



現代日本語における高頻度ヘッジ語形の抽出と使用 環境によるグルーピングの試み : ジャンル別のコー パス頻度調査に基づいて

肖, 錦蓮

(Citation)

統計数理研究所共同研究リポート, 444:103-122

(Issue Date)

2021-03-15

(Resource Type)

departmental bulletin paper

(Version)

Version of Record

(JaLCD0I)

<https://doi.org/10.24546/81012577>

(URL)

<https://hdl.handle.net/20.500.14094/81012577>



現代日本語における高頻度ヘッジ語形の抽出と 使用環境によるグルーピングの試み —ジャンル別のコーパス頻度調査に基づいて—

肖 錦蓮
神戸大学 (大学院生)

概要

本研究は話し言葉と書き言葉の差とジャンル差の違いによって、ヘッジ語形の使用頻度がどう異なるか調査し、また、これらの影響をあまり受けずに安定して出現する高頻度ヘッジ語形の抽出を試みた。分析の結果、(1) 話し言葉におけるヘッジ語形の出現頻度が書き言葉の 3 倍以上であること、また、情報提供を主たる目的とするテキストより、筆者のスタンスが自由に表出される小説系・創作系テキストのほうが、独話系テキストより、対話系テキストのほうが、ヘッジ語形が多用されること、(2) 話し言葉と書き言葉間、ジャンル間の高頻度ヘッジ語形の一致度が低く、ヘッジ語形の振る舞いは言語環境に強く影響されること、(3) 各種のヘッジ語形が、情報伝達の方角性と場面の特定性の 2 観点によって 5 カテゴリーに分類できること、(4) 異なるジャンルにおいて安定的に高頻度で使用されるヘッジ語形が 40 語抽出できたこと、という 4 点が確認された。

キーワード

ヘッジ語形, 言語環境, ジャンル差

1. はじめに

日本語においては、話者の陳述の度合いを緩和することがしばしば行われる。以下の例を見てみよう。これらは筆者の作例である。

- (1) 夜更かししたせいもあって、ちょっと眠たいかも。
- (2) 私の見るところ、あの学生は、最近さぼっているようだ。
- (1) については、「眠たい」は話者自身の感覚であるため、話者は断定しようと思えば、

「かも」を使わず「ちょっと眠たい」と文を終わらせることもできると考えられる。しかしながら、「眠たい」と言い切ると、自分が相手に対して攻撃的なニュアンスが込められているかもしれない。日本語では、自分のことであっても、主張が強くなりすぎないように調整するために、「かも」を入れている。

(2) については、教師は学生を評価する際に、「さぼっている」と断言することができなくもない。しかしながら、学生が最近さぼっていると分かっているにもかかわらず、「ようだ」をつけて文をやわらげることで、そうではないかもしれないという可能性を残している。

このように、日本語において、意味の陳述を緩和する機能は極めて重要である。一般にこれをヘッジと呼ぶ。しかしながら、このヘッジの習得は外国人学習者、特に中国語を母語とする日本語学習者にとって容易ではない。以下の例を見てみよう。

(3) 少子化は先進国で必ず起こる問題なのです。

(4) 台湾だけではなくて、世界のあらゆる所に少子化の問題があります。

(5) 一人一人がちょっとした努力をしたら、ゴミを減らすことができるかもしれないと思う。

(3) と (4) は少子化について書かれた作文から取り出された例である。「必ず」といった断定表現、「あらゆる」・「全」などの全称を表す表現が好んで使われる。これらの断定や強調を表す表現を使ってしまうと、言い切るニュアンスを読者に与えてしまう恐れがある。この場合、少子化の程度が小さい国も存在するため、このような文脈では、「少子化は多くの先進国で起こっている問題です」と言ったほうが妥当である。

(5) は、文章の最後で読者に対する呼びかけをするときに、ヘッジを使った例である。読者に共感してもらうために、より強く主張すべき文脈にもかかわらず、ヘッジを使ってしまうと、自信がなく、主張が弱すぎる印象を与えてしまう恐れがある。

例で示されたように、日本語学習者のヘッジの使用には、必要な場面での不使用、不必要な場面での使用、など様々な問題が存在する。よって、日本語学習者に対し、ヘッジを正しく使用するための指導が十分であると言えないだろう。では、学習者はどのようなヘッジを学習していけば良いのであろうか。筆者が知る限り、学習者が学ぶべきヘッジのリスト、言い換えれば、現代日本語において、高頻度に使われるヘッジのリストが存在しないのが実情である。このことを考えていくと、ヘッジの研究、さらに、ヘッジの教育を考える場合、まず現代日本語において、高頻度で使用されるヘッジを特定することが極めて重要であると言える。

ただし、ここで注意しないといけないことは、ヘッジの語形が同じであっても、実際ではヘッジとして機能していない場合もあるということである。その区分は筆者の後続する研究で扱うことにし、その予備調査として、本研究においては、語形に限って、重要なヘッジ

語形の特定を行っていききたい。

現代日本語において、重要なヘッジ語形を取り出そうとする場合、コーパス言語学の手法が極めて有効である。しかしここで注目すべきは、ヘッジ語形の使用は言語環境、つまり、産出モード（話し言葉・書き言葉）や内容ジャンルの影響を考慮することである。この点を踏まえれば、日本語コーパスで頻度調査を行う際にも、コーパスを全体的に調べるのではなく、書き言葉と話し言葉、および内容的なジャンルに分けて見ていくことが妥当であると言えるだろう。

統計学的に考えると、複数のジャンルから取り出された頻度をどのようにまとめるのかは興味深い研究テーマとなりうる。例えば、単純平均法、加重平均法、主成分法など様々な方法がある。先行研究の知見によると、単純平均よりも、主成分分析のほうが各ジャンルの特性をよりよく生かすと言われている。そこで、以下では主成分法を用いて、様々なジャンルの頻度を合成して、高頻度ヘッジ語形の特定を試みたい。

2. 先行研究

日本語のヘッジは多様なバラエティーを持つ。日本語のヘッジについて、主に使用実態と機能の 2 つの観点から研究が進んでいる。以下では、日本語のヘッジに関する先行研究を概観する。

Lauwereyns (2002) は、ヘッジを「不確かさや可能性、ためらい、近似性を言語的に表現する曖昧表現」と定義し、文末表現（「と思う」等）、副詞（「たぶん」等）、終助詞（「かな」等）、接辞（「～系」等）等 26 種の言語形式について、性差、年代差、場面の改まり度の 3 つの観点から考察した。分析の結果、(1) ヘッジの使用頻度について、上位 5 位は「とか」>「思う」>「かな／かしら」>「～たり～たりする」>「みたい／よう」となること、(2) 女性は男性よりも、若者は年配者よりも、改まり度の高い場面では低い場面よりも、多くのヘッジを使用したこと、が報告されている。女性は断定を避けるヘッジ表現を好む傾向があることは Lakoff (1990) においても指摘されている。

Tatiana (2003) は、「ようだ」、「めいた」、「げ」、「そうだ」、「かもしれない」、「っぽい」の 6 種類のヘッジを考察対象とし、それぞれの形式がどのように本来の意味によって「やわらげ」の効果を得るのかに注目した研究である。Tatiana は現代日本語において断定をきつい表現としてきらい傾向があると指摘し、ヘッジの使用の背後には陳述内容が断定に至っていない、もしくは断定に至ってはいるが表現をやわらげるために非断定的形式を用いている、という話者の心理が存在すると述べている。

入戸野 (2004) は、ヘッジを「陳述、質問、申し出、命令、拒絶といった命題の中、前、または後ろに置かれ、発話行為ないしは発話態度を緩和する機能を持つもの」と定義している。同研究は日本人の親しい友人同士の会話を調査し、「けど」、「ね」、「っていう」などを含める 138 種類の語彙的ヘッジを形式と機能の両面から考察している。その結果、(1) ヘ

ッジの使用頻度について、上位 10 位は「ね」>「とか」>「なんか」>「さ」>「けど」>「あの」>「の」>「な」>「かな」>「やっぱり」となること、(2) 品詞別に見ると、助詞が約 55%，間投詞が約 12%，副詞が約 10%を占めること、(3) 機能については、「正確な情報を伝達したい」という意思表示機能が約 20%，「人間関係を向上させる」機能が約 80%を占めたことが明らかになった。

李 (2008) はヘッジを「推量をしたり、または、断定や限定を避けたり、明言や断言をせずに、ぼかすことで発話内容の力を和らげる (弱めたり、緩和する) 表現」と定義している。対話データを用い、日本語と韓国語におけるヘッジの使用状況について、ベースの性別、対話相手の性別・年齢の 2 つの観点から考察した。その結果、(1) 日本語は男性ベースのほうヘッジの使用割合が高く、韓国語は女性ベースのほうヘッジの使用割合が高いこと、(2) 日本語の場合、異性の対話相手に対してヘッジ使用割合が高く、韓国語の場合は、同性の対話相手に対して使用割合が高いこと、(3) 韓国人男性は、年上に有意にヘッジを多く使っており、ヘッジが対話相手の年齢から生じる上下関係をマークしていることが明らかになった。

袁 (2018) はドラマのセリフに出現するヘッジがどのように中国語に訳されているかを調査した研究である。『ナオミとカナコ』、『お義父さんと呼ばせて』、『戦う！書店ガール』、『結婚しない』、『家族ノカタチ』の 5 本のドラマをデータとし、167 個のヘッジ用例を収集している。分析の結果、ドラマの中国語字幕では、ヘッジの訳出は全体の約 6 割であること、訳出されなかったヘッジとしては、「と思う」と「ちょっと」が目立つこと、などが報告されている。

以上で紹介した研究の多くは話し言葉を調査したのであったが、最近では、書き言葉におけるヘッジへの関心も高まってきた。Lee (2009) は Hyland (2005) の分類モデルに従い、「朝日新聞」と New York Times の社説に出現したスタンス表現 (ヘッジはスタンス表現の下位分類の 1 つである) を分析したものである。さらに、Lee・楊 (2013) では中国の新聞「新京報」の社説も比較対象に加えられ、3 言語の比較から、日本語の社説ではヘッジが多いことが観察された。

ヘッジに関する研究は、これまでに多くの観点から行われており、その中で様々な知見が示されてきた。しかしながら、いくつかの課題も認められる。

1 点目は異なる研究が、異なるヘッジを分析対象としており、得られた結果を比較して議論できないことである。この点について、新しい研究の前提として、分析対象とするヘッジを整理し、認定の基準を示すことが重要になろう。

2 点目は調査する資料が特定の場面のデータに限られ、ヘッジ語形の振る舞いへの言語環境の影響がほとんど考慮されていないことである。ヘッジは話し言葉においても、書き言葉においても重要な語彙項目とみなされている。また、内容ジャンルの違いによってヘッジ語形の使用状況が異なる可能性も存在する。よって、異なる内容ジャンルを収録した話し言葉および書き言葉コーパスを用い、話し言葉・書き言葉別およびジャンル別にヘッジ語形の使

用量や高頻度語形を調査する必要があるだろう。

3 点目は産出モード（話し言葉・書き言葉）や内容ジャンルと結びつきやすいヘッジ語形にはどのようなものがあるかは解明されていないことである。よって、言語環境を手掛かりとし、ヘッジ語形がどのようにグルーピングできるかを調査することが必要であろう。

4 点目は言語環境の影響をあまり受けずに安定して出現するヘッジ語形は明らかになっていないことである。ヘッジ語形を教える際に、多様なヘッジ語形の中からより重要なものを取り出し、それらを優先的に導入するのが望ましい。この点について、コーパスデータを調査し、幅広い言語環境で共通して多用される典型的ヘッジ語形の特定を行う必要があるだろう。

3. リサーチデザイン

3.1 研究目的とリサーチクエスチョン

以上を踏まえ、本研究では、言語環境（産出モード・内容ジャンル）がヘッジ使用に及ぼす影響を考慮しつつ、現代日本語における典型的ヘッジ語形の特定を目指す。その際、(1) 言語環境別のヘッジ語形使用量（総使用頻度）の調査、(2) 言語環境別の高頻度ヘッジ語形の抽出、(3) 言語環境を手掛かりとしたヘッジ語形のグルーピング、(4) 多様な言語環境で総合的に多用される典型的ヘッジ語形の抽出、の 4 点を行う。

こうした目的に沿い、以下の 4 つのリサーチクエスチョン（RQ）を設定した。

RQ1 言語環境が変わればヘッジ語形の使用量も変わるか？（使用量）

RQ2 言語環境ごとに多用されるヘッジ語形にはどのようなものがあるか？（高頻度語形）

RQ3 言語環境に着目すると各種のヘッジ語形はどのようにグルーピングできるか？（グルーピング）

RQ4 多様な言語環境で総合的に多用されるヘッジ語形にはどのようなものがあるか？（頻度合成）

3.2 データ

本研究では、話し言葉・書き言葉の両面を調査するため、以下の 3 種のコーパスを利用する。1 つ目は多言語母語の日本語学習者横断コーパス（International Corpus of Japanese as a Second Language：以下 I-JAS）である。I-JAS は、異なった言語を母語とする日本語学習者、および日本語母語話者の話し言葉および書き言葉を横断的に収集したコーパスである。ここでは、母語話者の発話データのみを利用する。母語話者の発話データには、イラストのストーリーを話すストーリーテリング（ST）、イラストについて説明する絵描写（D）、学習者と調査実施者が行う自然な対話（I）、設定された場面に応じて、与えられた役を演じて会話するロールプレイ（RP）の 4 種類のタスクがある。

2 つ目は、日本語日常会話コーパスモニター公開版 (Corpus of Everyday Japanese Conversation : 以下 CEJC) である。CEJC は 200 時間規模の母語話者による日本語の日常会話をバランスよく収録することを目標とするコーパスである。2021 年度の本公開に先立ち 2018 年 12 月に 126 会話 (50 時間) を対象とするモニター版が公開された。総語数はおおよそ 61 万語である。CEJC では個人密着法を中心に会話を収集している。個人密着法は、性別・年齢などの観点からバランスを考慮して選別された協力者に収録機器などを 2-3 ヶ月ほど貸し出し、日常生活で生じる会話を協力者自身に記録してもらうという収録法である。会話に関する情報として、(1) 会話形式、(2) 話者数、(3) 会話が行われた場所、(4) 活動、(5) 話者間の関係性の 5 種類の情報が提供される。また話者に関しては、(1) 年齢、(2) 性別、(3) 出身地、(4) 居住地、(5) 職業、(6) 協力者からみた関係性の 6 種類の情報が提供される。CEJC と I-JAS の対話 (I) はともに自然な会話であるが、CEJC は多様な場面の日常会話を収録しているのに対し、I-JAS の対話 (I) の内容はある程度統一された話題 (15 項目) に限られている。ゆえに、本研究は CEJC のデータと対話 (I) を区別し、それぞれ話し言葉の 1 つのジャンルとみなす。

3 つ目は現代日本語書き言葉均衡コーパス (Balanced Corpus of Contemporary Written Japanese : 以下 BCCWJ) である。BCCWJ には、13 ジャンル (出版・雑誌、図書館・書籍、出版・書籍、出版・新聞、ブログ、国会会議録、知恵袋、ベストセラー、韻文、教科書、広報紙、白書、法律) が含まれる。本分析では、このうち、「出版・書籍」「図書館・書籍」「ベストセラー」を同じ「書籍」ジャンルとみなして 1 つにまとめる。また、書籍については、話し言葉的な要素がより多く含まれる小説類をほかと区別するために、サンプル ID に表示された図書館資料の日本十進分類法 (NDC) の第 1 次区分の情報を利用して、「文学以外」(書籍 1) と「文学」(書籍 2) を分けて扱う。上記の 12 変種のテキストのうち、書き言葉の特徴が強い「新聞」、「雑誌」、「書籍 1」、「書籍 2」、「教科書」の 5 ジャンルを今回の調査対象とする。

以上の 3 種のコーパスは、すべて、国立国語研究所の「中納言」検索プラットフォーム上で検索することができる。

以上より、本研究は、話し言葉の 5 ジャンル (ST, D, I, RP, CEJC)、書き言葉 5 ジャンル (PN, PM, LB1, LB2, OT)、合計 10 種の言語環境別にヘッジ語形の頻度調査を行うことになる。

表 1

使用するコーパス

		ジャンル	語数
話し言葉	I-JAS 母語話者	ロールプレイ (RP)	2.2 万語
		ストーリーテリング (ST)	1.2 万語

	データ	絵描写 (D)	1.8 万語
		対話 (I)	20 万語
	CEJC	日常会話	61 万語
書き言葉	BCCWJ	新聞 (PN)	140 万語
		雑誌 (PM)	440 万語
		書籍 1 (LB1)	4160 万語
		書籍 2 (LB2)	2100 万語
		教科書 (OT)	90 万語

3.3 調査項目

すでに述べたように、研究により調査対象とされた言語形式がまちまちであるため、ヘッジの全体像は明確につかめていないものとなっている。そこで、本研究では、ヘッジを取り上げた先行研究を調査し、1 種以上でヘッジとされた語や表現を取り出した。ヘッジを取り上げた先行研究としては、Lauwereyns (2002), Tatiana (2003), 入戸野 (2004), 李 (2008), 山川 (2011), 堀江・堀田 (2012), Lee・楊 (2013), 堀田 (2013), 小森 (2016a), 小森 (2016b), 袁 (2018), 東泉・高橋 (2020) の 12 種がある。

ただし、調査対象の詳細を紹介していない研究もあるため、論文で提示された用例などを参考し、論文に言及されたすべての言語形式を集めた。12 種の先行研究から 1 種以上でヘッジとされた語はすべて 89 語あるが、他のヘッジを包含する「よね」、「かね」、「というか」や、他のヘッジに包含される「な」、は対象外とした。また、「わけ」は山川 (2011) ではヘッジと認定されるが、小森 (2016b) では表現を強める強意詞の範疇に入れているため、対象外とした。「よ」は「わけ」と同様な理由で除外した。以上により、コーパス頻度調査に用いたのは 83 種となる (表 2)。

表 2

調査対象とするヘッジ

品詞別	種類	対象ヘッジ
副詞	29	少し, ちょっと, しばらく, なんとか, けっこう, たぶん, おそらく, ただ, とりあえず, やや, 必ずしも, 少々, なんとなく, 概ね, もしかしたら, 万一, あいにく, やっぱり, あまり, なかなか, せっかく, まあ, 一応, 大抵, 大体, ほとんど, たしか, だいた, もちろん
助詞	10	ね, かな, とか, が, なんか, なんて, たり, けれど, だけ, くらい
名詞	10	ほう, 頃, 感じ, 可能性, 関係, 感覚, 方向, あたり, ほど, ふう (に), わけ

動詞類	10	と思う, 気がする, 恐れがある, と思われる, と考えられる, ように 思われる, ように考えられる, とは限らない, ～得る, 場合もある
連語	8	という, のではないか, んじゃないの, かも, っていう, ある程度, んです, 多くの, というか
接尾辞	7	～的, ～げ, ～っぽい, 風, 系, ～めく, ～にくい
助動詞	6	ようだ, みたいだ, らしい, だろう, でしょう, しそうだ
形容詞	1	少なくない
感動詞	1	あの (う)
連体詞	1	単なる
総計	83	

3.4 分析の手順

RQ1 では, ヘッジ語形の使用量に注目する。まず中納言アプリケーションを用い, ジャンル別に 83 種のヘッジの頻度を調査し, 10 万語あたり調整頻度に変換する。次に, 話し言葉と書き言葉, 話し言葉の 5 ジャンル, 書き言葉の 5 ジャンル, の 3 つの観点からヘッジ語形の使用量を比較する。また, 差があるかどうかを検証するため, カイ二乗検定を行う。

RQ2 では, 高頻度で使用されるヘッジ語形に注目する。話し言葉と書き言葉, 話し言葉の 5 ジャンル, 書き言葉の 5 ジャンル, の 3 つの観点から出現頻度の最も高いヘッジ語形上位 10 語を抽出し, その一致度を観察する。

RQ3 では, それぞれのジャンルにおいて, どのようなヘッジ語形が特徴的に使われるかを調査する。分析にはコレスポンデンス分析を行う。コレスポンデンス分析は, テータの関係性を視覚的に表現する分析手法である (石川・前田・山崎, 2010)。第 1 アイテムをジャンル, 第 2 アイテムをヘッジ語形とする。第 1 アイテムは 10 カテゴリー, 第 2 アイテムは 83 カテゴリーからなる。

最後に, RQ4 では, ジャンル影響を受けにくく, 共通して高頻度で使用されるヘッジ語形を特定する。10 ジャンルを変数, 83 種のヘッジ語形をケースとするクロス表に対して主成分分析を実施し, 第 1 主成分得点を手がかりに 83 語を並べ替える。石川・前田・山崎 (2010) によると, 「主成分分析を使えば, … (中略) それぞれの変数のばらつきや単語間の相関関係も加味しながら, 全体を最も効率よく代表する値を取り出すことができる」とされている。よって, 主成分分析を用いることによって, すべてのジャンルの情報を 1 つの変数に集約して扱うことができる。

4. 結果と考察

4.1 RQ1 ヘッジ語形の使用量

はじめに, 使用量に注目する。話し言葉と書き言葉におけるヘッジ語形の出現頻度は以下

のようになった。

図 1

話し言葉・書き言葉におけるヘッジの使用量

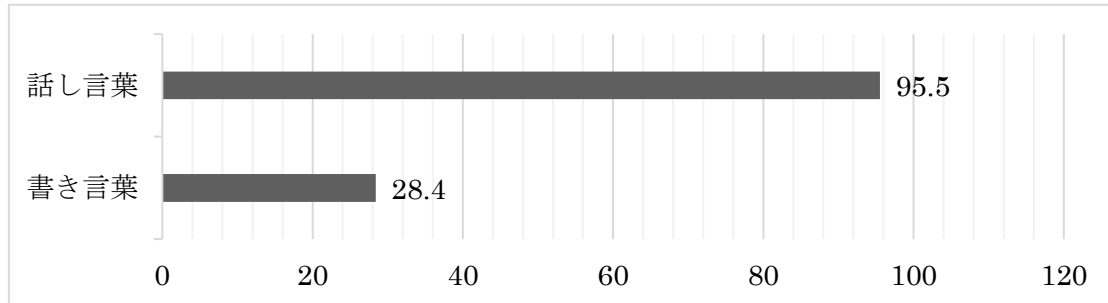
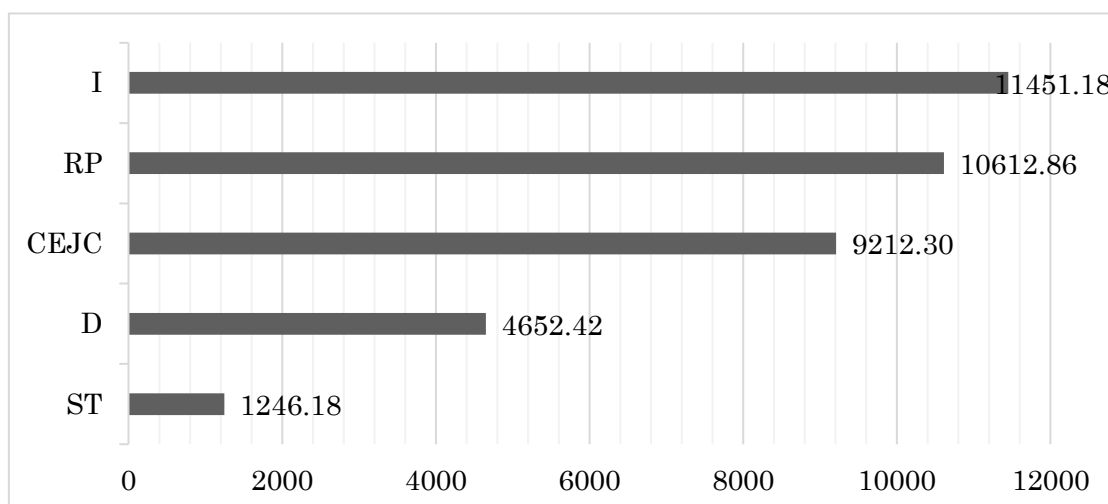


図 1 で示されているように、話し言葉におけるヘッジ語形の出現頻度が書き言葉を大きく上回っており、その差が 3 倍以上であることが確認された。カイ二乗検定を行ったところ、その差は統計学的にも有意であった ($\chi^2=135940$, $df=1$, $p<.001$)。この結果から、「もちろん」のヘッジ用法を調査した東泉・高橋 (2020) が示す「話し言葉>書き言葉」というヘッジ使用量の特徴はサポートされ、しかもその差は 3 倍以上であることが明らかになった。

続いて、話し言葉の 5 ジャンルそれぞれにおけるヘッジ使用量を調査したところ、以下の結果を得た。

図 2

話し言葉の 5 ジャンルにおけるヘッジの使用量



上図より、話し言葉におけるヘッジ語形の使用量について、隣接する 2 項目間の差の有意性をカイ二乗検定で確認した。なお、ここでは検定を 4 回繰り返すため、ボンフェローニ

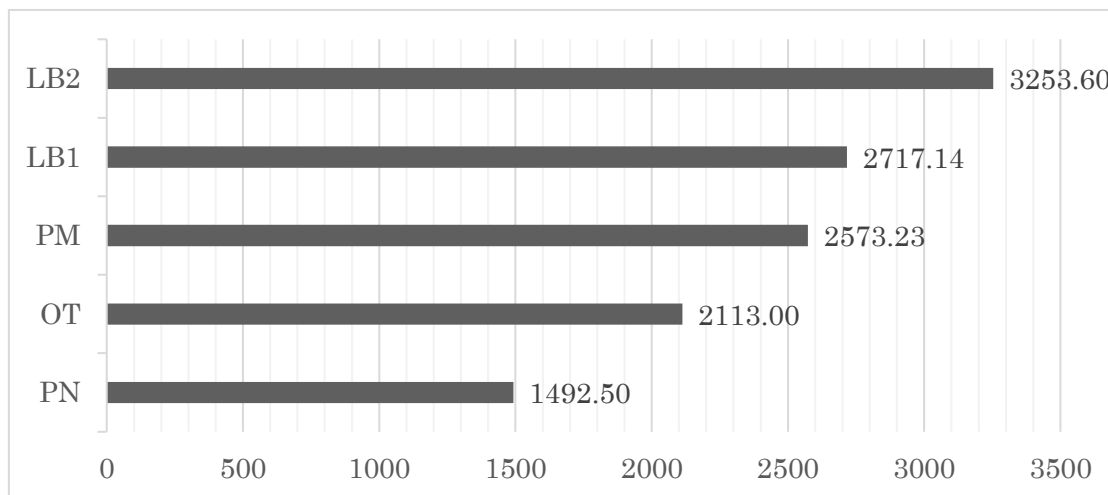
補正を行い、有意水準 $\alpha = 5\%$ を 4 で割り、 $\alpha_B = 1.25\%$ を有意判断の上限とした。この基準で判定を行ったところ、すべてのジャンル間の差があることがそれぞれ確認された (I・RP 間: $\chi^2=879$, $p<.001$, RP・CEJC 間: $\chi^2=50$, $p<.001$, CEJC・D 間: $\chi^2=449$, $p<.001$, D・ST 間: $\chi^2=265$, $p<.001$)。

以上より、I・RP・CEJC などの対話系テキストではヘッジ語形が多用されるのに対し、D・ST など独話系テキストではヘッジ語形があまり使用されないことが確認された。この結果は、「対話より独話でより多用される」という個別のヘッジの使用実態を調査した東泉・高橋 (2020) の指摘を支持しないものとなった。対話系テキストにおけるヘッジ語形の頻度の高さは、双方向的な談話では、相手の反応を踏まえて次の行動をするため、意見の衝突や相手の気分を害するような場面では、話し手ないし書き手の側があらかじめ相手の反応を想定し、よりやわらかい表現を使うことによって、発話態度を緩和する必要があるためではないかと考えられる。

次に、書き言葉の 5 ジャンルにおけるヘッジ使用量について見てみよう。

図 3

書き言葉の 5 ジャンルにおけるヘッジの使用量



上図より、書き言葉におけるヘッジ語形の総頻度について、隣接する 2 項目間の差の有意性をカイ二乗検定で確認した。その結果、すべてのジャンル間の差があることがそれぞれ確認された (LB2・LB1 間: $\chi^2=14276$, $p<.001$, LB1・PM 間: $\chi^2=313$, $p<.001$, PM・OT 間: $\chi^2=650$, $p<.001$, OT・PN 間: $\chi^2=1237$, $p<.001$)。

以上より、LB2, LB1 などの筆者のスタンスが自由に表出される小説系・創作系テキストではヘッジ語形が多用されるのに対し、PM・OT・PN など情報提供を主たる目的とするテキスト、特に公的な性質を持つ OT や PN などではヘッジ語形があまり使用されないことが確認された。

4.2 RQ2 高頻度ヘッジ語形

RQ1 の調査により、ヘッジ語形の使用は日本語の言語環境の影響を強く受けることがわかった。このことは、高頻度ヘッジ語形についても、言語環境ごとに異なっている可能性を示唆する。

そこで、話し言葉・書き言葉別、また、各内容ジャンル別に、高頻度ヘッジを調査した。まず、話し言葉・書き言葉の高頻度ヘッジ語形上位 10 語は以下の通りであった。

表 3

S/W 別高頻度ヘッジ語形上位 10 語

頻度順	話し言葉	書き言葉	頻度順	話し言葉	書き言葉
1	ね	ようだ	6	ほう	だろう
2	んです	という	7	が	ね
3	けれど	が	8	たり	たり
4	ちょっと	～的	9	と思う	けれど
5	とか	だけ	10	なんか	んです

注目すべきは、頻度順は大きく異なっているだけでなく、上位 10 語中 5 語しか一致していないことである。「ようだ・という・的・だけ・だろう」は書き言葉のみの上位 10 語に出現しており、「ちょっと・とか・ほう・と思う・なんか」は話し言葉のみの上位 10 語に入っている。ヘッジ語形の全体頻度を調査したときと同様、高頻度ヘッジ語形に限っていても、話し言葉・書き言葉間に大きな差異が存在し、特に、上位 6 語までのヘッジ語形においては、一致している語が 1 語もなく、これによってヘッジ語形の振る舞いは話し言葉と書き言葉間で大きな差異があることが示唆された。

つづいて、話し言葉の 5 ジャンルごとに高頻度ヘッジ語形を調べたところ、上位 10 語は以下の通りであった。

表 4

ジャンル別高頻度ヘッジ語形上位 10 語（話し言葉）

順位	ST	I	RP	D	CEJC
1	が	ね	んです	ね	ね
2	という	んです	ちょっと	たり	なんか
3	んです	けれど	けれど	んです	あの（う）
4	ね	とか	ね	ようだ	けれど

5	ようだ	ちょっと	ほう	けれど	とか
6	けれど	まあ	とか	が	ちょっと
7	しそうだ	やはり	と思う	ほう	まあ
8	ほう	たり	が	とか	みたいだ
9	かな	みたいだ	やはり	かな	かな
10	せっかく	と思う	かな	ちょっと	んです

まず、上位 3 語に限ってみても、すべて一致しているジャンルが見られないことが確認された。また、特定のジャンルにのみ高頻度で出現しているヘッジ語形も存在する。例えば、接続助詞の「が」I, CEJC では、上位 10 語に含まれていないが、ST, RP, D では上位 8 語に含まれている。かわりに I, CEJC では「が」と似たような意味を持つ「けれど」が上位 4 位に入っている。したがって、「が」と比べ、「けれど」のほうが自然に行われる双方向的な談話と結びつきやすいと考えられる。また、「しそうだ」と「せっかく」の 2 つのヘッジ語形は ST のみ高頻度で使用されることがわかった。以下はその例である。

- (8) 犬が、せっかく作った料理を食べてしまってるの気づいて、二人はとても残念そうでした。
- (9) マリは二階の窓からそれを不思議そうに見ています。
- (10) マリは起きそうにありません。

ST は提示されたイラストのストーリーを話すというタスクであり、例で明らかなように、特定のヘッジ語形が多用されるのはタスクの内容に関わると考えられる。様態の「しそうだ」が頻繁に使われるのは、話者はイラストに描かれている人物の表情や動きを確認し、外観から受けた印象（残念そう）や事態が起こる可能性（起きそうにありません）を表す場合が多いためであると考えられる。

つづいて、書き言葉の 5 ジャンル別に高頻度ヘッジ語形を調べたところ、上位 10 語は以下の通りであった。

表 5

ジャンル別高頻度ヘッジ語形上位 10 語 (書き言葉)

順位	PN	LB1	LB2	OT	PM
1	が	という	ようだ	ようだ	という
2	という	～的	が	～的	が
3	～的	ようだ	という	という	ようだ
4	ようだ	が	だろう	が	～的
5	だけ	だけ	ね	たり	ね
6	たり	たり	だけ	だけ	だけ
7	だろう	だろう	けれど	だろう	んです
8	頃	ね	～的	かな	けれど
9	しそうだ	ほう	ほう	ね	たり
10	可能性	でしょう	んです	でしょう	だろう

まず、順序は上下しているものの、LB1 と OT, LB2 と PM は上位 3 語が一致していることが確認された。また、特定のジャンルにのみ高頻度で出現しているヘッジ語形も存在する。例えば、「頃」と「可能性」の 2 つのヘッジ語形は PN, 「かな」は OT でのみ高頻度で使用されることがわかった。

以上より、RQ2 では、高頻度ヘッジ語形は話し言葉と書き言葉間で大きな差異があることや、特定のジャンルにのみ高頻度で出現しているヘッジ語形が存在することが確認された。

4.3 RQ3 ヘッジ語形のグルーピング

RQ3 では、言語環境を手掛かりとしてヘッジ語形をグルーピングする。以下は 10 ジャンルと 83 種のヘッジ語形の関係を視覚化したコレスポネンス分析の結果である。

		れがある, という, 方向
第二象限 (X:- / Y:+) (双方向・不特定場面)	CEJC	なんか, んじゃないの, みたいだ, あの (う)
第三象限 (X:- / Y:-) (双方向・特定場面)	RP	ちょっと, んです, もしかしたら, っ ていう, ほう
コア領域 (原点±1) (中核的)	I, D, ST	かな, かも, とか, 気がする, だいぶ, 感じ, 多分, 結構, 一応, なんとか, と 思う, けれど, くらい, ある程度, でし ょう, ただ, しばらく, おそらく, やは り
第四象限 (X:+ / Y:-) (単方向・特定)		少し, が

まず, 第 1 軸 (横軸) に注目する。図 4 で示されたように, 横軸のプラス方向に書き言葉の 5 ジャンルが固まって布置されている。これに対し, ST を除く話し言葉の 4 ジャンルが横軸のマイナス方向に布置されており, 離れた位置に点在していることが確認された。コレスポンデンス分析では, 性質の近いデータが近傍位置に布置されるため, このことから, 書き言葉より, 話し言葉のほうがジャンル差に影響されやすいということが統計的データから証明されたと言えよう。

また, 10 ジャンルの横軸上の位置を見ると, それらが大きく 3 つの領域に区分されている。左領域には, 個人間の対話を収録した CEJC, I, RP, 中間領域には, 独話系の D, ST, 右領域には, 公的に発行される LB2, PM, LB1, OT, PN が布置されている。ヘッジ語形に注目すると, 横軸のプラス方向において, 原点から最も離れた位置に教科書 (OT) や新聞 (PN) が布置されている。この 2 つのジャンルを特徴づけるヘッジ語形として, 「ように考えられる」, 「恐れがある」, 「とは限らない」など公的な立場から, 事実を中立的かつ客観的に論じるためのヘッジ語形が多くみられる。横軸のマイナス方向において, 「んじゃないの」, 「っていう」, 「みたい」など個人間の対話に多用されるくだけたヘッジ語形が布置されている。したがって, 右に近ければ近いほど, 伝達する情報の単方向性, パブリック性が強くなるとみられ, 横軸は伝達する情報のパブリック性・方向性を分ける軸となっていると言える。

次に, 第 2 軸に注目する。第 2 軸で, 上下に大きく二分される。上部領域には, 多様な場面の自然会話を収録した CEJC と BCCWJ の 5 ジャンルが含まれている。下部領域には, 特定の場面で行われる会話を収録した I-JAS の 4 ジャンルが含まれる。ヘッジ語形に注目すると, 上部には「頃」, 「だろう」, 「らしい」など汎用性の高いヘッジ語形が含まれている。一方, 下部領域には「なんとか」, 「もしかしたら」, 「んです」など, 依頼や事情説明などの

特定の場面に使われるヘッジ語形が布置されている。つまり、第 2 軸は特定場面と不特定多数場面を分ける軸となっているのではないかと考えられる。

以上より、各種のヘッジ語形が、情報伝達の方角性と場面の特定性によって大きく 5 カテゴリーに分類できることが確認された。

4.4 RQ4 典型的ヘッジ語形

これまでの調査で、話し言葉・書き言葉を含む 10 種のジャンル別の頻度データが得られた。RQ4 では、10 種すべてで共通して多用されるヘッジ語形を決定するために主成分分析を行ったところ、10 種の変数ごとの第 1 主成分の負荷量は以下の通りとなった。

表 7

主成分負荷量

ジャンル	第 1 主成分負荷量	ジャンル	第 1 主成分負荷量
PM	.963	ST	.764
LB2	.924	D	.700
LB1	.866	I	.514
PN	.855	CEJC	.385
OT	.780	RP	.466

表 7 に明らかなように、第 1 主成分に対し、すべてのジャンルの負荷量がプラスとなったため、この第 1 主成分は全ジャンルの頻度情報を統合した指標になっていると言える。そこで、この第 1 主成分得点を手がかりとして、83 種のヘッジ語形を並び変えたところ、以下の表 8 が得られた。また、上位 10 語から 80 語までそれぞれの頻度累計構成比を調べたところ、図 5 の結果を得た。

表 8

ヘッジ語形の主成分得点と頻度率

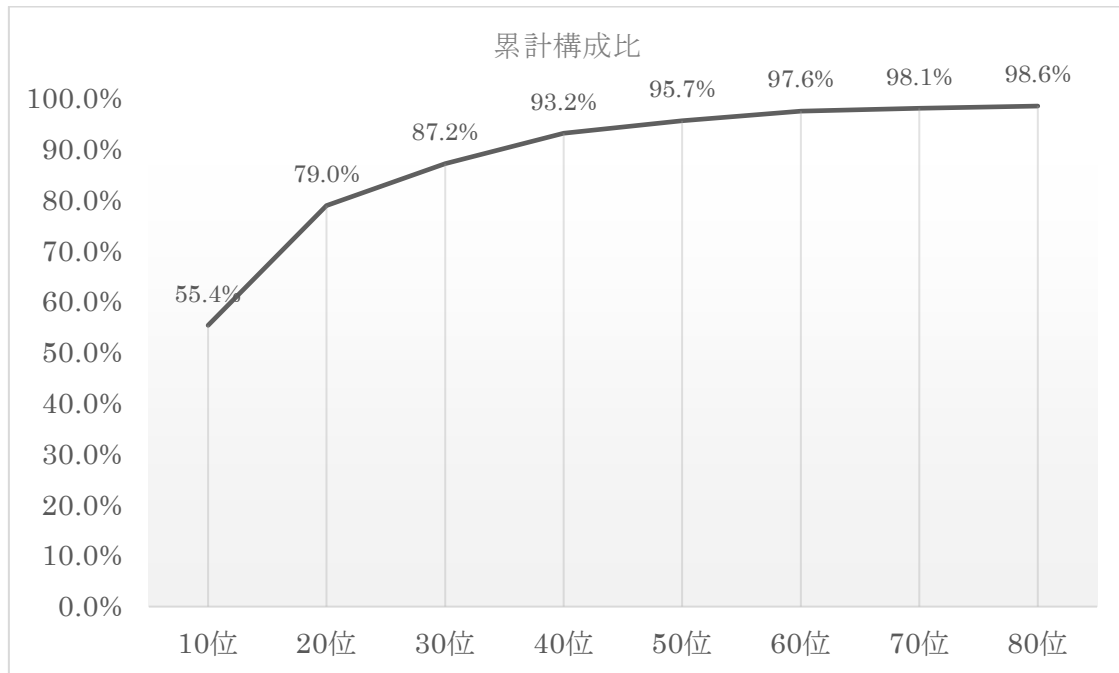
順位	ヘッジ語形	頻度率	主成分得点	順位	ヘッジ語形	頻度率	主成分得点
1	ようだ	5.40	326.5	43	ただ	0.18	6.5
2	が	5.32	306.0	44	のではないか	0.09	6.3
3	という	3.88	240.8	45	せっかく	0.14	6.3
4	ね	15.19	182.0	46	と考えられる	0.08	5.9
5	～的	2.92	168.1	47	ふう	0.32	5.9

6	たり	2.99	129.4	48	しばらく	0.16	5.9
7	んです	8.55	112.3	49	っていう	0.85	5.5
8	けれど	7.69	90.8	50	気がする	0.20	5.5
9	だけ	1.85	83.4	51	感覚	0.10	5.5
10	だろう	1.60	79.9	52	けっこう	0.95	5.2
11	ほう	3.13	75.7	53	もしかしたら	0.15	4.6
12	とか	4.86	57.8	54	風	0.07	4.5
13	でしょう	1.48	53.7	55	～げ	0.07	4.4
14	しそうだ	1.32	48.0	56	なんとか	0.24	4.1
15	と思う	2.52	47.4	57	やや	0.05	3.5
16	かな	1.97	45.7	58	大体	0.19	3.0
17	ちょっと	5.21	38.0	59	少々	0.03	2.4
18	かも	1.18	36.6	60	っぽい	0.08	2.2
19	感じ	1.35	31.4	61	～得る	0.03	2.1
20	頃	0.59	28.1	62	必ずしも	0.03	1.8
21	くらい	1.60	27.1	63	単なる	0.03	1.8
22	少し	0.51	26.8	64	ある程度	0.05	1.8
23	みたいだ	1.50	21.5	65	大抵	0.04	1.7
24	なんか	2.22	18.3	66	少なくない	0.02	1.6
25	系	0.29	15.7	67	場合もある	0.03	1.6
26	ほとんど	0.29	14.7	68	とりあえず	0.18	1.6
27	多くの	0.20	14.3	69	なんとなく	0.05	1.6
28	たぶん	1.01	14.0	70	だいふ	0.08	1.6
29	関係	0.26	13.7	71	一応	0.25	1.4
30	なんて	0.36	13.0	72	恐れがある	0.02	1.1
31	やはり	1.80	12.9	73	ほど	0.02	1.1
32	可能性	0.17	10.7	74	とは限らない	0.01	0.9
33	おそらく	0.22	10.3	75	たしか	0.05	0.9
34	と思われる	0.14	10.1	76	ように思われる	0.01	0.8
35	あたり	0.20	10.1	77	んじゃないの	0.08	0.8
36	もちろん	0.30	10.0	78	めく	0.01	0.7
37	あまり	0.89	9.4	79	概ね	0.01	0.6
38	らしい	0.25	9.1	80	万一	0.01	0.5

39	方向	0.17	8.9	81	あいにく	0.00	0.3
40	まあ	1.83	8.5	82	あの（う）	1.42	0.2
41	なかなか	0.29	7.6	83	ように考えられる	0.00	0.1
42	～にくい	0.13	6.8	計		100	

図 5

上位語の累計構成比



上図で明らかなように、上位 40 語まではカバー率が一気に上がっており、40 語を超えると、その傾きが非常に緩やかになる。頻度率に注目すると、上位 20 語までで 79.0%、上位 40 位までで 93.2%と、これらの 40 種のヘッジ語形で全体の頻度の大部分を占めていることが確認された。これら 40 種のヘッジ語形はヘッジ研究の際のサンプルとして使用できるのみならず、学習者のヘッジ学習の指針が提供できることにもなり、日本語学的にも日本語教育学的にも利点が多いと考えられる。

5. まとめ

以上、本研究では、話し言葉・書き言葉差とジャンル差の違いによって、ヘッジ語形の使用頻度がどう異なるか調査し、また、これらの影響をあまり受けずに安定して出現する高頻度ヘッジ語形の抽出を試みた。分析の結果、以下の 4 つのことが言える。

まず、言語環境とヘッジ語形使用量の関係については、話し言葉におけるヘッジ語形の出現頻度が書き言葉の 3 倍以上であることが明らかになった。また、ジャンル差におけるヘッジ語形使用状況の違いについて、話し言葉では $I > RP > CEJC > D > ST$ 、書き言葉では

LB2>LB1>PM>OT>PN の順に、ヘッジ語形の出現頻度が低くなっていることが確認された。特に、情報提供を主たる目的とするテキストより、筆者のスタンスが自由に表出される小説系・創作系テキストのほうが、独話系テキストより、対話系テキストのほうが、ヘッジ語形が多用されることが明らかになった。

つづいて、高頻度ヘッジ語形については、話し言葉と書き言葉それぞれにおいて高頻度に出現するヘッジ語形に関して、上位 10 語中半分しか一致しておらず、ヘッジ語形の振る舞いは話し言葉と書き言葉間で大きな差異があることが示唆された。また、特定のジャンルにのみ高頻度で出現しているヘッジ語形が存在することが確認された。

3 点目として、ジャンルとヘッジ語形の関係を見るコレスポネンス分析を通して、書き言葉より、話し言葉のほうがジャンル差に影響されやすいことや、各種のヘッジ語形が、情報伝達の方向性と場面の特定性の 2 観点によって 5 カテゴリーに分類できることが確認された。

4 点目は、異なるジャンルにおいて安定的に高頻度で使用されるヘッジ語形を 40 語抽出できたことである。それらの 40 種のヘッジ語形の使用割合は全体の 9 割あまりを占めているため、これからのヘッジ研究を行う上で、あるいはヘッジ語形を学習する上で重要な語が特定されたといえる。

もっとも、本研究には制約も残されている。1 点目は分析の対象がすべてのヘッジ語形をカバーしていないことである。従来の先行研究で取り上げられていないが、実際にヘッジとして使用される語が存在するため、今後は調査対象の範囲を広げ、ヘッジ語形と言語環境の関係についてさらに詳しく調査する必要があると思われる。2 点目はヘッジ語形の意味機能を議論していないことである。本研究ではヘッジとなりうる語形の使用実態を議論してきたが、ヘッジ語形が持つ意味機能の違いを考慮していない。例えば、「つまらない物ですけども、お受け取りください」の「けれども」がヘッジとして機能しているが、「登りは苦しいけれども、山頂は素晴らしい」の例では、「けれども」がヘッジの機能を実現していない。今後はヘッジ語形とその意味機能との関係について、内省調査・直観調査の必要性があると思われる。

引用文献

- 石川慎一郎・前田忠彦・山崎誠(2010). 『言語研究のための統計入門』 くろしお出版.
- 袁青(2018). 「日本語と中国語字幕に見られるヘッジ表現－ポライトネスの観点から－」『通訳翻訳研究への招待』(日本通訳翻訳学会) 19, 109-125.
- 小森万里(2016a). 「中級以上の日本語学習者に対するアカデミック・ライティングのための日本語文法教育－引用動詞を用いた意見述べの文末表現をめぐって－」『ヨーロッパ日本語教育：2015 日本語教育シンポジウム報告・発表論文集』 20, 303-308.
- 小森万里(2016b). 「日本語を母語とする大学教員と日本語学習者の意見文におけるメタディスコースの比較」『CAJLE 2016 Annual Conference Proceedings』 144-153.

- Kekidze, T.(2003). 「現代日本語における表現の :『やわらげ』～『そうだ』,『げ』,『ぼい』などの場合」『言葉と文化』(名古屋大学大学院国際言語文化研究科日本語文化専攻)4, 293-306.
- 入戸野みはる(2004). 「ヘッジの形とその機能－友人間の会話に見る－」南雅彦・浅野真紀子(編)『言語学と日本語教育Ⅲ』 (pp.289-304).くろしお出版.
- 東泉裕子・高橋圭子(2020). 「現代日本語における『もちろん』のヘッジ用法－現代語コーパスの用例より－」『アカデミック・ジャパニーズ・ジャーナル』 12, 26-34.
- 堀田智子(2013). 「中国人日本語学習者の『不同意』行為－中間言語語用論の観点から－」東北大学博士論文.
- 堀田智子・堀江薫(2012). 「日本語学習者の『断り』行動におけるヘッジの考察－中間言語語用論分析を通じて－」『語用論研究』 14, 1-19.
- 山川史(2011). 「学習者のヘッジ使用－OPI におけるレベル別会話分析－」『日本語教育研究』 57, 124-142.
- Lauwereyns, S. (2002). Hedges in Japanese conversation: The influence of age, sex, and formality. *Language Variation and Change*, 14, 239-259.
- 李恩美(2008). 「日本語と韓国語の初対面二者会話における対人配慮行動の対照研究 : ディスコース・ポライトネス理論の観点から」 東京外国語大学博士論文.
- Lee 風子・楊彩虹(2013). 「日本語・英語・中国語の新聞の社説に現れる Stance 表現と Engagement 表現－中国語の特徴を中心に－」 *Studies in language science* (立命館大学) 3, 75-91.