

## 会話コーパスの転記方式の相互変換 —言語・音響特徴を用いた会話分析方式の音調マーカ―の導出—

石本 祐一 (国立国語研究所言語資源研究系)<sup>†</sup>

土屋 智行 (国立国語研究所言語資源研究系)

小磯 花絵 (国立国語研究所理論・構造研究系)

伝 康晴 (千葉大学文学部/国立国語研究所言語資源研究系)

### **Towards Automatic Transformation between Different Transcript Conventions: Prediction of Intonation Markers from Linguistic and Acoustic Features**

Yuichi Ishimoto (Dept. Corpus Studies, NINJAL)

Tomoyuki Tsuchiya (Dept. Corpus Studies, NINJAL)

Hanae Koiso (Dept. Linguistic Theory and Structure, NINJAL)

Yasuharu Den (Faculty of Letters, Chiba University/Dept. Corpus Studies, NINJAL)

#### 1. はじめに

話し言葉コーパスでは音声収録・転記という初期段階にかかる大きな負担が課題となっている。特に大規模な会話コーパスの整備は未着手であり、会話研究の遅れの要因となっている。国立国語研究所独創・発展型共同研究「多様な様式を網羅した会話コーパスの共有化」(リーダー: 伝康晴・2011年11月~2014年10月)は、既存の会話コーパスを共有化することでこの課題を解決することを目的として立ち上げられた。

既存のコーパスを共有する上での問題のひとつとして、転記方式の不統一や基本アノテーションの欠如が挙げられる。土屋ほか(2012, 2013)では、『日本語し言葉コーパス』(CSJ)方式の言語・音響情報から会話分析(CA)方式の音調マーカ―を予測するための多変量モデルを構築し、CSJ方式からCA方式の音調マーカ―への自動変換を試みた。その結果、転記者ごと・データごとに音調マーカ―の特定に貢献する言語・音響特徴が異なり、また予測精度も音調ごとに異なることがわかった。

本研究では、先行研究で構築された多変量モデルにおいて音調マーカ―を正しく予測できなかった事例に対して、誤変換の音響的な要因を探る。具体的には、先行研究で特に予測誤りの多かった上昇の音調マーカ―であるクエスチョンに着目し、正しく予測された事例の音響特徴と誤って予測された事例の音響特徴の比較を行う。さらに、CSJ方式からCA方式の音調マーカ―へより精度の高い変換を行うために必要な音響特徴を検討する。

---

<sup>†</sup> yishi@ninjal.ac.jp

## 2. データ

### 2.1 談話資料

本研究で用いる会話コーパスは、先行研究(土屋ほか 2012, 2013)と同じ千葉大学3人会話コーパス(Den and Enomoto 2007)の2会話(chiba0232とchiba0432)、合計約20分である。本コーパスには、簡略版CSJ方式による転記テキストと、発話単位・形態論情報・韻律情報などの種々のアノテーションが与えられている。

### 2.2 転記・アノテーション

#### 2.2.1 CSJ方式

CSJ方式では、X-JToBI(五十嵐ほか 2006)に基づく韻律情報が提供されており、アクセント句の末尾に句末境界音調が付与される。句末境界音調として、具体的には

- (1) 下降調 (L%)
- (2) 単純な上昇調 (L%H%)
- (3) 上昇前に一定期間低ピッチが見られる上昇調 (L%LH%)
- (4) 上昇下降調 (L%HL%)
- (5) 上昇下降上昇調 (L%HLH%)

の合計5種類が認定される。ただし、上昇下降上昇調 L%HLH% は本データには出現しなかった。また、上昇下降調 L%HL% も下降調・上昇調に比べて極端に件数は少ない。下降調 L% は、複合境界音調が生じないアクセント句末に付与される音調であり、必ずしも明示的な下降が生じているわけではない。この点において、CA方式のピリオド‘.’とは若干異なる。また上昇調 L%H% および L%LH% は疑問上昇調だけでなく強調上昇調なども含まれており、クエスチョン‘?’とは必ずしも一致しない。

#### 2.2.2 CA方式

CA方式の転記に使われる種々の転記シンボルのうち、本研究では、

- (i) ピリオド‘.’ (per)
- (ii) クエスチョン‘?’ (ques)
- (iii) コンマ‘,’ (com)

の3つの音調マーカーに注目した。これらのマーカーはそれぞれ下降・上昇・継続の音調を表す。土屋ほか(2013)では平坦の音調を表すアンダーバー‘\_’ (ub) も取り上げていたが、上記の転記シンボルに比べて事例数が極端に少ないため、本研究の分析対象からは除外した。

本研究で用いるデータには、Gail Jeffersonの体系(Jefferson 2004)に準拠した転記が、X氏・Y氏・Z氏の3名の会話分析研究者によって作成されている。本研究ではこのうち、約6年の会話分析経験を有するZ氏によって作成されたデータを用いる。Z氏は、2004年から2007年までカリフォルニア大学ロサンゼルス校で会話分析を学び、2007年以降は日本国内のデータセッションや研究会に参加しており、大学院博士課程在籍時より主要な分析手法の1つとして会話分析を採用している。また、会話分析以外に、認知言語学と談話機能主義言語学の知識を有している。

## 2.3 言語・音響特徴

分析の基本単位として、土屋ほか(2013)による先行研究と同様に、強い切れ目で区切られたアクセント句 (Break Index が 3 または 2+b, 2+p, 2+bp) を用いた。土屋ほか(2013)では、CSJ の言語・音響情報から CA 方式の音調マーカを予測する多変量モデルの構築に際し、分析対象アクセント句から以下の言語・音響特徴が抽出され用いられている。<sup>\*1</sup>

### ■言語特徴

句末境界音調 (tone) アクセント句末の句末境界音調。L%, H%, HL%, LH%。

末尾単語の品詞 (lastPOS) アクセント句末尾の単語の品詞。品詞は以下の 7 種に分類した。体言・用言・助動詞・終助詞・接続助詞・その他の助詞・その他の品詞。

次末単語の品詞 (penultPOS) アクセント句の最後から 2 番目 (次末) の単語の品詞

発話中の位置 (loc) 当該アクセント句の発話の先頭からの距離 (アクセント句数)

発話末からの位置 (revLoc) 当該アクセント句の発話の末尾からの距離 (アクセント句数)

### ■音響特徴

アクセント句の最小 F0 (f0MinAP) アクセント句中の F0 の最小値 (標準化得点)

アクセント句の最大 F0 (f0MaxAP) アクセント句中の F0 の最大値 (標準化得点)

句末単語の最大 F0 (f0MaxWord) 末尾単語中の F0 の最大値 (標準化得点)

アクセント句の最大パワー (pwrMaxAP) アクセント句中のパワーの最大値 (標準化得点)

句末単語の最大パワー (pwrMaxWord) 末尾単語中のパワーの最大値 (標準化得点)

アクセント句の平均モーラ長 (amdAP) アクセント句の継続時間をモーラ数で除したもの (標準化得点)

最右 F0 抽出点の値 (lastF0Val) アクセント句中で最後に抽出できた F0 点の値 (標準化得点)

最右 F0 抽出点の位置 (lastF0Loc) 最右 F0 抽出点の句末から計った時間 (対数値)

F0 と平均モーラ長は対数変換後、パワーはそのまま、話者ごとに標準化得点に変換した。

## 3. 分析

### 3.1 先行研究の言語・音響特徴を用いたモデル

2.3 節の言語・音響特徴から Z 氏によって転記された chiba0232 と chiba0432 の音調マーカを予測する多変量モデルを構築した。多変量モデルとしてランダムフォレスト法 (Breiman 2001) を用い、統計解析ソフト R 言語の randomForest パッケージを利用した (mtry = 4 とした)。2 つのデータのうち一方を学習データとし、他方のデータの音調マーカの予測結果と人手による音調マーカを比較したところ、chiba0232 を学習データとしたモデルの正解率は 75.9%、chiba0432 を学習データとした正解率は 72.0% となり、比較的高い精度で予測ができていた。

<sup>\*1</sup> 音響特徴として、アクセント句の平均 F0・句末単語の平均 F0・句末単語の最小 F0・アクセント句の平均パワー・句末単語の平均パワー・句末単語の平均モーラ長も抽出されたが、これらの特徴との相関が高いため用いられていない。

表1 先行研究の言語・音響特徴によるモデルの予測結果 (上昇調 H% のみ)

学習 = chiba0432, テスト = chiba0232 (正解率 = 54.4%, $\kappa = .35$ )					学習 = chiba0232, テスト = chiba0432 (正解率 = 56.6%, $\kappa = .36$ )				
予測値	観測値				予測値	観測値			
	none	per	ques	com		none	per	ques	com
none	14	1	3	1	none	16	0	4	4
per	6	38	32	0	per	7	21	9	0
ques	0	1	10	0	ques	2	9	10	1
com	2	1	0	0	com	0	0	0	0

しかし、句末境界音調が上昇調 H% の結果だけを抜き出すと、表1に示すように正解率はそれぞれ 55% 前後でかなり低い。特に、H% の典型的な機能と考えられるクエスチョン (ques) をピリオド (per) と誤って予測する例が多い。

### 3.2 本研究の方法

前節に示したように、先行研究の言語・音響特徴では句末境界音調が H% である事例に対しクエスチョン (ques) の音調マーカを精度良く予測することができない。以下では ques の予測に焦点をあて、誤った予測へ導く音響特徴の調査を行う。また、ques とそれ以外の音響マーカを区分できる音響特徴を探る。

### 3.3 誤った予測へ導く音響特徴

H% の特徴として、句末の F0 の高さが考えられる。人手による音調マーカが ques の場合における、最右 F0 抽出点の値 (lastF0Val) を予測値ごとに分類した結果を図1に示す。図1からわかるように、lastF0Val が高い場合は正しく ques として予測できている。一方、lastF0Val が低い場合はピリオド (per) と誤って予測されている。すなわち、H% であってもアクセント句の最終 F0 が高くない場合があり、そのようなアクセント句は正しく ques とは判別されないと考えられる。

しかし、転記者によって句末境界音調が H% と転記されていることから、これらは句末で F0

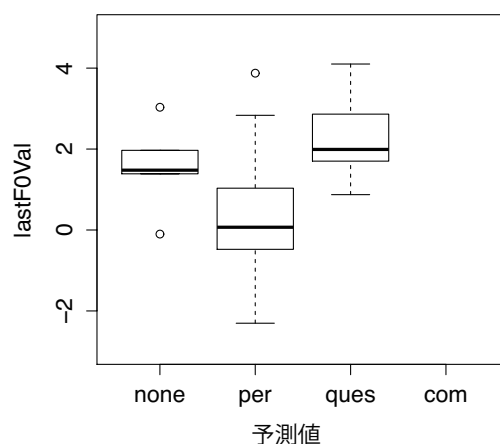


図1 観測値が ques のときの最右 F0 抽出点の値 (lastF0Val)

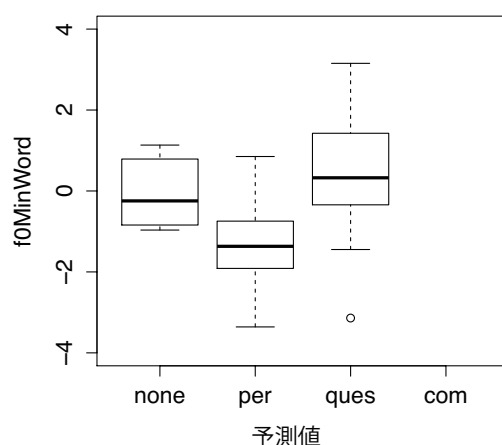


図2 観測値が **ques** のときの句末単語の最小 F0 (**f0MinWord**)

の上昇をとまなうはずである。そこで、図2に示すような、人手による音調マーカが **ques** の場合の句末単語の最小 F0 (**f0MinWord**) を観察してみる。すると、**lastF0Val** と同様に予測値が **per** となったデータでは **ques** と比べて **f0MinWord** が低くなっている。つまり、句末単語（もしくはアクセント句）全体が低い F0 になっており、結果として句末の F0 が上昇しても **lastF0Val** の絶対的な値は通常の H% よりも低い値となっていると思われる。

### 3.4 新たな音響特徴の追加

ここで、2.3 節の音響特徴を利用して、句末の F0 上昇を表す音響特徴を導出することを考える。句末の F0 上昇は簡易的に次の音響特徴で表現できる。

**末尾 F0 上昇幅 (lastF0Rise)** 最右 F0 抽出点の値 (**lastF0Val**) から末尾単語の最小 F0 の値 (**f0MinWord**) を引いたもの

人手による音調マーカと **lastF0Rise** の関係を図3に示す。図3から明らかなように、**ques** でのみ **lastF0Rise** は大きな値となり、それ以外の音響マーカでは 0 に近い値となる。すなわち、**ques** で起こる句末の F0 上昇をうまく表現できているといえる。

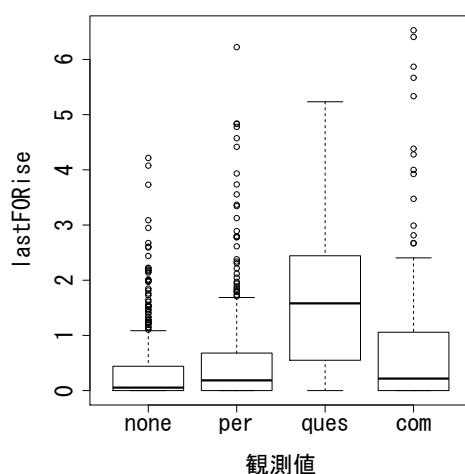


図3 各観測値における末尾 F0 上昇幅 (**lastF0Rise**)

表2 句末のF0上昇を加えたモデルによる予測結果(上昇調H%のみ)

学習 = chiba0432, テスト = chiba0232 (正解率 = 57.8%, $\kappa = .36$ )					学習 = chiba0232, テスト = chiba0432 (正解率 = 61.4%, $\kappa = .43$ )				
予測値	観測値				予測値	観測値			
	none	per	ques	com		none	per	ques	com
none	<b>14</b>	1	4	1	none	<b>17</b>	0	3	4
per	6	<b>38</b>	30	0	per	6	<b>21</b>	7	0
ques	0	1	<b>11</b>	0	ques	2	9	<b>13</b>	1
com	2	1	0	<b>0</b>	com	0	0	0	<b>0</b>

### 3.5 本研究のモデル

2.3 節の言語・音響特徴に **lastF0Rise** を加えて、ランダムフォレスト法による多変量モデルを構築した。chiba0232 を学習データとして chiba0432 の音調マーカを予測した場合と、chiba0432 を学習データとして chiba0232 を予測した場合の、H% に対する結果を表2に示す。句末のF0上昇を表す音響特徴が用いられているにも関わらず、**ques** の予測はあまり改善されず、chiba0232 の予測の正解率は3.4ポイントの上昇、chiba0432 の予測では4.8ポイントの上昇にとどまった。

予測に対する各言語・音響特徴の貢献度を図4に示す。今回導入した **lastF0Rise** は、chiba0232 では全体として4位、音響特徴としては **lastF0Loc** に次ぐ高い貢献度であった。このモデルにおいては **lastF0Rise** が予測に有効な音響特徴として働いているといえる。一方、chiba0432 では **lastF0Rise** は8位であり、**lastF0Val** と同程度の貢献しかしていない。

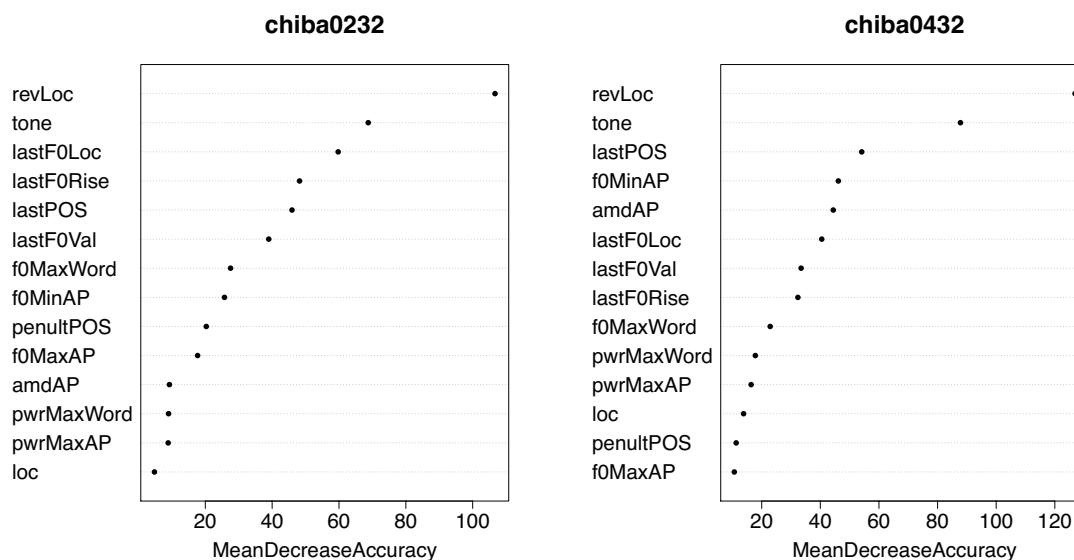


図4 言語・音響特徴の重要度

#### 4. 考察

句末の F0 上昇を表す音響特徴 **lastF0Rise** を用いたにもかかわらず **ques** の予測があまり改善されなかった要因のひとつとして、モデルの学習データにおける **ques** の件数の少なさが挙げられる。H% に限れば **ques** の件数は他の音調マーカーとあまり差がないが、L% のアクセント句は H% の約 3~4 倍の件数がある上、L% では多くのアクセント句が **per** またはラベルなし (**none**) となっている。そのため、単純にモデルを構築すると L% の **none** および **per** を予測する精度が高くなるように学習され、比較的少数の H% の **ques** の予測の精度を上げることができないと考えられる。

実際のところ、ランダムフォレスト法の学習時のブートストラップサンプルの作成の際に、**ques** の件数と同じ件数となるよう他の音調マーカーの事例をサンプリングして学習させると、**ques** の予測の正解率が高くなるモデルが構築される。しかし、**ques** の件数が少ないことから全体のデータ数の減少につながり、**none** や **per** の正解率は大きく低下してしまう。よって、より高精度な予測を行うモデルを構築するためには、予測に役立つ新たな言語・音響情報を見いだすだけでなく、データ数の増加が必要であると思われる。

#### 5. おわりに

本研究では、CSJ 方式の言語・音響情報から CA 方式の音調マーカーを予測するための多変量モデルを構築した先行研究において特に予測精度が悪かった上昇調 H% におけるクエスチョン (**ques**) に焦点をあて、誤った予測になる要因の解明を試みた。その結果、H% にも関わらず、句末付近の F0 値が低い事例が存在することがわかった。さらに、**ques** に関わる句末の F0 上昇を表すことのできる音響特徴を導出した。導出した音響特徴を用いて新たに多変量モデルを構築したところ、わずかに予測精度は向上したがまだ十分な精度には達しなかった。会話コーパス間の転記方式の相互変換に向けて、今後もさらなる言語・音響情報の検討や会話データの増加が必要である。

**謝辞** 会話分析方式の転記を作成していただいた遠藤智子・黒嶋智美・横森大輔の各氏に感謝します。本研究は国立国語研究所独創・発展型共同研究「多様な様式を網羅した会話コーパスの共有化」(リーダー: 伝康晴) による成果である。

#### 参考文献

- Breiman, Leo (2001). "Random forests." *Machine Learning*, 45, pp. 5–32.
- Den, Yasuharu, and Mika Enomoto (2007). "A scientific approach to conversational informatics: Description, analysis, and modeling of human conversation." Toyoaki Nishida (Ed.), *Conversational informatics: An engineering approach*. Hoboken, NJ: John Wiley & Sons. pp. 307–330.
- 五十嵐陽介・菊池英明・前川喜久雄 (2006). 「韻律情報」 『国立国語研究所報告 124: 日本語話し言葉コーパスの構築法』 pp. 347–453.

Jefferson, Gail (2004). “Glossary of transcript symbols with an introduction.” Gene Lerner (Ed.), *Conversation analysis: Studies from the first generation*. Amsterdam/Philadelphia: John Benjamins. pp. 13–31.

土屋智行・伝康晴・小磯花絵 (2012). 「会話コーパスの転記方式の相互変換に向けて—イントネーションに着目して—」 『第2回コーパス日本語学ワークショップ予稿集』 pp. 117–126.

土屋智行・伝康晴・小磯花絵 (2013). 「会話分析方式への転記変換におけるデータ間・個人間のゆれに関する分析」 『第3回コーパス日本語学ワークショップ予稿集』 pp. 417–424.

#### 関連 URL

「会話コーパス」 ホームページ : <http://www.jdri.org/kaiwa/>