

日本語話し言葉コーパスを用いた「発話」の韻律的特徴の分析 —イントネーション句を切り口として—

小磯 花絵 (国立国語研究所理論・構造研究系)[†]

石本 祐一 (国立情報学研究所音声メディアグループ)

Prosodic Features of Utterances in the Corpus of Spontaneous Japanese: Intonational Phrase–Based Approach

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

Yuichi Ishimoto (Speech Media Group, NII)

1. はじめに

本研究は、『日本語話し言葉コーパス (*Corpus of Spontaneous Japanese*:以下 CSJ)』(前川 2004, 2006) を用いて、自発音声における「発話」の韻律的な傾向を探ることを目的とする。Pierrehumbert & Beckman(1988) では、アクセント句やイントネーション句より上位に「発話 (utterance)」という単位を設定している。「発話」は、F0 declination (発話に要する時間の関数として単純に F0 が低下する現象) が見られる範囲であり、その末尾で final lowering (平叙文末尾で F0 が局所的に下降し発話の終了を示す現象) が生じるとされる。ニュースのように、ある程度発話内容が準備されたスピーチであれば、「発話」は文末表現で区切られた文に相当する単位ということになる。しかし自発性の高い話しことばでは、次の例に見られるように、複数の節をつなげて長々と発話を続けることも少なからず見られる。このように、途中で強い統語境界をはさみうる長い「発話」であっても、全体的に declination が生じ末尾に final lowering が観察されるのか、という素朴な疑問が生じる。

で (エ) 絶対音感群に比べますと 多少半音から一音のエラーが見られますが (エ) こちらの条件あるいはこちらの条件の時はさほどオクターブエラーがたくさん見られる訳ではないんですが (アイアル) I RN条件の時には (エ) 1 オクターブエラーがかなり見られてることが分かります (談話 ID: A01F0067)

そこで本研究では、F0 の下降現象に着目し、「発話」の長さや統語構造との関係について調べる。手続きとしては、CSJ に付与された節単位情報を用いて暫定的に「発話」を定義した上で、イントネーション句に基づき節単位の韻律的特徴を調べる。Maekawa (2009)・前川 (2011) も同種の分析をアクセント句を単位に行っているが、イントネーション句の中でダウンステップの生じるアクセント句よりも、イントネーション句の方が長い発話の全体的な F0 の推移をとらえる上で適切であると考え、分析単位として採用した。アクセント核の有無など細かな統制のもとで分析を実施した前川の結果と合わせて考察を進める。

[†] koiso@ninjal.ac.jp

2. 方法

2.1 データ

分析には CSJ を用いた。CSJ は自発性の高いモノログを中心に構成された話し言葉コーパスであり、学会における口頭発表（以下「学会講演」と、一般話者による主に個人的な内容に関するスピーチ（以下「模擬講演」）を主対象としている。CSJ 全体は 661 時間の音声から構成されるが、本研究ではこのうち「コア」と呼ばれるデータ範囲の中から学会講演 70（約 29 時間）・模擬講演 107（約 20 時間）を分析対象とした。実際の分析には CSJ 第 3 刷に基づき作成された RDB（小磯ほか 2012）を用いた。

2.2 節単位

「発話」に相当する単位を認定するにあたり、CSJ に付与されている節単位情報を利用した。節単位情報は原則「節 (clause)」の境界によって得られる文法的・意味的なまとまりを持った単位であり、CSJ において構文・談話レベルの情報を付与するための基本単位として設計されたものである（丸山ほか 2006）。節単位は、節境界の構造的な切れ目の大きさの観点から以下の 3 つに分類される。

絶対境界： いわゆる文末に相当する境界。明示的な文末表現の直後。

強境界： 後続の節に対する従属度の低い、つまり切れ目の度合いが強い節境界。

弱境界： 後続の節に対する従属度の高い、つまり切れ目の度合いが弱い節境界。

これらの境界は形態素解析結果に基づき自動で判別され、人手による修正・操作を経た上で、絶対境界、強境界のいずれかで区切られる単位が「節単位」と認定される。

ここで絶対境界は文末に相当するため、絶対境界と絶対境界で区切られる範囲を「発話」と解釈するという考え方もあろう。しかし大石（1971）が指摘するように、話しことばでは「…ケド」「…ガ」「…ノニ」「…テ」「…タラ」などの接続表現がある種のイントネーションやポーズを伴い文末を表わすのに転用されることがある点に注意しなければならない。例えば「今日はもう帰らないといけないんだけど」という発話の場合、末尾の音調が句末を強調するような上昇調や上昇下降調で発話された場合には、そのあとにまだ発話が続くことを予感させるのに対し、末尾に final lowering を伴い一定以上のポーズが後続した場合には、必ずではないものの相対的にそこで発話が終了したと感じられる傾向にある。つまり、強境界の末尾は「発話」の末尾であることもあれば「発話」の途中であることもある、ということであり、節単位のレベルからそれを判断することはできないのである。

そこで分析 1 では、直前が絶対境界である（強境界ではない）絶対境界を最後に持つ節単位（図 1 の節単位 4 と 8）を対象に、節単位の長さ と F0 の下降現象との関係を見る。この条件で分析対象となるデータは、その内部に強境界など強い統語境界を持たないケースである。

次に分析 2 では、発話内部に強い統語境界が置かれる場合を対象に、そこで F0 の下降が継続するか否かを見る。具体的には、強境界の節単位と絶対境界の節単位の二つの連鎖（図 1 の節単位 2・3 と 6・7）を対象とする。

この条件に相当する事例を幾つか聴取したところ、大石（1971）が指摘するようなある種のイントネーション・ポーズを伴い「発話」が終了したと感じられる事例はあまり無く、「発話」が継続し最後の絶対境界で終了するものが多くを占めた（この点については今後改めて確認する必要がある）。この観察が正しいならば、分析2で対象とするデータ（の大半）は一つの発話の中に強い統語的境界（強境界）が（少なくとも）一つ存在するケースということになる。

2.3 イントネーション句

「発話」の韻律的特徴を探るために、本研究ではイントネーション句（Intonational Phrase; 以下 IP）を用いる。IP は、アクセント核が後続するアクセント句（Accental Phrase; 以下 AP）の F0 ピークを反復的に低下させるダウンステップの生じる領域であり、IP 境界でピッチレンジのリセットが生じる。

CSJ にはラベリングスキーム X-JToBI（五十嵐ほか 2006）に基づき韻律情報が付与されているが、この中に、Break Index (BI) という、韻律境界の切れ目の強さに関する情報が存在する。BI=2 は AP 境界、BI=3 は IP 境界、BI=F はフィラー境界（例：「えっと」「あー」）、BI=D は言い淀み境界（例：「す 全ての」）に対応する。ここでは、BI=3, D, F で区切られる範囲を IP と認定した。ただし、フィラーを狭んでダウンステップが続く場合はフィラーを内包する形で IP を認定した。

このように IP を認定した上で、X-JToBI に付与されたトーンの情報に基づき、IP の F0 最大値、F0 最小値、ピッチレンジを次のようにもとめた（図 2）。

F0 最大値： IP 冒頭の AP の句頭音調 (H-) あるいはアクセント核 (A) のうち高い方の F0 値

F0 最小値： IP 末尾の AP の下降音調 (L%) の F0 値

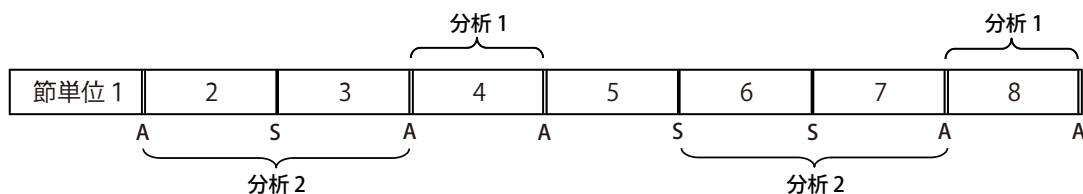


図 1 分析対象とする節単位 (A:絶対境界, S:強境界)

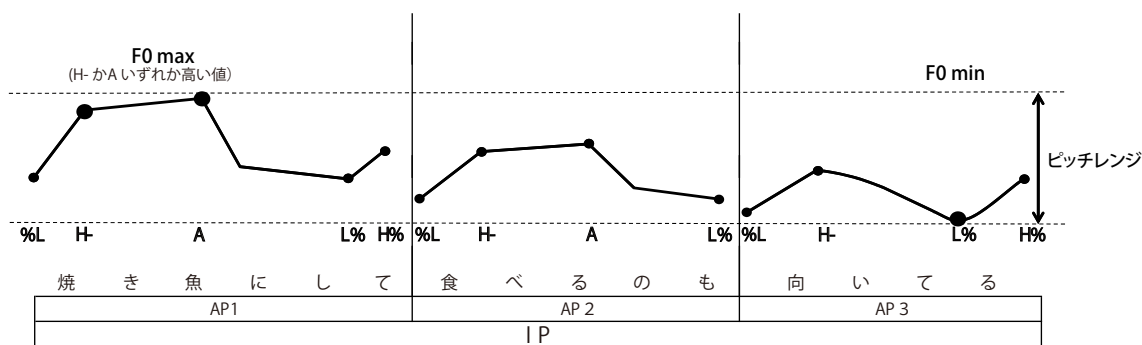


図 2 分析に用いる IP の F0 パラメータ：F0 最大値 (F0 max)・F0 最小値 (F0 min)・ピッチレンジ

表 1 節単位に含まれる IP 数とその頻度

IP 数	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21~
頻度	570	546	656	569	544	478	364	285	228	192	127	90	68	49	42	30	19	17	14	7	19

ピッチレンジ： F0 最大値と F0 最小値の差

概ね節単位の方が IP よりも長い単位であり前者が後者を内包するが、まれに交差することもある（例えば接続詞の前にピッチレンジがおかれず前の発話と続けて発話される場合など）。このような交差が生じた箇所は分析対象外とした。

3. 結果と考察

3.1 分析 1

「発話」内部に絶対境界など強い統語境界を持たないケースを対象に、節単位の長さで F0 の推移（特に下降現象）との関係を見る。具体的には、直前が絶対境界である（強境界ではない）絶対境界を最後に持つ節単位を対象とする（図 1）。節単位の長さの指標として、節単位に含まれる IP 数（フィラー・言い淀みを除く）を用いる。

分析 1 で用いるデータは、節単位数 4914、IP 数 26763 であり、1 節単位の平均 IP 数は 5.4 である。表 1 は、節単位に含まれる IP 数ごとの頻度である。以下に節単位中の IP 数が 1 の場合、6 の場合、12 の場合の発話例を記す。

IP 数 1 の場合 長調のメロディーです

IP 数 6 の場合 次に強さの影響を検討する強さパターン制御条件では弱強型強弱型共に高さは一定にしており強さのみを弱強にした音もう一つは強弱にした音です

IP 数 12 の場合 そして (エー) 照合システムから (エー) テキストが提示されましたらこの音素 HMM を (エー) このテキストに従って接続しパラメーター (ワ) (ン) (エー) スペクトルパラメーターおよびピッチを生成して (エー) 合成フィルターにより音声を作成します

図 3 は、節単位中の IP 数ごとに IP の F0 最大値・最小値・ピッチレンジの推移を図示したものである。例えば六つの IP から構成される節単位の場合（図中 len=6）、冒頭から数えて 1～6 番目の IP の各 F0 パラメータの分布がこの順に示されている。また F0 は談話ごとに Z 値に変換している。

図を見ると、節単位に含まれる IP 数に関わらず、節単位冒頭から末尾にかけて、F0 最大値・最小値が徐々に下降する傾向にあることが分かる。特に F0 最大値（IP 冒頭の AP の H-あるいは A の F0 値）と比べて F0 最小値（IP 末尾の AP の L% の F0 値）の下がり方が大きいので、ピッチレンジも徐々に下降する。紙面の都合で省いたが、ここに載せていない長さの節単位についても同様の傾向が見られる。またこの傾向は男性・女性に分けた場合にも概ね見られる（図 4・5）^{*1}。

^{*1} 冒頭二つの IP のピッチレンジが同じあるいは増加していることもあるが、これは冒頭に接続詞が置かれることが多く、その他の IP と異なる性質を有するためと推察される。接続詞の特殊性については Maekawa (2009) でも指摘されている。

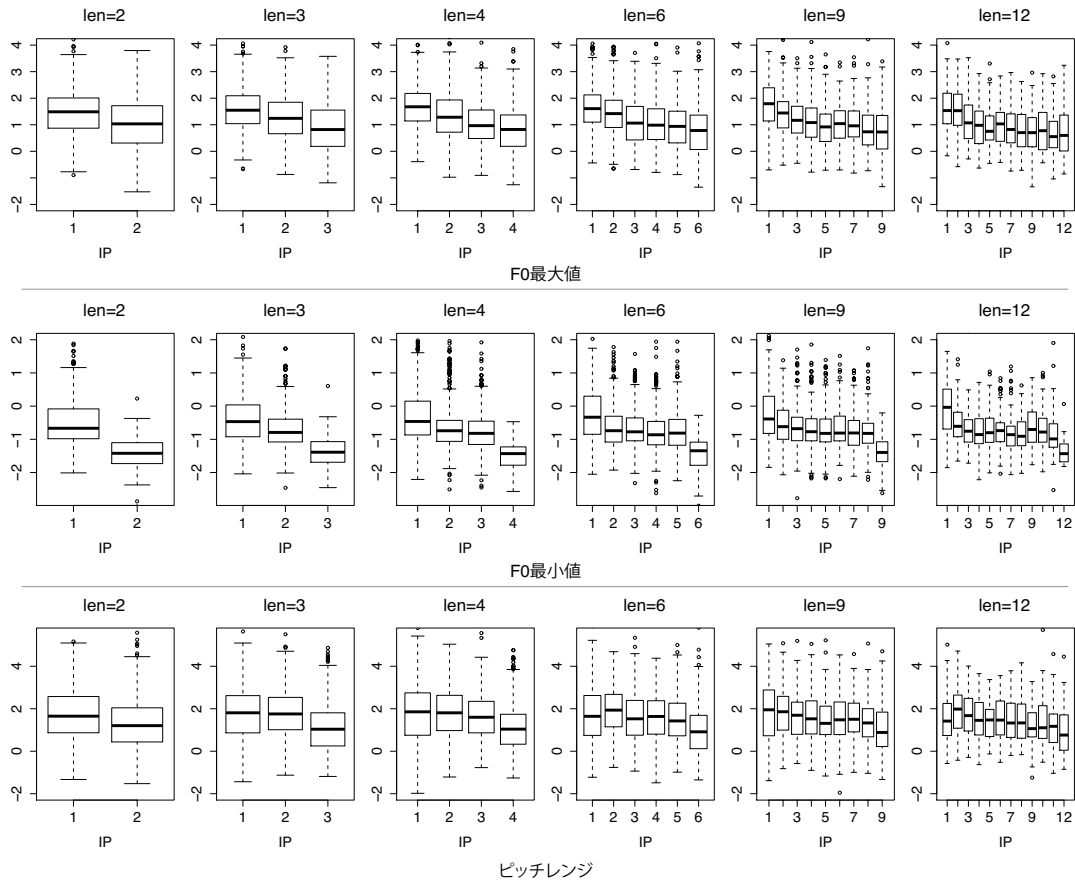


図3 節単位（絶対境界）内のイントネーション句のF0パラメータ（Z値）の推移

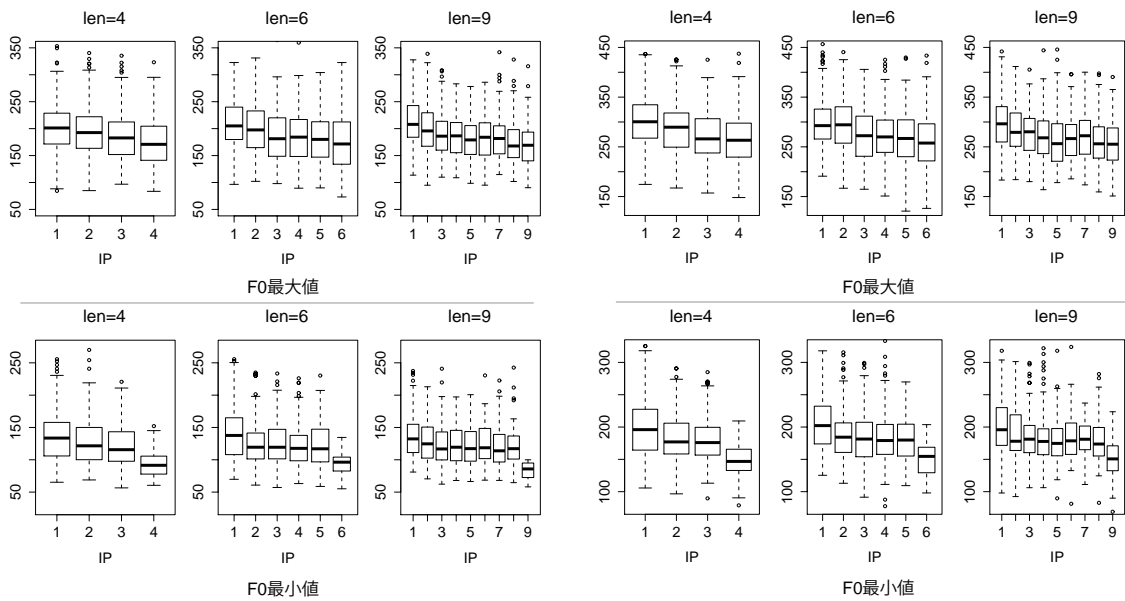


図4 男性の場合 (Hz)

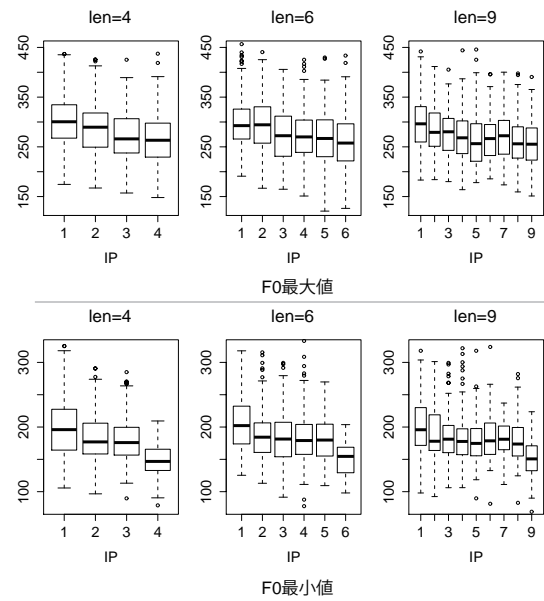


図5 女性の場合 (Hz)

本分析では IP を分析の単位として見ているため、この下降はダウンステップではない。仮に declination の効果であるならば、declination は節単位の長さに関わらず生じるということになる。また、節単位最後の IP の F0 最小値（図 3 F0 最小値の各図の一番右）がその前の IP と比べてより大きく下降するという傾向が節単位の長さに関わらず観察されるが、これは final lowering の効果と考えられる。以上二つの傾向は、AP を対象に分析した Maekawa(2009)・前川 (2011) でも報告されており、その妥当性が改めて確認されたことになる。

ここで、図 3 を細かく見ると、節単位冒頭（特に接続詞の影響のない 2 番目の IP）の F0 最大値と節単位末尾の F0 最小値が、節単位の長さに関わらずほぼ一定であるという点に気付く。つまり、短い発話をするにせよ長い発話をするにせよ、ほぼ同じ高さから始まりほぼ同じ高さで終わるということである。この点を確認するために、節単位中の IP 数が 2~13（いずれも頻度 50 以上）の範囲で、節単位頭から 2 番目の IP の F0 最大値と節単位末尾の F0 最小値だけをまとめてプロットした（図 6）。節単位最初の IP ではなく 2 番目の IP を比較するのは接続詞の影響を除くためである。図 6 を見ると、確かに節単位末尾の F0 値はほぼ一定であるのに対し、節単位 2 番目の F0 最大値は節単位中の IP 数が増えるにつれてわずかであるが高くなる傾向が見られる*2。

前川 (2011) は、節単位末尾の AP のピッチレンジがほぼ一定範囲に納まる傾向が見られることから、発話長を考慮に入れた F0 制御が行われている可能性があることを指摘している。上記の節単位末尾の F0 最小値がほぼ一定であるという結果は、前川の節単位末尾の AP のピッチレンジの結果とほぼ同じことを指しているが、今回更に、節単位の IP 数の増加に伴い節単位冒頭の F0 最大値が徐々に高くなる傾向にあることが明らかになった点は興味深い。というのも、発話長を考慮し、発話末尾の F0 最小値に向けて発話の途中途中で F0 制御をするというだけでなく、長い発話をする場合には発話の冒頭時点で既に少し高めに発話を開始するといった調整が行われている可能性があるためである。またこの結果は、発話のピッチレンジの下限がかなり固定的であるのに対し、上限は若干幅があるということも示唆する。

ところで図 3 を詳細に見ると、実は節単位冒頭からの IP 数が 5 を越えたあたりで F0 最大値・最小値ともに下降が停滞し、それより長く節単位が続く場合、F0 は一旦上昇した上で下降する傾向にあることに気付く。ただし、上昇に転じる位置（途中で下がり切った位置）の F0 最小値は節単位末の F0 最小値ほど低くはならず、また上昇ののち再び下降に転じる位置（途中で上がり切った位置）も節単位冒頭の F0 最大値ほど高くはならない。つまり発話がかなり長い場合、少し高い位置から発話を開始し下降の傾きを小さくするといった調整をしたとしても、それには限界があり、F0 の下限に達する前に一度 F0 を上昇させ、発話の立て直しをしている可能性があるということであろう。勿論この位置で final lowering は観察されない。

3.2 分析 2

本節では、一つの発話の中に強い統語的境界（強境界）が少なくとも一つ存在するケースを対象に分析を行う。具体的には、強境界の節単位と絶対境界の節単位の二つの連鎖を分析対象とする（図 1）。ここでの着目点は、発話内部に強い統語境界が置かれる位置において、前節

*2 数が少ないため明確なことは言えないが、IP 数が 14 以上の場合、このまま上昇し続けることはないようである。

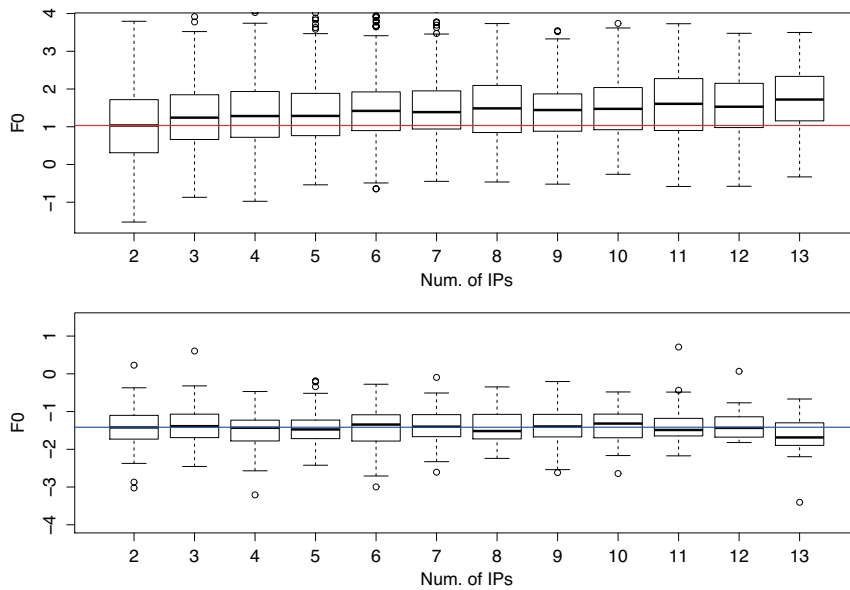


図6 節単位冒頭・末尾の F0: (上) 節単位 2 番目の IP の F0 最大値, (下) 節単位末尾の IP の F0 最小値

表2 先行節単位（強境界）と後続節単位（絶対境界）の IP 数とその頻度

	絶対境界の IP 数										
	1	2	3	4	5	6	7	8	9	10～	
1	75	84	65	51	25	20	22	8	2	18	
2	127	136	148	86	58	36	35	19	10	24	
3	116	164	126	98	64	33	33	18	12	31	
4	105	152	107	71	54	46	29	31	17	18	
強境界の IP 数	5	90	121	93	105	49	36	15	21	14	26
	6	76	82	66	67	63	28	22	18	4	27
	7	65	88	45	66	28	28	22	13	5	26
	8	33	45	38	46	43	31	22	10	17	16
	9	40	41	31	20	18	8	8	6	2	10
	10～	87	124	91	81	65	44	46	26	26	55

で見たような F0 の下降が継続して見られるかどうかということである。表 2 に、先行節単位（強境界）と後続節単位（絶対境界）に含まれる IP 数ごとの事例数を示す。

前節と同じ分析を、先行節単位と後続節単位に含まれる IP 数ごとに行った。組合せが多いため、典型的な事例に限定して IP の F0 最大値と最小値の結果を図 7 と図 8 に示す。強境界 (S) と絶対境界 (A) の位置を縦線で示している。

図 7 と図 8 から、先行節単位と後続節単位の範囲でそれぞれ下降の傾向は見られるのに対し、強境界 (S) で下降がリセットされていることが分かる。また後続節単位の末尾、つまり絶対境界 (A) の位置の F0 最小値が、いずれのケースでもおよそ -1.5 にまで急激に下がっており (図 7・8 (下)), final lowering が確認できる。またこの値は分析 1 の図 6 (下) で確認したものと一致する。それに対し先行節単位の末尾である強境界 (S) では、-1 程度にまでしか下がっておらず、final lowering も見られない。仮に -1.5 が「発話」のレンジの下限であるとするならば、強境界の末尾ではその下限にまで落ち切らないところで (final lowering を伴わず) リ

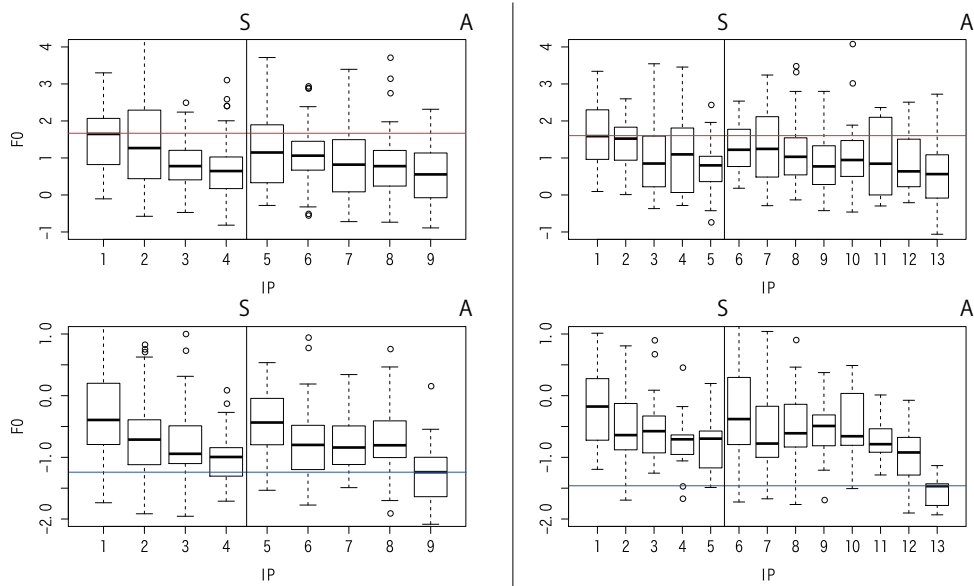


図7 強境界後の F0 上昇幅が小さい例: (上)F0 最大値, (下)F0 最小値

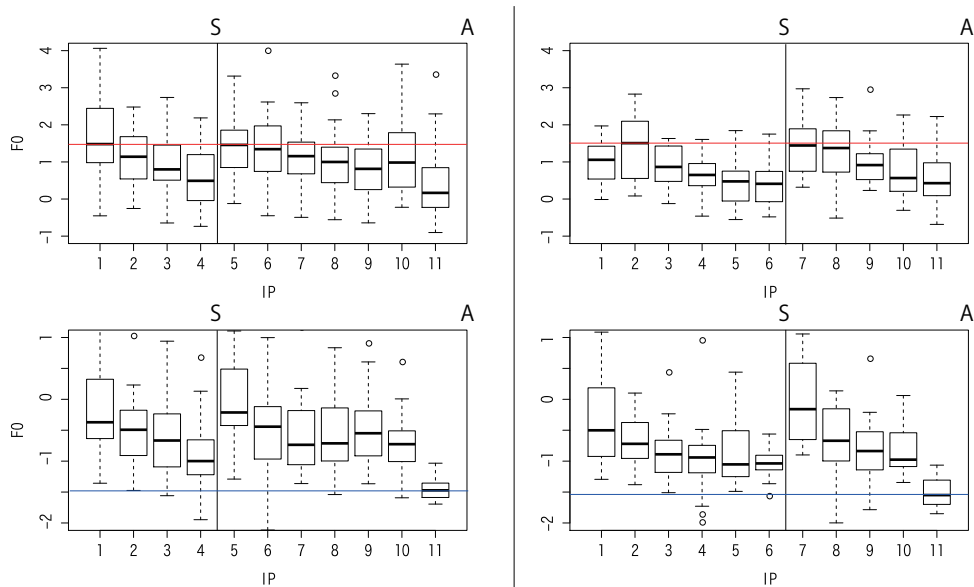


図8 強境界後の F0 上昇幅が大きい例: (上)F0 最大値, (下)F0 最小値

セットしているということである。

仮に、ここで対象とする強境界が概ね「発話」の途中であり、節単位全般に見られる下降現象が declination (だけ) の効果であり、declination が「発話」全体に見られる現象であるとするならば、強境界の位置で下降は継続することになるが、この位置でリセットされる傾向が見られるということは、この仮定のいずれかに問題があるということになる。勿論、強境界が「発話」末になることがないわけではないが、平均値としてここまで強いリセットが見られることにはならないだろう。また「発話」の末尾に観察される final lowering が後続節単位の絶対境界

にのみ見られるという事実からも、強境界の仮定（のみ）が問題ということはなさそうである。

単純に考えるならば、「発話」の内部であっても強い統語境界があると declination がリセットされるという可能性が指摘できる。強境界の位置では final lowering を伴わないことから、final lowering の生起範囲と declination の範囲が必ずしも一致しないことになる。また declination は本来、発話に要する時間の関数として単純に F0 が低下する現象であるのに対し、本分析で見た下降も Maekawa(2009)・前川(2011)で観察された下降も、いずれも発話長に応じて下降の傾きが変化するというものであり、declination だけで説明することは難しい (Maekawa 2009) という点にも注意する必要がある。つまりここで見られる下降には declination 以外の要因も関わっており、その別の要因が強い統語境界の範囲に影響する可能性もありうるということである。

なお分析 1 で対象とした、「発話」内部に絶対境界など強い統語境界を持たないケースでも、IP 数が 5 を越えたあたりで F0 の下降の停滞（場合によってはリセット）が観察されることを指摘した。同じ傾向がここで対象とした節単位（特に後統節単位）にも観察される。実は節単位の内部であっても、例えば「ここはちょっとデータが足りない ので (アノ) 欠損値になっておりますが」の「ので」の位置のように、いわゆる「弱境界」に相当する統語的な切れ目が存在することもあり、これが丁度 5 番目の IP 前後に置かれる傾向にある、という可能性も存在する。今後、弱境界まで含めた詳細な分析を行う必要がある。

最後に、図 7 と図 8 の結果をもう少し細かく見ておこう。上述の通り、強境界の F0 最小値が -1 程度のところで上昇に転じるという点で共通した傾向を示すのに対し、後統節単位冒頭の F0 最大値は、ここに示していない組合せも含め二つのケースに分類される。一つは図 7 に示すパターンで、先行節単位冒頭の F0 最大値と比較して後統節単位冒頭の F0 最大値があまり高くない（先行節単位末からの上昇幅が小さい）もの、もう一つは図 8 に示すパターンで、先行節単位冒頭の F0 最大値と比較して後統節単位冒頭の F0 最大値がほぼ同じ高さにまで戻る（先行節単位末からの上昇幅が大きい）ものである。この上昇幅が何によって決まるのかは分からないが、「発話」の途中でであってもレンジの上限に戻ることがありうるということであり、「発話」の途中で下限には達しないということと対称的である。

4. おわりに

本稿では、F0 の下降現象に着目し、「発話」の長さや統語構造との関係について分析した。その結果、分析 1 では、「発話」のレンジがある程度固定されておりその範囲で全体的に F0 が下降する傾向にあること、「発話」の長さに応じて下降の傾きが変わる傾向にあること、「発話」が長い場合には若干高い位置で発話を開始する傾向にあることなどが分かった。また分析 2 では、「発話」の内部に強い統語境界があると下限に達する前に F0 の下降がリセットされる傾向にあること、その場合でも final lowering は「発話」末にしか現れないことが分かった。今後、統語的な条件を細かく設定した上で F0 との関係を見ていく必要がある。またポーズなど統語以外の要因についても考慮する必要がある。

参 考 文 献

- 前川喜久雄 (2004) 「『日本語話し言葉コーパス』の概要」『日本語科学』, 15, pp. 111–133.
- 前川喜久雄 (2006) 「概説」『日本語話し言葉コーパスの構築法』(国立国語研究所報告 124), pp. 1–21.
- Kikuo Maekawa(2009) “Contributions to corpus phonetics” (国語研究所 NINJAL サロン).
- 前川喜久雄 (2011) 『コーパスを利用した自発音声の研究』(博士論文) .
- 丸山岳彦・高梨克也・内元清貴 (2006) 「節単位情報」『日本語話し言葉コーパスの構築法』(国立国語研究所報告 124), pp. 255–322.
- Pierrehumbert and Beckman(1988) *Japanese Tone Structure*, Cambridge: MIT Press.
- 大石初太郎 1971. 『話しことば論』秀英出版.
- 五十嵐陽介・菊池英明・前川喜久雄 (2006) 「韻律情報」『日本語話し言葉コーパスの構築法』(国立国語研究所報告 124), 347–453.
- 小磯花絵・伝康晴・前川喜久雄 (2012) 「『日本語話し言葉コーパス』RDBの構築」『第1回コーパス日本語学ワークショップ予稿集』

※ 本研究は萌芽・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。