

研究ノート

表現の「ゆれ」をレジスターから考える試論  
—「と思う」を例に—

東出 朋<sup>1)\*</sup>, 王 琪<sup>2)</sup>

(<sup>1)</sup> 長崎国際大学人間社会学部国際観光学科、<sup>2)</sup> 杭州師範大学、\*連絡対応著者)

A Preliminary Study on Variations in Expression from the  
Perspective of Register: Focusing on “to omou” (I think) as an Example

Tomo HIGASHIDE<sup>1)\*</sup>, Wang Qi<sup>2)</sup>

(<sup>1)</sup> Faculty of Human and Social Studies, Department of International Tourism, Nagasaki International  
University, <sup>2)</sup> Hangzhou Normal University, \*Corresponding author)

**Abstract**

This paper focuses on the four expressions related to “と思う (I think that)” and uses corpora to statistically analyze the actual usage variations regarding the presence or absence of the copula. As a result, although these expressions are used across various registers, certain trends were observed. In the “Diet minutes”, expressions including copula are common, while the ones without copula are common in “Newsletter” and “Blogs”. We considered that expressions including the copula tend to be used in registers where the subject of the expression (speaker or writer) is easily brought to the fore.

**Key words**

“I think”, variations, corpus, register

**要 旨**

本稿は「と思う」という表現に焦点を当て、コーパスを利用していくつかのレジスターにおけるコピュラの有無に関する「ゆれ」の実態を分析した。その結果、「と思う」に関する表現は様々なレジスターで用いられるものの、いくつかの傾向も観察された。「国会会議録」ではコピュラを含む「(こと) だと思う」が多く、「広報誌」と「ブログ」ではコピュラを含まない「(こと) のと思う」が多い。表現主体 (話し手/書き手) を前面に出しやすいレジスターではコピュラを含む表現も用いられやすい傾向があると考えた。

**キーワード**

と思う、ゆれ、コーパス、レジスター

## 1. はじめに

言語使用には、Aとも言えるしBとも言える、どちらでもいいのではないか、というバリエーション、または「ゆれ」がしばしば観察される。「ゆれ」は音声、表記、語彙、構文レベルなどの異なるレベルで存在する。「ゆれ」に直面した際、辞書や文法書を参考にしたり周囲の母語話者に尋ねたりすることも可能であるが、結局のところ人によって判断に違いが生じ、「正解」を見つけるのは困難である。本研究は、「と思う」という構文を例に、使用実態のデータベースであるコーパスを用いて「ゆれ」の様相を分析する。変化しつつある言語において、言語使用の唯一の「正解」を求めるのではなく、「ゆれ」を積極的に認める態度も重要だと考える。

「と思う」構文において引用節を導く「と」の前にナ形容詞や名詞が共起する場合、規範文法によればコピュラ「だ」が必要である(1a)が、実際にはコピュラが出現しない例がしばしば見られる(1b)。また、バリエーションとして「ことと思う」という連語化した表現(阿部 2015)もあり(1c)、これに対するコピュラを含む表現も可能である(1d)。これら4つの表現にはコピュラの有無の違いがあるが、意味的にはほぼ同じように受け取られる。

- (1) a. 確かに親にとっては大変だと思います。(BCCWJ)
- b. チップという習慣に対する慣れの問題だと思います。(金城 2012: 21)
- c. 皆さまもご先祖を偲んで、お墓参りに行かれたことと思います。(阿部 2015: 65)
- d. 人に会うにあたって、そのように気を使うことはとても重要なことだと思います。(BCCWJ)

本稿では、「と思う」という構文に焦点を

当て、コーパスを利用して「ゆれ」の実態を分析する。

## 2. 先行研究

### 2.1. コーパスを用いた文法研究

近年、コーパスを用いてレジスターと文法の関係进行分析の研究が盛んに行われている(丸山 2015、滝島 2020、馬場 2018, 2021、中俣 2020, 2022)。コーパスは言語現象とその使用域/レジスター/文体の関係を調査するのに利用価値が高い。

丸山(2015)は、書き言葉(BCCWJ)と話し言葉(CSJ)およびそこに含まれる各レジスターにおける助詞の使われ方の違いをコレスポネンス分析によって分析した。また、滝島(2020)は話し言葉(CEJC)と書き言葉(BCCWJ)における敬語名詞の出現傾向を比較した結果、話し言葉ではできるだけきれいな言葉遣いをする「美化意識」があると主張した。馬場(2018, 2021)と中俣(2020, 2022)は副詞、接続詞など1つの品詞グループとされる各詞の出現傾向とレジスターとの関連性を統計的に明らかにしている。たとえば、中俣(2022)は、「中納言」の「まとめて検索KOTONOHA」を用いて、並列を表す接続詞とレジスターとの関係を主成分分析によって可視化し、その結果「一方向的VS双方向的」と「自己表現的VS公共的」という対立軸を抽出した。

これらの研究は、コーパスを用いて計量的に文法とレジスターを分析することで、動的・流動的な言語使用の実態を明らかにするものである。

### 2.2. 引用句におけるコピュラに関する研究

「と思う」という表現に関する先行研究は2種類に分けられる。連語表現「ことと思う」と「と思う」の意味的な違いに関する研究と、「だと思う」と「〇と思う」、つまりコピュラの有無に関する研究である。

まず、いわゆる連語表現「ことと思う」について、高橋（2021）は形式名詞「こと」を含む様々なモダリティ表現を分析するなかで「ことと思う」も取り上げ、「ことと思う」という表現と「こと」を削除した例を比較している。次の例について、「「こと」が使われている場合は、「思う」の内容が確実ではなく、それを推測している」（p.7）と説明している。

- (2) a. 帰らない娘を待って、六十過ぎの  
き志はただ一人、東京の仮住まいで  
不安な日夜を送ったことと思われる。  
(BCCWJ)
- b. 帰らない娘を待って、六十過ぎの  
き志はただ一人、東京の仮住まいで  
不安な日夜を送ったと思われる。  
(高橋 2021 : 6-7、下線引用ママ)

つまり、「こと」の有無と思考内容の確実性の関係を述べている。ただし、これ以上の記述はなく、分析は不足している。

次に、「と思う」という表現におけるコンピュータの非出現について分析した研究には金澤（2008）、金城（2012）、阿部（2015）東出・王（2022）が挙げられる。

金澤（2008）は新聞の意見文を観察し、「思う」は受動態のほうが能動態よりコンピュータの脱落が起きやすいことを指摘した。金城（2012）は、母語話者の「と思う」「と考える」という2つのコンテキストについて分析した結果、「と思う」でのコンピュータの非出現傾向は形式名詞>形容動詞>名詞の順であると述べた。また、非出現と自発表現「と思われる」の共起率が高いことから、コンピュータの非出現は断定を避ける、つまり発話の力の軽減という語用論的な要因があると考察した。阿部（2015）は、コンピュータの非出現の条件を「思う」の形式と文内での位置という文法的観点から分析し、文末の場合は能動態でコンピュータが出現しやすいのに対し受動態で出現と脱落が同

等であること、連体修飾（文内）の場合は文末より脱落が起きやすいことを指摘している。また、コンピュータの非出現の容認性は引用句が「思考内容そのもの」を表す環境にあるか否かに依存すると主張した。東出・王（2022）は、コーパスを用いてコンピュータの有無に関する母語話者と学習者の違いを分析した。母語話者の運用については、コンピュータの有無は書き言葉と話し言葉、また前接名詞タイプと関わっている傾向が見られた。

「と思う」表現をめぐる以上の先行研究は、バリエーションの意味的、語用論的、使用者による違いを議論している。これらの違いに着目することも有用だが、同時に、現実の運用の総体を観察し、使用傾向をつかむことも重要だと考える。

### 2.3. リサーチクエスチョン

先行研究を踏まえた上で、本稿はレジスターに着目し、構文とレジスターとの関係について分析する。リサーチクエスチョンは以下の3点である。

1. 「だと思う」と「〇と思う」は、どのレジスターで頻繁に用いられているのか。
2. 「ことだと思う」と「こと〇と思う」は、どのレジスターで頻繁に用いられているのか。
3. 「だと思う」と「〇と思う」のレジスターの傾向と「ことだと思う」と「こと〇と思う」のレジスターの傾向は類似しているのか。

## 3. データと分析手順

### 3.1. レジスターの選定

本稿では、検索アプリケーション「中納言」に記載されている4つのコーパス（「現代日本語書き言葉均衡コーパス中納言版（以下、BCCWJ）」、「日本語話し言葉コーパス（以下、CSJ）」、「名大会話コーパス（以下、NUCC）」、「現日研・職場談話コーパス（以下、

CWPC)) を利用した。これらのコーパスにはレジスターや場面情報が付与されている。レジスターの選定は次のように行った。BCCWJについて、「図書館・書籍」「出版・書籍」には幅広い書籍が収録されているため1つのレジスターとして扱うのは妥当ではないと考え、対象から除外した。文体的特徴を比較するには鮮明な特徴を持つレジスターを選ぶのが重要であるため、特定目的のサブコーパス(「ベストセラー」、「Yahoo!知恵袋」、「Yahoo!ブログ」、「国会会議録」など、「韻文」を除く7つ)を利用することとする。また、CSJからは「学会講演」と「模擬講演」という2つのレジスターを抽出できる。NUCCは自然な会話のコーパスであり、「雑談」というレジスターにタグづけた。CWPCからは「会議」と「休憩」という2つのレジスターを抽出した。

### 3.2. 分析手順と結果

分析手順は以下の3ステップである。まず、中納言を用いて各コーパスから「名詞/形状

詞/代名詞+だ/〇+と+思う」及び「名詞/形状詞+こと+だ/〇+と+思う」という文字列を抽出し、13のレジスターにおける出現回数を集計する。なお、「思う」は全ての活用形を含む。次に、13のレジスターの総語数を確認した上で、各構文の出現回数の調整頻度(pmw)を算出する。分析では粗頻度(出現頻度)が5以上を分析対象とする。この条件に合致するレジスターは、「(だ)と思う」に関して10、「こと(だ)と思う」に関して7つであった。それから、各構文のレジスター別の分析を行う。

各構文のレジスター別出現頻度と調整頻度、及び割合を表1と表2に示す。

### 4. 分析

表1及び表2の調整頻度及び割合を見ると、コンピュータを含む構文「(こと)だと思ふ」のほうがより頻繁に使われていることは明らかである。レジスター間において頻度に有意な差があるかどうか検定したところ、下の結果を得た。

表1 「だと思ふ」と「〇と思ふ」のレジスター別出現頻度及び調整頻度、割合

レジスター	だと思ふ			〇と思ふ			総語数
	出現頻度	pmw	割合	出現頻度	pmw	割合	
知恵袋	6338	623.64	84.73%	1142	112.37	15.27%	10,162,945
国会会議録	2329	456.45	89.99%	259	50.76	10.01%	5,102,469
ブログ	1974	195.42	72.52%	748	74.05	27.48%	10,101,397
模擬講演	1713	475.08	91.46%	160	44.37	8.54%	3,605,729
ベストセラー	679	181.44	74.29%	235	62.8	25.71%	3,742,261
学会講演	497	151.55	79.78%	126	38.42	20.22%	3,279,364
雑談	323	285.34	85.90%	53	46.82	14.10%	1,131,971
広報誌	128	34.09	64.65%	70	18.64	35.35%	3,755,161
教科書	29	31.23	76.32%	9	9.69	23.68%	928,448
白書	15	3.2	8.82%	155	33.08	91.18%	4,685,801
会議・休憩	51			1			186,906
法律	0			0			1,206,077

表2 「ことだと思う」と「こと〇と思う」のレジスター別出現頻度及び調整頻度、割合

レジスター	ことだと思う			こと〇と思う			総語数
	出現頻度	pmw	割合	出現頻度	pmw	割合	
国会会議録	474	92.9	91.68%	43	8.43	8.32%	10,101,397
知恵袋	308	60.36	88.00%	42	8.23	12.00%	3,742,261
模擬講演	260	35.65	90.91%	26	3.56	9.09%	5,102,469
ブログ	109	21.36	78.42%	30	5.88	21.58%	10,162,945
ベストセラー	75	14.7	86.21%	12	2.35	13.79%	7,293,246
学会講演	69	9.46	84.15%	13	1.78	15.85%	7,293,246
広報誌	18	3.53	54.55%	15	2.94	45.45%	3,755,161
雑談	4			1			1,135,329
会議・休憩	2			0			186,906
教科書	1			0			928,448
白書	0			0			4,685,801
法律	0			0			1,206,077

表3 「だと思う」と「〇と思う」の検定結果

レジスター	だと思う		〇と思う	
	出現頻度		出現頻度	
知恵袋	<b>6338</b>	<b>6.541**</b>	1142	-6.541**
国会会議録	<b>2329</b>	<b>10.790**</b>	259	-10.790**
ブログ	1974	-15.115**	<b>748</b>	<b>15.115**</b>
模擬講演	<b>1713</b>	<b>10.732**</b>	160	-10.732**
ベストセラー	679	-6.801**	<b>235</b>	<b>6.801**</b>
学会講演	497	-1.886+	126	1.886+
雑談	323	1.715+	53	-1.715+
広報誌	128	-6.696**	<b>70</b>	<b>6.696**</b>
教科書	29	-1.021 ns	9	1.021 ns
白書	15	-25.490**	<b>160</b>	<b>25.490**</b>

表4 「ことだと思う」と「こと〇と思う」の検定結果

レジスター	ことだと思う		こと〇と思う	
	出現頻度		出現頻度	
国会会議録	<b>474</b>	<b>3.273**</b>	43	-3.273**
知恵袋	308	0.075 ns	42	-0.075 ns
模擬講演	260	1.745+	26	-1.745+
ブログ	109	-3.592**	<b>30</b>	<b>3.592**</b>
ベストセラー	75	-0.494 ns	12	0.494 ns
学会講演	69	-1.067 ns	13	1.067 ns
広報誌	18	-5.935**	<b>15</b>	<b>5.935**</b>

#### 4.1. カイ二乗検定の結果

まず、「だと思う」と「〇と思う」の頻度についてレジスター間でカイ二乗検定を行った結果、「雑談」を除く9つのレジスターで有意な差が認められた ( $\chi^2(9)=1177.725$ ,  $p<.01$ )。表3に結果を示す。「だと思う」は「模擬講演」「国会会議録」「知恵袋」で多く、「〇と思う」は「白書」「ブログ」「ベストセラー」

「広報誌」が多い。

次に、「ことだと思う」と「こと〇と思う」の頻度についてレジスターでカイ二乗検定を行った結果、「国会会議録」と「ブログ」においては有意な差が認められた ( $\chi^2(6)=56.923$ ,  $p<.01$ )。表4に結果を示す。「ことだと思う」は「国会会議録」で多く、「こと〇と思う」は「広報誌」と「ブログ」が多い。

## 4.2. まとめ

分析の結果、リサーチクエスチョンに対する答えは次の通りに得られた。

1. 「だと思う」と「〇と思う」は、どのレジスターで頻繁に用いられているのか。

「だと思う」は「模擬講演」「国会会議録」「知恵袋」で多く、「〇と思う」は「白書」「ブログ」「ベストセラー」「広報誌」が多い。

2. 「ことだと思う」と「こと〇と思う」は、どのレジスターで頻繁に用いられているのか。

「ことだと思う」は「国会会議録」で多く、「こと〇と思う」は「広報誌」と「ブログ」が多い。

3. 「だと思う」と「〇と思う」のレジスターの傾向と「ことだと思う」と「こと〇と思う」のレジスターの傾向は類似しているのか。

類似しているレジスターは「国会会議録」「広報誌」「ブログ」である。「国会会議録」ではコンピュータを含む「(こと) だと思う」が多く、「ブログ」と「広報誌」ではコンピュータを含まない「(こと) 〇と思う」が多い。

## 5. 考察

まず大前提として、「と思う」の4つの表現は様々なレジスターで出現している。表1及び表2からは、名詞や形容動詞、代名詞に接続する際にコンピュータ「だ」が脱落する現象が十分に存在しており、この現象を単純な誤用であると考えたり、安易に無視したりすることができないことは明らかである。なお、「法律」は基本的に一個人の意見を表明する媒体ではないことから、出現数も0である。

そのうえで、リサーチクエスチョン1、2、3に対する結果からは、表現主体（話し手/書き手）を前面に出しやすいレジスターではコンピュータを含む構文も用いられやすく、一方

で表現主体（話し手/書き手）を前面に出にくい、または出す必要性が低いレジスターではコンピュータが脱落しやすいという傾向があると考えられる。そもそも、コンピュータ「だ」は断定を表す助動詞である。金城（2012）も指摘するように、コンピュータの非出現には断定を避けたいという語用論的要因もある。

### 5.1. リサーチクエスチョン1

まず、「模擬講演」は「人生で一番嬉しかった/悲しかったこと」「私の住んでいる街」などの簡単なテーマに基づいた一般話者によるスピーチで、発話スタイルは比較的くだけている。内容的にも形式的にも、自分の意見・感想を主張しうる場である。「国会会議録」は国会で開かれる常任委員会や特別委員会で、専門的かつ詳細に「議論」する場である。予算を獲得したり法案を通したりするなど、自分(達)の主張を強く訴える必要が生じる。「知恵袋」は、質問に対して回答を提示する場であるが、回答を提供する際にはその根拠を述べる必要が生じる。意見を強く主張する場面が発生しやすく、それが容易に可能であるレジスターでは、コンピュータを伴う「だと思う」が観察される。

一方、「白書」「ブログ」「ベストセラー」「広報誌」における「〇と思う」の出現頻度及び相対頻度は十分に小さい値であるものの、それでも「と思う」が用いられる場合を考察しよう。これら4つは全て、文字による情報伝達を目的としているレジスターである。「白書」は情報をまとめて報告する媒体で、そこに出現した「〇と思う」はそのほぼ全てが「ものと思われる」であり（125/155）、「白書」特有の表現と考えられる。「ブログ」は自分の経験や意見を書き発信する媒体、「ベストセラー」は様々な内容について書き手による熟考を経た、技巧的な表現を含む媒体、そして「広報誌」は何らかの情報を広報する媒体である。これら4つは、一定の時間をかけて

まとまった内容を練り、不特定多数に向けて情報を発信するものである。「白書」と「広報誌」は書き手である一個人が前面に出る必要性が非常に低いと言えるだろう。「ブログ」と「ベストセラー」もカジュアルに書かれる傾向があり、表現主体が強く意見を主張する必要性が低いと言える。

## 5.2. リサーチクエスチョン2、3

「ことだと思う」は「国会会議録」で多く、「こと〇と思う」は「広報誌」と「ブログ」が多い（リサーチクエスチョン2）。これら3つは「だと思う」と「〇と思う」の特徴的なレジスターと重なるため、結果として「国会会議録」では「（こと）だと思う」が、「広報誌」と「ブログ」では「（こと）〇と思う」が多い（リサーチクエスチョン3）。

これら3つのレジスターについて、前述の通り、「国会会議録」は実際に対面で質問者と回答者が交互に意見を述べ合っていく場であるのに対し、「広報誌」と「ブログ」は文字によって一方的に情報を伝達・発信する場である。前者はその場で意見に反対されたり批判されたりするが、後者はそのようなことはほぼなく、あっても時差がある。意見を主張する必要性、そして表現主体が前面に出る必要性に大きな違いがあるだろう。

## 5.3. 「こと〇と思う」

表2及び表4からは、「こと〇と思う」の出現回数が非常に少ないことが分かる。「ことと思う」は連語化した表現であるという意見（阿部 2015）があるものの現実にはそれほど多くないと同時に、前述の通り、「白書」において「ものと思われる」が定型表現として出現していることから、この2つの表現については質的な分析が必要だと言える。

## 6. おわりに

本稿は「と思う」という表現に焦点を当て、

コーパスを利用して「ゆれ」の実態を分析した。その結果、「と思う」は様々なレジスターで用いられるものの、いくつかの傾向も観察された。「国会会議録」ではコンピュータを含む「（こと）だと思う」が多く、「ブログ」と「広報誌」ではコンピュータを含まない「（こと）〇と思う」が多い。そして、表現主体（話し手/書き手）を前面に出しやすいレジスターではコンピュータを含む表現が用いられやすい傾向があると考えた。

どの表現も使用可能でありつつ、同時にある種の傾向も存在する。表現の「ゆれ」とその傾向を知ることは、学習という点でも示唆を含んでいる。教科書に書いてあることだけが現実に用いられているわけではないことは、学習者の言語運用能力が上がるにつれて徐々に理解されてくる。学習者に対して、音声や語彙のバリエーションのみならず、語法的なバリエーションも提示する必要があるだろう。

本稿が用いたコーパスはテナーやモードが限定的であるため、「ゆれ」の全貌を掴んだとはもちろん言えない。今後は、レジスター別に好まれる形式を量的・質的に分析する必要があるだろう。

## 参考文献

- 阿部二郎（2015）「引用句内におけるコンピュータの非出現について——「～だと思う」と「～と思う」——」『文法・談話研究と日本語教育の接点』57-78、くろしお出版
- 小椋秀樹（2019）「書き言葉コーパスに見る尊敬表現——「お（ご）～になる」「お（ご）～なさる」をめぐって——」『日本語学』38(2)、22-31、明治書院
- 金澤裕之（2008）『留学生の日本語は、未来の日本語——日本語の変化のダイナミズム』ひつじ書房
- 金城克哉（2012）「コーパスに基づく引用句内のコンピュータ（「だ」）の顕在と潜在に関する研究」『留学生教育：琉球大学留学生センター紀要』9、21-33、琉球大学留学生センター
- 高橋雄一（2021）「「もの」「こと」を含む複合辞の広がり」『東アジア国際言語研究』第2号、1-12、東アジア国際言語学会

- 滝島雅子 (2020) 「話し言葉と書き言葉における敬語名詞の語彙比較 CEJCとBCCWJのデータを用いて」『計量国語学』32(6)、315-330、計量国語学会
- 中俣尚己 (2020) 「主成分分析を用いた副詞の文体分析」『計量国語学』32(7)、419-435、計量国語学会
- 中俣尚己 (2022) 「並列を表す接続詞と文体「まとめて検索KOTONOHA」を利用して」『計量国語学』33(7)、422-434、計量国語学会
- 馬場俊臣 (2018) 「『BCCWJ図書館サブコーパスの文体情報』を利用した語の文体差研究の可能性」『言語資源活用ワークショップ発表論文集=Proceedings of Language Resources Workshop』Vol. 3、241-256、国立国語研究所
- 馬場俊臣 (2021) 「副詞の文体の計量について——中俣論文の主成分分析結果との比較——」『語学文学』60、26-35、北海道教育大学語学文学会
- 丸山直子 (2015) 「助詞の使用実態—BCCWJ・CSJにみる分布—」『第8回コーパス日本語学ワークショップ予稿集』179-188、国立国語研究所
- 東出朋, 王琪 (2022) 「关于「と思う」引用句前接的判断助动词消隐现象的考察——学习者与日语母语者的对比分析」『高等日語教育』10、30-40、外研社综合语种教育出版分社