

日本語話し言葉コーパスにおける句末音調のバリエーション

菊池英明 (早稲田大学人間科学学術院)

宮島崇浩 (早稲田大学人間科学学術院)

Variation of Tones at the Accentual Phrase Edge in the Corpus of Spontaneous Japanese

KIKUCHI Hideaki (Faculty of Human Sciences, Waseda University)

MIYAJIMA Takahiro (Faculty of Human Sciences, Waseda University)

1. はじめに

表現豊かな音声伝える様々な情報について、科学的解明や工学的応用の関心が高まっている(Erickson(2005), Schuller(2009))。発話の速さや大きさ、イントネーション、声質など、音声表現を豊かにする音響特徴は多数あるが、その中でもアクセント句末の音調が様々な非言語的情報を伝達することがわかっている。Venditti et al.(1998)は、アクセント句末に生じるピッチの変動を”BPM: Boundary Pitch Movement”と表現して、日本語東京方言における句末音調(ピッチの変動のない音調は含まない)について、生成・知覚の双方の観点で5種類の音調が独立して存在することを明らかにした。日本語話し言葉コーパス(CSJ: Corpus of Spontaneous Japanese)にはX-JToBIのスキーム(前川ら(2001))に基づいてラベリングがなされており、付与されたラベル系列のパタンからは、日本語(の主に東京方言)の話し言葉においては主に7種類の句末音調(ピッチの変動のない音調を含む)が存在するといえる(前川(2011))。筆者らは、表現豊かな音声の特性を調べることを目的に、声優や俳優などに多様な状況設定を与えて演技音声を収集することにより多様な音声表現コーパスを構築している(Miyajima et al. (2011))。これまでに収集した3000発話以上もの音声においても、上述のX-JToBIによって8種類のパタンで句末音調を表現できる見通しを得ている。

しかしながら、CSJにおけるイントネーションラベリングにおいて、BPMの種類やラベル付与位置に困難をおぼえたケースは数多くあり、実際の基本周波数(F0)変動のパタンがCSJにおいてどのように分布しているかを詳細に観察する必要がある。本研究では、CSJにおける句末音調のバリエーションについて、まずX-JToBIラベルに基づくBPMの出現頻度分布分析結果を示し、句末モーラ区間でのF0変動のパタンを上述の多様な音声表現コーパスとの比較とともに観察する。

2. データ

2.1 CSJ

2011年にリリースされたCSJ第三版(CSJ(2011))のコア(分節単位ラベルとイントネーションラベルが付与された201講演)と、それを用いて作成されたアクセント句単位XMLを用いる。アクセント句単位XMLは、原則としてX-JToBIのBIラベルの値が2以上の位置を境界とする単位でアクセント句を構成し、論理的に包含するトーンラベルの情報を下位要素として表現する形で記述される。これによって、アクセント句ごとにBPMの種類と句末モーラの始端・終端時間位置、句末モーラの無声化の有無などの情報を容易に取得できる。

なお、CSJにはモノログを中心に、学会講演(APS)、模擬講演(SPS)、再朗読(R)、対話(D)の4種類のスタイルの談話が存在する。以降はこれらの種別を略称で表記する。

2.2 多様な音声表現コーパス

筆者らは、声優や俳優に指示を与えて多様な音声表現を収集してコーパス(通称「千の声コーパス」、以降“SEN”の略称を用いる)を構築する試みを2008年より続けている。指示の具体的な例を表1に示す。

以下では、こうした指示を受けて1名の40代女性声優が発声した発話内容「あーそうですか」の100発話のデータを用いる。収集方法の詳細やこのデータにおける物理的・心理的多様性の検証についてはMiyajima(2011)を参照されたい。

CSJと同様に、分節単位ラベルとX-JToBIラベルを付与しており、以降の分析ではこれらのラベルを用いる。

表1 表現豊かな音声表現を得るための指示の例

共通	発話時の場所・状況	大家族を取り扱った特集において(テレビ番組)
	発話者と聞き手の関係	親子
聞き手	年齢/性別	10歳未満/男
	職業・役柄	小学生
	人物像	典型的なやんちゃな小学生。元気があり待っている状態
発話者	年齢/性別	30代/女
	職業・役柄	主婦
	人物像	元ヤンのヤンママと言った感じ。言葉遣いはキレイではない。
	発声時の背景	子供のだらしなさに対し、思わず声を張って叱る様子

3 句末音調タイプの出現割合

本章では、句末音調の種類の出現割合を、コーパスごとおよび談話のスタイルごとに観察する(図1参照)。なお、“L%>”と“H%>”は、X-JToBIにおいてそれぞれ“L%”(BPMを伴わない音調)と“H%”にトーンの引き延ばしをあらわす“>”(エクステンダー)が付与されたものを示し、いずれも“L%”, “H%”の内数である。図1より、CSJにおいては再朗読(R)を除いて概ね出現割合が似通っていることがわかる。再朗読に出現しないHL%やL%>, H%>は話し言葉の特徴づける音調であるといえる。CSJとSENの比較においては、多様な音声表現の収集を意図したコーパスSENに比べてCSJは“HLH%”と“LH%”, “L%>”, “H%>”などの音調の出現割合が著しく低いことがわかる。

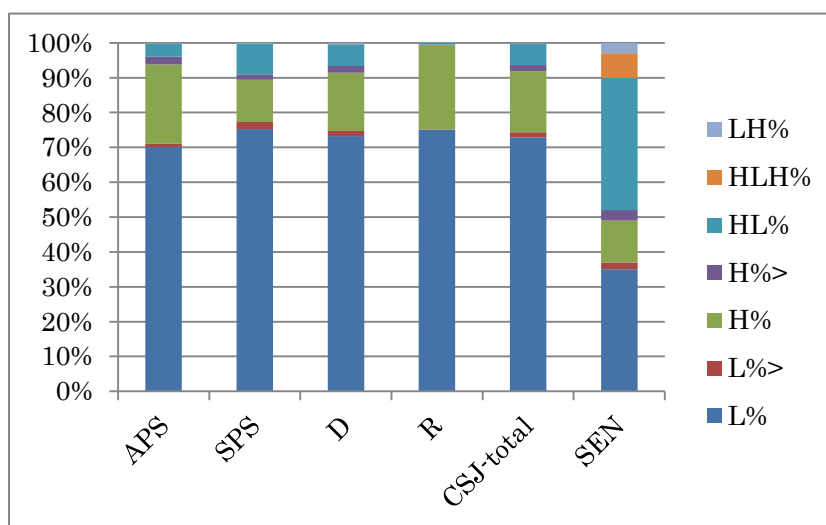


図1 CSJとSENにおける談話のスタイルごとの句末音調タイプ出現割合

4 句末モーラにおける F0 変動

本章では、アクセント句末のモーラにおける F0 変動のパターンを CSJ と SEN で比較観察する。いずれも F0 についてはセミトーンで話者正規化した。まず図 2 に、CSJ の講演種別ごとにランダムに抽出した 50 か所の F0 変動の分布を示す。図 3 に SEN のランダムに抽出した 50 音声の F0 変動の分布を示す。なお、図 3 には、「怒り」「喜び」などの典型的な感情表現のみを指示として与えた際の同一話者による音声表現(Miyajima(2011))における分布を比較のために示す。視認による判断ではあるが、APS と R が比較的分布が狭いものに対して、SPS と D は高低および時間方向のいずれも広がりを見せており、図 3 の SEN と同様に多様な F0 変動パターンが出現していることがわかる。

今後はこれらの分布の違いを定量的に計測するとともに、BPM の分類と F0 変動の形状との関係を調べる予定である。

謝辞

本研究は国立国語研究所（言語資源研究系）基幹型共同研究「コーパス日本語学の創成」（リーダー：前川喜久雄）および萌芽・発掘型共同研究「会話の韻律機能に関する実証的研究」（リーダー：小磯花絵）による成果である。

日本語話し言葉コーパスのアクセント句単位 XML は、小磯花絵、西川賢哉、野口広彰の 3 氏との共同作業によって構築した。ここに記して感謝の意を表す。

参考文献

- D. Erickson (2005). "Expressive speech: Production, Perception and Application to Speech Synthesis", *Acoust. Sci. & Tech.*, vol.4, no.26, pp.317-325.
- B. Schuller, S. Steidl, A. Batliner (2009). "The INTERSPEECH 2009 Emotion Challenge", *Proc. of INTERSPEECH 2009*, pp.312-315.
- J. Venditti, K. Maeda, and J. P. H. van Santen (1998). "Modeling Japanese boundary pitch movements for speech synthesis." *Proc. of the 3rd ESCA Workshop on Speech Synthesis*.
- 前川喜久雄, 菊池英明, 五十嵐陽介 (2001). 「X-JToBI: 自発音声の韻律ラベリングスキーム」, *電子情報通信学会技術報告(NLC2001-71, SP2001-106)*, pp.25-30.

前川喜久雄 (2011). 「コーパスを利用した自発音声の研究」, 東京工業大学大学院博士論文.
 CSJ(2011). 「日本語話し言葉コーパス」, 国立国語研究所, <http://www.ninjal.ac.jp/csj/>
 T. Miyajima, H. Kikuchi, K. Shirai (2011). "Collection and analysis of emotional speech focused on the psychological and acoustical diversity", Proc. of ICPhS2011, pp.1394-1397.

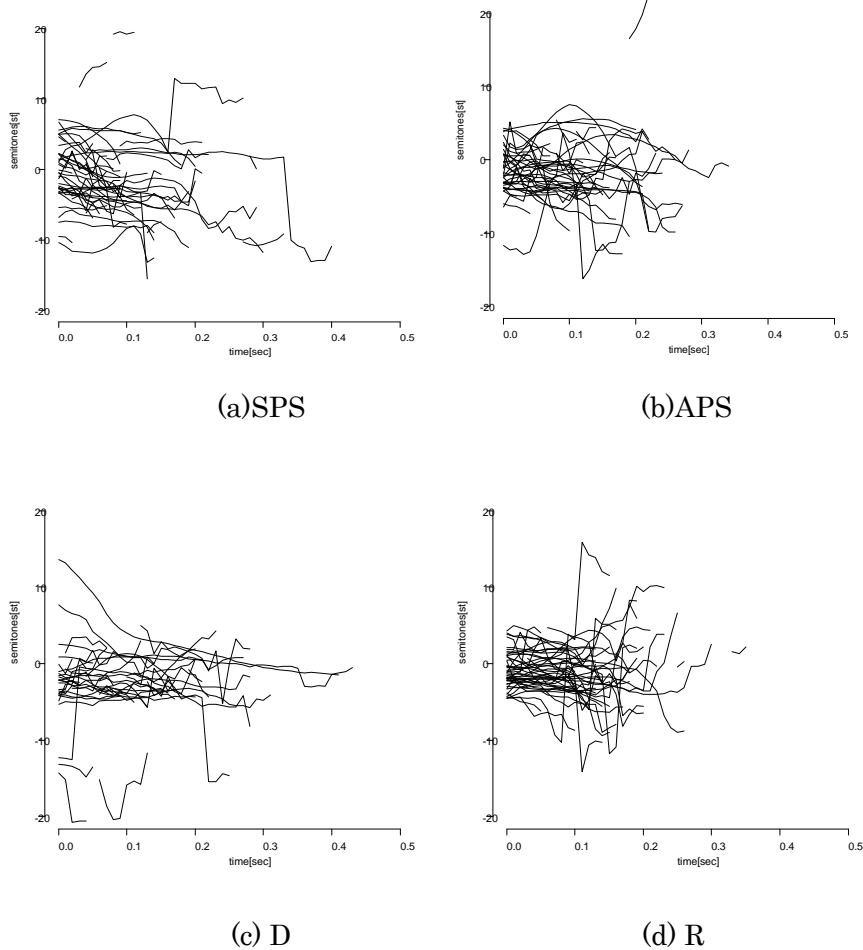


図 2 CSJ における句末モーラの F0 変動

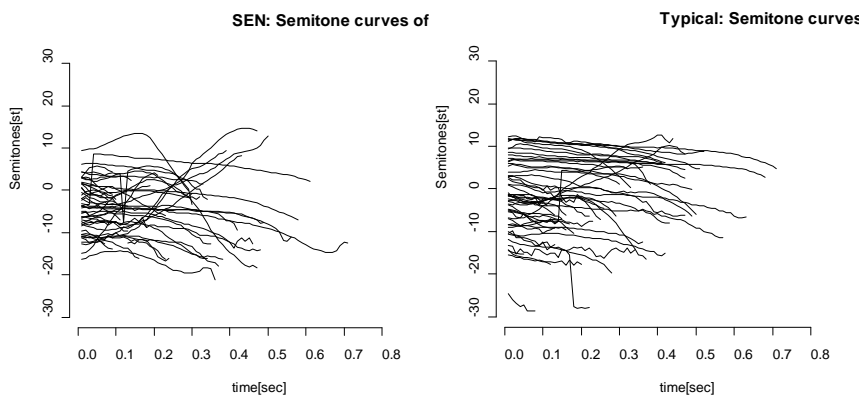


図 3 SEN における句末モーラの F0 変動
 (左は SEN の 50 音声、右は典型的な感情指示語に対する音声表現 50 個)