「かもしれません_助動詞」だけでなく、「か_副助詞 / も_係助詞 / しれ_動詞 / ませ_助動詞 / ん_助動詞」と解析されている場合があり得る。これらの異なりを可能な限りカバーするように検索式を調整した。

さらに、形態素解析に伴う必然的な問題として、一定の割合で誤解析が含まれる、という点に注意する必要がある。例えば、「もっと長くいたかったのだが」というガ節の例には、願望を表すモダリティ形式「たい」が現れているが、実際の解析結果は「もっと_副詞 / 長く_形容詞 / いたかつ_形容詞 / た_助動詞 / のだ_助動詞 / が_接続助詞」となっており、「いたい」の部分が形容詞「痛い」として誤解析されてしまっている。このような場合は「たい」の用例としては抽出できないことになる。BCCWJ全体の解析精度は98%という高い精度を達成しているが(国立国語研究所, 2011)、それでも一定の割合で誤解析が含まれること、それに伴うモダリティ形式の抽出漏れがあることは不可避的な前提として、以降の分析を進める。

5 分析：連用節とモダリティ形式の分布

5.1 連用節とモダリティ形式の出現数

以下では、BCCWJから抽出された連用節の各形式、およびモダリティ形式の分布について、分析を行なう。まずはじめに、分析対象データから抽出された連用節全体の出現数、および連用節中に現れたモダリティ形式の出現数を、表3、4に示す。

表 3: 連用節の出現数

| 分類 | 形式 | 出現数 |
|-----|------|-----------|
| 並列節 | ガ | 383,523 |
| | ケレドモ | 21,014 |
| | ケドモ | 695 |
| | ケレド | 10,741 |
| | ケド | 58,607 |
| 理由節 | カラ | 159,187 |
| | ノデ | 137,430 |
| 条件節 | タラ | 99,886 |
| | タラバ | 136 |
| | ト | 237,910 |
| | ナラ | 22,997 |
| | ナラバ | 6,253 |
| | レバ | 169,252 |
| 総計 | | 1,307,631 |

表 4: 連用節内部のモダリティ形式の出現数

| 分類 | 形式 | 出現数 |
|-----|----------|--------|
| 認識的 | だろう | 9,731 |
| | かもしれない | 8,641 |
| | はずだ | 2,581 |
| | にちがいない | 393 |
| 証拠的 | (し) そうだ | 2,754 |
| | ようだ | 7,162 |
| | みたいだ | 1,525 |
| | らしい | 2,442 |
| | (する) そうだ | 1,936 |
| 根源的 | なければならない | 2,932 |
| | たい | 12,891 |
| | てもいい | 1,659 |
| | べきだ | 630 |
| 総計 | | 55,277 |

表3のうち、並列節について見ると、ガの出現数がケレドモ類(ケレドモ、ケドモ、ケレド、ケド)の合計数を大きく上回っている。両者とも、基本的には逆接的または対比的な関係を表す並列節であるが、書き言葉の中ではガが圧倒的に多く選ばれているということになる。これは、ナロック(2006)で報告された結果と同様である。また、ケレドモ類に含まれる4形式の出現数を比較すると、

ケドが最も多く、ケレドモ、ケレドの順に続き、ケドモの出現数は極めて少ない。次に理由節について見ると、カラの方がノデよりも若干多いものの、出現数の上では、両者に大きな開きはない。これもナロック (2006) での報告と符号する。さらに条件節について見ると、ト、レバ、タラ、ナラの順に出現数が多くなっている。条件節を導くこれら4形式が持つ文法的性質の異同については、従来の文法研究の中で盛んに取り上げられてきたテーマであるが (益岡, 1993)、書き言葉における出現数という点から見ると、4つの形式に対してこのような順列を与えることができる。なお、ナラに対するナラバの出現数に対して、タラに対するタラバの出現数は極めて少なくなっている。

一方、表4に示したモダリティの各形式の出現数については、ナロック (2006) での報告と大きく異なる分布となっている。ナロック (2006) では、各連用節におけるモダリティ形式の出現数として、「カモシレナイ → ソウ2 → ヨウ・ミタイ → ハズ → テモイ → ニチガイナイ → ラシイ → ダロウ → ナケレバナラナイ → タイ → ソウ1 → ベキ」という順列が挙げられているが、今回 BCCWJ から抽出した結果とはかなりの差異がある。このような出現数の分布の違いが、使用したコーパスの性質によるものかどうかは、現段階では不明である。

5.2 各レジスターにおける連用節の分布

次に、BCCWJを構成する各レジスターに連用節がどのように分布しているかを見る。ここでは、各レジスターにおける連用節の出現数を100万語あたりの出現数に正規化した上で、比較を行なった。各レジスターにおいて100万語あたりに出現する並列節・理由節・条件節の出現数を、図3に示す。

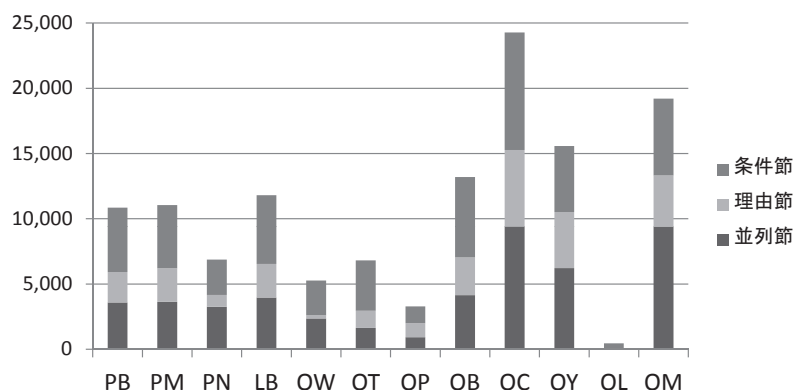


図3: 各レジスターにおける連用節の分布 (100万語あたり)

図3を見ると、Yahoo!知恵袋 (OC)、国会会議録 (OM)、Yahoo!ブログ (OY) の順に、連用節全体の出現数が多くなっていることが分かる。これらは、一般人の書いた比較的くだけた書き言葉 (OC、OY)、または話し言葉を転記した書き言葉 (OM) である。一方、連用節の出現数が少ないのは法律 (OL)、広報紙 (OP)、白書 (OW) であり、これらは公共性の高い書き言葉である。ここから、書き言葉のスタイルの違いが連用節の出現に影響を与えていることが推測される。すなわち、比較的くだけたスタイルの書き言葉では連用節が多用されるのに対し、硬いスタイルの書き言葉では連用節はあまり使用されない、という傾向を見て取ることができる。この結果は、テキストの特徴的な傾向として、前者では多重的な節連鎖構造が発生することによってダラダラと続く長文や接続助詞で終わる「言いさし文」が現れやすいことを、後者では比較的短い文の連続が現れやすいことを、それぞれ予測させる。

5.3 連用節内部のモダリティ形式の分布

以下では、連用節の各形式の述語句部分に現れたモダリティ形式の分布について分析を行なう。はじめに、各連用節の全数のうち、各モダリティ形式を含む用例を集計した結果を、図4に示す。

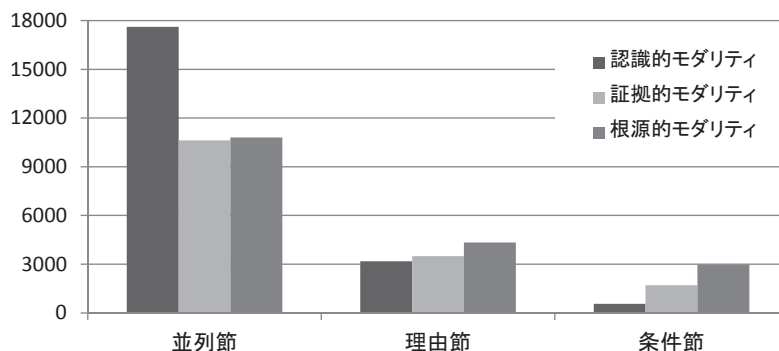


図4: モダリティ形式を含む連用節の分布 (1)

図4を、図1で示したナロック (2006) による調査結果をと比べると、大まかには似た分布が観察される。ただし、(1) 並列節において根源的モダリティが証拠的モダリティを上回っている点、(2) 理由節の出現率が全体的に低く押さえられている点、という2つの差を見出すことができる。これは、5.1節の最後に述べた、モダリティ形式の分布そのものがナロック (2006) と本稿との間で大きく異なるという点に起因している可能性がある。

次に、連用節の各形式について、その内部にどのような種類のモダリティを取るかを集計した。各モダリティ形式を含む割合の高い順に連用節を並べ替えた結果を、図5に示す。

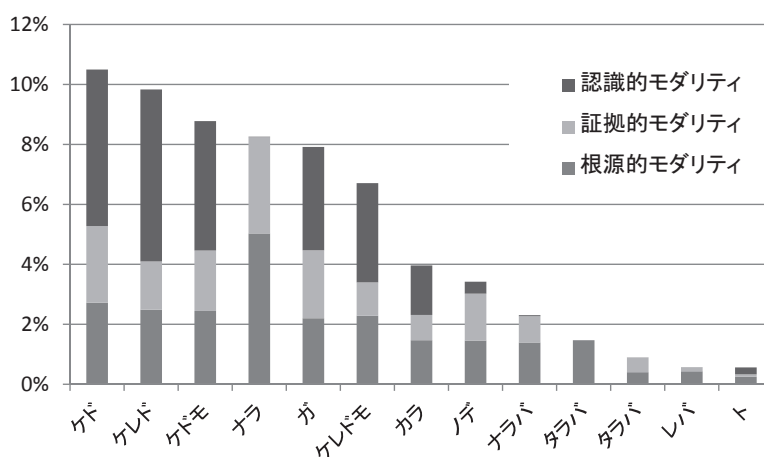


図5: モダリティ形式を含む連用節の分布 (2)

図5を見ると、モダリティ形式を含む比率の高いから低いものに移るにしたがって、並列節、理由節、条件節へと移行していくことが分かる。これは、南モデルにおけるC類に分類される並列節が、より広い範囲のモダリティ形式を取り得るという統語的な特徴を反映しているものと考えられる。

唯一、条件節のナラが並列節の中に混じって高い比率を示しているが、これはグラフを見れば分かるように、ナラが根源的モダリティを突出して多く取っていることに起因している。元データを確認したところ、ナラ節が根源的モダリティを取る1157例のうち、約93%の1075例が「たい」で占められていた。さらに、そのうち約45%はYahoo!知恵袋(OC)からの例であった。すなわち、投稿された質問に対して回答が寄せられる知識検索サービスの文脈において、「～したいなら」という言い

回しが定型的に多用されていることに原因があると考えられる。

なお、ナロック (2006) には、モダリティ形式を内部に含む割合の高い連用節として、「カラ > ケレド > ノニ > ガ > ノデ > ニモカカワラズ」という順列が与えられているが、今回 BCCWJ を分析した結果からは、この順列を再現することはできなかった。

5.4 各レジスターにおけるモダリティ形式の分布

最後に、図4と同様、各連用節がどのようなモダリティ形式を含むかについて、レジスターごとに集計を行なった。結果を図6に示す。なお、ここでは縦軸を出現数とする。

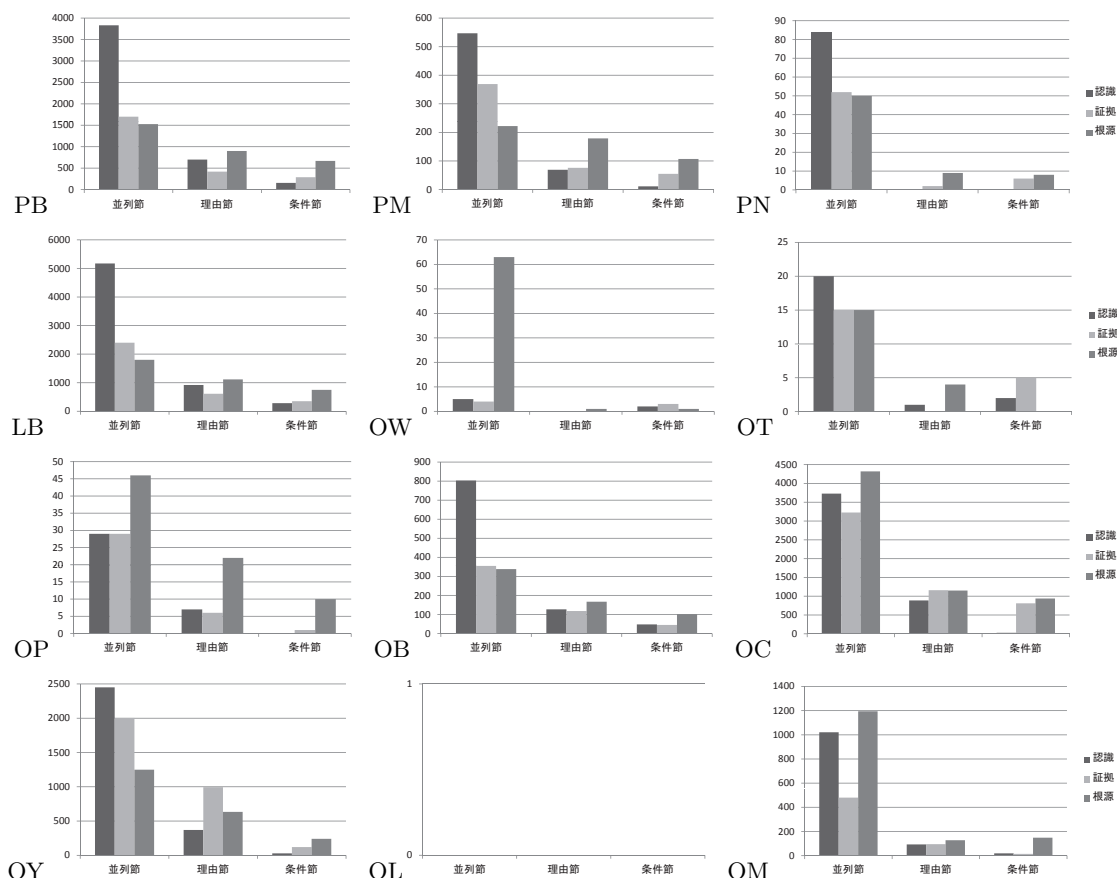


図6: 各モダリティ形式を含む連用節の割合 (レジスターごと)

図6からは、レジスターによって連用節が含み得るモダリティ形式の分布に違いがあることが分かる。出版サブコーパスの書籍(PB)、図書館サブコーパスの書籍(LB)、ベストセラー(OB)はほぼ同じ分布を示していることから、母集団の性質は異なるものであっても、同じ書籍として共通した文法的特徴を備えていると見てよい。雑誌(PM)は図1で示したナロック(2006)の調査結果と最も近似した分布になっている。雑誌には、文芸小説やファッション誌、さまざまな情報誌、対談記事など、多彩なテキストが含まれていることから、多くの電子ブックやさまざまな電子テキストをブレンドして使用したナロック(2006)のデータと、テキスト集合としての性質に近いのかもしれない。白書(OW)は他のレジスターと大きく分布が異なり、「義務・必要・許可」などを表す根源的モダリティが並列節に突出して多く見られる一方、その他のモダリティ形式はほとんど見られない。特に認識的モダリティのように、書き手・話し手の不確かな捉え方を表すための文法形式は、客観的な事実を記述するスタイルを取る白書のテキストとしてふさわしくないだろう。また、法律(OL)では、連用

節中にモダリティ形式が一切現れない(連用節の数そのものも非常に少ない)。そもそも法律とは、ある行為を命令・禁止したり、ある権利を保障したりすることを明示的かつ曖昧性のないように述べるためのテキストであるため、話し手の判断や態度を表すモダリティ形式は、法律文というテキストが持つ機能にそぐわないものと考えられる。

6 考察

ここまでの分析で明らかになったのは、BCCWJのレジスターごとに、使われる連用節の形式と数、およびモダリティ形式の種類と数が大きく異なるという点である。言語の使用場面によって使用される文法形式は異なる、という直感的には理解される言語の使用実態に伴う特徴が、BCCWJのレジスターの違いを利用して実証的に明らかにできたことになる。母語話者の(少なくとも筆者の)内省に基づくだけでは、異なるレジスター間で連用節・モダリティ形式の使用傾向にどのような差があるかを正確に把握することは難しい。

本稿での分析の立脚点としたナロック(2006)では、図1のような結果をもとに、対比節と理由節におけるモダリティ形式の使用傾向に逆の関係があることを結論として述べていた。これは重要な知見の一つであり、BCCWJの一部のレジスターでもほぼ同じ分布と傾向を確認することができた。しかしながら、それがどのような場面でも成り立つ使用傾向であるかと言えば、本稿での分析でも明らかかなように、必ずしもそうではない。コーパスからある文法的特徴を抽出できたとき、それが現代日本語に広く観察される特徴であるのか、ある特定の使用場面に依存した特徴であるのかという区別に、分析者は常に意識的である必要がある。

性質の異なるテキストの間で文法形式の分布に差が観察されるとき、それを裏付けるための説明の仕方については、さまざまな観点があると考えられる。例えば、本稿でも一部で触れた書き言葉のスタイル差、すなわち、テキストの硬さ・軟らかさ、客観的か主観的か、改まった書き言葉かくだけた書き言葉か、などといった差は、文法形式の分布の違いを裏付ける指標として考えられるだろう。さらに、書き手の性別・年齢、読者対象、執筆年代といったメタ情報の違いや、あるテキストが執筆された目的やテキスト自体が担う機能の違い、さらに名詞率、MVR、TTR、品詞構成比率などといったテキスト自体から得られる統計的指標なども、文法形式の分布の違いを説明するための手がかりになり得ると考えられる(小磯, 田中, 小木曾, 近藤, 2012)。レジスターごとに観察される言語的特徴と、それを裏付けるためのさまざまな指標・観点を組み合わせて考えることにより、社会における言語活動のどのような場面にどのような実態が観察されるのか、その現実的なありさまを記述していくことが可能になると思われる。

7 おわりに

本稿では、現代日本語における複文を構成する連用節、およびモダリティの諸形式が、実際の書き言葉の中でどのように分布しているかについて分析を行なった。ナロック(2006)が実施した調査・記述を立脚点として、BCCWJを用いてさまざまなレジスターにおける連用節・モダリティ形式の分布を見た。その上で、ナロック(2006)による観察結果と同様の結果を、一部のレジスターで確認することができた。一方、別のレジスターではそれとは異なる分布を示していることを示した。

文献

- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman Grammar of Spoken and Written English*. Pearson Education.
- Carter, R. & McCarthy, M. (2006). *Cambridge Grammar of English*. Cambridge University Press.
- Huddleston, R. D. & Pullum, G. K. (2002). *The Cambridge Grammar of the English Language*. Cambridge University Press.
- 日本語記述文法研究会 (編) (2010). 『現代日本語文法 1』. くろしお書房.
- 北原保雄 (1970). 「助動詞の相互承接についての構文論的考察」. 『国語学』, **83**, 32-59.
- 小磯花絵, 田中弥生, 小木曾智信, 近藤明日子 (2012). 「テキストの多様性をとらえる分類指標の体系化の試み (2)」. 『言語処理学会 第 18 回年次大会 発表論文集』, pp. 739-742.
- 益岡隆志 (1991). 『モダリティの文法』. くろしお出版.
- 益岡隆志・田窪行則 (1992). 『基礎日本語文法 —改訂版—』. くろしお出版.
- 益岡隆志 (編) (1993). 『日本語の条件表現』. くろしお出版.
- 益岡隆志 (2007). 『日本語モダリティ探究』. くろしお出版.
- 三上章 (1953). 『現代語法序説』. 刀江出版.
- 南不二男 (1974). 『現代日本語の構造』. 大修館書店.
- 南不二男 (1991). 「現代日本語の従属句についての小調査」. 『日本語学』, **10** (12).
- 南不二男 (1993). 『現代日本文法の輪郭』. 大修館書店.
- ナロックハイコ (2006). 「従属節におけるモダリティ形式の使用」. **6** (1), 21-37.
- 仁田義雄 (編) (1991). 『日本語のモダリティと人称』. ひつじ書房.
- 野田尚史 (1995). 「文の階層構造からみた主題ととりたて」. 益岡隆志, 野田尚史, 沼田善子 (編), 『日本語の主題と取り立て』, pp. 1-35. くろしお出版.
- Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1972). *A Grammar of Contemporary English*. Longman.
- Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A Comprehensive Grammar of the English Language*. Longman.
- 田窪行則 (1987). 「統語構造と文脈情報」. 『日本語学』, **6** (7), 37-48.
- 渡辺実 (1953). 「叙述と陳述 —述語文節の構造—」. 『国語学』, **13/14**, 20-34.
- 国立国語研究所コーパス開発センター (2011). 『『現代日本語書き言葉均衡コーパス』利用の手引 第 1.0 版』. 国立国語研究所.