琉球大学学術リポジトリ

コーパスに基づく引用句内のコピュラ（「だ」）の顕在と潜在に関する研究

メタデータ	言語: Japanese 出版者: 琉球大学留学生センター 公開日: 2012-04-19 キーワード (Ja): 「だ」, 潜在, コーパス分析, 自発表現, 発話の力 キーワード (En): da, copula, corpus analysis, pragmatic motivation 作成者: 金城, 克哉, Kinjo, Katsuya メールアドレス: 所属:
URL	http://hdl.handle.net/20.500.12000/24137

研究論文（査読あり）

留学生教育 第９号，2012
（琉球大学留学生センター紀要）

コーパスに基づく引用句内のコピュラ（「だ」）の顕在と潜在に関する研究

金 城 克 哉

要 旨

本稿では引用の「と」が導く引用句内でのコピュラ「だ」の顕在と潜在について「Yahoo!知恵袋」のデータに対し以下の観点から調査・考察をおこなった：(1)「と思う」と「と考える」というコンテクストでの名詞・形容動詞・形式名詞他における「だ」の潜在率，(2)「と思う」と「と考える」という2つのコンテクストの比較における名詞と形容動詞の潜在率に関する相関関係，(3) 潜在例と自発表現の共起，(4) 潜在の語用論的解釈。結果，「と考える」での潜在率が「と思う」よりも高いこと，形容動詞では相関が見られること，潜在と自発表現の共起率が高いこと，発話の力の軽減のための方策として「だ」が省略されている可能性があることが明らかとなった。

【キーワード】：「だ」，潜在，コーパス分析，自発表現，発話の力

0. はじめに

引用の「と」が導く引用句の中でのコピュラ（助動詞）の「だ」は次のデータ例が示すように現れる場合(a)と現れない場合(b)がある：

a. こういうのって，二者択一には出来ない問題だと思いますよ。
b. 日本と比べると，タクシー料金が安いので，少し多めにチップを渡すこともあります。チップという習慣に対する慣れの問題と思います。

(Yahoo!知恵袋)

このような引用句の中でのコピュラの顕在と潜在については先行研究が限られており(田野村2006 ，田野村2008)，現時点でのコピュラの潜在の状況が十分に明らかになっているとは言えない。またこの問題は日本語教育ともかかわりを持つが，現時点ではコピュラが現れない可能性があるということにまで触れた日本語テキストは管見ではほとんどない。本稿では大規模コーパスを用いて分析を行い，「と思う」・「と考える」を後続のコンテクストとする引用句の中でのコピュラの顕在と潜在ついて考察を加える。

1. 先行研究と本稿のリサーチ・クエスチョン

引用（句）・話法については様々な先行研究がある。従来の議論（引用・話法・「と」の扱い・引用表現の習得等）に関しては鎌田（2000）や杉浦（2002），また藤田の一連の研究（2010他）に詳しい。本稿では引用句内の「だ」の顕在・潜在に焦点を絞る。まず，本稿で取り上げる「だ」についてみてみよう。

野田（2001）では「だ」は様々な成分を述語にすることがその主たる働きであるとされ，「『だ』が現れるのは，主に，述語のない文が嫌われる書きことばである」（2001：51）と指摘されている。三枝（2001）では「だ」全般について考察を加えており，特に引用句の「だ」については野田（同上）の言う述語性に関し，「引用句の中にあることで述語性はあるが，概念化している」とし，モダリティ性を欠いているために「です」には置き換わらないとする。さらに三枝（同上）は上記(a)，(b)のような「だ」の顕在と潜在に関し次のように述べている：

> 「引用句」の場合は，言い切りの場合にもそうであるように，そもそもこの述語の「だ」はなくてもいいものなので，省くことができる。ただ，書き言葉では省略しないことも多い。（三枝2001：16）

メイナード（2005）は談話という視点から「だ体」と「です・ます体」を比較し，特に書き言葉で「だ体」を基調とするスタイルでは，「書き手が現象文で直接描写したり，従属節的な表現に用いられることが多い」とする。また「だ体」で助詞などのつかない「裸のだ体」については，「感嘆表現，内的疑問文，乱暴な言い回しなどに使われることが多く，たとえば『何だ！』や『勝手にしろ！』などの乱暴な言いまわしがある」と指摘されている（さらに詳しい議論についてはMaynard（1991），メイナード（1997）も参照のこと）。

これらの研究は「だ」についての論考ではあるが，必ずしも引用句における省略（もしくは潜在化）を扱ったものではなかった。引用句内の「だ」の潜在化とコーパスを用いた経年変化を分析したものに田野村（2006，2008）がある。田野村（2006）では「あれはタヌキか？」という文を取り上げ，この文における「タヌキ」は「意味的に『タヌキである』というコピュラ述語と等価」であり，「『タヌキ』の後ろには無形の―すなわち，音形を持たない―コピュラが潜在する」と捉える。またこの「タヌキか？」の部分には「何かを端折った」という印象もなく硬い文章でも用いられる

ため，いわゆる口語的な省略（例：「お前（は）タヌキ（を）見た？」）とは一線を画す性質のものであるとされ，「省略」ではなく「潜在化・顕在化」という用語が用いられている。

さらに、田野村（2008）では国会会議録（1940年代～2000年代）のデータを用いて引用句におけるコピュラの顕在化傾向の経年変化を調査し，「引用句におけるコピュラの顕在率が全体に徐々に高まっており，コピュラの省略化が進行しているわけではない」と結論づけている。また，「～べきφとの意見」のような表現にみられるような「～べき｛だ／である／φ｝と～」における顕在率を4種類の後続文脈「（～と）思う」，「（～と）いう～」，「（～と）考える」，「（～と）の～」の場合について調査し，「（～と）の～」と「（～と）考える」の場合に「べき」に続くコピュラが省略される傾向が強まっているとしている（田野村2008：63）。

上述の先行研究を踏まえ，本稿のリサーチ・クエスチョンを次の通りとする。(1)「と思う」と「と考える」というコンテクストでの名詞・形容動詞・形式名詞他における「だ」の潜在率はどの程度なのか（最も潜在率が高い品詞の特定，品詞同士の潜在率の比較，高頻度語の特定），仮に「任意」に起こる現象であるならば必要十分な用例調査（大規模コーパスにおける出現率調査）において顕在率と潜在率は同程度もしくはそれほどの差はみられない結果となると考えられるため，一定の偏りが見られるならばこの現象は一概に「任意」であるとは言えない，(2)「と思う」と「と考える」という2つのコンテクストの比較で名詞と形容動詞での潜在に相関はあるか，(3)潜在のコロケーションと自発表現には関わりがあるのか，(4)「だ」の潜在が任意のものであり文法の適格性に関与しないとすれば，この潜在現象は文法以外の要因に左右される（例えば語用論的に説明可能な側面がある）のではないか。

2. データと分析方法

本稿では，国立情報学研究所がヤフー株式会社との契約に基づき，研究者に対して提供している「Yahoo! 知恵袋データ（第2版）」（質問数：約1,600万，回答数：約5,000万，期間：2004年4月～2009年4月）を用い，この中から，無作為に選んだans_4.tsvファイル（約6,200MB，総語数508,785,918語）のデータを分析した。解析には立命館大学の樋口耕一氏によって開発されたKH Coderを使用した（形態素解析器は茶筌）。今回は田野村にならい，「～｛だ／である／φ｝と思う」および「～｛だ／である／φ｝と考える」という6通りの言い回しの用例中に占める潜在率を求めた。

田野村では名詞と「べき」の調査がなされていたが，本稿では名詞，形容動詞，（「べき」などをふくむ）形式名詞の3つのカテゴリーに分けてそれぞれの差を比較する。その一方で，「と思われる」・「と考えられる」のような自発表現と潜在例の共起についても調査を行った。

3. 結果

調査の結果「～｛だ／である／φ｝と思う」の総データ例数は540,712，そのうち「～｛だ／である｝と思う」のコピュラ顕在データ例数は487,711，「～φと思う」のコピュラ潜在データ例数は53,001，潜在率は9.80%であった。一方，「～｛だ／である／φ｝と考える」の総データ例数は18,418，そのうち「～｛だ／である｝と考える」のコピュラ顕在データ例数は6,945，「～φと考える」のコピュラ潜在データ例数は11,473，潜在率は62.29%であった。また，「と思う」と「と考える」双方の潜在データ例について，どのような語で潜在率が高いかを確認するために品詞別に頻度を調査し，表2と表3に示すような結果が得られた（名詞と形容動詞は上位10，形式名詞その他は5位まで）。次にそれぞれのデータ例を示す：

表1　「＿と思う」・「＿と考える」でのコピュラの潜在率

	と思う	と考える
顕在例	487,711	6,945
潜在例	53,001	11,473
計	540,712	18,418
潜在率	9.80%	62.29%

表2　「φと思う」の潜在率に関する統計

	名詞	度数	形容動詞	度数	形容名詞	度数
1	原因	374	大丈夫	1,503	もの	5,263
2	友達	304	必要	1,313	こと	2,480
3	人間	150	当たり前	582	べき	1,349
4	状態	106	可能	566	から	710
5	方法	95	無理	558	だけ	684
6	一つ	86	十分	326		
7	チャンス	82	ラッキー	268		
8	常識	78	普通	244		
9	感じ	64	不要	216		
10	女性	57	幸せ	205		

表3　「φと考える」の潜在率に関する統計

	名詞	度数	形容動詞	度数	形容名詞	度数
1	別物	199	必要	203	もの	1,862
2	原因	192	無理	91	同じ	484
3	一つ	133	当たり前	73	べき	286
4	手段	74	不要	64	物	191
5	目安	54	可能	64	こと	152
6	一種	54	重要	63		
7	寿命	47	不可能	57		
8	状態	43	妥当	41		
9	程度	35	同等	40		
10	基本	28	大丈夫	34		

(1) 呼吸が安定していれば大丈夫だとおもいます。原因と思われる薬の名前をしっかり覚えておいて，今後処方されないように毎回病院に行くたびに話しをする必要があります。

(2) 私はかれこれ40年前に，黄色のバイエルを卒業しましたがなんとか弾けましたので，きっとあなたなら大丈夫と思いますよ

(3) もし貸したとしたら，１度は催促します。それでダメなら諦めます。貸すときは，返ってこないものと思って貸してやってください。

(4) あなたの離婚とその方の離婚は別物と考えてはいかがでしょうか？？親友ならば，そういうことで友人関係はこわれないはずですよ！

(5) 足が不自由な人が車から降りて車椅子に乗り換えるのに，あれくらいのスペースが必要と考えて作ってあるのに台無しです。

(6) 車は事故と切り離せない関係です。自分のお金で自己責任において維持していくものと考えています。

「と思う」では潜在率が10%を下回ったものの，「と考える」では潜在率が60%を上回った。双方のケースに共通しているのは形式名詞「もの」の場合に潜在率が高いこ

とである。その一方，潜在率はそれほど高くないものの，「と思う」では名詞よりも形容動詞で潜在率が高くなっている。しかしながら，「と考える」では「別物」・「原因」・「一つ」という特定の名詞で潜在率が高くなっており，形容動詞でも「必要」という特定の語での潜在率が突出している。また同じ形容動詞での比較では，「と思う」で潜在率が1位であった「大丈夫」が「と考える」では第10位となっている点が対照的である。

次に名詞頻度と形容動詞頻度における「φと思う」と「φと考える」の相関がどの程度か調べるために，表4の度数表に対してSeagull-Statの単回帰分析マクロの付属機能を利用して相関分析を行った（注1)。「と思う」と「と考える」という異なるコンテクストにおいて高頻度語にどの程度の関係性が存在するのかを調べた。仮にいずれのコンテクストであっても，高頻度語の使用パターンが安定しているのであれば，コンテクスト間に高い相関が得られる。一方，コンテクストによって使用頻度が大きく変化するのであれば，相関係数は低くなるはずである。結果，名詞では相関が見られなかったものの，形容動詞では5%水準で高い相関が確認された（表5参照)。

表4　名詞・形容動詞度数表

名　詞			形容動詞		
語	φと思う	φと考える	語	φと思う	φと考える
原因	374	192	大丈夫	1,503	34
友達	304	11	必要	1,313	205
人間	150	11	当たり前	582	73
状態	106	43	可能	566	64
方法	95	6	無理	558	91
一つ	86	133	十分	326	25
チャンス	82	14	ラッキー	268	23
常識	78	19	普通	244	16
感じ	64	3	不要	216	64
女性	57	2	幸せ	205	11
親友	53	4	妥当	197	41
程度	47	35	大事	190	10

表5　単相関行列表

＊：5%有意(>0.576)

	名　詞		形容動詞	
	φと思う	φと考える	φと思う	φと考える
φと思う	1	0.564	1	0.588＊
φと考える	0.564	1	0.588＊	1

表6　自発形「思われる」と潜在率

顕　在		潜　在	
ものだと思う	10,854	ものと思う	5,263
ものだと思われる	275	ものと思われる	2,377
率	2.53%	率	45.16%

表7　自発形「考えられる」と潜在率

顕　在		潜　在	
ものだと考える	421	ものと考える	1,824
ものだと考えられる	66	ものと考えられる	614
率	15.68%	率	33.66%

また，高頻度で潜在が見られる形式名詞「もの」に焦点を絞り，動詞の形に着目して調査したところ，潜在データ例の場合に「思われる」という自発形が用いられる割合は約45%であり，顕在データ例に比べて約17倍であることがわかった。同様の現象は「考えられる」という自発形でも起こっている（約2倍）（表6・表7参照）。(注2)

4. 考察

4−1. 潜在の恣意性

「と思う」よりも「と考える」というコンテクストで潜在率が高くなっていることが分かったが，このように潜在性に一定の傾向があるという事実が指し示しているのは，三枝が指摘していたようにこの述語の「だ」は（あっても）なくてもいいもの，言い換えれば恣意的に省略されるものとは言えないということである。また，品詞別の潜在傾向の強さは概ね次のように捉えられる：

(7)　「と思う」　：名詞 ＜ 形容動詞 ＜ 形式名詞（もの・こと）
　　「と考える」：形容動詞　＜ 名詞　＜ 形式名詞（もの・同じ）

田野村（2008）は名詞のケースについて国会会議録の資料をもとに経年変化を調査し，それほど潜在化が進んでいるわけではないという結論を導いていたが，本稿の調査では，経年変化こそ調べることはできなかったが，共時的な傾向として，「と思う」では名詞よりも形容動詞で潜在が観察されること，さらに形容動詞よりも形式名詞「もの」で多く見られることが示された。一方「と考える」では「別物」・「原因」・

「一つ」という特定の名詞での潜在率が高いことが確認された。これらはそれぞれ弁別・因果・カテゴリー化という思考活動の結果述語として選択されたものであると言えるだろう。それ以外の名詞ではほぼ形容動詞と同じような割合であることが明らかとなった。また形式名詞では「と思う」の場合同様，「もの」での潜在率が他を大きく上回ることが分かった。

潜在例において，引用符「と」の直前の名詞では相関が見られず，形容動詞で高い相関が見られたことから，「と思う」の環境で潜在が起こる名詞には一定の傾向があるが，それはそのまま「と考える」というコンテクストにはあてはまらないということがわかった。一方，形容動詞において高い相関が見られたことから，「と思う」「と考える」双方のコンテクストにおいて潜在（もしくは顕在）例に似通った点があるということがわかった。

4－2. 潜在の誘因

「と思う」よりも「と考える」というコンテクストで，さらに名詞／形容動詞よりも形式名詞の「もの」というように潜在率が高くなっているのはなぜなのであろうか。結論を先取りして言えば，それは引用句内部の命題が真偽に関わるものだからである。すなわち，「だ」や「である」を用いて述語を形成するということは文に「XはYだ」もしくは「XはYである」という構造を与え，形式の整った命題（主語述語がきちんとそろった1つの命題）を完成させることになる。見方を変えれば，「だ」や「である」を用いないという選択をすることにより，話し手（書き手）は命題を不完全な形式，述語として提示されるはずの部分が述語の体を成していないまま叙述を行うということになる。

この「命題としての形式を整えないこと」（＝潜在）の理由として，本稿では語用論的な配慮が働いたということがあげられると考える。高橋（2003）は「考える」の基本的意味を「＜（結論を導くという）目的意識を持って＞＜知力を働かせる＞」としているが，そうであるならば引用の「と」に導かれる引用句の内容はその「知力の働き」を用いて導かれた命題を聞き手や読み手に開示するものとなる。書き手や話し手はillocutionary force（発話の力）を軽減させるための方策（hedge）の一つとして「形式を整えない」という選択をしたのではないだろうか。このことを以下の作例を通して考えてみよう：

(8) a.　これらの状況証拠からAが犯人である と考える。
　　b.　これらの状況証拠からAが犯人だ と考える。
　　c.　これらの状況証拠からAが犯人φ と考える。

(8a)「である」体は論文などフォーマリティの高いスタイル，(8b)「だ」はフォーマリティの高いスタイルでも一般的なスタイルいずれでも用いられるが，そのような文体差を一旦保留にして，「断定性」という点からみると，「である」と「だ」が用いられる引用句は命題として1つのまとまりを持ち，話し手（書き手）の断定判断が強く表れているのに対し，潜在例ではそのような断定性が表わされていないという対照的な構図が見えてくる。知力を働かせて導いた結果の表明としての引用句において「である」もしくは「だ」を用いるということは話し手（書き手）が自ら判断し断定したこと（＝命題）に責任を持つということになろう。このことは裏を返せば「だ・である」の潜在は書き手や話し手が自らのfaceを保つための方策であると言える。

4－3. 自発形と潜在の結びつき

「だ・である」の潜在が発話の力の軽減のための方策であるという上記議論の一つの傍証として，自発形と潜在の関係を見てみよう。浅野（1996）は文形式の選択において日本語では「他者の反対意見や他の可能性が低い場合でも，十分な証拠のないことは言わないこと」というルールが働いているとし，「と思われる」には「発言を緩和する働きや，客観性を暗示したり，責任を回避したりする働きがある」とする（1996：45）。前節の命題完成の議論を継承すれば，表6で示したように「～ものと思われる」というように「もの」の直後に潜在が起こるのは命題完成を避けることによって発話の力を軽減させようとする書き手の意図が働いていると見ることができるのではないだろうか。これは顕在例である「ものだと思われる」のデータ数が潜在化の約17分の1にとどまっていることからも裏付けられる。「だ」を顕在させることが命題の完成を意味し，それによって書き手（話し手）は命題に対する責任を負うことになる（ひいてはそれが自身のfaceを脅かすこととなる）。命題完成と浅野の指摘する「責任回避」の表現である「と思われる」は相反する働きをするために共に用いにくい（または違和感のある表現となる）ということになる（自発形「思われる」については安達(1995)も参照のこと）。

5. おわりに

本稿では（1）「と思う」と「と考える」というコンテクストでの名詞・形容動詞・形式名詞他における「だ」の顕在率および潜在率を調査し，「だ」の潜在が「任意」に起こる現象ではないことを明らかにした。また，（2）「と思う」と「と考える」という2つのコンテクストの比較で名詞と形容動詞での潜在の相関を調べ，名詞では相関が見られなかったが形容動詞では高い相関が見られることが分かった。（3）「思われる」や「考えられる」という自発表現の際に潜在率が高くなることから，（4）「だ」の潜在は発話の力の軽減という語用論的な誘因によって導かれるものであることを論証してきた。

本論で取り上げることはできなかったが，「と」によって導かれる引用句内の「だ」の潜在が必須となるケースもある。次の例を見てみよう：

(9)　a.　拝啓 時下ますます御健勝のことと存じます。

時候の挨拶などで用いられる「ことと存じます」はコロケーションとして一般化しており，「だ」を顕在させて「ことだと存じます」とすることはできない。また，「と」の引用句を独立させ，「だ」を顕在させた次の文は非文である：

(9)　b.　＊時下ますます御健勝のことだ

このような例では，本論で議論したように形式名詞「こと」のあとに潜在が起こっているとするよりも，「の」以下をコロケーションとして考えなければならないと考える。（注3）また，本論で触れた「ものと思われる」という自発表現での潜在率の高さはこのようなコロケーション化の方向へつながっていると見ることもできるかもしれない。将来的にはこのような点を含めた包括的な議論が必要となろう。

註

(1) 本稿で用いた「多変量解析システムSeagull-Stat」は，早狩進氏（元青森県環境保健センター，現NPO法人グリーンシティ常務理事）の開発によるマクロ・プログラムである（石川他2010）。

(2) 加藤（1997）によると「ものだ」には（1）一般的傾向，（2）当為，（3）希望，（4）回想，（5）驚きなどの意味があるが，「思う」にはこれら全てが引用句内に用いられるものの，「希望」と「驚き」の意味では「ものだと考える」では用いられないとされる。

(3) この点については高橋（2009）の「と思う」を文法化の進んだ形式として捉えるという議論も参考となる。

参考文献

浅野裕子（1996）「『思われる』にみる日英の語用論的原則」『日本語教育』88，35-47

安達太郎（1995）「思エルと思ワレル」宮島達夫・仁田義雄編『日本語類義表現の文法（上）単文編』くろしお出版，121－130

石川慎一郎・前田忠彦・山崎誠編（2010）『言語研究のための統計入門』くろしお出版

加藤理恵（1997）「『と』節を含む文について」『名古屋大学人文科学研究』26，115-127

鎌田修（2000）『日本語の引用』ひつじ書房

三枝令子（2001）「『だ』が使われるとき」『一橋大学留学生センター紀要』4，3-17

杉浦まそみ子（2002）「日本語の引用表現研究の概観 ： 習得研究にむけて(第１章 文法形式と機能の習得と使用)」『言語文化と日本語教育』増刊特集号，120-135

高橋圭介（2003）「引用節を伴う「思う」と「考える」の意味」『言葉と文化』4，99-114

高橋圭介（2009）「『思う』の多義構造再考--文法化の進んだ『と思う』の位置付けをめぐって」福島工業高等専門学校『研究紀要』50，167-174

田野村忠温（2006）「コピュラ再考」藤田保幸・山崎誠編『複合辞研究の現在』和泉書院，249-270

田野村忠温（2008）「大規模な電子資料に見る現代日本語の動態」『待兼山論叢』42，55-76

野田尚史（2001）「うなぎ文という幻想—省略と『だ』の新しい研究を目指して」『国文学 解釈と教材の研究』 46(7)，122-128

藤田保幸（2010）「引用研究の『今』をめぐって」『龍谷大学国際センター研究年報』19，33-47

メイナード，泉子・K（1997）『談話分析の可能性：理論・方法・日本語の表現性』くろしお出版

メイナード，泉子・K（2005）『談話表現ハンドブック』くろしお出版

Brown, P. & S, C. Levinson (1987) *Politeness: Some Universals in Language Usage*. Cambridge:Cambridge University Press.

Maynard, Senko K. (1991) Pragmatics of discourse modality: A case of *da* and *desu/masu* forms in Japanese. *Journal of Pragmatics,* 15, 551-582

（琉球大学法文学部）

A Corpus Based Study of Latent Copula *da* in Quoted Clauses

KINJO Katsuya

Keywords: *da*, copula, corpus analysis, pragmatic motivation

Abstract

It seems that not much attention has been given to latent copula *da* (or, the deletion of the copula) in literature. Couple of previous studies mentioned that such a copula can be arbitrarily deleted. This paper's research questions are (1) to clarify whether the deletion (or latency) of *da* occurs arbitrarily, (2) in comparison of two contexts, *to omou* and *to kangaeru*, whether or not latency of *da* following nouns and adjectives correlate, (3) the frequency of collocation with spontaneous forms (*omowareru* and *kangaerareru*), (4) if *da* can be deleted arbitrarily and such latency is not related to grammaticality, it might be possible to give explanation from pragmatic point of view. This study of a large corpus "Yahoo! Chiebukuro (second edition)" offered by National Institute of Information (NII) shows that (1) latent copula phenomenon does not occur arbitrarily, (2) latent copula can be found more in *to kangaeru* than *to omou*, (3) spontaneous forms are strongly related to latency of *da* and (4) mitigation of FTA (to take responsibility in what is said in the quoted clause) is one possible motivation for latent copula phenomena.

(University of the Ryukyus)