

話し言葉に現れる 2 種類のカラ節の判別

丸山 岳彦[†]伝 康晴[‡]小磯 花絵[†][†] 国立国語研究所研究開発部門 [‡] 千葉大学文学部

1 はじめに

音声コミュニケーションにおける統語解析とはいかなるものであろうか。発話の結果を統語的に分析するという見方に立てば、発話を構成する各要素は、発話が完了した時点で、文法的に適格な統語構造の中に位置付けられることになる。しかしながら、発話の産出と理解は実時間に沿った漸進的な営みであり、発話が終了した時点で統語解析が始まるわけではない。ここで、音声コミュニケーションの過程において、発話を構成する要素の統語的な位置付けが、どのような手がかりによってリアルタイムに認識されているのかという疑問が生じる。

本研究では、日本語の接続助詞「から」で導かれる節(カラ節)の多義性に着目し、この問題に迫りたい。次節で詳述するように、カラ節が表す意味には 2 種類あり、それぞれ統語的な位置が異なる。音声コミュニケーションにおいて、我々が発話の完了を待たずにこの違いをある程度把握していると仮定するならば、カラ節の出現時点までに現れる統語的・韻律的要素がこの判別に関わっているであろう。そこで本研究では、『日本語話し言葉コーパス』を対象に、カラ節が出現した時点および出現直後までに得られる種々の統語・韻律素性を用いて、2 種類のカラ節を判別するためのモデルの構築を試みる。この試行を通し、カラ節の漸進的判別の可能性を探ると同時に、判別にどのような要素に関わるかを検討する。

2 2 種類のカラ節

田窪(1987)によると、カラ節が表す意味には「事態の原因・行動の理由」「判断の根拠」という 2 種類があるとされる。次の (1) は「行動の理由」を表すカラ節、(2) は「判断の根拠」を表すカラ節の例である。

(1) 彼が来たから、彼女は帰ったのだろう。

(2) 彼が来たから、彼女は帰っただろう。

(1) と (2) のカラ節は、日本語文の階層構造を論じた南(1974)による分類のうち、「B 類の従属節」「C 類の従

属節」にそれぞれ対応する。以降では、(1) のカラ節を「カラ B」、(2) のカラ節を「カラ C」と呼ぶ。

カラ節の統語的な位置付けを考えると、カラ B は主節で表される「事態」を修飾する従属節であるのに対し、カラ C は主節で表される「判断」を修飾する従属節であり、統語的に異なる位置を占める。スコープの違いを示せば、以下ようになる。

(3) [[彼が来たから 彼女は帰った] のだろう]

(4) [[彼が来たから] 彼女は帰っただろう]

統語的な位置の違いに応じて、カラ C の内部には「だろう」のようなモダリティ要素が生じ得るがカラ B の内部には生じ得ないなど、内部に取り得る文法カテゴリの範囲に違いがある。また、カラ B が文脈上焦点化することができるのに対し、カラ C は焦点化できないなど、情報構造上の違いもある。

3 方法

3.1 データ

『日本語話し言葉コーパス (CSJ)』の「コア」に含まれる模擬講演を分析対象とした。模擬講演は、一般話者による主に個人的な内容に関するスピーチであり、計 107 講演、話者 82 人、約 20 時間分のデータである。

3.2 カラ節の分類

CSJ に付与されている節単位情報(国立国語研究所, 2006) から、節境界ラベルに「理由節カラ」と記された 374 例を抽出し、当該節単位および後続する節単位を見て、カラ B とカラ C の別を人手で分類した。この際、「～からです」のようにカラ節が述語句になっている場合、カラ節の係り先がないと判断された場合は対象外とした。その結果、カラ B が 173 例、カラ C が 162 例得られた。カラ B とカラ C の例を、(5) と (6) に示す。

(5) その先生のことは (F まー) あまり思い出したくもない から 思い出せないんでしょうけど (S02M0161)(6) その点犬は撫でられると尻尾振って喜んじゃいます から 全くかわいいもんです (S00F0031)

3.3 説明変数の選定

カラ節が出現した時点および出現直後までの情報を用いて、2 種類のカラ節の判別を行なうために、図 1 の統語・韻律素性を用いた*1。値の数が明記されているものはカテゴリ変数、それ以外はすべて連続変数である。

カテゴリの分類は、分布を考慮して設定した。カテゴリ変数のうち、「から」の直前の短単位の (5b) Break Index と (5c) 句末音調は一方のカテゴリ値しか取らなかったため、分析に使用しなかった。連続変数のうち、いくつかのものは互いに強い相関が見られたため、以下のように対処した。

■節全体の文節数・アクセント句数・短単位数 互いに .9 以上の強い相関があった。主成分分析の結果、第 1 主成分の寄与率が 95% 以上に達しており、かつ、第 1 主成分に対する各変数の負荷量がほぼ等しいことから、「文節数」のみを分析に用いた ((1a) と (1b) は使用せず)。

■節全体・文節・アクセント句・「から」自身の平均モーラ長 文節とアクセント句の平均モーラ長に .9 以上の強い相関があり、これらと「から」自身の平均モーラ長との間にも .7 以上のやや強い相関が見られた。主成分分析の結果、第 1・第 2 主成分までの累積寄与率がそれぞれ 73.2%, 90.5% であることから、第 1・第 2 主成分得点を分析に用いた ((1d), (2b), (3b), (4a) の代わりに (7a) 第 1・(7b) 第 2 主成分得点)。第 1 主成分は、どの変数も同符号で同程度の負荷量を持つことから全体的な発話の遅さ、第 2 主成分は、文節と「から」自身の平均モーラ長が逆符号で比較的大きな負荷量を持つことから節全体に対する節末部分の遅さを表すと解釈できた。

以上の処理の結果、全 16 個の素性を説明変数として用いた。

3.4 解析手法

基準変数がカラ B/C の 2 値であることから、ロジスティック回帰分析が適用できる。ただし、82 人の話者から 335 件の事例を抽出しており、事例は互いに独立でない (1 話者あたりの事例数は最小値 1・中央値 3・平均値 4.1・最大値 15)。そこで、話者によるクラスターを考慮した一般化線形混合効果モデルを用いた。分析には、R の lme4 パッケージの lmer を使い、二項分布とロジットリンクを指定した。

*1 これらの素性は、CSJ のコアダータに対して人手で付与された種々の研究用情報から抽出した。国立国語研究所 (2006) 参照。

1. 「から」を含む節全体に関する素性
 - (a) 文節の数
 - (b) アクセント句の数
 - (c) 短単位の数
 - (d) 平均モーラ長^a
 - (e) 「は」でマークされた文節 (以下の 4 値)
 - 無: なし
 - 係: 「から」文節に係る
 - 前: 「から」以前の文節に係る
 - 後: 「から」以後の文節に係る
2. 「から」を含む文節に関する素性
 - (a) 短単位の数
 - (b) 平均モーラ長
 - (c) この文節に係る文節の数
 - (d) 丁寧体か否か (0/1 の 2 値)
3. 「から」を含むアクセント句に関する素性
 - (a) 短単位の数
 - (b) 平均モーラ長
 - (c) Break Index^b (以下の 2 値)
 - 弱: BI=1, 2 および語断片
 - 強: BI=3 および休止・句末音調を伴う BI=2
 - (d) 句末音調 (L%, H%, HL% の 3 値)^c
4. 「から」自身 (と後続助詞) に関する素性
 - (a) 平均モーラ長
 - (b) 後続助詞の有無 (0/1 の 2 値)
5. 「から」の直前の短単位に関する素性
 - (a) 平均モーラ長
 - (b) Break Index (上記の 2 値)
 - (c) 句末音調 (上記の 3 値)
6. 「から」節の直後の短単位に関する素性
 - (a) 平均モーラ長
 - (b) 「から」節との間の休止の有無 (0/1 の 2 値)
 - (c) フィラーか否か (0/1 の 2 値)
 - (d) 品詞 (以下の 6 値)^d
 - 体: 名詞・代名詞・記号
 - 用: 動詞・形容詞・形状詞
 - 相: 連体詞・副詞
 - 感: 感動詞 (フィラーを含む)
 - 接: 接続詞 (「でも」「だから」などを含む)
 - 淀: 言い淀み

^a モーラ (概ね個々の仮名に相当) の平均継続時間長

^b 韻律境界の切れ目の強さに関する情報。基本は BI=1 (単語境界相当)・2 (アクセント句境界相当)・3 (イントネーション句境界相当) であり、この順に切れ目の強さが増す。

^c アクセント句末の音調。下降調 (L%)・上昇調 (H%)・上昇下降調 (HL%) が大半を占める。少数範疇の下降上昇調 (LH%) は H% に編入した。

^d 「で/も」「だ/から」などの長単位接続詞が、冒頭短単位のみでは「助動詞」となってしまうため、長単位の品詞を使用した。

図 1 統語・韻律素性

表1 ランダム切片モデルに対するモデル選択（変数増加法）

追加した変数	自由度	AIC
（ヌルモデル）	2	449.0
節全体の短単位数	3	439.9
文節の丁寧体	4	437.0
AMD の第 1 主成分	5	435.6
直前短単位の AMD	6	434.7

表2 ランダム切片モデルのパラメタ推定値

パラメタ	推定値	標準誤差	z 値	p 値
（切片）	-3.97	1.45	-2.74	< .01
節全体の短単位数	0.70	0.20	3.38	< .001
文節の丁寧体	-0.69	0.29	-2.35	< .05
AMD の第 1 主成分	0.22	0.09	2.45	< .05
直前短単位の AMD	-0.99	0.56	-1.76	< .08

切片の分散 = 1.15

4 結果

4.1 ランダム切片モデル

まず、ランダム切片のみを考慮し、AIC に基づくモデル選択（変数増加法、主効果のみ）で最適なモデルを求めた（表 1）。ランダム切片モデルでは、線形予測式の切片のみが話者によって変動すると仮定する。16 個の説明変数のうち、4 変数が AIC の減少に貢献する変数として選択された。これらは、(1c)「から」節全体の短単位数、(2d)「から」文節の丁寧体、(7a) 平均モーラ長（AMD）の第 1 主成分得点（全体的な遅さ）、(5a)「から」の直前の短単位の平均モーラ長（AMD）であった。このモデルに対するパラメタ推定の結果を表 2 に示す。正の推定値はカラ C の確率を高め、負の推定値はカラ B の確率を高める。

節全体の短単位数は、正の係数を持ち、短単位数が増えるほどカラ C の確率が高まる。AMD の第 1 主成分はやはり正の係数を持ち、発話速度が全体的に遅いほどカラ C の確率が高まる。一方、文節の丁寧体は負の係数を持ち、「から」文節が丁寧体であると、カラ B の確率が高まる。直前短単位の AMD もやはり負の係数を持ち、直前短単位が遅いほどカラ B の確率が高まる。ただし、最後の効果は 5% 水準では有意でない。

4.2 ランダム係数モデル

次に、選択された説明変数について、ランダム傾きを持つモデルを検討した。ランダム傾きモデルでは、線形

表3 ランダム係数モデルのパラメタ推定値

パラメタ	推定値	標準誤差	z 値	p 値
（切片）	-4.65	1.61	-2.89	< .005
節全体の短単位数	0.69	0.21	3.24	< .005
文節の丁寧体	-0.61	0.40	-1.52	= .13
AMD の第 1 主成分	0.29	0.10	2.93	< .005
直前短単位の AMD	-1.26	0.62	-2.05	< .05

切片の分散 = 2.22, 傾きの分散 = 4.71

切片と傾きの相関係数 = -.82

表4 判別結果

観測値	予測値	
	B 類	C 類
B 類	145	28
C 類	34	128

予測式の各説明変数の係数が話者によって変動することを許容する。選択された 4 変数の各係数に変量効果を導入するランダム係数モデルを検討した結果、(2d)「から」文節の丁寧体の係数をランダムとしたモデルのみ、AIC がさらに減少した（AIC = 428.8）。このモデルに対して、残りの 12 変数の効果をモデル選択（変数増加法、主効果のみ）を用いて検討したが、他に有効な変数はなかった。そこで、このモデルを最適なモデルとした。このモデルに対するパラメタ推定の結果を表 3 に示す。

直前短単位の AMD の係数の絶対値がやや大きくなり、5% 水準で有意になった点を除いて、表 2 の結果に類似している。しかし、文節の丁寧体の固定効果（全話者に対する平均的な効果）はもはや有意でない。この変数の変量効果（傾きの分散）は 4.71 と大きく、丁寧体がかラ B/C いずれの確率を高めるかには、話者による大きな揺れがある。

4.3 正解率

最適モデルによる判別結果を表 4 に示す。正解率は 81.5% であった。

5 考察

分析より明らかになったことは、以下の 3 点である。

1. カラ節が長くなるほど、カラ C の確率が高まる。
2. カラ節全体の発話速度が遅くなるほど、カラ C の確率が高まる。
3. カラ節の直前短単位が遅くなるほど、カラ B の確率が高まる。

1 は、比較的長い説明の後で「から」が現れた場合、その後方で示される話し手の判断の根拠（カラ C）として認識されやすい、と解釈できる。2 節で述べたように、カラ C はカラ B に比べて内部に取り得る文法カテゴリの範囲が広いという統語的な特性を持つ。この特性、つまりカラ節全体の（潜在的な）節長の違いが、カラ B/C の判別に寄与している可能性が指摘できる。

このようにカラ C の方が内部に取り得る文法カテゴリの範囲が広いという統語的な事実はあるものの、個々に見た場合、事例に基づく検討が必要と考えられるものもある。例えば南 (1974) では、C 類の従属節内部には主題の「ハ」が現れ得るが、B 類では現れないとされている。しかし少なくとも本分析では、(1e) カラ節内部の主題の有無はカラ B/C の判別に有効でなかった。

2 は、カラ節全体がゆっくり発話されるほど、話し手の判断の根拠（カラ C）として解釈されやすい、ということである。南 (1974) によれば、カラ節直後の統語的な切れ目の大きさは、カラ B よりもカラ C の方が大きい。つまり、比較的ゆっくりと発話されたカラ節は、その直後に統語的に大きな切れ目があることを示していると考えられる。ここから、カラ C の直後には休止が置かれやすい傾向があると予測することができるものの、実際には、(6b) カラ節直後の休止の有無はカラ B/C の判別に影響を与えなかった。個々の事例を観察すると、カラ B の場合でも焦点化されていなければ直後に休止が置かれるケースが多く、統語的な切れ目の大きさと休止の有無とは必ずしも対応していないと考えられる。

3 に挙げた直前短単位（述語句の末尾）の発話速度の効果については、現時点で意味のある解釈を与えることは難しい。この係数の推定値は、モデルの変更（表 2 → 表 3）による変動も大きく、不安定であるため、現時点で決定的なことは言えない。

最後にカラ節の述語の丁寧度について触れたい。丁寧度の係数をランダムにしたモデルから、丁寧度がカラ節の判別に与える影響には個人差があることが分かった。つまり、丁寧体のカラ B の確率を高める話者もいれば、カラ C の確率を高める話者もいる。この結果は、一般にカラ C の方が内部に取り得る文法カテゴリ（丁寧さを含む）の範囲が広い、という南 (1974) や田窪 (1987) の見解と矛盾する。しかし事例を観察する限り、以下のように、カラ節が丁寧体であっても事態の原因・行動の理由（カラ B）としてしか解釈できない例が多く存在する。

(7) いわゆる航空会社です から (F ま) ただみたいなもので海外行けるんですけども (S021F0038)

(8) うちにあたしの実家に子供を置いてあります から (F あのー) 電話したんですね (S025F1517)

田窪 (1987) は、「～ますから」がカラ C であることとして、(9) のように焦点化できないという事実を挙げているが、

(9) 田中さんがいますから、北海道に行くのでしょうか。「「ますから」を C 類の接続助詞とするのは、すこし問題もある (p. 47)」とも指摘しており、丁寧体の述語句を持つカラ節が C 類に限定されるわけではないことを示唆している。本分析結果はこの見解を支持するものである。少なくとも (7), (8) のようにカラ節が文脈上焦点化されておらず、発話全体が丁寧体で話され、更にカラ節直後で多少逡巡するような場合、カラ B 節の内部であっても丁寧体が共起できると考えてよいと思われる。

それでは、そもそも丁寧体に見られた個人差をどう解釈したら良いだろうか。事例を観察すると、話者によって「～するものですから」などの特定の言い回しをカラ B として多用する話者が見られた。このような話し方の特徴が個人差に結び付いた可能性が考えられる。

6 おわりに

本研究では、カラ節が現れた時点までの統語・韻律素性を用いて、2 種類のカラ節の判別を試みた。その結果、カラ節の節長や発話速度などの要因がカラ節の判別に関与するという結果が得られた。この結果は、我々がこれらの特徴を用いてカラ節の判別をリアルタイムで処理している可能性があることを示唆するものと言える。

本分析ではカラ節の多義性の判別問題を取り上げたが、同様の分析手法により、さまざまな言語現象を検討することができる。音声コミュニケーションにおける漸進的な発話理解において、種々の言語的・非言語的情報がどのように用いられているか、今後検討を積み重ねていくことが求められる。

参考文献

- 国立国語研究所. (2006). 国立国語研究所報告書 124: 日本語話し言葉コーパスの構築法. 国立国語研究所.
南不二男. (1974). 現代日本語の構造. 大修館書店.
田窪行則. (1987). 統語構造と文脈情報. 日本語学, 6(5), 37–48.