日本語日常会話コーパスにおける推量副詞とモダリティ形式の働き

イレーナ・スルダノヴィッチ（ユライ・ドブリラ大学プーラ）

The Use of Suppositional Adverbs and Modal Forms in the Corpus of Everyday Japanese Conversation (CEJC)

Irena Srdanović (Juraj Dobrila University of Pula)

要旨

推量副詞と文末モダリティの呼応および遠隔共起関係の現象については、従来の研究で既に議論されている（南 1974、工藤 2000、Bekeš 2006、Srdanović ら 2008、スルダノヴィッチら 2009a,b,c、ホドシュチェックら 2010）。複数の日本語コーパスを用いた実証的な研究では、コーパス・ジャンル別の副詞分布および副詞と文末モダリティ形式の共起傾向が異なり、モダリティタイプの程度で表されることが確認された（スルダノヴィッチら 2009a,b,c、ホドシュチェックら 2010）。本研究の目的は、その現象を日本語日常会話コーパス（CEJC）（Koiso ら 2022、小磯ら 2022）内で観察し、他のコーパスと比較しつつ、推量副詞の分布を示し、推量副詞とモダリティの振る舞いを明らかにすることである。日本語日常会話コーパスから抽出したモダリティ形式を提供し、その多様性および書き言葉とは異なる文体的特徴を確認した。推量副詞が出現した発話およびその長い前後文脈におけるモダリティ形式を考察した結果、発話における高頻度の体系的なモダリティタイプとの呼応関係と同時に、談話における重要な役割を果たす異なるモダリティ形式およびタイプの出現が明確になった。得られた研究結果は、会話における副詞とモダリティの振る舞いをさらに明確にし、日本語教育、談話分析、語用論などの分野に応用できる。

Abstract

The phenomenon of the correspondence between suppositional adverbs and sentence-final modalities, as well as their distant collocational relations, has already been discussed in previous research (Minami 1974, Kudo 2000, Bekeš 2006, Srdanović et al. 2008, Srdanović et al. 2009a, b, c, Hodošček et al. 2010). Empirical studies using multiple Japanese corpora have confirmed that the distribution of adverbs and their co-occurrence with sentence-final modality forms vary by corpus genre and are characterized by different degrees of modality types (Srdanović et al. 2009a, b, c, Hodošček et al. 2010). The aim of this study is to investigate these phenomena within the Corpus of Everyday Japanese Conversation (CEJC) (Koiso et al. 2022), and, by comparing it with other corpora, to show the distribution of suppositional adverbs and clarify their behavior along with modality. This study provides an analysis of modality forms extracted from the CEJC and confirms their diversity and stylistic features distinct from written language. Examination of modality forms in utterances containing suppositional adverbs, including their extended preceding and following contexts, reveals a clear correspondence with typical high-frequency modality types and the emergence of various modality forms and types that play significant roles in discourse. The results further clarify the behavior of adverbs and modality in conversation, with implications for fields such as Japanese language education, discourse analysis, and pragmatics.

１．はじめに

推量副詞は、話し手の主観的な態度を表す陳述副詞の一つのグループである。従来の様々な研究では、推量副詞と文末モダリティの強い呼応および遠隔共起関係の現象について議論されている（南 1974、工藤 2000、Bekeš 2006、Srdanović ら 2008、スルダノヴィッチら 2009a,b,c、ホドシュチェックら 2010）。南（1974）は、日本語の四階層の入れ子構造において、副詞と文末モダリティ表現（例えば、「どうやら～らしい」や「多分～のだろう」）が C 層にあり、意味的に呼応していると述べている。工藤（2000）は、特定のモダリティを示す副詞群とモダリティ形式との間の連続性を表していると指摘している。新聞や近代文学のデータを用いて、モダリティ副詞が文末や発話のモダリティ形式と強い関連を示す傾向が明確に示されている。コーパス言語学において、これらの関係は共起の確率として観察され、「遠隔共起」として扱われている（Bekeš 2006、Srdanović ら 2008、スルダノヴィッチら 2009a,b,c）。

この副詞群は四つに区分され、「確信」「推測」「推定」「不確定」の順で、事態の実現の確実さおよび話者の確信の度合いが低くなることを連続的な関係で示している（工藤 2000、スルダノヴィッチ 2009）。推量の副詞と共起するモダリティ形式は、四つのモダリティタイプ、すなわち「に違いない」「はずだ」のような確信（NEC, necessity）、「だろう」「と思う」のような推測（EXP, expectation）、「らしい」「みたい」のような推定（CON, conjecture）、「かもしれない」「かな」のような不確定（POSS, possibility）に分けられる（工藤 2000, p.203）。例えば、「多分」は「だろう」「のだろうか」「と思う」「のではないか」などの推測（EXP）タイプに属するモダリティ形式と強く呼応する。陳述副詞と述部モダリティの呼応の関係は、コーパスにおける共起の統計的な傾向として捉えることができる。

Srdanović ら（2008）、スルダノヴィッチら（2009a,b,c）の実証的な研究では複数の日本語コーパスを用いて、コーパス・ジャンル別の副詞分布および副詞と文末モダリティ形式の共起傾向が異なり、モダリティタイプの程度で表していると確認した。さらに、副詞の代表的なモダリティタイプによって、コーパスのタイプ・ジャンルを判定できることを発見した。その後、ホドシュチェックら（2010）は、計量的に副詞と文末モダリティを様々なコーパスのジャンルごとに分析し、レジスターの区別について論じた。

本研究では、国立国語研究所で新しく構築された日本語日常会話コーパス（Koiso ら 2022、小磯ら 2022）における推量副詞の分布および推量副詞とモダリティ形式の振る舞いを検討することを目的とする。扱う副詞は「きっと」「必ず」「絶対」「絶対に」「恐らく」「たぶん」「さぞ」「大方」「大抵」「大概」「どうやら」「どうも」「余程（よっぽど・よほど）」「あるいは」「もしかすれば」「もしかすると」「もしかしたら」「ひょっとしたら」「ひょっとすると」「ことによると」「案外」の 22 語の推量的副詞である。副詞の分布を複数の日本語コーパス（13 種）で比較するために、「KOTONOHA」というコーパス検索ツール（Oka ら 2020）を用いる。推量副詞とモダリティ形式を抽出するために、中納言などの CEJC のツールを用いる。

副詞が出現している会話において、対象の副詞から長い前後文脈に現れるモダリティの振る舞いを検討し、モダリティ形式が会話における話し手と聞き手のコミュニケーションにおける様々なアイディア、感情、態度などを関連付けて理解するための不可欠な役割を持つことを示す。

２．推量副詞分布

２.１　複数コーパスにおける推量副詞

図１は複数の日本語コーパス（「NINJAL コーパス」[1]）における 22 語の推量副詞の分布を示している。国立国語研究所で開発された「KOTONOHA」というコーパス検査ツールを用いて分析し、各コーパス内の形態素解析された短単位（SUW）(Den et al. 2008)の総数が異なるため、分析には調整された頻度 PMW (per million words) でまとめている。平均以上の値および高頻度の値はピンクで強調表示する。

分布のデータから以下のようなことが分かった。

i)　日本語コーパス内で最も広く出現する推量副詞は、「たぶん」、「ぜったい」、「どうも」、「かならず」、「きっと」、「おそらく」である。一方で、きわめて出現が少ないものは,「もしかすれば」、「ことによれば」、「ことによると」、「ひょっとする」などである。

ii)　推量副詞が最も多く出現するコーパスは，話し言葉コーパス（CEJC、NUCC、C-JAS、CWPS、I-JAS）である。最も高頻度の副詞は話し言葉コーパスに最も広く分布し、コミュニケーションにおける副詞の重要な役割と一致している。

iii)　一部の副詞は書き言葉のデータ（SHC、CHJ、BCCWJ における「おそらく」、「かならずしも」、「どうやら」など）およびフォーマルな話し言葉（CSJ や SSC における「おそらく」）でより頻繁に現れる傾向がある。

iv)　最も頻度の低い副詞群（「さぞ」、「ひょっとしたら」、「ひょっとすると」、「ことによると」、「ことによれば」、「もしかすれば」）は、主に話し言葉ではなく書き言葉のコーパスで使用されている。具体的には、特に BCCWJ や SHC はさまざまな副詞をバランスよくカバーしていることが分かる。副詞の「かならず」は特に最もバランスのとれたコーパス別、時代別の分布となっている。

v)　一部の副詞は、主に昭和時代の使用として、特筆されるが、それ以前の時代にも見られる（「よっぽど/よほど」、「たいがい」、「おおかた」、「かならずしも」、「さぞ」、「かならず」など)。BCCWJ でカバーされている現代の文書も含まれる。

vi)　学習者コーパスにおける副詞の分布は母語話者の分布とだいぶ異なる。「多分」は I-JAS に過剰に使用され、「ぜったい」と「かならず」は C-JAS に多い。学習者の推量副詞の全体の使用頻度はすべてのコーパスの副詞の平均頻度を上回っている一方で、使用していない推量副詞が多い。 C-JAS で 11 個の副詞および I-JAS で 6 個の副詞が出現しないことから、副詞の分布の不均衡があると言える。それがデータ自体の特徴によるものなのか、学習者の日本語使用の特徴によるものなのか、さらに検討する必要がある。

vii)　推量副詞のモダリティの種類に関しては、「推測」(EXP) と「確信」(NEC) は最も広く出現し、「不確定」(POSS) と「推定」(CON) より頻繁に使用されている。この観察は、異なる日本語データを分析した Srdanovic (2009) の研究結果と一致している。コミュニケーションにおいて、「推測」と「確信」の方が「不確定」と「推定」より機能的に優先することを暗示している。

---

[1] 使用した日本語コーパスは、「NINJAL コーパス」と呼ばれ、国立国語研究所（NINJAL）のホームページに記述されている。

表 1 複数コーパスにおける推量副詞の分布

モダリティ	WPMの合計 / 副詞	コーパス BCCWJ	CEJC	CHJ	C-JAS	COJADS	CSJ	CWPC	I-JAS	NUCC	SHC	SSC	合計	平均	
EXP	たぶん	75	1195	18	1465	10	222	465	2303	878	49	79	6759	614	
NEC	ぜったい	62	386	20	448	50	100	273	71	469	77	117	2073	188	>300
CON	どうも	53	112	89	15	281	107	407	51	107	74	484	1778	162	>100
NEC	かならず	107	74	289	159	133	130	139	48	77	121	106	1382	126	>平均
NEC/EXP	きっと	73	205	61	22	36	99	107	37	259	51	114	1063	97	
EXP	おそらく	75	16	63	0	41	63	59	6	17	113	108	561	51	
EXP	たいてい	30	17	70	7	89	17	32	31	18	37	81	431	39	
NEC	ぜったいに	52	26	12	26	28	31	21	14	49	64	42	366	33	
CON	よっぽど・よほど	23	27	74	22	49	9	16	1	44	36	55	355	32	
NEC	かならずしも	35	2	80	0	3	37	11	0.3	8	51	38	266	24	
POSS	もしかしたら	16	50	1	7	0	32	37	13	41	8	4	209	19	
NEC/EXP	たいがい	4	3	23	0	57	10	16	0	16	8	68	205	19	
POSS	あんがい	9	14	8	0	9	4	11	1	27	18	47	149	14	
CON	どうやら	27.4	2.5	13.6	0	9.0	9.1	0	1.2	13.2	29.0	13.2	118	11	
EXP	おおかた	2.2	0.8	30.5	0	9.0	2.2	0	0	0	2.6	1.9	49	4.5	
POSS	もしかすると	6.1	4.1	0.5	0	0	5.8	16.1	0.3	3.5	4.9	0	41	3.8	
EXP	さぞ	3.8	0	24.4	0	4.9	0	0	0	0	7.3	0	40	3.7	
POSS	ひょっとしたら	5.2	2.9	0.2	0	2.5	8.6	10.7	0	4.4	3.9	0	38	3.5	
POSS	ひょっとすると	3.5	0.4	0.5	0	3.3	1.8	0	0	0	3.1	0	13	1.2	
POSS	ことによると	1.3	0	1.3	0	0	0.8	0	0	0	2.0	0	5.4	0.5	
POSS	ことによれば	0.1	0	0.1	0	0	0	0	0	0	0.1	0	0.3	0.0	
POSS	もしかすれば	0.1	0	0	0	0	0	0	0	0	0.1	0	0.2	0.0	
	合計	665	2137	879	2171	814	889	1621	2578	2031	760	1358	15903	1446	

2.2 日本語日常会話コーパスにおける推量副詞の分布および特徴

日本語日常会話コーパス（CEJC）における対象にした 22 語の推量副詞を調べた結果を表 2 に示す。会話に最も多く現れている「たぶん」の次に、「ぜったい」、「きっと」、「どうも」、「かならず」、「もしかしたら」が現れている。一方で、「さぞ」、「ことによると」、「ことによれば」、「もしかすれば」は会話に出現されていない。

表 2 日本語日常会話コーパス（CEJC）における推量副詞の分布

副詞	件数	WPM
たぶん	2890	1194.6
ぜったい	934	386.1
きっと	496	205.0
どうも	270	111.6
かならず	178	73.6
もしかしたら	122	50.4
よっぽど・よほど	65	26.9
ぜったいに	63	26.0
たいてい	41	16.9
おそらく	38	15.7
あんがい	34	14.1
もしかすると	10	4.1
たいがい	7	2.9
ひょっとしたら	7	2.9
どうやら	6	2.5
かならずしも	5	2.1
おおかた	2	0.8
ひょっとすると	1	0.4
さぞ	0	0.0
ことによると	0	0.0
ことによれば	0	0.0
もしかすれば	0	0.0

コーパスから推量副詞の例文を抽出し、様々な特徴を検討する。表 3 は CEJC に現れる「たぶん」推量副詞のバリエーション、つまり、タグ付き書字形の件数、パーセンテージと説明を頻度 5 まで示しているものである。表 4 は発話における「たぶん」の繰り返し、および「たぶん」の前後 5 語以内の他の副詞の出現を示している。

分析した結果から、「たぶん」を例にして、以下の特徴①～④が見られた。

① 述語後置（post-predicate position）

表 3 に見られるように、発話単位末の「たぶん。」は、他のタグ付きの組み合わせを加えると、すべての出現の 16%を超えている。その表示は、話し言葉の特徴であり、日本語の代表的な主語・目的語・述語（SOV）の標準的な順序とは異なっている。もっと詳しく調べると、そのうち 13%は単独の発話となっている「たぶん。」で、残りの 3%ぐらいは例⑴のように発話の中の文章の終わりに使われている。話し言葉の方が書き言葉より表現を計画する時間が少なくなっていることと、話し手が不確実な情報を伝えていることと、聞き手への配慮を持ちつつ、話し手が伝えたい情報の内容を柔らかくすることなどの関係があると言える。質的な分析を今後の課題にする。

例⑴ **まあ(D ツ) (F その:) あれなので 変わらないので たぶん。**
（CEJC, T010_007）

② 重複（repetition, duplication）

表 4 に見られるように、推量副詞は発話の中では、何回も繰り返す例文がある。キーワードから 5 語前後の文脈だけを見ると、合わせて全体の例文の 6%に近付いている。さらに長いコンテクストの方はそれ以上となっているであろう。例⑵、例⑶はそのような例である。例⑵は発話の中で「たぶん」が 2 回重複し、例⑶はもっと長いコンテクストで「たぶん」の 4 回の重複となっている。そのうち、2 件は連続で繰り返す。重複の役割は、話しての意図、態度、感情を強調すること、聞き手の注目を引くこと、談話の中の様々な働きだと考えられる。

例⑵ たぶんうちが買ったようなテントはたぶん だいじょぶなんじゃない?。
（CEJC、IC02）

例⑶ 左右|対称|に|絶対|なら|ない#たぶん|な|ずれ|てん|ねん#たぶん|こう|感覚|が#あの#うん|うん|うん#あっ# たぶん |たぶん|ね|目|が|悪い|から|だ|と|思う|よ#そう|そう#だ|歩い|て|て|も|な|右|に （CEJC、K001）

③ 他の副詞との共起（cooccuring with other adverbs）

近い距離で、副詞は他の副詞と出現していることが表 4 からわかる。「たぶん」は、完全にや全くないことの意味を持つ、主に否定文で使われる副詞である「全然」と出現する傾向がある（例⑷）。推量副詞の中、「全然」とも現れる傾向がある（例⑸）。他の副詞との出現は、コミュニケーションの内容を弱くするか、強くするために利用される。あるモダリティから別のモダリティへ移動し、会話の参加者および状況に応じたコミュニケーションの流れともつながる。場合によっては、話し手の直し、発話の中の態度のスイッチなどの役割を持つ。

例⑷　(R れい)ちゃん(.)全然 たぶん気にしないと思う とかってゆって。（CEJC、K005_033）

例⑸　で 絶対みんな たぶん(D ア) あのまんまだと京都奈良(L 行っちゃうんだろうなと思って)。（CEJC、K005_033）

④ 発話のバリエーション（母音の引き延ばし、小さい声の発話、笑いが生じるなど）(variations in speech)

表 3 に示しているように、順序・発音の面から見た普通の「たぶん」の出現は 74.2%で、発話単位の出現は 15.3%となり、残りは様々な発音のバリエーションとなっている。

非語彙的な母音の引き延ばし「たぶん：」は頻繁に表れる。「たぶん」は小さい声で発話している傾向と笑い声で発話している傾向が見られる。それらは主に話者の不安定な気持ちを表していると言える。例⑹の（T たぶん）は小さい声で発話していることを表し、恥ずかしい気持ち・不安定ともつながる可能性がある。

例⑹　(T たぶん)すごい作業効率落ちそう。（CEJC、T016_006）

表 3　CEJC における副詞「たぶん」のタグ付き書字形の頻度（上位 5 件まで）

副詞	タグ付き書字形（件数）	（%）	説明
たぶん	2143	74.2%	普通の代表的な出現
たぶん。	441	15.3%	発話単位末
たぶん:	96	3.3%	非語彙的な母音の引き延ばし
(U たぶん)	47	1.6%	聞き取りや語の判断が不確かな箇所
(U たぶん)。	21	0.7%	聞き取りや語の判断が不確かな箇所
(T たぶん)。	18	0.6%	小さい声で発話している箇所
(W タブ|たぶん)	18	0.6%	言い誤り・発音の怠け等の一時的な発音エラー
(L たぶん	11	0.4%	笑いが生じている箇所，あるいは単独の笑い
(L たぶん)。	11	0.4%	（タグ付きの組み合わせ）
たぶん:。	11	0.4%	（タグ付きの組み合わせ）
(W ダブン|たぶん)	5	0.2%	（タグ付きの組み合わせ）
(T たぶん	5	0.2%	（タグ付きの組み合わせ）

表 4　「たぶん」の前後 5 語以内に現れる副詞の出現（左：前脈、右：後脈）（上位 10 項目まで））

副詞(前脈)	件数	副詞(後脈)	件数
たぶん	**67**	そう	103
もう	51	**たぶん**	**67**
ま	50	もう	61
まあ	43	こう	43
そう	23	一番	24
こう	12	ちょっと	21
ちょっと	9	結構	13
全然	8	そ	11
もともと	7	まだ	11
まだ	6	全然	10

３． 日本語日常会話コーパスにおける推量副詞とモダリティ形式

３.１　推量副詞とモダリティ形式—高頻度の共起およびその特徴の考察—

従来の研究（スルダノヴィッチら 2009c）では、日本語書き言葉均衡コーパス（BCCWJ）およびウェブコーパスである JpWaC における推量副詞と文末モダリティを分析し、様々な文末モダリティのバリエーションを抽出した。本研究では、日本語日常会話コーパスにおける推量副詞とモダリティの呼応に焦点を当て、CEJC の発話から抽出した推量副詞とモダリティ形式を表 5 にまとめる。従来の研究結果と特に異なる点は、話し言葉特有のモダリティ形式である。例えば、書き言葉での「のだろう」は話し言葉では「んだろ、んだろう」と表れ、書き言葉の「かもしれません」は話し言葉で「かもしれない、かも、かもしんない」と出現する。

最も高頻度のモダリティタイプは推測（EXP）で、その代表的な形式には「と思う、と思うんだ、んじゃない、んだろう、でしょう、んだと思う、と思います、んでしょう、んじゃないか、だろうな、じゃない」などが含まれる。次に高頻度で見られるのは確信（NEC）のモダリティ形式で、「んだ、はず、わけ」などが該当する。その次が推定（CONJ）の「みたいな、みたい、気がする、らしい、みたいなの」などのモダリティで、最後に不確定（POSS）のモダリティ形式として「かもしれない、のかな、かもしんない、のかもしれない、のかもしんない」などが挙げられる。

書き言葉のコーパスに比べて、よく現れる形式のバリエーションが異なるが、モダリティタイプの頻出順は同様である（EXP→NEC→CONJ→POSS）。

表 5　CEJC における高頻度の副詞とモダリティ形式

モダリティ／副詞	おそらく	きっと	たぶん	どうも	もしかしたら	もしかすると	よほ(っぽ)ど	大抵	必ず	絶対	絶対に	合計
と思う	3	16	213	1			1		2	34		270
んだ		9	117	6	1		3	1	11	21	5	174
と思うんだ	2	5	123							14		144
んじゃない	2	6	81		2					6		97
んだろう		6	63				1		2	4		76
みたいな		2	26	10	2	1		1	6	14		62
でしょう	1	2	25					1	3	21		53
かもしれない			28		20	1	1					50
んだと思う			44							1		45
と思います	1	1	41							1		44
のかな	1	4	32		2					3	1	43
かもしんない		1	26		8	1						36
はず		2	12			1		3	3	8		29
みたい			16	2			1		1	4		24
んでしょう		6	15									21
わけ	1		4	1	1		2		1	8		18
んじゃないか			16	1	1							18
だろうな	1	2	9				1			5		18
気がする	1	1	7							6		15
のかもしれない			10		4	1						15
らしい			3	8						1		12
のかもしんないん			6		4							10
かもしれません		1	2		5							8
のか		2	4		1	1						8
じゃない		2	2		1					1	1	7
みたいなの			6							1		7
わけじゃない	1			1					1	1	2	6
かな		2	4									6
んだと思います			4									4
ってゆう			4									4
じゃないか			3									3
ことでしょう		1	1								1	3
らしいんだ			1	1						1		3
のかも			1		2							3
わけない			1							2		3
ようと思う			1							1		2
ような			2									2
ってゆうか			1						1			2
気がします			2									2

推測（EXP）
確信（NEC）
推定（CONJ）
不確定（POSS）

３.２　モダリティ形式のバリエーション

推量副詞と共起するモダリティ形式には多くのバリエーションがあることがコーパスのデータからわかった。そのバリエーションは、以下の異なる単位から成り立っている。

① ある形式の異なる発音
　例：　「だろ、だろう、だろう：」、　「かもしんない、かもしれない」

② 文体・丁寧さの程度
　例：　「んだと思う、のだと思います」、「んだろう、のでしょう」、「かもしれない、かもしれません」

③ モダリティ形式・意味・機能の積み重ね・連鎖（複数のモダリティ（態度や形態）が連続して使用され、意味や機能を構築する）。
　例：　「だろうと思うんじゃないか（だろう＋と思う＋んじゃない＋か）」

「だろうと思うんじゃないか」というモダリティ表現はよく「たぶん」という推量副詞と共起され、話し手が自分の推測や意見を述べた後、それに対する相手の反応や意見を確認するために利用される。会話相手に同意を求めることで相手に配慮し、コミュニケーションをスムーズに行う役割を持つ。

以下はコーパスからのモダリティ形式の積み重ねの例である。例(7)には推量副詞の「多分」と呼応しているモダリティ形式が「んじゃなかろうかと思うんですが:」。この表現は、話し手が自分の推測や意見を述べた後、それに対する確認や相手の反応を求めるニュアンスを持っている。

例(7)　去年は[たぶん](0.428)んー これが届いた日の(0.257)四時から(0.882)だった<u>んじゃなかろうかと思うんですが:</u>。（CEJC、T020_020 IC03）

３.３　広い文脈に現れるモダリティ形式

本節では、推量副詞が現れる会話を日本語日常会話コーパスから抽出し、ランダムサンプリングで各推量副詞の 100 出現およびその 300 語前後のコンテクストおよびアノテーションデータを分析する。このデータを用いることで、話し手の発話だけではなく、広い文脈におけるモダリティ形式やタイプの振る舞いを検討することができる。結果は表 6 に示す。

300 語以内の前文脈を分析した結果、4 種類のモダリティタイプおよび数多くのモダリティ形式が現れていることが明らかになった。それは、会話の中で、会話に参加している話し手と聞き手のコミュニケーションの相互作用の中で、感情・意図・態度を表しているためである。

副詞とモダリティ形式は代表的なモダリティタイプに分類されるが、それ以外のタイプや役割を持つこともある。特に、モダリティ形式「のだ」は複数のモダリティの意味を持ち、関連する副詞によって、モダリティの役割を強めたり弱めたりする働きを持つ。

対象となっている 4 種類のモダリティタイプのうち、推測モダリティタイプが最も高頻度である。その理由としては、推測モダリティが人間のコミュニケーションで頻繁に現れること、また、示しているデータが推測の意味を持つ推量副詞「たぶん」をキーワードにしているため、その代表的な形式が最も多く含まれていることが挙げられる。

表 6　「たぶん」の広い文脈に現れるモダリティ形式
（ランダム 100 例、前後文脈 300 語から取り出したモダリティ）

EXPACTATION（推測）	だろ	6	10	65	66	131	404
	だろう	20	23				
	でしょ	18	13				
	でしょう	21	20				
	のだろ	0	0	25	31	56	
	ことだろ	0	1				
	んだろ	2	7				
	んだろう	13	13				
	のでしょうか	1	0				
	んでしょう	5	7				
	んでしょうか	0	1				
	んでしょ	4	2				
	と思う	23	35	57	65	122	
	と思うんだ	9	7				
	と思うんです	8	4				
	んだろうなと思う	2	2				
	のだと思う	0	0				
	んだと思う	5	8				
	かなと思う	10	9				
	じゃないかな	4	3	52	43	95	
	じゃないか	6	5				
	じゃない	42	31				
	じゃなかった	0	3				
	のじゃない	0	1				
	のではないかと	0	0				
NECCESSITY（確信）	に違いない	0	0	18	8	26	134
	はず	2	1				
	わけ	16	6				
	べき	0	1				
	のだ	2	0	52	56	108	
	んだ	50	56				
CONJECTURE（推定）	よう	16	20	66	58	124	124
	のよう（だ）	1	1				
	らしい	3	2				
	みたいな感じ	7	3				
	みたいな#	22	17				
	みたい	17	15				
POSSIBILITY（不確定）	のか	7	9	22	32	54	123
	のかな	15	23				
	かもしれない	8	11	30	39	69	
	かもしれないか	0	0				
	かも	7	7				
	かもしんないか	0	1				
	かもしんない	8	11				
	んかもしれない	0	1				
	のかもしんない	1	2				
	のかもしれ	2	2				
	のかも	4	4				

４．おわりに

本稿では、日本語日常会話コーパス（CEJC）における推量副詞の分布および推量副詞とモダリティ形式の共起関係の振る舞いを検討した。22 の推量副詞の分布を複数のコーパスで調査した結果、推量副詞は書き言葉コーパスよりも話し言葉コーパスの方が出現頻度が高いことが分かった。最も高頻度の推量副詞は、話し言葉のコーパス（CEJC、CWPS、NUCC、SSC での「たぶん」、「ぜったい」、「どうも」、「きっと」）で主に表現されており、話し言葉における重要な役割を持つことと一致している。一方、一部の副詞は書き言葉のデータ（「おそらく」、「かならずしも」、「どうやら」、「もしかすると」などの SHC、CHJ、BCCWJ）やフォーマルな話し言葉の文脈（CSJ や SSC での「おそらく」）でより頻繁に現れる傾向がある。

推量副詞のコーパス例を分析した結果、話し言葉における副詞の特徴として、①述語後置、②重複、③他の副詞との出現、④発話のバリエーションが明らかになった。

モダリティの種類に関しては、「推測」と「確信」は最も広く表現されており、コミュニケーションにおいて「推測」と「確信」の方が「推定」と「不確定」よりも機能的に優先することを示唆している。この結果は、副詞の分布や副詞とモダリティ形式の代表的な共起からも確認された。

CEJC コーパスの分析結果では、副詞とそのモダリティ形式の組み合わせの強弱および様々なモダリティ形式のバリエーション（①異なる発話、②異なる文体・丁寧さ、③形式・意味・機能の積み重ね・連鎖）が示された。推量副詞が出現した発話およびその長い前後文脈におけるモダリティ形式を考察した結果、発話における高頻度の体系的なモダリティタイプとの呼応関係とともに、談話における重要な役割を果たす異なるモダリティ形式およびタイプの出現が統合性の面から明確になった。得られた研究結果は会話における副詞とモダリティの役割をさらに明確にし、日本語教育、談話分析、語用論などの分野に応用できる。

謝　辞

本研究は、国際交流基金による日本研究フェローシッププログラム助成金（番号：10148587）の支援を受けています。また、本研究成果は、国立国語研究所の機関拠点型基幹研究プロジェクト「学習者辞書用語彙資源の構築」（リーダー：柏野和佳子）、JSPS 科研費 23H00072（リーダー：松下達彦）、および、ユライ・ドブリラ大学プーラ人文学部アジア研究科の機関研究プロジェクト「日本語と社会：基本概念と実証分析（Japanski jezik i društvo: temeljni pojmovi i empirijska analiza）」（リーダー：Irena Srdanović）によるものです。

文　献

工藤 浩 (2000). 副詞と文の陳述のタイプ『日本語の文法３モダリティ』（森山卓郎，仁田義雄，工藤浩）岩波書店, pp. 161–234.

小磯花絵・天谷晴香・石本祐一・居關友里子・臼田泰如・柏野和佳子・川端良子・田中弥生・伝康晴・西川賢哉・渡邊友香 (2022).『『日本語日常会話コーパス』設計・構築・特徴』（国語研究所「日常会話コーパス」プロジェクト報告書 6）

スルダノヴィッチ・イレーナ、ベケシュ・アンドレイ、 仁科喜久子(2009a)「BCCWJ における推量副詞とモダリティ形式の共起」「日本語コーパス」平成 20 年度公開ワークショップ（研究成果発表会）予稿集, 237-244.

スルダノヴィッチ・イレーナ、ベケシュ・アンドレイ、 仁科喜久子(2009b).「コーパスに基づいた語彙シラバス作成に向けて—推量的副詞と文末モダリティの共起を中心にして—」『日本語教育』142 号, pp. 69-79.
スルダノヴィッチ・イレーナ、ホドシュチェック・ボル、ベケシュ・アンドレイ、仁科喜久子(2009c). 「ウェブコーパスと検索システムを利用した推量副詞とモダリティ形式の遠隔共起抽出と日本語教育への応用」『自然言語処理』16 巻 4 号, pp. 29-46.
南不二男 (1974).『現代日本語の構造』大修館書店
ホドシチェク ボル, 仁科 喜久子, ベケシュ アンドレイ(2010).「レジスターに基づく日本語のモダリティ形式の分類」『特定領域研究「日本語コーパス」平成 21 年度公開ワークショップ予稿集』 251-256.
Bekeš, Andrej (2006). “Japanese suppositional adverbs in speaker-hearer interaction”. *Proceedings of the third conference on Japanese language and Japanese language teaching*, Rome 2005. Venezia: Libreria editrice cafoscarina. pp. 34–48.
Den, Yasuharu, Nakamura, Junpei, Ogiso, Toshinobu, & Ogura, Hideki (2008). “A proper approach to Japanese morphological analysis: Dictionary, model, and evaluation”. *Proceedings of the Sixth International Conference on Language Resources and Evaluation* 26(1-2), 129-148.
Koiso, Hanae, Haruka Amatani, Yasuharu Den, Yuriko Iseki, Yuichi Ishimoto, Wakako Kashino, Yoshiko Kawabata, Ken'ya Nishikawa, Yayoi Tanaka, Yuka Watanabe, and Yasuyuki Usuda (2022). “Design and Evaluation of the Corpus of Everyday Japanese Conversation”, *Proceedings of LREC2022*, pp.5587-5594.
Oka, Teruaki, Ishimoto, Yuichi, Yagi, Yutaka, Nakamura, Takenori, Asahara, Masayuki, Maekawa, Kikuo, Ogiso, Toshinobu, Koiso, Hanae, Sakoda, Kumiko & Kibe, Nobuko (2020). “KOTONOHA: A Corpus Concordance System for Skewer-Searching NINJAL Corpora”. *Proceedings of the Twelfth Language Resources and Evaluation Conference*, 7077–7083.
Srdanović, Irena, Bekeš, Andrej, Nishina, Kikuko (2008). “Distant Collocations between Suppositional Adverbs and Clause-Final Modality Forms in Japanese Language Corpora”. *LKR 2008, LNAI 4938* (Eds.T. Tokunaga and A. Ortega). Springer-Verlag Berlin Heidelberg. 252–266.

関連 URL

- コーパス検索アプリケーション『中納言』
  https://chunagon.ninjal.ac.jp/
- コーパス一覧 (NINJAL)
  https://clrd.ninjal.ac.jp/index.html
- まとめて検索『KOTONOHA』
  https://chunagon.ninjal.ac.jp/integrated/
- 『日本語日常会話コーパス（CEJC）』
  https://www2.ninjal.ac.jp/conversation/cejc.html