T2R2 東京科学大学 リサーチリポジトリ
Science Tokyo Research Repository

論文 / 著書情報
Article / Book Information

題目(和文)	コーパスを利用した自発音声の研究
Title(English)	
著者(和文)	前川喜久雄
Author(English)	Kikuo Maekawa
出典(和文)	学位:博士(学術), 学位授与機関:東京工業大学, 報告番号:乙第4051号, 授与年月日:2011年3月31日, 学位の種別:論文博士, 審査員:古井 貞熙
Citation(English)	Degree:Doctor (Academic), Conferring organization: Tokyo Institute of Technology, Report number:乙第4051号, Conferred date:2011/3/31, Degree Type:Thesis doctor, Examiner:
学位種別(和文)	博士論文
Type(English)	Doctoral Thesis

Powered by T2R2 (Science Tokyo Research Repository)

学位論文

コーパスを利用した自発音声の研究

前川 喜久雄

2011 年 3 月

目次

第 I 部 1
第 1 章．導入 3
1.1 本研究の目的 3
1.2 自発音声と朗読音声 3
1.3 自発音声研究の必要性 4
1.4 自発音声研究の方法 5
1.5 本研究の構成 6
第 2 章．『日本語話し言葉コーパス』 8
2.1 はじめに 8
2.2 CSJ の設計と実装 9
2.2.1 設計の基本方針 9
2.2.2 CSJ の構造 10
2.2.3 対象とする音声 11
2.2.4 音声のレジスターと量 12
A. 学会講演と模擬講演 12
B. 対話と朗読 13
2.2.5 音声の自発性についての覚書 14
A. 自発性の序列 14
B. 印象評定 15
2.3 研究用付加情報（アノテーション） 15
2.3.1 転記テキスト 15
2.3.2 形態論情報 17
2.3.3 節単位情報 18
2.3.4 印象評定データ 20
2.3.5 その他のアノテーション 21
2.4 CSJ の特性 22
2.4.1 話者の分布 22
2.4.2 内省と行動の乖離 25
2.4.3 発話速度の分布 26
2.4.4 語形の変異 28
2.5 結論 30
第 3 章．X-JToBI 31
3.1 問題 31
3.2 X-JToBI の構造 31

3.2.1 単語層 31
3.2.2 分節音層 31
3.2.3 トーン層 32
3.2.4 BI 層 32
3.2.5 プロミネンス層 32
3.2.6 注釈層 32
3.3 単語層のラベリング 34
3.4 トーン層のラベリング 34
3.4.1 アクセント 35
3.4.2 句頭音調 36
3.4.3 句頭境界音調 36
3.4.4 句末境界音調 37
A. 単純境界音調 37
B. 複合境界音調（BPM） 38
3.4.5 ポインター 38
3.4.6 エクステンダー 38
3.4.7 フィラーのトーンラベル 39
3.5 BI 層のラベリング 39
3.5.1 中間値 39
3.5.2 非流暢性 40
3.6 プロミネンス層のラベリング 43
3.7 注釈層のラベリング 46
3.8 非流暢性の BI ラベリングについての補遺 49
3.8.1 フィラー 49
3.8.2 語断片 50
3.9 分節音ラベル 51
3.10 ラベリング精度 54
3.11 結論 55
第Ⅱ部 57
第 4 章．母音の無声化 59
4.1 問題 59
4.2 母音無声化の概観 60
4.3 典型的環境における無声化の問題点 63
4.3.1 子音調音様式の影響 64
4.3.2 促音の影響 65
4.3.3 無声化連鎖の影響 66

4.3.4 韻律の影響 68
4.4 議論 68
4.4.1 先行研究 68
4.4.2 モーラ境界知覚の容易性 69
4.4.3 促音と無声化連鎖の説明 70
4.5 コーパスの再分析 70
4.6 結論 72
第 5 章. /z/の調音様式 74
5.1 問題 74
5.2 先行研究 74
5.3 分析 75
5.3.1 データ 75
5.3.2 X-JToBI ラベルによる/z/の変異の表現 75
5.3.3 音素の位置に関する分析 76
A. 短単位中のモーラ位置の分析 76
B. 長単位冒頭の短単位の分析 76
C. アクセント句冒頭の短単位の分析 77
5.3.4 近傍分節音に関する分析 78
A. 後続母音の影響 78
B. 拗音の影響 78
C. 直前の分節音の影響 78
5.3.5 韻律的特徴に関する分析 79
A. アクセント 79
B. 直前韻律境界の BI 80
C. 直前のポーズ 80
5.3.6 言語外的特徴に関する分析 80
A. 講演の種別 81
B. 話者の性別 81
C. 話者の年齢 81
D. 発話速度 81
5.4 議論 81
5.4.1 分析のまとめ 81
5.4.2 仮説 82
5.4.3 Time allotted for consonant articulation 83
5.5 TACA による再分析 85
5.5.1 位置 85

5.5.2 分節音 88
A. 後続母音 88
B. 先行母音 88
C. 拗音 88
5.5.3 アクセント 89
5.5.4 言語外的特徴 89
A. 話者の性別とレジスター 89
B. 話者の年齢 91
C. レジスター 91
D. 発話速度 92
5.5.5 話者の個人差 92
5.6 統計的予測 93
5.6.1 決定木 93
5.6.2 交互作用の分析 95
5.7 結論 96
第 6 章．有声破裂音における閉鎖調音の弱化 97
6.1 問題 97
6.2 破裂音弱化の判定 97
6.3 分析 99
6.3.1 位置との関係 99
6.3.2 先行分節音との関係 100
6.3.3 先行 BI 及びポーズとの関係 101
6.3.4 外的要因 102
6.4 議論 103
6.4.1 TACA の導入 103
6.4.2 /g/の特異性 105
6.4.3 TACA による分析 106
6.4.4 例外の検討 107
6.5 統計的予測 108
6.6 音韻体系との関係 110
6.7 結論 112
第Ⅲ部 115
第 7 章．助詞のアクセントの独立性 117
7.1 問題 117
7.2 データ 118
7.2.1 『日本語話し言葉コーパス』 118

7.2.2 助詞の選択 118
7.2.3 音声環境の選択 119
7.2.4 助詞アクセントの聞きとり 119
7.3 分析 120
7.3.1 助詞の意味特性の影響 120
7.3.2 核間距離の影響 122
7.3.3 句末音調の影響 123
7.3.4 先行語の品詞の影響 124
7.3.5 レジスターの影響 126
7.4 考察 127
7.4.1 とりたての助詞 128
7.4.2 核間距離 128
7.4.3 句末音調 129
7.4.4 あらたまり度 129
7.5 結論 129
第 8 章．Final Lowering 131
8.1 問題 131
8.2 データ 133
8.2.1 Fo 値 133
8.2.2 節単位情報(CBL) 135
8.3 分析 137
8.3.1 節単位情報とレジスター 137
8.3.2 アクセント句の位置 138
8.3.3 Final Lowering の領域 140
8.4 結論 143
第 9 章．Boundary Pitch Movements 144
9.1 問題 144
9.2 日本語の BPM 144
9.2.1 日本語 BPM の概観 144
9.2.2 上昇及び上昇下降 BPM の変異形 146
9.3 データ 148
9.3.1 Fo データ 148
9.3.2 印象評定データ 148
9.4 発話スタイルに関する分析 149
9.4.1 レジスターによる BPM の分布の相違 149
9.4.2 印象評定データとの関係における H%と HL%の相違 149

9.4.3 印象評定データとの関係における HL%と PNLP の相違 152
9.5 談話機能に関する分析 153
9.5.1 ABS 区分発話 153
9.5.2 生起頻度の計算 154
9.6 音声的形態の分析 157
9.6.1 上昇下降タイミングの分析 157
9.6.2 ピーク Fo の分析 158
9.7 議論 159
9.7.1 スタイル依存性 159
9.7.2 境界表示機能 160
9.8 結論 162
第 10 章．韻律的レジスター 163
10.1 問題 163
10.2 データ 165
10.2.1 X-JToBI データ 165
10.2.2 追加データ 166
10.2.3 データの正規化 167
10.3 データ全体の分析 167
10.3.1 X-JToBI ラベルデータを用いた判別分析 167
10.3.2 X-JToBI データと追加データによる判別分析 169
10.3.3 データ量と発話中の位置の効果 170
10.4 個々の変数の分析 172
10.4.1 統計的検定 172
10.4.2 変数選択付判別分析 176
10.5 決定木 176
10.6 議論 177
10.6.1 レジスターを特徴付ける韻律特徴 177
A. 学会講演（APS） 178
B. 対話 178
C. 模擬講演（SPS） 178
D. 再朗読 179
10.6.2 X-JToBI ラベルの有効性 179
10.6.3 韻律的レジスターと発話スタイル 180
10.7 結論 181
第Ⅳ部 183
第 11 章．結語 185

11.1 本研究のまとめ 185
11.1.1 コーパスの設計と実装 185
11.1.2 分節的特徴の分析 185
11.1.3 韻律的特徴の分析 186
11.2 得られた成果の分類 186
11.3 CSJ の評価 187
11.4 今後の課題 188
謝辞 190
参考文献 193
発表文献 201

図目次

図 2-1: CSJ における研究用情報の付与方式.....10
図 2-2: 模擬講演の転記テキストの例.....16
図 2-3: 学会講演話者と模擬講演話者の生年による分布（延べ）..... 22
図 2-4: 学会講演話者と模擬講演話者の生年による分布（異なり）..... 23
図 2-5: 話者の出生地の分布（延べ）..... 24
図 2-6: 話者の出生地の分布（異なり）..... 24
図 2-7: 生年代による「コラレル」と「コレル」の分布..... 25
図 2-8: 朗読音声と自発音声の発話速度の比較..... 27
図 2-9: CSJ 内部での発話速度の変動..... 27
図 2-10: CSJ データのレジスターと変異形平均出現率.....30
図 3-1: X-JToBI ラベリングの例..... 33
図 3-2: 頂点と屈曲点..... 35
図 3-3: PB の使用例..... 43
図 3-4: EUAP の適用例..... 46
図 3-5: 句頭が高句始まるアクセント句の例（C の場合）..... 47
図 3-6: QQ の適用例.....48
図 4-1: 無声化連鎖環境中の 2 母音の無声化率..... 67
図 4-2: 子音セグメント間の音響距離と無声化率の関係..... 72
図 5-1: 破擦音(左)及び摩擦音(右)として実現された/z/..... 76
図 5-2: 3 種類の TACA と破擦率の関係..... 84
図 5-3 短単位の語頭及び語中における TACA と破擦率の関係..... 86
図 5-4: 長単位の語頭及び語中における TACA と破擦率の関係..... 87
図 5-5: アクセント句の句頭及び句中における TACA と破擦率の関係..... 87
図 5-6: TACA と破擦率の関係における性差.....90
図 5-7: 破擦率, TACA, 発話速度とレジスターの関係.....91
図 5-8: TACA と破擦率の関係に対する発話速度の影響..... 92
図 5-9: 平均破擦率の個人差..... 93
図 5-10: CART アルゴリズムによる決定木..... 94
図 5-11: 話者グループごとの平均破擦率..... 95
図 6-1: 音素/b/の変異と X-JToBI ラベル..... 98
図 6-2: TACA 及び子音持続時間長と RSA..... 104
図 6-3: /g/における TACA-RSA 関係.....105

図 6-4: 発話速度と TACA 107
図 6-5: 話者の生年代による平均発話速度の変化 108
図 6-6: 日本語有声破裂音・有声破擦音・鼻音の分布.....111
図 6-7: /b/, /d/, /g/及び/z/における TACA-RSA 関係.....111
図 7-1: 判定結果の助詞ごとの比率（KM と YI の平均値） 121
図 7-2: 核間距離との関係 122
図 7-3: 助詞ごとにみた核間距離の影響（判定 1 の百分率） 123
図 7-4: 学会講演と模擬講演の差 126
図 8-1: Final lowering の模式図..... 131
図 8-2: Poser と P&B における FL の扱い.....132
図 8-3: F0 測定点の名称.....133
図 8-4: 節単位情報とアクセント核の有無による平均 FLT 値の変動 136
図 8-5: レジスターと節単位情報による平均 FLT の変動 137
図 8-6: 発話冒頭からの距離による FLT と節単位情報の関係の再分析.....139
図 8-7: 有核アクセント句連鎖中の F0 測定点の平均値.....142
図 9-1: 日本語（東京方言）の BPM 145
図 9-2: 通常の HL%と PNLP の比較 147
図 9-3: 印象評定データと H%及び HL%生起率 150
図 9-4: 印象評定データと HL%及び PNLP 生起率 150
図 9-5: ABS 区分発話の概念図..... 153
図 9-6: 発話長と句末イントネーションの平均生起数(頻度 1 以上)の関係 154
図 9-7: ABS 区分発話における PNLP の位置別生起率(5AP から 10AP まで)..... 155
図 9-8: ABS 区分発話における PNLP の位置別生起率(11AP から 15AP まで)..... 155
図 9-9: ABS 区分発話における HL%の位置別生起率 156
図 9-10: ABS 区分発話における H%の位置別生起率..... 156
図 10-1: CSJ-Core の全 201 講演の判別空間上の分布 169
図 10-2: 累積的な判別分析の結果 170
図 10-3: 60 秒幅のデータ窓を利用した判別分析の結果..... 171
図 10-4: 主要な独立変数のレジスターによる変動 174
図 10-5: 変数選択後の 9 変数を用いて構成された決定木..... 177

表目次

表 2-1: 模擬講演のテーマ 13
表 2-2: CSJ に格納された音声のレジスターとその内訳 13
表 2-3: 形態論的単位数 18
表 2-4: CBAP-csj が検出する 49 種類の節境界ラベル 19
表 2-5:　話者の性別の分布（延べ） 23
表 2-6: 話者の性別の分布（異なり） 23
表 2-7: 変異形（非代表形）の比率が高い語彙素上位 20 語 28
表 2-8: 上位 3 位までの変異形によるカバー率 29
表 3-1: X-JToBI のトーンラベル 34
表 3-2: X-JToBI のトーン層補助記号 34
表 3-3: 句末境界音調のラベリングに使用されるラベル 37
表 3-4: X-JToBI の BI ラベル（その１：中間値） 40
表 3-5: X-JToBI の BI ラベル（その２：非流暢性など） 40
表 3-6: X-JToBI のプロミネンスラベル 43
表 3-7: X-JToBI の注釈層ラベル 46
表 3-8: X-JToBI 分節音層母音ラベル 51
表 3-9: X-JToBI 分節音層子音ラベル 52
表 3-10: X-JToBI 分節音層補助ラベル 53
表 3-11: 現代日本語のモーラと CSJ 分節音ラベルの対応表 53
表 4-1: 本章で利用する分節音ラベルの整理 61
表 4-2: データにおける各種母音の頻度と無声化率 62
表 4-3 : 先行子音(C1)と後続子音(C2)と単母音無声化率の関係 62
表 4-4: /CoVnCo/環境における/i/の無声化率 63
表 4-5: /CoVnCo/環境における/u/の無声化率 64
表 4-6: /CoVnCo/環境における/i/の無声化率(%)に及ぼす子音の調音様式の影響 65
表 4-7: /CoVnCo/環境における/u/の無声化率(%)に及ぼす子音の調音様式の影響 65
表 4-8: 後続促音の有無が無声化に及ぼす影響 66
表 4-9: 無声化連鎖環境中の最初の２母音における声の分布 67
表 4-10: 無声化連鎖中の第 1 モーラと通常の無声化の無声化率(%)の比較 68
表 4-11: トーンの影響 68
表 4-12: 母音/i/の無声化率(%) 71
表 4-13: 母音/u/の無声化率(%) 71

表 5-1: 本章で利用したデータにおける話者の分布 75
表 5-2: 短単位中のモーラ位置の影響 77
表 5-3: 長単位冒頭の短単位におけるモーラ位置の影響 77
表 5-4: アクセント句冒頭の短単位におけるモーラ位置の影響 77
表 5-5: 直後の分節音の影響 78
表 5-6: 拗音化（音韻的口蓋化）の影響 78
表 5-7: 拗音における後続母音の影響 78
表 5-8: 直前の分節音の影響 79
表 5-9: アクセントの影響 79
表 5-10: 直前の韻律境界の影響 80
表 5-11: 直前のポーズの影響 80
表 5-12: 話者の生年代の影響 81
表 5-13: 発話速度の影響 81
表 5-14: TACA による破擦率の予測 85
表 5-15: 話者の性別による変動 90
表 5-16: 話者の性別とレジスターによる変動 90
表 5-17: レジスターによる分節的特徴の分布の差 91
表 5-18: 交互作用項を含むロジスティック回帰分析の結果 95
表 6-1: 各種言語単位中の位置による RSA（閉鎖率）と TACA の変動 100
表 6-2: 先行分節音による RSA と TACA の変動 101
表 6-3: 先行短単位境界 BI による RSA と TACA の変動 101
表 6-4: 先行ポーズの有無による RSA と TACA の変動 102
表 6-5: 発話速度による RSA と TACA の変動 102
表 6-6: 話者の性別による RSA と TACA の変動 103
表 6-7: 話者の生年代による RSA と TACA の変動 103
表 6-8: ロジスティック回帰分析の結果 109
表 6-9: 有声破裂音弱化の予測精度 110
表 7-1: 分析対象データのサンプル数 118
表 7-2: 判定者２名の判定結果 120
表 7-3: 判定結果の一致度 120
表 7-4: 句末音調との関係 123
表 7-5: 先行語の品詞との関係 125
表 7-6: 助詞と先行語品詞のクロス集計 125
表 7-7: 助詞の意味特性と助詞連鎖の関係 126

表 7-8: 助詞の意味特性とレジスターの関係 127
表 8-1: 節単位情報 134
表 8-2: FLT の回帰分析（女声データ） 140
表 8-3: FLT の回帰分析（男声データ） 140
表 8-4: 有核アクセント句のみからなる発話の数 141
表 9-1: CSJ-Core における主要な BPM の生起頻度 146
表 9-2: 印象評定データ全評価語対の H% ないし HL% 生起率との相関 151
表 9-3:印象評定データ全評価語対の H% ないし PNLP 生起率との相関 152
表 9-4: FLT と FHT のタイミング 158
表 9-5: FHT の平均 F0 値 159
表 10-1: CSJ-Core における講演と話者の分布 165
表 10-2: 主要な X-JToBI ラベルとその頻度 166
表 10-3: X-JToBI ラベルデータだけを利用した判別分析の結果 167
表 10-4: X-JToBI ラベルの単独及び組み合わせによる判別分析の結果 168
表 10-5: 追加データの効果 168
表 10-6: 個々の変数の Kruskal-Wallis 検定結果 173
表 10-7: 決定木による分類の結果 177

概　要

自発音声は音声研究の本来的な研究対象であるにもかかわらず、従来の音声学は自発音声の研究を避けて通ってきた。本研究の目標は『日本語話し言葉コーパス』（CSJ)の分析を通して、音声情報のアノテーションが施された大規模コーパスがあれば、自発音声の研究が可能になることを示すことにある。第 1 章で自発音声研究の問題点を指摘し、第 2 章で CSJ の設計と実装について説明した後、第 3 章では第 4 章以降の分析で重要な役割を果たすことになる日本語自発音声用の韻律ラベリング体系である X-JToBI を詳しく説明した。

その後、第 4, 5, 6 章では日本語の分節的特徴について、また第 7, 8, 9, 10 章では日本語の韻律的特徴について、CSJ の分析によって得られた知見を報告した。これらの章では、朗読音声を対象とした従来の音声研究において定説とされてきた分析が自発音声の分析によって覆された事例や、朗読音声の研究では未解決であった問題が自発音声の分析によって解決された事例が多数提示されている。

以上の研究によって、CSJ のように豊富なアノテーションが施された自発音声コーパスがあれば自発音声の音声学的研究が実際に可能となり、音声学研究の新しい可能性が拓けることを示した。

第Ⅰ部

第 1 章. 導入

1.1 本研究の目的

人文科学の領域で行なわれる音声学においても、また自然科学的な研究をふくめた音声研究全般においても、20 世紀の音声研究はいくつかの研究対象を慎重に避けて通ってきたように思われる。そのなかには、話し手の発話意図や心的態度などについての研究、いわゆるパラ言語情報の研究(Fujisaki 1996, 前川・北川 2002, Maekawa2004a)と、発話の形式が発話時に即時的に生成される自発音声の研究（前川 2001, 2005）とが含まれる。

このうちパラ言語情報は、近年、感情音声の研究の一環として盛んに研究されるようになってきた。一方、自発音声の研究は、少なくとも量的に見るかぎり、現在も音声研究の主要なテーマとはなっていない。このことは音声研究の代表的な国際会議(ICPhS, INTERSPEECH 等)のセッション構成をみるとよくわかる。Paralanguage, emotional speech などを標榜するセッションはどの会議にも設けられており、最近では複数のセッションが設置されていることも稀でない。一方、spontaneous speech のセッションははるかに稀である。

本研究は、このような現状を打開するために、日本語を対象として自発音声の音声学的研究の可能性を示そうとするものである。

1.2 自発音声と朗読音声

自発音声(spontaneous speech)とは、発話の言語形式が発話生成時に即時的に生成されるタイプの音声である。発話の内容は事前に決まっていてもいなくてもよい。発話内容が決まっていても、それを具体的な言語形式にエンコードする方法には非常に多く―おそらくは無限―の可能性が存在するので、その選択が発話生成時に行なわれるならば、それは自発音声であると考えることができる。

自発音声の対立概念は朗読音声(read speech)もしくは実験音声(laboratory speech)である。発話の言語形式が発話開始時に予め決定されていれば、それは朗読音声である。当然ながら朗読音声では、発話内容も予め決定されている。従来行なわれてきた音声研究は、ほぼすべてが朗読音声ないし実験音声の研究である。上記の規定に従えば、言語学者や方言学者が隣地調査で記録してきた音声もまた大部分が朗読音声である。

自発音声と朗読音声とでは発話のプラニングに要する時間が大幅に異なると考えられている(Blaauw, 1995)。また自発音声にはフィラー（filled pauses 言い淀み）や言い直し(false start)、言い直しによって生じる語断片などの非流暢性要素(disfluency)が多数生じることが知られている。自発音声にこれらの要素が生じる原因の一端は、発話の進行中に発話プランが生成されたり、変更されたり、あるいは破綻をきたすことなどに求められる。朗読音声の場合は、発話形式が事前に定まっているために、発話プランの変更や破綻は基本的に生じにくい。

さらに、自発音声が特定の聴き手の存在を前提として生成され、その結果、音声特徴中に聴き手の特性が反映されている（大人に対する発話と子供に対する発話を比較せよ）のに対し、朗読音声は聴き手を特定することなしに生成することが可能であることも両者の基本的な相違点として指摘できる。

自発音声と朗読音声は現実においては常に範疇的に対立しているわけではない。俳優が脚本に記されたセリフにアドリブのセリフを混ぜることがあるように、両者の中間形態の発話が存在する。しかし従来の音声研究（特に実験研究）が対象としてきたのはほぼ純粋な朗読音声であるから、本研究では、多少とも自発的である音声をまとめて自発音声と呼ぶことにする（自発音声については 2.2.5 節の議論も参照）。

1.3 自発音声研究の必要性

自発音声と朗読音声のどちらが音声研究の本来的な対象かと言えば、それが自発音声であることには論をまたない。量的な観点からすれば、我々が日常生成したり知覚したりしている音声の大部分は自発音声である。また工学的応用の観点から考えても、自発音声を認識できない音声処理システムや音声対話システムの実用性は極めて限られているといわざるを得ない。

それにもかかわらず、従来の音声研究は、音声学的な研究にあっても音声情報処理的な研究にあっても、朗読音声を対象として行なわれてきた。それは何故か。詳しく検討すれば、様々な理由があっただろうが、基本的には、いきなり複雑な対象にとりくむのではなく、まずは理想化された単純な対象に取り組むという科学研究の常識に従った判断だったと考えられる。この判断が正しかったことは、20 世紀後半における音声研究の著しい発展によって明らかである。しかし、そこにはやはり問題もあった。

一体、朗読音声の研究をひたすら充実させてゆけば、自発音声の特性も解明されるかと言えば、そうではない。朗読音声の研究成果を外挿して自発音声を理解することに困難が生じる事例は以下のように類型化できると思われる。

第一に、自発音声にはごく自然に生じるが、朗読音声を用いて研究することが非常に困難な現象がある。その代表は上述の非流暢性要素であるが、それに準ずるものに、本研究の第 9 章でとりあげる句末イントネーション（BPM）がある。BPM は日本語の韻律的特徴のうち最も耳につきやすい特徴であるが、川上(1963)が夙に明らかにしたように、上昇調に限っても少なくとも 5 種類の変異形があり、それぞれ異なる語用論的意味を伝達している。しかし、そのような差異の存在が話者の意識にのぼることはほとんどなく、これを実験環境下で再現することは、一般の話者はおろか、言語学を専門とする研究者にあっても非常に困難である。第 10 章で扱う韻律レジスターの問題、すなわち一般に「口調」として理解されているような、発話のタイプに相関した韻律特徴の変化もこのタイプに属する現象である。

もちろん話者に種々の指示を与えて、BPM や韻律レジスターを再現させることは不可能ではなく、実際にそのような研究も一部で行われている。しかし、そのようにして収録され

た音声が、言語運用の本来の姿を伝えているとは限らず、最終的には、やはり自然に生成された自発音声にあたって確認することが欠かせない。つまりこのタイプに属する音声行動は無意識的に実行されるものであり、それを意識的に実行すると「わざとらしい」音声になってしまうということである。

第二に、自発音声にも朗読音声にも生じるが、両者で生じ方が異なっている音声現象がある。特に問題となるのは、朗読音声においては規則的に生じると見える現象に自発音声において何らかの逸脱が生じる場合である。本研究の第 4 章で扱う母音の無声化現象と第 7 章で扱う 2 モーラ有核助詞の独立性の問題とがこのタイプに属する問題である。自発音声における「規則」からの逸脱は、自発音声が統制を受けていないために生じる誤差の増大として処理されてしまうことが多いが、本稿で紹介する現象についてはその様な解釈は不適当であり、むしろ従来の規則が朗読音声に束縛されたものであることが示唆される。

第三に、朗読音声に生じる現象であるが、朗読音声を研究しているだけでは正しく理解することが困難な現象がある。本研究では第 5, 6 章で扱う子音調音の変異と第 8 章で扱う final lowering と呼ばれる韻律現象がこのタイプに該当する。この種の現象に共通しているのは、研究の前提となる基本概念（例えば「語」や「発話」）の吟味が不十分であったり、あるいは概念的にすぎて言語運用の実情にそぐわないものであったりしていることである。

また、問題となる変異に本来複数の要因が影響を及ぼしているにもかかわらず、従来の研究ではそのうちのひとつだけに焦点が当てられてしまっていることも、このタイプの現象に共通した問題である。そのような場合、特に想定される要因数が非常に多い場合には実験を設計することが困難であるから、やはり大量の自発音声を分析することが有効な研究手段になる。

以上を要するに、自発音声の研究によらなければ解明することができない音声現象が少なくない。音声による情報伝達の全体像を理解するためには、それらの現象の解明が欠かせない以上、音声研究の健全な発展のために我々は自発音声の研究法を考案しなければならない。

1.4 自発音声研究の方法

20 世紀の音声学は、種々の言語研究のなかで最も早く自然科学の思考法を採用して実験科学を志向してきた。しかし、自発音声には実験計画を適用することができない。自発音声の収録においては、自発音声の定義上、刺激の統制が許されないからである。そのため、自発音声は実験研究の対象から除外されてしまうのである。

このような対象に対しては、大量のデータを統計的に分析してデータに内在する構造を把握する、いわゆるデータマイニング的なアプローチが有効だと考えられる（前川 2005）。ただし、そのような研究を実行するためには、二つの要件が満たされなければならない。

そのひとつは、大規模な自発音声データが存在しており、そのデータに分析遂行上必要なアノテーションが施されていること、つまり自発音声コーパスの存在である。自発音声を録

音することは簡単だが、単に録音されただけの音声は、たとえそれがデジタル化されていても検索することができず、従って大量に分析することもできない。

もうひとつの要件は、大規模音声データを効率的に分析することのできるコンピュータ環境である。実際上すべてのデータマイニング手法はコンピュータの利用を前提としているし、また、アノテーション情報を活用して特定の条件を満たすデータセットを抽出する作業においても SQL などで複雑なプログラムを組める高機能なデータベースが必要になる。

このような、自発音声研究を支えるインフラの整備が実現したのは、比較的近年のことである。わけても日本語のアノテーション付自発音声コーパスが構築されたのは、筆者らによる『日本語話し言葉コーパス』(CSJ)の構築（Maekawa et al.2000, Maekawa 2003,前川 2004）が最初である。

CSJ は 2004 年 6 月に一般公開されたが、それ以来、本稿執筆時までに、少なくとも 500 件以上の研究論文が発表されている。これまでのところ、CSJ による音声研究の成果は専ら音声情報処理の領域において蓄積されてきており、（Furui 2005, Furui et al. 2005, 南條・加藤・李・河原 2003, 秋田・河原 2005, Akita and Kawahara 2010 等）、その成果は音声ドキュメント処理（秋葉他 2008, 伊藤・堀 2009, 中川 2009）や国会会議録の自動作成（秋田・三村・河原 2010）などに応用されている。

これらの研究によって、CSJ のようなアノテーション付コーパスがあれば自発音声が科学的研究の対象となりうることは既に証明されていると言ってよい。しかし、現代の音声認識技術は統計的なパタン認識の技術であるため、その成果は、分節音や韻律の特徴を個別に、また主に記号ベースで記述しようとする音声学にとって、直截的に解釈することが難しい。音声学のためには、やはり音声学の観点から、伝統的な音声学の成果と対比する形で自発音声を検討することが必要であり、本研究のめざすところもそこにある。

1.5 本研究の構成

以下、本研究では、CSJ を活用して、日本語音声の様々な問題を音声学の立場から検討する。コーパスの分析によって、自発音声の研究が可能になることを示すことが本研究のひとつの目的であるから、意図的に、多岐にわたる現象をとりあげている。

本研究は本章を含めて 4 部、11 章から構成されている。第 I 部を構成する第 1 章から第 3 章までは、研究の前提を扱っており、第 2 章は、本研究全体が立脚することになる『日本語話し言葉コーパス』(CSJ)の設計と実装についての解説である。本研究では CSJ のうち特に豊富なアノテーションが施された CSJ-Core を主に分析することになるが、第 2 章では CSJ の全体を解説している。第 3 章では、CSJ-Core における最も重要なアノテーションである X-JToBI の仕様の解説である。

第Ⅱ部を構成する第 4 章から第 6 章は、日本語の分節的特徴についての研究であり、本研究の第２部として位置づけられる。第 4 章では、母音の無声化にかかわる問題を検討して、無声化が例外的に回避される現象のメカニズムを解明する。

第 5 章ではザ行子音/z/が有声摩擦音として実現されるか有声破擦音となるかの変異を分析して、現在広く受け入れられている定説を覆す。第 6 章は、第 5 章と同一の手法で、有声破裂音/b, d, g/における閉鎖調音の弱化現象を分析することによって 5 章の結論を強化することを狙った研究である。

第Ⅲ部にあたる第 7 章から第 10 章は、韻律的特徴についての研究であり、本研究の第 3 部として位置づけられる。第 7 章はアクセントを扱っており、教科書や辞典類に広く記載されている助詞アクセントに関する法則が自発音声では必ずしも成立していないことを明らかにする。

第 8 章と第 9 章はイントネーションの研究である。第 8 章では final lowering と呼ばれる現象を分析する。従来、この現象が自発音声に存在するか、また存在するとして、その作用域の広さがどの程度かについて研究者の意見に一致がみられていない。本研究はこれらの問題に対する明確な解答を提供する。

第 9 章は、日本語のイントネーションのうち最も顕著な特徴である句末イントネーション(BPM)の分析である。BPM の選択は発話のスタイルと強く相関していることと、BPM のなかには談話上の単位の境界を表示するものもあることを明らかにする。

第 10 章では、話し言葉のレジスターの相違が韻律的特徴にどのように反映しているかを X-JToBI ラベルの統計分析によって検討し、あわせて J_ToBI から X-JToBI への拡張の有効性を検討する。従来、発話の韻律は発話の状況や内容によって大きく変化することが広く信じられてきているが、それを客観的に解明した研究は行なわれていなかった。

最後に第Ⅳ部で本研究の結論を述べる。

以上の構成からわかるように、本研究では非流暢性要素の分析は行っていない。これは本研究が全体として従来の音声研究の成果と自発音声研究の成果とを対比することを意図しているためであり、非流暢性要素の重要性を低く評価しているわけではない。本研究で非流暢性要素をとりあげていないのは、従来の朗読音声研究の成果と CSJ を用いた自発音声研究の成果を対比することが研究全体の眼目となっているからである。CSJ に記録された非流暢性要素については Watanabe (2009)が言語心理学的な観点から組織的な研究を実施しており、参考になる。

第 2 章. 『日本語話し言葉コーパス』

2.1 はじめに

書き言葉と話し言葉の研究を比較すると、話し言葉の研究には何かと制約が多い。書き言葉のテキストは、電子的手段で作成されたものであれば、ほぼそのまま研究の一次資料として利用できる。さらに、テキストを語に分割して品詞情報を付与することも、現在ではかなり高い精度で自動実行することができるようになっている。

これに比べて話し言葉の研究では、録音した音声を文字に転記する作業がまず大変な手間を要求する。しかも－この点が重要なのだが－、ただ単に音声を文字に転記しただけではイントネーションやポーズなどの韻律的特徴が脱落してしまうので、理想的にはこれらの情報まで含めた転記が必要になる。そうしないと、或る発話が断定なのか質問なのか、発話のどこに強調が置かれているかといった基本的な情報すらも分明でなくなることがありうる。韻律情報（正確にはパラ言語情報）は、話し言葉と書き言葉の本質的な相違点にかかわる情報である（前川 1999, 2008, 前川・北川 2002）。

さらに、言い誤りや言い淀みのような現象も転記が必要である。これらの現象は会議録などの書き起こしでは省略されるのが普通であるが、言語心理学的な研究のためには、こうした非流暢性の要素が重要であることがわかっている。そのため転記テキストは一層複雑化し、作成コストが増大する。話し言葉研究用データのコスト高は、話し言葉の研究が書き言葉にくらべて著しく立ち遅れていることの最大の理由のひとつであろう。

国立国語研究所は 1948 年の創立以来多くの調査研究を実施してきているが、やはり、その大部分は書き言葉を対象とした調査であった。そのなかで『談話語の実態』（国語研 1955）と『話しことばの文型』（国語研 1960, 1963）の報告書にまとめられた調査は、自発音声を含む話し言葉を正面きってとりあげた研究として異彩をはなっており、現在でも引用されることが少なくない。しかし『話しことばの文型』以降は、自発音声ないし話し言葉の特色を解明するための研究は国立国語研究所の公式な研究課題から姿を消してしまうことになった 。本章で紹介する『日本語話し言葉コーパス』は、この話し言葉調査の系譜を現在に蘇らせる試みである。

『日本語話し言葉コーパス』（以下ではその英語正式名称である Corpus of Spontaneous Japanese を略して CSJ と呼ぶ）は国立国語研究所、情報通信研究機構（旧通信総合研究所）、東京工業大学の三者が共同開発した現代日本語の話し言葉研究用データベースであり、プロジェクトの総括責任者は東京工業大学の古井貞熙教授であった。開発費用の多くは科学技術振興調整費開放的融合研究制度補助金に拠った。研究課題名は「話し言葉の言語的・パラ言語的構造の解明に基づく『話し言葉工学』の構築」、研究期間は 1999-2003 年度であった（古井・前川・井佐原 2000）。

CSJ には時間にして約 660 時間、語数にして 750 万語以上の話し言葉が格納されている。上述した『談話語の実態』で分析された録音資料が約 9 時間分であることと較べれば、CSJ

のサイズが理解できる。CSJ は日本語の音声データベースとして最大であるだけでなく、世界の主要音声データベースと比較しても遜色がない。研究用に付加された情報の多様性と精度の高さにおいては、むしろ諸外国のデータベースを凌駕している。データベースの価値が、そのデータ量と付加情報の多様性の積で決まるとすれば、明らかに現時点における世界最高の音声言語データベースである。CSJ は、2004 年 6 月に一般公開されている。CSJ の設計と実装については多くの文献が発表されているが（Maekawa et al, 2000, 前川他 2000, 前川 2004）、最終的な仕様の大部分は国語研究所(2006)に記述されている。以下では本研究との関係において CSJ の仕様を紹介する。

2.2 CSJ の設計と実装

2.2.1 設計の基本方針

音声言語データベースそのものはこれまでにも世界各地で構築されてきている。それらは二種に大別できる。ひとつは、1980 年代から世界中で盛んに構築され始めた音声情報処理用のデータベースである。これは、大量の学習データを用いて音声の自動認識や合成を行なおうとする工学的研究に用いられたもので、その内容は、ほとんどの場合、単語ないし文章を多数の話者が読み上げた朗読音声である。

朗読音声の話者は職業的な朗読者（ナレーターやアナウンサー）であることが多く、当然ながら、誤りのない理想化された音声になっている。音声の他に提供されるのは、朗読用テキストとその音素表記程度であり、韻律情報が提供されることは稀である。

もうひとつは音声学や言語学のために構築されたデータベースである。英国で 1959 年に開始された Survey of English Usage (SEU)のデータがその嚆矢となった（現在は London-Lund Corpus の名で知られている。Svartvik and Quirk, 1980 参照）。SEU は書き言葉と話し言葉の双方を対象とした調査であり、全体の半分、約 50 万語分が話し言葉データにあてられていた。そのうち 76%が独話音声、24%が対話音声である。話し言葉データの大半は、一般話者による、練習無しの自発音声(spontaneous speech)であり、さらに韻律特徴や言い淀み等の情報も付与されているので、非常に付加価値の高いデータであるのだが、残念なことに肝心の音声そのものは提供されていない。そのため、ユーザーは転記テキストに埋め込まれた複雑な音声記号群から音声を想像しなければならい。もちろん音声情報処理に利用することもできない。

もうひとつ、BNC(British National Corpus)の例を挙げておこう。BNC は英語の辞書学やコーパス言語学で広く活用されているデータベースであり、1 億語のうち 1000 万語を話し言葉に充てているが、やはり音声は公開されておらず、音声記号によるアノテーションも与えられていない（Aston and Burnard 1998）。

我々は上に述べた二種類の音声言語データベースそれぞれの特長を CSJ で同時に実現しようと考えた。これは研究立案当初からの方針であった。具体的には、対象を自発音声とし、自発音声の音声認識技術を開発するために必要なデータ量を確保しながら、一方で音声・言

語研究のための付加情報も豊富に提供しようというよくばった設計方針である（Maekawa et al. 2000, 前川他 2000, Maekawa 2003）。

2.2.2 CSJ の構造

一定の研究コストの制約内で上記の設計方針を実現するためには、それなりの工夫が要る。CSJ ではデータベースに一種の階層構造を導入して付加情報に濃淡をつけるという方策を採用した。

図 2-1 に CSJ の階層構造と、研究用付加情報の濃淡を示した。最初に CSJ 全体のサイズを 700 万語（短単位, 2.3.2 節参照）と見積もった 。これは音声認識研究に最低限必要なデータ量を朗読音声の認識研究での知見に依拠して推定したものである。この 700 万語に対しては、音声の他に精密な転記テキストと形態論情報（つまりテキストを語に区切って品詞をつけた情報）を提供する。これらは音声認識研究を実施するために最低限必要な情報である。また、講演音声が聴き手にどのような印象を与えたかを主観的に評定したデータ（印象評定データ）と、転記テキストにおける節(clause)境界の情報（節単位情報）も提供し、さらに話し手に関する種々の属性情報（性別、年齢など）も提供する。

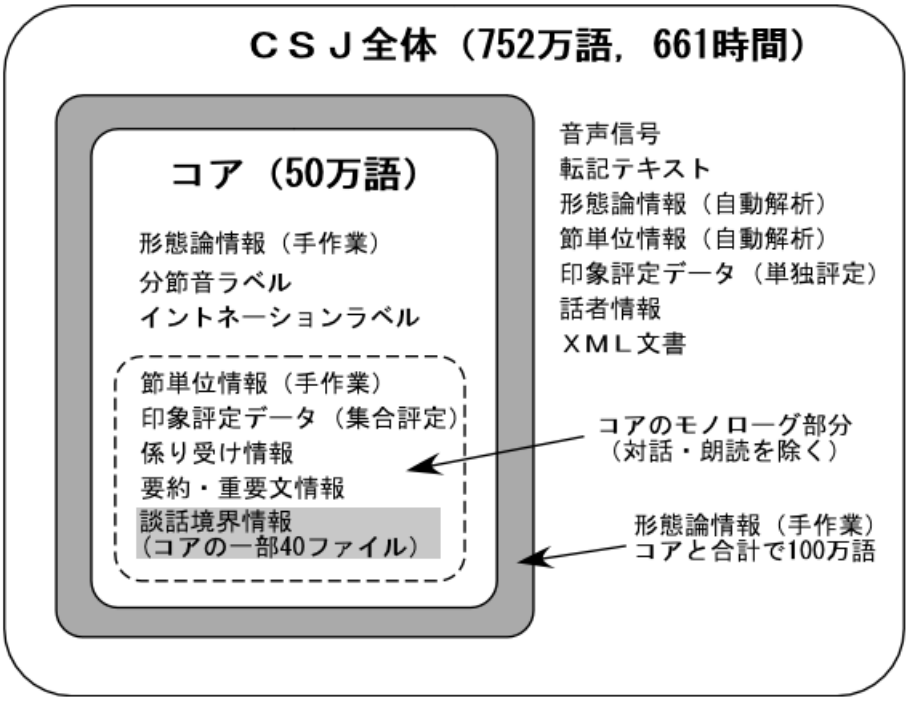


図 2-1: CSJ における研究用情報の付与方式

一方、CSJ の一部約 50 万語に限っては上よりもはるかに豊富な研究用情報を提供することにした。この 50 万語は、データベースの中核部分という意味で「コア」(CSJ-Core)と呼ばれている。50 万語というサイズは、研究コストから逆算して処理可能な最大なデータ量を見積もることによって決定した。コアだけに付与される研究用付加情報は以下のものである。

― 分節音ラベル
― イントネーションラベル
― 印象評定データ（集合評定）
― 係り受け情報
― 要約・重要文情報
― 談話境界情報

形態論情報と節単位情報は CSJ の全体に提供される情報であるが、コアに対しては手作業で綿密な分析を実施しており、コア以外に対するものよりも情報の精度が高い。コアの形態論情報は、コア以外のデータを自動解析するための学習データとして利用されている。

最後に図 2-1 中の網掛けを施した部分は、コア以外であるが、手作業による形態論情報が付与されている部分である。手作業による高精度形態論情報は、コアを含めて約 100 万語に対して付与されている。

2.2.3 対象とする音声

CSJ の対象は自発音声である。しかしひとくちに自発音声といっても実際には多種多様である。まず問題となるのが、独話（モノローグ）と対話（ダイアローグ）の別であるが、CSJ では独話を中心に据えることにした。その理由は、現在の音声認識研究が基本的に独話を対象としているからである。言語研究者のなかには対話にしか興味がないという人もいるようだが、実は日本語の場合、対話のデータベースは少量であっても或る程度整備されているのに対して、自発的な独話のデータベースが存在していないことを考えると、言語研究の観点からも独話データは価値が高いと考えられる。

次に自発性には高低さまざまな段階がある。CSJ では親しい間柄での雑談のように極端に自発性の高い発話は対象とせず、従来研究されてきた朗読音声よりは自発性が高いが、音声だけを聞いても内容が十分に理解できる、まとまった内容をもつ発話を対象に据えることにした。これも工学的応用を考えての選択である（自発性の問題については 2.2.5 節参照）。

また CSJ では、いわゆる標準語を対象とすることにした。標準語という概念を正確に規定することは難しいが、我々は「高校卒業程度の教育をうけた現代人が多少とも公的な場面で用いる日本語で、分節音の音韻特徴及び語彙・文法上の特徴が東京方言に酷似したもの」というやや大雑把な規定によってデータを選別することにした。

この規定は、韻律特徴については何も言及していないので、アクセントが明らかに東京方言とは異なる発話も CSJ には収録されている。ただしコアには韻律特徴のラベルを付与する関係上、韻律特徴が東京式と判断された音声だけを格納している。

2.2.4 音声のレジスターと量

A. 学会講演と模擬講演

2.2.3 節に述べた方針に適う音声のレジスター（発話タイプ）として学会講演と模擬講演を収録することにした。学会講演は、人文、理工、社会の各領域にまたがる様々な学会での研究発表を実況録音した音声である。学会講演は内容が論理的であると期待できるから、上述の音声認識・要約技術が最初に対象としてとりあげるべき種類の音声である。

各学会から承諾をいただいた後に講演者に連絡をとり、データベースが公開されることを承知のうえで承諾書を提出した講演者の口頭発表を収録した。1999 年から 2001 年にかけて収録した学会講演は延べで 987 件に達している（後掲の表 2-2 参照）。

しかし、学会講演の話者には言語学的に見て強い偏りがある。どの学会でも講演者には大学院生が多いため年齢が 20 代半ばから 30 代前半に集中しており、理工系学会では大半が男性である。さらに専門領域ごとに使用語彙の著しい偏りがあることも想像に難くない。さらに学会講演は一般にスタイルの高い発話が多く、少数ではあるが原稿を朗読しているに近い講演もある。

これらの問題を解決するために企画されたのが模擬講演である。人材派遣会社に依頼して年代（20 代から 60 代まで）と性別に偏りのない話者を派遣してもらい、当方で指定した一般的テーマにそった 10 分程度のスピーチを各人に三種類語ってもらった（ただし最初期に収録した一部のデータに関しては話者のバランスがとれておらず、テーマも指定していない）。表 2-1 に指定したテーマのリストを示す。人材派遣会社にはできるだけ首都圏出身の話者を手配するよう依頼したが、この要望は完全には叶えられていない（2.4.1 節参照）。

話者には収録の二日ほど前にテーマを連絡した。話者は収録までに各テーマについて具体的なスピーチを考え、その概要を簡単なアウトラインにまとめてタイトルをつける。例えばテーマ１であれば「大学に合格したこと」、テーマ２であれば「母の死」などである。講演用の朗読原稿を準備することは禁止した 。模擬講演の話者からもデータ公開の承諾書をもらっている。

模擬講演の総数は 1715 件である。初期に収録した一部を除いて、国立国語研究所内の音声スタジオで収録した。模擬講演を収録する目的のひとつは、学会講演よりもくだけたスタイルの発話を収録することにあったから、可能なかぎりリラックスした状態で講演してもらうために工夫をこらした（2.2.5 節参照）。後述する印象評定値及び収録されたデータの予備的分析結果をみると、統計的には模擬講演のスタイルが学会講演よりも低下していることがわかる（前川 2002, Maekawa 2009）。

表 2-1: 模擬講演のテーマ

0 .（指定なし）
1 . 人生を振り返って嬉しかった・楽しかった出来事
2. 人生を振り返って悲しかった・つらかった出来事
3. 住んでいる町や地域について
4. よく知っていること、興味・関心のあることの客観的説明
5. 人生を振り返って印象に残っていること
6. 過去数年の間にマスコミで扱われたニュース
7. 無人島に持っていくもの 3 つ
8. ～のやり方、作り方*
9. ～の歴史*
10. 自分にとっていちばん大事なもの・人
11. 21 世紀に残したいもの・残したくないもの
* ～は話者が選択する

表 2-2: CSJ に格納された音声のレジスターとその内訳

音声のレジスター	タイプ	話者数(異なり)	ファイル数	時間
学会講演	独話	819	987	274.4
模擬講演	独話	** 594	1,715	329.9
その他	独話	*** 16	19	24.1
学会講演インタビュー	対話	* (10)	10	2.1
模擬講演インタビュー	対話	* (16)	16	3.4
課題指向対話	対話	* (16)	16	3.1
自由対話	対話	* (16)	16	3.6
再朗読	朗読	* (16)	16	5.5
朗読	朗読	* (248)	507	15.5
	計	1,417	3,302	661.6

* ( )内は全員が学会講演話者もしくは模擬講演話者としてカウントされている
** 10 名は学会講演話者としてもカウントされている
*** 2名は学会講演話者としてもカウントされている

B. 対話と朗読

表 2-2 に CSJ に収録された音声のレジスターとその内訳を示す。CSJ の 92%（605 時間）は学会講演と模擬講演であるが、それ以外に約 55 時間の音声が収録されている。そのうち「その他」に分類されているのは、一般向けに開催された博物館の連続講演会、専門学校における日本語学の講義、国立国語研究所が開催した一般向け講演会などの独話である。いずれも学術的な講演であるが、話し手が専門家、聴き手が一般人という点で、学会講演とは相違している。

「学会講演インタビュー」から「自由対話」までは合計約 12 時間分の対話音声である。また「朗読」（新書から抜粋した自然科学に関するテキスト 2 種類を模擬講演話者が朗読し

たもの）と「再朗読」（収録済の自発音声の転記テキストを同一話者が朗読した音声）も合計約 21 時間分が収録されている。これらは、CSJ の中心を占める学会講演ないし模擬講演と比較対照して CSJ に格納された独話の性質を評価するために収録したものである。16 名分と量は限られているが、同じ話者が学会講演（10 名のみ）、模擬講演、4 種類の対話、再朗読をおこなったデータも提供されているので、独話と対話の違いなど、音声の種類による影響を同一の話者グループにおいて比較できる。CSJ の 8%程度とはいえ、これらの音声も従来の水準からすれば少なからぬ量が収録されているので、目的によっては独話と切り離して単独で分析することが可能である。

2.2.5 音声の自発性についての覚書

CSJ には種々様々な話し言葉音声が収録されている。これが、工学的応用のために独話（学会講演と模擬講演）を中心にすえつつも、話し言葉の多様性を或る程度幅広く記録するための措置であることも既に指摘したとおりである。

ところで、話し言葉の多様性を考える際に重要な観点となるのが音声の自発性(spontaneity)の問題である。音声研究では、音声を「朗読音声」と「自発音声」に二分することが多い（2.2.1 節参照）。しかし、音声ないし言語の自発性という概念を明確化することは、実は簡単でない。自発性について用いられる説明のひとつに「発話の時点において、あらかじめ発話の形式が決定されていない発話」というものがある。CSJ に収められた音声は、朗読及び再朗読音声を除外すれば、総じてこの規定に適う。しかし、2779 個におよぶ講演ないし対話音声を比較すると、そこには自発性の程度差が存在していることが明らかである 。データベースに存在する自発性の程度差は、擾乱要因ととらえるにせよ、あるいは積極的に利用するにせよ、それを何らかの方法である程度客観的に評価できることが望ましい。CSJ で採用した方策を以下に説明する。

A. 自発性の序列

CSJ に格納された種々の音声レジスターは、その種類によって、かなりの程度まで自発性の程度が組織的に異なっており、全般的な傾向としては、音声種別間で以下のような序列を想定してよいものと考えられる（記号‘＜’はその左側よりも右側の方が自発性が高いことを示し、‘＜＜’はその差が顕著であることを示す）。このような自然の序列は、話し言葉の研究にとって積極的に利用すべき性質の情報である。

自発性低 ⟵⟶ **自発性高**

再朗読ないし朗読　＜＜　学会講演　＜　模擬講演　＜＜　インタビュー　＜　課題指向ないし自由対話

ただし、このうち学会講演と模擬講演との間の差異については、他のカテゴリ間の差よりも小さい可能性があり、また、量的に CSJ の大部分を占める音声でもあるために、両者間の差異をきわだたせる対策をとった。具体的には、先述のとおり模擬講演では原稿の利用を禁止したことの他に、講演話者ができるだけリラックスした状態で録音に臨めるよう、収録に先立って収録スタッフと雑談を交わす時間を設ける、収録中も収録スタッフは積極的にうなずき等の反応をかえす等の対策である。

B. 印象評定

上に示した序列は、しかしながら、絶対的なものではない。特に学会講演と模擬講演の間では、前節に述べた収録上の対策にも関わらず、序列の逆転が生じていることが少なくないと思われる。そこで、個々の講演についても、その自発性をある程度客観的に評価する手段があるとよい。そのような評価の一助として、CSJ のデータ収録作業では、原則としてすべての講演音声に対して音声収録記録票を作成し、その一部を音声が聴き手に及ぼす印象の主観評定に充てた。これを印象評定データと呼ぶ。印象評定には、このようにして収集したコーパスのほぼ全体に対するデータ（単独評定データ）の他に、コアの独話だけを対象としてより詳細な評定をおこなったデータ（集合評定データ）がある。印象評定データについては 2.3.4 節で紹介する。

2.3 研究用付加情報（アノテーション）

本節では CSJ に付与された研究用付加情報について概観する。紙幅の関係で細部には触れることができないので、詳細な情報は、国立国語研究所(2006)を参照のこと。

2.3.1 転記テキスト

収録された音声は、そのままでは検索することができないので、これを文字に書き起こした転記テキストを作成する必要がある。この作業の精度によってデータベースの価値が決まると言ってよい重要な作業である。音声認識に用いる言語モデルの精度もこの作業に強く依存する。図 2-2 に CSJ の転記テキストの例を示す。

話し言葉を転記しようとする際に必ず遭遇する重要な問題は、転記の単位をどう定めるかという問題である。CSJ では文法的ないし統語的な基準は採用せず、長めのポーズ（原則として 200ms 以上）位置で音声を転記基本単位に分割している。各転記基本単位には開始時刻と終了時刻の情報が提供されている。ポーズという物理的な指標によって転記の単位を定めたのは、CSJ のように大量のデータを扱う場合、その全体に一貫して適用可能な「文末」の客観的特徴を規定することが、実際上不可能であると判断されたからである。そのため CSJ の転記基本単位は統語的な文末と一致しているとは限らない。

0087 00187.217-00193.684 L:	
(F あーの)	& (F アーノ)
オーストラリアに	& オーストラリアニ
行くと	& (? イク)ト
大概	& タイガイ
(F あの)	& (F アノ)
ビーチの	& ビーチノ
あるところに	& アル<H>トコロニ
滞在したりとか	& タイザイシタリトカ
住んだりっていう	& スンダリッテユー
経験が	& ケーケンガ
あるんですが	& アルンデスガ
0088 00194.417-00194.918 L:	
で	& (? デ)
(F えー)	& (F エー)
0089 00195.255-00195.979 L:	
(F ま)	& (F マ)
どうしても	& ドーシテモ
こう	& コー
0090 00196.284-00197.702 L:	
ビーチに	& (W ピーチ;ビーチ)ニ
行くと	& イクト
嬉しいので	& ウレシーノデ<H>
（以下略）	

図 2-2: 模擬講演の転記テキストの例

「0087」等の数字で始まる行は、転記基本単位の通し番号、開始時刻、終了時刻を示している。転記基本単位中は文節で改行されており、&で区切られた左側が基本形、右側が発音形である。図中で用いられているタグについては本文参照。

CSJ の転記テキストには、発話を漢字仮名まじりで表記した基本形と片仮名だけで表記した発音形の２種類がある。基本形は主として情報検索のための利用を想定しているので表記にゆれを生じさせないことを徹底して追及した（小磯・西川・間淵 2006）。

一方、発音形の役割は、基本形の漢字の読みを確定させると同時に、発音上の変異を正確に表記することにある。「私」が「ワタクシ」か「アタクシ」か、「本当」が「ホントー」か「ホント」か、「前川」が「マエカワ」か「マエカー」か、「国語研」が「コクゴケン」か「コッゴケン」か等々が、人間の耳で聞き分けられ仮名文字で表現できる範囲で、可能なかぎり正確に表記されている。発音形は、日本語話し言葉の音声変異について貴重な情報を

提供する。また近年の音声認識技術で重要性を増している発音辞書の構築にも利用できる（秋田・河原 2005）。

転記テキストには多くのタグが挿入されている。代表的なタグに「エー」「アノー」等の言い淀みを表す(F)、言いさしによって断片化された語を示す(D)、聞き取りが困難な箇所を示す(?)、発音の転訛ないし不正確な発音を示す（W）、非語彙的な母音の延長（「あれが」が「アーレガ」と発音されるようなケース）を示す<H>などがある。タグの多くは当該文字列を囲む形で転記テキスト中に挿入されている（タグの詳細は小磯・西川・間淵 2006 参照 ）。図 2-2 に CSJ の転記テキストの例を示した。

転記テキストは CSJ に対するアノテーション全体の基礎となるアノテーションである。本研究の第 10 章では上述の（D）タグ、（F）タグの統計情報を、発話のレジスターと音声の韻律的特徴の相関関係を分析するために利用している。

2.3.2 形態論情報

形態論情報とは発話を語に分解して品詞分類を施した情報である。その際、当然ながら、語をどう規定するかによって結果が異なってくる。この問題はあらゆる言語に存在するが、日本語のように造語法上の自由度が高く、正書法が確立されていない言語では殊に重要である。理論上は、漢字のひとつひとつが単位となってしまうような短い単位から、いわゆる臨時一語（例えば「国立国語研究所外部評価委員会報告書」）が一単位となるような長い単位までを考えることができる。

CSJ では、国語辞典の見出し語に該当するような短めの単位と、それよりも長めの単位との二種類を採用して二重の形態論情報を提供している。これらをそれぞれ短単位、長単位と呼ぶ。一例を示せば「これからディズニーワールドについてお話しいたします」というテキストは、短単位では「これ|から|ディズニー|ワールド|に|つい|て|お|話し|いたし|ます」と 11 単位に、長単位では「これ|から|ディズニーワールド|について|お話しいたし|ます」と 6 単位に分解される。

短単位及び長単位の設計及び認定基準については、小椋(2006）に詳しいが、二種類の形態論情報を同時に提供することによって、日本語の造語法についての貴重な知見を得ることができる。また、語と韻律特徴との関係を吟味する研究のためにも、二重の形態論情報は有益に利用できると考えられる。

CSJ の形態論分析では、まず、コアの全体を含む短単位で 100 万語相当のテキストを国語研究所の研究員が手作業で分析した。このデータは通信総合研究所に渡されて、形態素自動解析ソフトウェアの学習用データとして利用された。CSJ のうち上記 100 万語を除外した残り 650 万短単位は、このソフトウェアによって自動解析されたものである 。自動解析結果には国立国語研究所において可能な限りの手修正をくわえている。

手作業による短単位形態論情報の精度はランダムサンプリングによって約 99.9%と推定されている。これを 1000 語にひとつも誤りがあると考える方もあるかもしれないが、実際に

話し言葉のデータを分析してみると、語境界や品詞を一意に決定しがたいケースが 1000 語にひとつ程度は出現するので、この数字は人知の限界であると考えている。自動形態素解析の精度は手作業に較べると若干低下することは避けられず、おそらく 98%前後である。そのためコアを含む 100 万短単位とそれ以外とでは形態論情報の精度に相違がある。CSJ の活用にあたって注意が必要となろう。

表 2-3 は CSJ に含まれる長短単位数を音声の種類ごとに示している。言い誤りによって生じた語の断片などは除外されている。表の最終列は短単位数に占める長単位数の百分率である。この率は、学会講演で最低値をとり、模擬講演と対話で最大値をとっている。これは学会講演には専門語が多く用いられるために相対的に多くの複合語が含まれていることによると考えられる。

本研究では第 5 章と第 6 章で、分節音の変異を様々な言語単位中の位置との関係で分析する際に、短単位と長単位をともに利用している。

表 2-3: 形態論的単位数

音声のレジスター	短単位数	長単位数	%長単位
学会講演	3,279,364	2,654,823	81.0
模擬講演	3,605,729	3,115,302	86.4
その他	282,728	239,989	84.9
朗読と再朗読	207,478	172,216	83.0
対話	149,826	131,544	87.8
全体	7,525,125	6,313,874	83.9

2.3.3 節単位情報

独話音声では、形態論的に典型的な文末特徴が生じることなく発話が連綿と続いてゆくことがある。「みんなで相談したんですけど、賛成しようということになって、私は反対だったんだけど、それでもみんなは賛成なんで、一応賛成しようということになったんだけど、やっぱり私は...」というような発話である。

書き言葉を基準にしてこの種の発話を分析すると大変な長文が生じてしまう。しかし、話し言葉として見た場合、「節」(clause)が情報処理上の単位として機能している可能性が高い。上例に読点を挿入した箇所である。

CSJ には、転記テキストを解析して節境界の位置を検出し分類した節単位情報（節境界ラベル）が付与されている。この情報付与作業には、ATR 音声言語コミュニケーション研究所で開発された節境界解析プログラム CBAP（丸山他 2003）を CSJ 用に改良した CBAP-csj を利用した。また、既に述べたようにコアに関しては自動解析結果を人手で修正した高精度情報を提供している。表 2-4 に CBAP-csj によって検出される 49 種類の節境界ラベルを示す。

表 2-4: CBAP-csj が検出する 49 種類の節境界ラベル

絶対境界：［文末］,［文末候補］,［と文末］
強境界： /並列節ガ/，/並列節ケレドモ/，/並列節ケレド/，/並列節ケドモ/，/並列節ケド/，/並列節シ/，/ヨウニ節/
弱境界：<条件節タラ>，<条件節タラバ>，<条件節ト>，<条件節ナラ>，<条件節ナラバ>，<条件節レバ>，<理由節カラ>，<理由節カラニハ>，<理由節カラ-助詞>，<理由節ノデ>，<タリ節>，<タリ節-助詞>，<テ節>，<テハ節>，<テモ節>，<テカラ節>，<テカラ節-助詞>，<テ節-助詞>，<トカ節>，<トカ節-助詞>，<ノニ節>，<連用節>，<引用節>，<引用節-助詞>，<引用節トノ>，<トイウ節>，<間接疑問節>，<間接疑問節-助詞>，<連体節テノ>，<並列節ダノ>，<並列節デ>，<並列節ナリ>，<フィラー文>，<感動詞>，<接続詞>，<接続詞 C>，<接続詞 L>，<接続詞 CL>，<接続詞 M>

大部分のラベルはラベル名称が対象となる形態論単位を明示しているが、接続詞関係の名称は抽象的なので、説明が必要であろう。丸山・高梨・内元(2006) p.291 より引用すると、

「<接続詞 C>は接続詞「けれど」+ 副助詞「も」のように短単位の連接から 1 語相当の接続詞と見なすことができるもの，<接続詞 L>は代名詞「それ」+ 格助詞「で」のように短単位では 2 単位だが長単位では 1 単位の接続詞「それで」になるもの，<接続詞 CL>は接続詞 L 「ですけれど」+ 副助詞「も」のように長単位と短単位の連接から 1 語相当の接続詞と見なすことができるもの，<接続詞 M>は強境界直後に現れる副助詞「って」+動詞の終止形「言う」+ 終助詞「か」を 1 語相当の接続詞と見なしたものである」。

表 2-4 中の「絶対境界」「強境界」「弱境界」は、節境界ラベルを、その節境界直後の構造的な切れ目の深さによって順序づけたクラスである。絶対境界は、書き言葉の文法におけるいわゆる文末表現に話し言葉固有の文末表現（と文末）を加えたもので、その直後には発話の完全な切れ目があると考えられる。

強境界は、南(1974) の従属節の分類における C 類の従属節にほぼ相当する境界である。書き言葉では文末表現になることはないか、きわめて稀であるが、話し言葉では、強境界で発話が終了することが珍しくない。弱境界は，もっとも弱い境界で、通常は発話末となることはない（ただしごく稀に発話の終端となることもある）。以下に節境界ラベルの実例を示す。いずれも丸山・高梨・内元(2006)からの引用である。

- 簡単に最初に復習をしておきたいと<引用節>思います[文末]
- 一番左が最初千文書を取得した段階での初期検索の精度ですね[文末候補]
- で<接続詞>総論では賛成なんだけれども/並列節ケレドモ/ 少し煮詰める必要があるんではないだろうかと[と文末]

- で<接続詞>結果ですが/並列節ガ/ まず絶対音感群の結果から見ていきたいと<引用節>思います[文末]
- まずその教室の様子ですけれども/並列節ケレドモ/ 教室の名前は俳句文法教室と言います[文末]
- よく知られますように/ヨウニ節/ 線形予測法というのがございまして/テ節/. . .
- 翌日朝食を済ませると<条件節ト>私達は旅行代理店へ向かいました[文末]
- で<接続詞>ついでにこれもいい日本語に替えてくれれば<条件節レバ>いいなと<引用節>そんな気がしてます[文末]
- どうしてもこの遊びの方が先に入ってたから<理由節カラ>そっちを優先しちゃったっていう<トイウ節>ことでした[文末]
- 縮むことがないので<理由節ノデ>安心して<テ節>何度でも洗います[文末]
- いわゆる制作進行助監督みたいなこともやったりも<タリ節-助詞>しました[文末]
- 十一月頃イラン革命が始まって<テ節>まず戒厳令と言って<テ節>夜の最初は八時以降は外出しちゃ<テハ節>いけない[文末]
- 洗濯機の音ですとか<トカ節>テレビの音ですとか<トカ節>色々あります[文末]
- 聞くところによると<条件節ト>この地方を治めていたスペイン人はとても温厚な人柄だったらしく<連用節>先住民と争うようなことはあまりなかったのだそうです[文末]
- 何となく雰囲気は分かったと<引用節>思います[文末]
- もう私としては裏切られたとしか<引用節-助詞>言いようがありません[文末]
- 横軸が時間で<並列節デ>縦軸が周波数になっています[文末]
- けれども<接続詞 C>実際の会話の中で起こる相づちは常に同じ形式のものが義務的に使われてる訳ではありません[文末]
- だから<接続詞 L>そういう意味では歴史のある町なんですね[文末]
- だからこそ<接続詞 CL>今そういう状況なってるのかもしれないけれども/並列節ケレドモ/. . .

説単位情報は、以下に述べる談話境界情報や係り受け情報を作成する際の単位の切り出しに利用することができるし、それ以外にも多くの利用が可能であると考えられる。本研究では、第 8 章で final lowering 現象の実在性を検討するために節単位情報を利用している。さらに第 8 章と第 9 章ではイントネーション研究における発話の単位を操作的に規定するために節単位情報を利用している。

2.3.4 印象評定データ

既に述べたように、印象評定には単独評定データと集合評定データの二種類がある。単独評定データは、ほぼすべての講演音声を、その収録の現場において、収録スタッフ中の 1 名が評価したデータである。例えば発話の自発性に関する項目としては「自発性」と「発話スタイル」があり、いずれも五段階で評定されている。前者は「講演に際してどれだけ原稿を読みあげているか」（原稿への依存度が低いほど自発性の評定値が高い）、後者は発話が「どれだけ改まっているか」（改まり度が高いほどスタイルの評定値が高い）を評定してい

る。上記以外の 5 段階評定項目には「難関な専門用語の多少」「発話スピード」「発音の明瞭さ」「方言の多少」がある。

5 段階評定項目以外に、講演の印象にあてはまる形容語句をリスト中から自由に選択する形式の評定も実施した。リストには以下の語句が含まれており、評定者は複数の語彙を選択できる：「たどたどしい、流暢な、単調な、表情ゆたかな、自信のある、自信の無い、優しい、落ち着いた、落ち着きのない、いらいらした、緊張した、リラックスした、大きい声、小さい声、かすれた声、裏返った声、こもった声、重厚な、軽薄な、若々しい、年寄じみた、元気のある、元気のない、聞き取りやすい、聞き取りにくい、生意気な、尊大な、鼻にかかった、高い、低い、きっぱりした」。

単独評定データには、ひとつの講演に評定者が 1 名だけであること、データベース全体では多数の評定者が参加していること、講演の前半を聴いた時点で評定を行っているため、講演中のどの部分が特に印象形成に影響したかが明らかでない場合がありうること等の問題がある。また、上に示した形容語句のリストも慎重に検討して作成したものではなかった。これらの問題を解決するために作成したのが集合評定による印象評定データである。

集合評定印象評定データは、全データの収録を終えた後に、コアに含まれる独話音声を対象として作成した（山住他 2005,2007, 籠宮他 2007）。1 講演の冒頭、中程、終盤からそれぞれ 1 分程度の連続した音声サンプルを抽出し、それを 20 名の評定者が独立に評定した。CSJ には、そのうち、評定の再現性が高い 10 名分のデータが格納されている。また、単独評定の評定が心理学的な厳密性に欠けるのに対して、集合評定では、実験心理学的に厳密な手順で構成した 5 種類の評定尺度を用いている点も特徴である。それぞれの尺度は 4 種類の下位尺度から構成されている。データベースの全体にわたる評価が必要な場合は単独評定データを、対象とする講演は少なくとも信頼性の高い評価値を利用したい場合は集合評定データを、それぞれ利用すればよい。

本研究では第 9 章でイントネーションと発話スタイルの関係を分析する際に、集合評定印象評定データを利用している（表 9-2, 9-3 参照）。

2.3.5 その他のアノテーション

先に 2.2.2 節で言及したように、CSJ のコアにはこのほかにも多くのアノテーションが施されているが X-JToBI による分節音ラベルと韻律ラベル、係り受け構造、談話構造、要約・重要文情報などである。

このうち分節音ラベルについては第 3 章にラベルの一覧表を示すほか、第 4 章、第 5 章、第 6 章で個別の研究に係るラベルについて解説する。イントネーションラベル(X-JToBI)については、第 3 章をその解説にあてているほか、第 8 章から第 10 章にかけて各種イントネーション現象の分析に利用する。また第 10 章では X-JToBI の有効性を全体的に検証する。CSJ の音声ラベリング全般については菊池他(2003)参照。

その他のラベルの解説は本研究では省略する。詳細は CSJ のマニュアル（http://www.ninjal.ac.jp/products-k/katsudo/seika/corpus/releaseinfo/）ないし国立国語研究所(2006)参照。

2.4 CSJ の特性

CSJ に記録された自発音声の特性を理解してもらうために、いくつかの分析結果をトピックとして報告する。本研究の第 4 章以下の音声学的分析と関係するものだけでなく、CSJ の全体的特性をよく示すと思われるものは対象に含めている。

2.4.1 話者の分布

話し言葉の多様性の一部は、性別、出生地、居住歴、学歴、講演経験の有無など、話者の社会的属性に起因している。そのため、話し言葉の研究では話者の属性への配慮が欠かせない。CSJ では、話者のプライバシーを侵害しないと判断された範囲で話者の属性情報を公開している。ここでは、CSJ 公開版を対象として、最も代表的な話者属性である話者の生年代、性別、出生地の分布を概観する。

まず、図 2-3、図 2-4 に学会講演と模擬講演における話者の生年代の分布を示した。CSJ のデータでは、話者の生年を西暦で 5 年刻みに区分して公開しているが、図 2-3、図 2-4 ではこれを 10 年ごとに区分しなおして集計した。図 2-3 は生年代ごとの延べ話者数、図 2-4 は同じく異なり話者数の分布を示している。延べと異なりの区別が必要となるのは、模擬講演だけでなく、学会講演においても同一話者の音声が複数回収録されていることがあるからである。これを重複してカウントしたのが延べ話者数、何回講演しても 1 名としてカウントしたのが異なり話者数である。

図 2-3 においても図 2-4 においても、学会講演話者数は生年代が下がるにつれ単調に増加している。一方、模擬講演話者は、学会講演に較べれば相対的にバランスのとれた分布を示している。なお、学会講演話者のうち 9 名については生年が不明であるために集計から除外している。

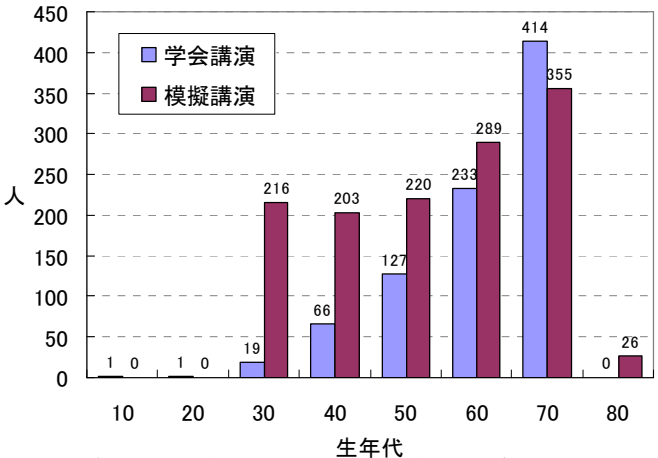


図 2-3: 学会講演話者と模擬講演話者の生年による分布（延べ）

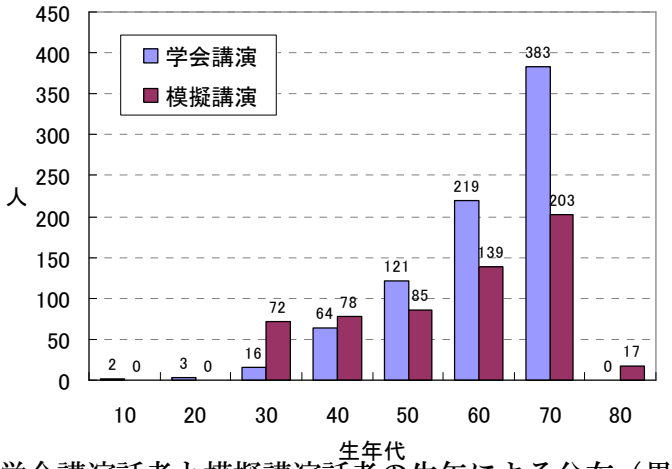


図 2-4: 学会講演話者と模擬講演話者の生年による分布（異なり）

次に、表 2-5、表 2-6 に話者の性別と音声の種類によるクロス集計を示す。表 2-5 が延べ話者数、表 2-6 が異なり話者数である。表 2-5 では学会講演から対話までの合計が「全体」欄の数字と一致していない。これは同一話者が複数の種類にまたがってデータを提供している場合に重複してカウントしているためであり（ひとつの種類内部での重複はカウントしていない）、再朗読と対話の話者を学会講演ないし模擬講演の話者から選択していることと「その他」の話者のうち 2 名が学会講演話者でもあることが、その原因である。

先にも述べたように、学会講演話者の大多数は男性である。これは学会発表の多くが大学院生によっておこなわれており、その大部分が男性であることによる。この傾向は特に理工系大学院において著しい（ちなみに図 2-3、2-4 の学会講演において 70 年代生まれの話者数が突出しているのも大学院生の多さによる）。一方、模擬講演以下では、男女がほぼ均等に分布している。

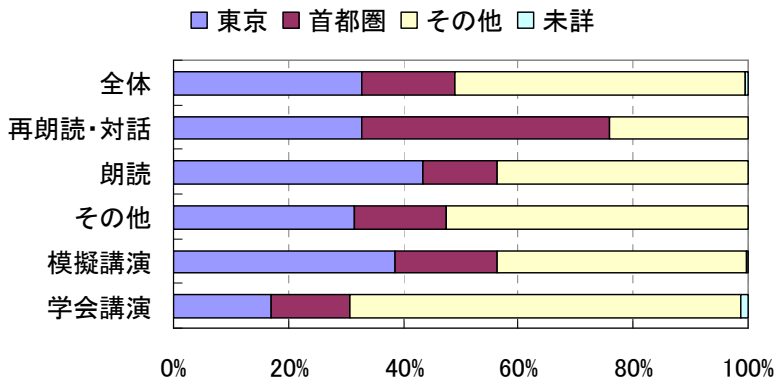
表 2-5: 話者の性別の分布（延べ）

性別	学会講演	模擬講演	その他	朗読	再朗読	対話	全体
女	173	910	9	252	8	29	1381
男	814	805	10	255	8	29	1921
計	987	1715	19	507	16	58	3302

表 2-6: 話者の性別の分布（異なり）

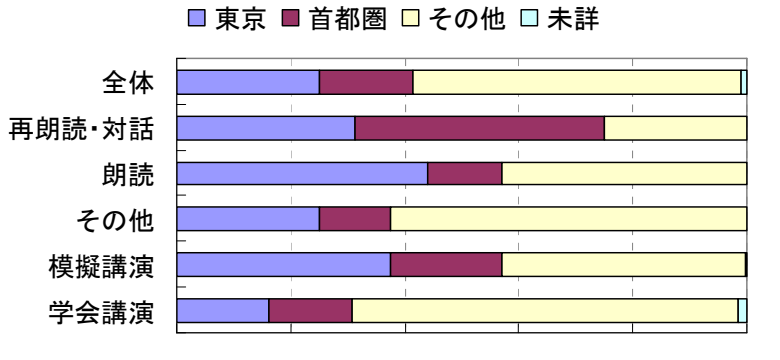
性別	学会講演	模擬講演	その他	朗読	再朗読及び対話	全体
女	138	*331	6	(122)	(8)	****47
男	681	**263	***10	(124)	(8)	947
計	819	594	16	(246)	(16)	1417

( )内の数字は学会講演もしくは模擬講演と重複、* 5 名が学会講演と重複、 **5名が学会講演と重複、***2名が学会講演と重複、****インタビュワーを含めると 471 名



	学会講演	模擬講演	その他	朗読	再朗読・対話	全体
未詳	12	3	0	0	0	15
その他	672	747	10	221	14	1664
首都圏	135	305	3	66	25	534
東京	168	660	6	220	19	1073

図 2-5: 話者の出生地の分布（延べ）



	学会講演	模擬講演	その他	朗読	再朗読・対話	全体
未詳	12	1	0	0	0	13
その他	555	255	10	106	4	816
首都圏	120	115	2	32	7	233
東京	132	223	4	108	5	355

図 2-6: 話者の出生地の分布（異なり）

最後に、図 2-5、図 2-6 に話者の出生地による分布を示す。図 2-5 が延べ話者数、図 2-6 が異なり話者数である。いずれの図においても「東京」「首都圏」「それ以外」に分類し百分率で示している。「首都圏」とは千葉、埼玉、神奈川の 3 県をさす。なお、ここで、出生地とは文字通り話者が生まれた土地であって生育地ではない。社会言語などの研究においてさらに詳しい履歴が必要とされるであろうが、その種の情報も、プライバシーを侵害しない範囲で公開されている。

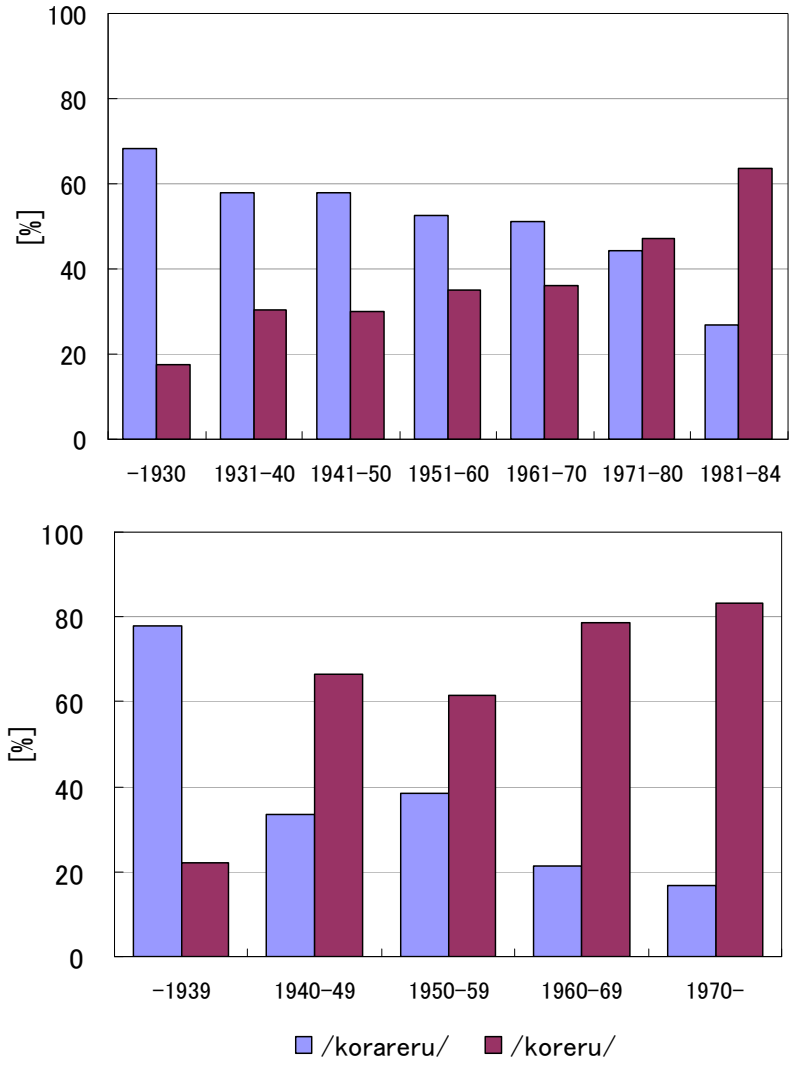


図 2-7: 生年代による「コラレル」と「コレル」の分布
上は文化庁によるアンケート調査、下は CSJ に記録された語形

2.4.2 内省と行動の乖離

話者の年齢差は、現在進行中の言語変化を研究するための手段として、社会言語学ではさかんに利用される。CSJ においても、それが本来的な利用目的はないものの、話者の生年代の情報を活用することで、一部の言語現象について、社会言語学的な研究を行うことができる。一例として、一段動詞の可能動詞化現象の分析例を示す（前川 2007）。

一段動詞の「見る」「起きる」、カ行変格動詞の「来る」などに助動詞を付与して「～することができる」の意味を表わす場合、伝統的には「ミラレル」「オキラレル」「コラレル」の形を用いたが、現代の話し言葉では、「ミレル」「オキレル」「コレル」などの、いわゆる「ラ抜き言葉」が優勢である。この現象は言葉の乱れとして意識されることが多く、マスコミでもしばしばとりあげられてきたものであるが、この変化の背景に日本語動詞の活

用体系の歴史的な変化が存在していることは、金田一春彦が指摘しているとおりである(金田一 1966)。

ここでは CSJ に記録された変化と、従来の社会言語学的なアンケート調査による結果を比較してみる。図 2-7 上部のパネルは文化庁が 2001 年に実施した全国規模の世論調査で「来ることができる」を話し言葉でどう言うかを質問した際の回答である（文化庁 2001）。図の横軸は回答者の生年代であるから、若い世代ほどラ抜きを使うという回答が増えていることがわかる。

一方、同図下部のパネルは、CSJ に記録された語形を話し手の生年代によって集計した結果である。ここでも年齢の低下とともにラ抜き言葉の使用率が上昇しているが、伝統的語形をラ抜きが逆転するタイミングに注目すると、両パネル間には顕著な相違がある。世論調査（アンケート調査）では 1970 年代生まれの話者で逆転が生じているが、CSJ に記録された実際の行動では 1940 年代生まれの話者において、既にラ抜きが伝統的な語形を逆転している。アンケートで把握された話者の意識と現実の行動との間に約 30 年のギャップが存在しているのである。

この結果は、アンケート調査には話し手の実際の言語行動が正確に反映されているとは限らないことを示している。その原因としては、言語の規範性に対する回答者の意識に影響している可能性なども考えられるが、対象とする言語現象によっては、そもそも内省によって言語行動を把握しようとすることが無理な現象もあるのではないかと思われる。

2.4.3 発話速度の分布

朗読音声と自発音声を比較する際に、両者の差異が際立つ特徴のひとつが発話速度である。図 2-8 は朗読音声データベースの代表として ATR 音声データベース・セット F (Sagisaka et al. 1990 参照）を選び、発話速度(mora/sec)を CSJ と比較した結果である(Maekawa 2004b)。ATR データベースでは文単位に、また CSJ では転記ファイルにおける転記基本単位（2.3.1 節参照）ごとに発話速度を計算した。

図 2-8 からは自発音声の平均発話速度(8.01)が朗読音声(7.11)よりも速いことがわかるが、おそらくそれよりも本質的な差異は、自発音声の分布の標準偏差(2.07)が朗読音声(0.96)よりもはるかに大きい点にある。毎秒 15 モーラに達するような発話速度の音声が一部に存在することは、工学的にも、音声認識システムにおける特徴抽出窓の設計に深刻な影響を及ぼしたが（奥田・河原・中村 2001）、本研究でも第 5 章、第 6 章、第 10 章などの分析で、発話速度が重要な説明要因となる。なお、図 2-8 は構築途上の評価実験の結果なので、完全に CSJ 全体を利用したものではない。

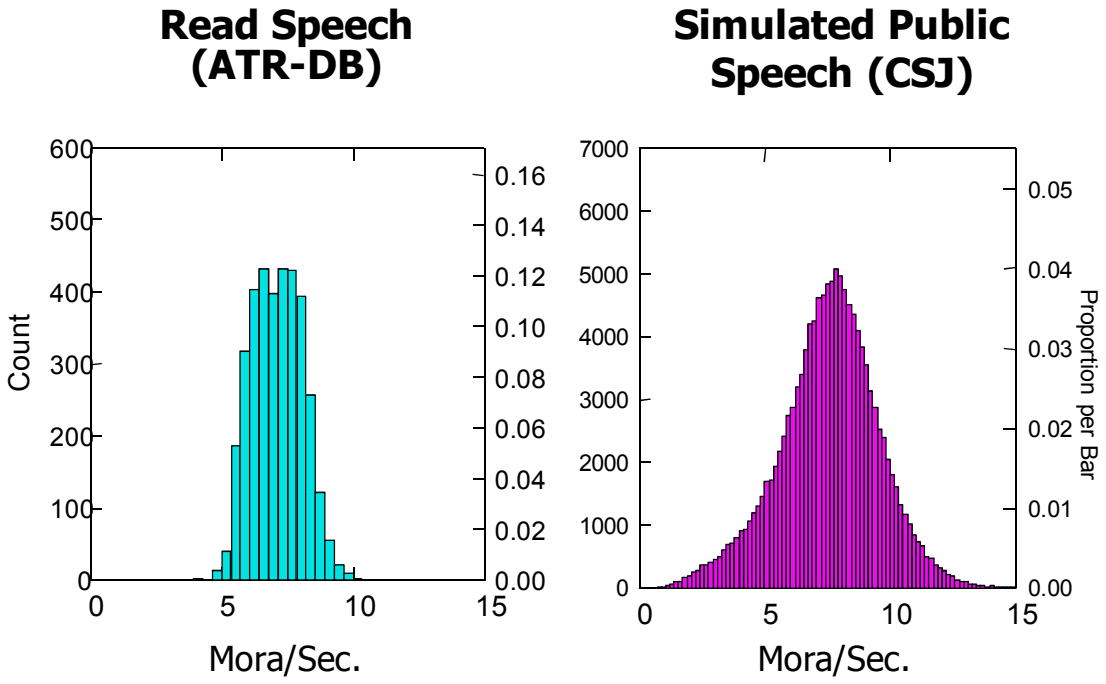


図 2-8: 朗読音声と自発音声の発話速度の比較

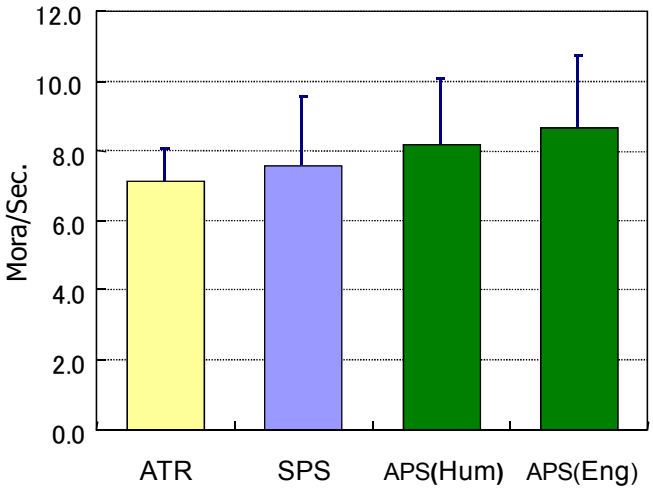


図 2-9: CSJ 内部での発話速度の変動

次に図 2-9 は、CSJ データを模擬講演（SPS）、人文系学会での学会講演(APS-Hum)、工学系学会での学会講演(APS-Eng)に三分したうえで、それぞれの平均発話速度をもとめ、ATR データベース全体の平均発話速度と比較した結果である(Maekawa 2004b)。CSJ の学会講演内部では工学系学会が文科系学会よりも速いことがわかるが、これは一般に工学系学会の発表時間(15~20 分)が文科系学会(25~35 分)よりも短いことに起因していると思われる。また模擬

講演の平均発話速度が学会講演よりも遅いのは、講演時間を厳密には管理していないことの反映だと考えられる。

2.4.4 語形の変異

最後に CSJ に記録された語形の変異を観察しておく（Maekawa 2005a,b, 前川 2006）。CSJ の形態論情報には、種々の国語辞典を参照して決定した「代表形」の情報が含まれている。代表形はその語彙素の代表的な発音を示すもので、例えば名詞「日本」の代表形は「ニッポン」である。しかし、この語には「ニホン」という非代表語形（変異形）があり、こちらで発音されることが多いので、代表形の実現率は 100%よりも低い値をとる。

表 2-7 は、すべての語彙素について、代表形と実際の発音の食い違いを集計し、非代表形の生起頻度が最も高い 20 語形を示したものである（前川 2006）。ここで「語形」とは、活用形のレベルで区別した語の音声形（発音）を意味している。動詞「歩く」であれば、終止形と連体形の語形は「アルク」であるが、連用形の語形は「アルキ」もしくは「アルイ」（イ音便形）である。

表 2-7: 変異形（非代表形）の比率が高い語彙素上位 20 語

語彙素	代表形	品詞 (活用形)	語彙素頻度	変異形頻度	% 非代表形	異なり話者数
{言う}	イウ	動詞 (連体形)	132,818	132,332	99.6	1,411
{の}	ノ	準体助詞	153,521	79,829	52.0	1,326
{けれど}	ケレド	接続助詞	47,032	26,534	56.4	1,092
{何}	ナニ	代名詞	23,067	17,140	74.3	1,054
{言う}	イウ	動詞 (終止形)	9,155	7,991	87.3	1,031
{って}	ッテ	副助詞	50,704	7,834	15.5	956
{日本}	ニッポン	名詞	8,242	8,045	97.6	849
{くらい}	くらい	副助詞	8,947	7,758	86.7	951
{に}	に	格助詞	206,614	7,568	3.7	1,097
{矢張り}	ヤッパリ	副詞	11,746	7,022	59.8	706
{それ}	ソレ	代名詞	44,000	6,016	13.7	767
{良い}	ヨイ	形容詞 (連体形)	5,950	5,177	87.0	934
{良い}	ヨイ	形容詞 (終止形)	4,446	4,026	90.6	866
{もう}	モウ	副詞	18,501	3,669	19.8	674
{です}	デス	助動詞 (終止形)	141,084	3,431	2.4	624
{で}	デ	接続詞	55,717	3,290	5.9	756
{皆}	ミナ	名詞	4,309	2,634	61.1	593
{もの}	モノ	名詞	31,794	2,373	7.5	593
{私}	ワタクシ	代名詞	15,749	2,367	15.1	395
{そう}	ソウ	副詞	29,698	2,327	7.8	619

表 2-8 は、表 2-7 と同じ語形について、変異形の異なり数（その語形にいくつの変異形が観察されたか）、頻度 1 位の変異形による語彙素頻度の（表 2-7 参照）のカバー率、同じく 3 位までの変異形によるカバー率、そして上位 3 位までの変異形リストを示している。多くの語彙素は頻度 1 位の変異形のカバー率が 90%を超えている。なかには「の」「矢張り」のように頻度 1 位の変異形のカバー率が低い語彙素もあるが、上位 3 位までの変異形が与えられれば、累積カバー率は 90%を超え、多くの場合 99%に達している。比較的少数の変異形で発音変動の大部分をカバーできることがわかる。

表 2-8: 上位 3 位までの変異形によるカバー率*

語彙素	変異形の数	頻度 1 位の変異形によるカバー率 (%)	頻度 3 位まででの累積カバー率(%)	上位 3 位までの語形
{言う} (連体形)	31	90.3	99.6	/yuH/, /yu/, /yuu/
{の}	15	52.2	99.7	/N/, /no/, /do/
{けれど}	53	53.2	98.5	/kedo/, /keredo/, /keHdo/
{何}	25	73.9	97.4	/naN/, /nani/, /naNni/
{言う} (終止形)	11	90.3	99.0	/yuH/, /yu/, /yuu/
{って}	22	82.6	99.2	/Qte/, /te/, /Qti/
{日本}	6	96.8	99.6	/nihoN/, /niQpoN/, /nion/
{くらい}	6	88.7	99.7	/gurai/, /kurai/, /gura/
{に}	33	96.3	99.8	/ni/, /N/, /i/
{矢張り}	56	49.3	91.9	/yaQpari/, /yahari/, /yaQpa/
{それ}	98	85.8	93.8	/sore/, /soe/, /soi/
{良い} (連体形)	5	86.0	99.7	/iH/, /yoi/, /i/
{良い} (終止形)	7	91.1	99.4	/iH/, /yoi/, /i/
{もう}	20	80.1	99.3	/moH/, /mo/, /mu/
{です} (終止形)	60	97.4	99.2	/desu/, /esu/, /su/
{で}	33	91.6	98.9	/de/, /Nde/, /te/
{皆}	6	63.3	99.3	/miNna/, /mina/, /miNHna/
{もの}	25	92.3	99.4	/mono/, /moN/, /moH/
{私}	34	83.5	98.0	/watasi/, /atasi/, /tasi/
{そう}	28	92.0	99.0	/soH/, /so/, /soQ/

*語形表記中の/H/は長母音の後半要素、/N/は撥音、/Q/は促音を表している。

最後に図 2-10 は CSJ に含まれる 5 種類のレジスターごとに変異形の平均出現率を計算した結果である。つまりひとつの講演に含まれる全語数のうち、代表形以外の語形（変異形）の占める百分率の平均である(Maekawa 2009)。CSJ を構成する種々のレジスターが、音声に組織的な影響を及ぼしていることがわかる。ちなみに、図では、対話(dialogue), 模擬講演(SPS), その他 (PL,表 2-2 参照), 学会講演(APA), 朗読(reading,再朗読を含む)の順に変異形出現率が低下しているが、この序列は、先に我々がこれらのレジスター間に存在すると予想した発話の自発性の序列（2.2.5A 参照）に正確に対応している。

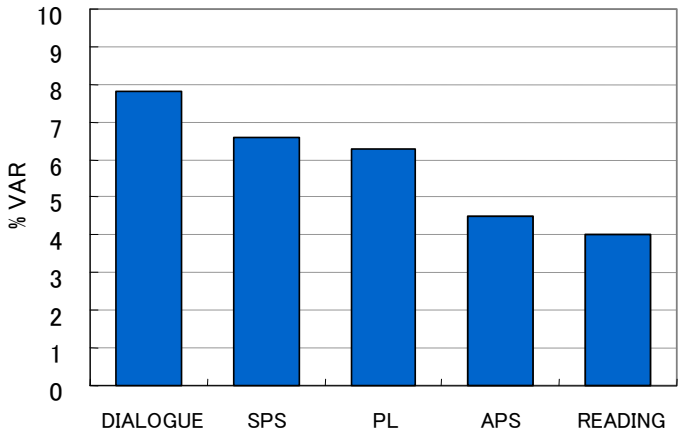


図 2-10: CSJ データのレジスターと変異形平均出現率

2.5 結論

本章では、前半で本研究の土台となる『日本語話し言葉コーパス』(CSJ)の設計を説明し、主要なアノテーションについても説明した後、後半では、構築された CSJ の特性を示す分析結果をかいつまんで報告した。CSJ には広い生年代にわたる多数の話者による、様々なレジスターの音声が収録されており、そこには従来の朗読音声データでは把握することが不可能であったり、困難であったりする音声現象が多数記録されているものと期待される。

第 3 章. X-JToBI

3.1 問題

本章では、CSJ-Core に施された韻律ラベリングについて詳しく解説する。日本語の韻律ラベリング体系としては Venditti らが開発した J_ToBI システム(Venditti, 1995=1997)が有名である。しかし、CSJ が主要な対象としている自発音声に対して J_ToBI を適用すると種々の困難が生じる (菊池他 2003)。そのため、J_ToBI を大幅に拡張した X-JToBI(eXtended J_ToBI)を提案した（Maekawa et al. 2002; 前川・菊池・五十嵐 2001; 五十嵐・菊池・前川 2006）。X-JToBI は、日本語（東京方言）の音韻的構造分析に基づくラベリング体系である点においては、J_ToBI と同一である。しかし、1) 分節音層の導入、2) Tone の時間情報の正確な表現、3) Tone 要素の拡張、4) BI 要素の拡張、5) 種々の非流暢性(disfluency)への対処など、さまざまな拡張が施されている。

これらの拡張は、ひとつには CSJ に格納された自発音声の韻律には朗読音声に較べてはるかに大きな多様性があるため、それに対処する必要から実施したものであるが、他に、自発音声の問題とは独立に、J_ToBI 固有の問題点を解消するために実施した拡張もあり、上記 1) と 2)がこれに該当する。

3.2 X-JToBI の構造

X-JToBI のラベル情報は、J_ToBI 同様、複数の層(tier)から構成されている。単語層(word tier)、分節音層(segment tier)、トーン層(tone tier)、BI 層(break index tier)、プロミネンス層(prominence tier)、注釈層(miscellaneous tier)の 6 層であるが、このうち分節音層とプロミネンス層は X-JToBI において新規に導入したものである。本節では各層の性格について簡単な説明を加え、各層におけるラベリングについては 3.3 節以下で詳しく説明する。

3.2.1 単語層

単語層は、発話を構成する形態論的単位（語）の境界を表示する。CSJ では長短二種類の形態論的単位を採用しているが、X-JToBI では、このうち「短単位」とよばれる単位の境界を表示している（小椋 2006）。CSJ の短単位には、韻律情報（アクセント）は含まれていないが、X-JToBI の単語層ではアクセント核の位置がアポストロフによって表示されている。これはアクセント辞書等に記載されている規範的なアクセント位置ではなく、実際の発音を反映した位置である。

3.2.2 分節音層

分節音層は、発話を構成する分節音のラベルを表示する。CSJ で用いた分節音ラベルは、基本的には音素ラベルであるが、一部は音声レベルの情報も表現している。分節音層をもつこ

とによって、X-JToBI では各種トーンと分節音との時間関係の詳細な検討が可能となっている。分節音ラベルについては 3.9 節参照。

3.2.3 トーン層

トーン層は、イントネーションを構成するトーンの種類と時間軸上の位置を表示する。J_ToBI においては、トーン層のラベルは、アクセント、句頭上昇の頂点を示す H-、三種類の境界音調(boundary tones)にほぼ限られており、これらは純粋に音韻論的な要素であったが、X-JToBI では、イントネーションの表現力を向上させるために、句末境界音調の要素を拡張し（3.4.4 節）、さらに、ポインター（3.4.5 節）、エクステンダー（3.4.6 節）等の新要素を導入した。これによって X-JToBI は音韻的情報だけでなく、音声的な情報も提供することになり、複雑化したが、ラベリング精度は J_ToBI よりも向上している（3.10 節参照）。また、X-JToBI におけるトーン層ラベルは、可能な限り、ラベルに対応する F0 イベントの時刻に付与されているので（J_ToBI は必ずしもそうなっていない）、トーンにかかわる時間情報も J_ToBI より豊富である。

3.2.4 BI 層

BI(break index)とは、韻律特徴からみた発話の階層構造を表現する情報である。基本的には 1,2,3 の整数によって表現され、この順により深い階層の境界を表現する。おおまかには、１は語境界、２はアクセント句境界、３はイントネーション句境界及び発話境界に該当すると考えてよい。単語層における語境界はすべて最低でも１の BI を与えられている。

ここまでは J_ToBI における BI と共通である。しかし X-JToBI では、自発音声の韻律的な多様性に対応するために、BI を大幅に拡張している。これについては 3.5 節で解説する。

3.2.5 プロミネンス層

プロミネンス層は X-JToBI で新設された層である。イントネーションを解釈するうえで必須な情報のうち、トーンや BI には分類できない情報がこの層に記入されている。その多くはトーンの実現位置の変異に関する情報である。例えば、ピッチの上昇と下降を↑と↓で表すと、東京方言には「デー↑タ↓が」「ディズニーワール↑ド↓で」のように、アクセントとは無関係にアクセント句の句末から二番目のモーラを選択的に卓立する話し方がある。X-JToBI ではこれを通常の句末境界音調のひとつである L%HL%の時間軸上の変異形とみなして、トーン層にはただ L%HL%としての指定だけがなされている。しかし、同時にプロミネンス層に PNLP というラベルが付与されるので、それによって通常の L%HL%（つまりアクセント句の最終音声ないしモーラ内部で上昇下降する句末イントネーション）と区別できるようになる。PNLP は本研究の第 9 章で詳しく分析する。

3.2.6 注釈層

X-JToBI の注釈層は、J_ToBI の miscellaneous tier と同様、種々雑多な情報を記入するために設けられている。自発音声の韻律的特徴はきわめて多様である。そのために、現在の X-JToBI の仕様でも完全には対応できないことがある。そのような現象が生じている場合に、利用者の注意を喚起するために、注釈層にラベルを記入している。

X-JToBI によるラベリングのサンプルを図 3-1 に示す。「エこれは エー 同じ図なんですが」という発話の音声波形と F0 曲線が上部に示されており、その下に、上から順に分節音層、単語層、トーン層、BI 層、プロミネンス層のラベルが表示されている（紙幅の関係で注釈層を省略している）。プロミネンス層の情報（3.6 節参照）からは、発話の中ほどに強調を受けた無核アクセント句があること(EUAP)、また、発話末尾に「浮き上がり調」と呼ばれる特殊な句末上昇イントネーションが生じていること(FR)が分かる。

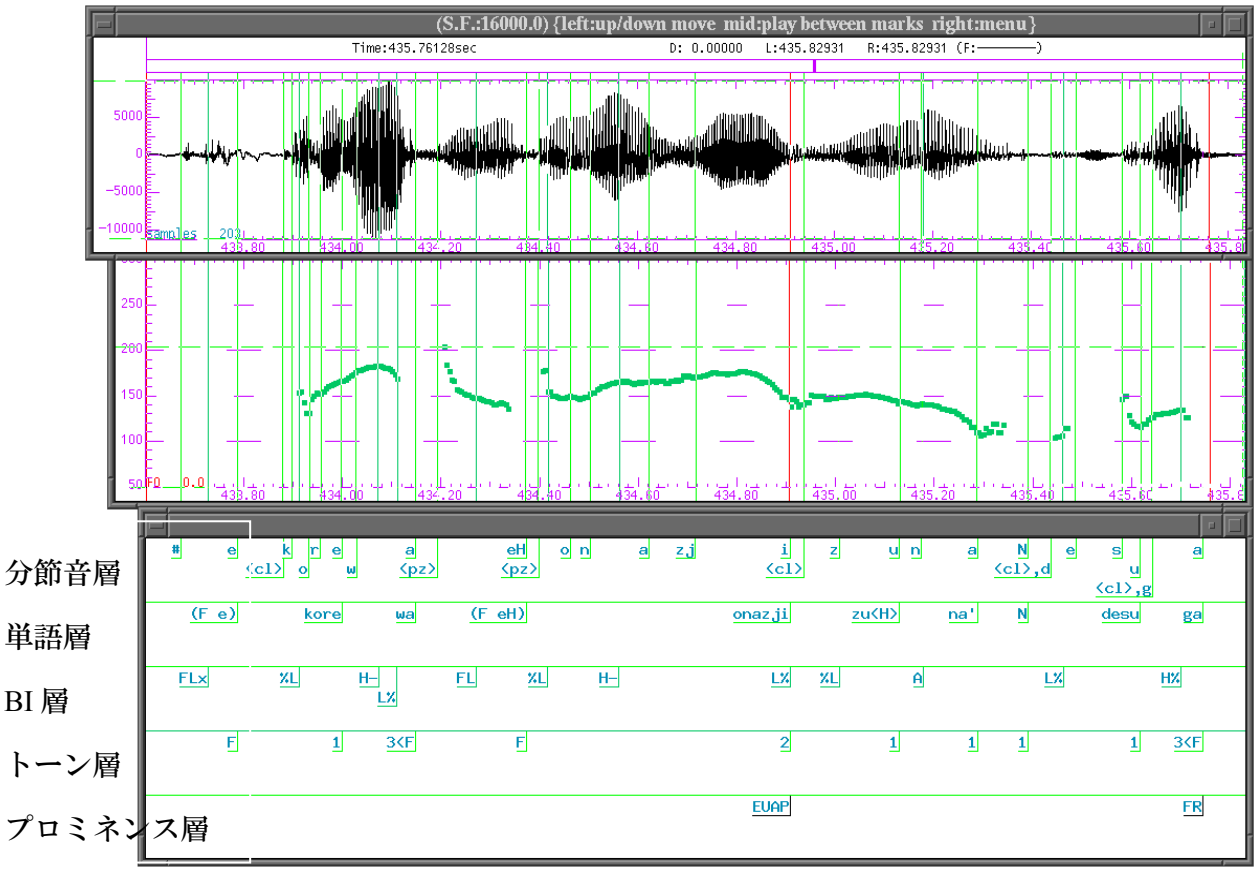


図 3-1: X-JToBI ラベリングの例

3.3 単語層のラベリング

単語層には、発話を構成する単語がアクセント核記号とともに表示している。単語の単位としては短単位(小椋 2006)が用いられており、転記テキストの音声形（カタカナによる音声表記）を分節音ラベル列に変換したものが表示される。単語層ラベルの付与位置は、各短単位の終端である。この位置は同時に BI ラベルの付与位置ともなっている（3.5 節参照）。

3.4 トーン層のラベリング

トーン層に生じるラベルの一覧を表 3.1 に示す。またトーンラベルと組み合わされて利用される補助記号の一覧を表 3.2 に示す。

表 3-1: X-JToBI のトーンラベル

ラベル	分類	出現位置
A	語	アクセント
H-	句頭音調	アクセント句頭からアクセントまでの位置
%L	句頭境界音調	アクセント句頭
(H%)	仮想的句末境界音調	例外的に高いピッチで始まるアクセント句の冒頭
L%	句末境界音調	アクセント句末
H%	句末境界音調	アクセント句末、L%の後
HL%	句末境界音調	アクセント句末、L%の後
LH%	句末境界音調	アクセント句末、L%の後
HLH%	句末境界音調	アクセント句末、L%の後
pH	ポインター	HL%ないし HLH%の前
pL	ポインター	LH%ないし HLH%の前
>	エクステンダー	
FH	フィラー音調	フィラー区間
FL	フィラー音調	フィラー区間

表 3-2: X-JToBI のトーン層補助記号

記号	使用例	意味
x	Ax, L%x 等	トーン時刻の F0 情報が存在しない、もしくは信頼できない
?	A?, HL%?	要素の存在（あるかないか）ないし選択（どれか）に自信がない
!	A!	位置が不明確*

* E_ToBI（English ToBI）における“!”の用法とは異なっていることに注意

3.4.1 アクセント

東京方言の語彙的に定まったアクセント核をラベル A で表示する。アクセント辞書に有核語（アクセントを持つ語）として記載されている語であっても、無核語（アクセントをもたない語）として発音されていればラベル は付与していない。同じくアクセント辞書に無核語として記載されていても、有核語として発音されていればラベル A を付与している。A が付与された語に対応する単語層のラベルでは、アクセント核の位置がアポストロフで示されている。

ラベル A の付与位置は、アクセントによる F0 の下降が始まる時刻である。F0 曲線に明瞭な頂点（ピーク）が存在していれば、その位置がラベル位置になる。F0 が下降しつづけている場合は、下降の加速度が急激に変化する時刻（屈曲点と呼ぶ）にラベルを付与している。ここで頂点とは、F0 曲線が上昇から下降に転ずる（ピークをなす）時刻をさし、屈曲点とは F0 曲線の変化率が符号を変えずに顕著に減少ないし増大する時刻をさす（図 3-2 参照）。

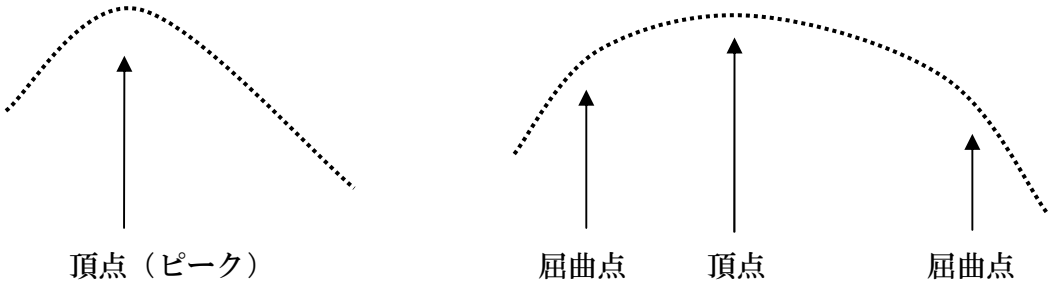


図 3-2: 頂点と屈曲点

F0 にはピークも屈曲点も見あたらず、しかも、聴覚的には明瞭にアクセントが知覚されることがある。その場合は、分節音情報を参照して、アクセントを担う母音領域の近傍にラベル A!を付与している。このラベルが生じやすいのは主に以下のふたつの環境である。

1) H%で終わるアクセント句に頭高型（第 1 モーラにアクセントをもつ）アクセント句がポーズ無しに後続する場合、
2) 先行語のピッチレンジが拡大されており対象とする語のピッチレンジが狭められている場合（先行語が強調されている場合）。

ラベル Ax は、F0 抽出アルゴリズムの問題等によって、アクセント核が存在すると想定される母音近傍に F0 値が存在しない、もしくは信頼できない場合に使用する。ラベルの位置はアクセントを担うと想定される母音の終端である。ただし、アクセント核を有すると知覚される母音に無声化が生じている場合の処理は特別である。この場合、無声化母音に後続する有声区間の F0 頂点（始端が最高値であればそこ）にラベル A を付与している。

この処理は、アクセント核を有する母音が無声化した場合、後続する有声区間の F0 が代償的に上昇するという音声学的な観察を根拠としている。

最後に、ラベリング作業者がアクセントの有無の判断に迷い、どうしても自信を持って判断できない場合は、ラベル A?を付与している。アクセントの連続などによって、ピッチレンジが極めて狭くなった発話の句末ではこのラベルを用いることが多い。

3.4.2 句頭音調

東京方言のアクセント句頭では通常、第１モーラから第２モーラにかけてピッチの上昇が知覚される。この上昇に該当する F0 の頂点ないしプラスの屈曲点にラベル H-を付与している（なお、X-JToBI ではアクセント句境界には必ず２以上の強さの BI が付与される）。

頂点ないし屈曲点の位置は、分節音からみると、必ずしも第２モーラの内部とは限らない。第１モーラにも第３モーラにも生じることがある。X-JToBI では、句頭の上昇が知覚される限りにおいて、実際に頂点ないし屈曲点が観察された位置にラベル H-を付与している。第２モーラが促音であったり、母音に無声化が生じたりした場合は、第３モーラにラベルが付与されることが多い（注：第３モーラも無声化している場合は、第２モーラの終端にラベル H-x を付与している）。また稀に、上昇が知覚されるにもかかわらず F0 曲線がほとんど平坦であったり下降しつづけていたりすることがある。その場合にはラベル H-!を第２モーラの中心部に付与している。

ただし、すべてのアクセント句に H-が付与されているわけではない。句頭におけるピッチの上昇が知覚されず、F0 も上昇していないアクセント句にはラベルは付与されていない。第１ないし第２モーラにアクセント（核）を有するアクセント句において H-が省略されやすい。

3.4.3 句頭境界音調

アクセント句の始端にはラベル%L が付与されている（唯一の例外はラベル(H%)%L に関するものである。3.7 節の HBP についての記述参照）。このラベルは、前節で言及した句頭の上昇の起点を示すものであり、句頭の上昇は、%L とそれに後続する H-ないしアクセント(A)の組合せによって表現されている。%L は F0 の局所的最低値の時点に付与するのが原則である。最低値が確定できない場合の処理は以下のとおりである。

1) アクセント句頭の F0 が全体的に信頼できない場合は、ラベル%Lx をアクセント句頭モーラの母音始端に付与する（母音が無声化している場合も同様）。
2) F0 の局所的最低値がみつからない場合は、ラベル%L!を付与する。
3) アクセント句頭の F0 に先行子音の影響などで乱れが生じている場合は、アクセント句頭モーラの母音区間のうち F0 が安定して抽出されている箇所に付与する。
4) 頭高型のアクセント句では、F0 曲線の開始時刻に F0 が最高値をとることがある。この場合、%L と A は同じ箇所（最高値）に付与する。

5) 先行するアクセント句末に句末境界音調 H%, LH%, HLH%（3.4.4 節参照）が生じており、アクセント句間にポーズが介在していない場合、当該アクセント句の冒頭が高い F0 値ではじまり、そのまま単調に下降し続けることがある。このような場合、下降区間中の屈曲点を探してその位置に%L を付与する。F0 が途切れなく下降していて屈曲点を確定できない場合は先行句の H%, LH%, HL%, HLH%と同時刻に%L を付与する。
6) アクセント句頭に音声の延長が観察され、F0 も低い値のまま延長されている場合は、%L とエクステンダー記号 > を組み合わせることによって、F0 の延長区間を表現している（3.4.6 節参照）。
7) 自発音声には、稀にではあるが、アクセント句の冒頭が高いトーンで始まると解釈すべき発話も生じている。その場合、ラベル”(H%)”を用いたラベリングをおこなう。この現象については 3.7 節の HBP 参照。

3.4.4 句末境界音調

日本語のアクセント句末には多彩なイントネーションが生じる。これを句末境界音調と呼ぶ。表 3-3 に X-JToBI における句末境界音調ラベルの一覧を示す。句末境界音調は２種類に大別できる。アクセント句末が単純に下降して終わる場合とそれ以外の場合、すなわち句末に上昇をともなう複雑なピッチ変動が観察される場合である。前者を単純境界音調、後者を複合境界音調(boundary pitch movement; BPM)と呼ぶ。複合境界音調の第１要素は必ず L%である。BPM は第 9 章で詳しく分析する（BPM へのラベル付与例は図 9-1 参照）。

表 3-3: 句末境界音調のラベリングに使用されるラベル

ラベル	使用するラベル列	特 徴
L%	L%	下降調（単純境界音調）
H%	L%, H%	上昇調１（BPM）
LH%	L%, pL, LH%	上昇調２（BPM）
HL%	L%, pH, HL%	上昇下降調（BPM）
HLH%	L%, pH, pL, HLH%	上昇下降上昇調（BPM）
pH	HL%,HLH%参	BPM における上昇の頂点
pL	LH%,HLH%参	BPM における上昇の開始点

A. 単純境界音調

L%（下降調）： 句末に複合境界音調が生じていなければ（すなわち下降調であれば）、アクセント句末にラベル L%を付与している。なお、アクセント句末にはきしみ発声(creaky phonation)が生じるなどの原因でしばしば F0 抽出が不安定になり、母音が無声化することも多い。このような場合にはラベル L%x（対応する F0 が存在しない）を句末モーラの終端に付与している。

B. 複合境界音調（BPM）

L%H%（上昇調１）：句末境界音調のうち上昇調１は、単純な上昇調である。上昇の始端に L%、頂点に H%を付与している。

L%LH%（上昇調２）：上昇が始まる前に一定期間低いピッチが持続されるタイプの上昇調であり、川上(1963)の「反問の上昇」に該当する。低ピッチ区間の始端に対応する屈曲点に L%を、低ピッチ区間の終端に対応する F0 上昇の始端に pL を、そして上昇の終端に LH%を付与している。上昇調１と２は、異なるパラ言語的意味を伝達するが、ときに両者の区別が困難なことがある。その場合には L%H%に分類している。

L%HL%（上昇下降調）：これはピッチが上昇したのち下降するタイプの BPM である。上昇の始端に L%、上昇の頂点に pH、BPM 全体の終端に HL%を付与している。

L%HLH%（上昇下降上昇調）：主に自発性の高い対話音声に生じる BPM である（CSJ-Core における頻度は表 9-1 参照）。最初の上昇の始端に L%、その上昇の頂点に pH、下降の谷に pL を付与し、BPM 全体の終端に HLH%を付与している。

3.4.5 ポインター

以上のラベリングで利用されたラベル pH と pL をポインターと呼ぶ。ポインターは、BPM における F0 変化の時刻を明示すために付与されるラベルである。BPM のタイプそのものは、アクセント句末に与えられたラベルが L%、H%、LH%、HL%、HLH%のいずれであるかによって識別できる。

朗読音声では、句末境界音調は原則として発話の最終音節（ないしモーラ）内部に生じる。しかし CSJ に記録されているような自発音声では、上昇の始端や頂点が最終音節以前に位置することがある。その場合は、プロミネンス層に特別なラベルを付与して、通常の BPM と区別する（3.6 節参照）。

3.4.6 エクステンダー

自発音声では、様々な要因によって、音声の延長（引き伸ばし）が生じることがある。すなわち、発音すべきテキストの一部が言語的要因（長母音など）とは無関係に、知覚可能な程度に延長される現象である。音声の延長はアクセント句の句頭ないし句末に生じることが多い。音声の延長が生じ、なおかつ延長部分のピッチもほぼ一定値を保ったまま延長されている場合、トーン記号と記号 >との組合せによって、延長の区間を表示する。この記号をエクステンダーと呼ぶ。

例えばアクセント句頭で分節音の延長が生じており、かつ当該箇所がほぼ一定の低いピッチで発音されていれば、延長区間（低ピッチ区間）の始端にラベル%L を付与し、終端にエクステンダーを付与している。

同様の現象（つまり分節音の延長と一定のピッチの持続）は、アクセント句末でも生じる。その場合は、延長区間の始端に句末境界音調ラベルに含まれる最後のラベルを付与し、終端にエクステンダーを付与している。なお、CSJ のラベリングでは、句末においてエクステンダーを適用しうるトーンラベルを、L%と H%に限定して作業をおこなった。

3.4.7 フィラーのトーンラベル

フィラー（filled pauses）にはフィラー専用のトーンラベル FH と FL を付与している。フィラーのラベリングについては、別途 3.8.1 節 で解説する。

3.5 BI 層のラベリング

BI ラベルは、J_ToBI と同様、すべての語境界（短単位境界）に付与される。J_ToBI の BI ラベルは、韻律境界の深さに応じて基本的には 1 以上 3 以下の整数値をとる。しかし、自発音声においては種々の原因によって韻律境界が曖昧化するために、従来の整数値の中間値を認め、さらに種々の非流暢性(disfluency)に対処するために、数値を用いない特殊な BI を考案した。以下、X-JToBI における拡張部分について説明する。

3.5.1 中間値

中間値を示すラベルは、整数 1,2,3 を N とし、アルファベットを X とすると、N+X の形で与えられる。ここで、X は２文字となることもある。N＋は、その韻律境界の深さが N と N+1 の中間値であることを示しており、X はその境界を中間値と判断する音声学的根拠を示している。

表 3-4 に BI 層のラベリングに用いられるラベルのうち従来の BI ラベルと中間値ラベルの一覧を示す。表中の「イントネーション句」はアクセント句よりも深い韻律上の境界をさす（後掲する図 8-1 の IP ノードはイントネーション句境界を導入するノードである）。

イントネーション句の特徴は、そこでピッチレンジがリセットされることである。ダウンステップ（アクセント核によって引き起こされるピッチレンジの狭め）はイントネーション句境界でリセットされる。

表 3-4: X-JToBI の BI ラベル（その１：中間値）

ラベル	用法	備考
1	語境界	CSJ では短単位境界
1+p	アクセント句内のポーズ	ポーズ位置は語と語の境界
2	アクセント句境界	
2+p	アクセント句境界にポーズが後続	
2+b	アクセント句境界に BPM が存在するが、ピッチレンジのリセットがない	ピッチレンジがリセットされていれば３。
2+bp	アクセント句境界に BPM が存在し、ポーズが後続するがピッチレンジのリセットがない	ピッチレンジがリセットされていれば３。
3	イントネーション句境界	

1+p：アクセント句境界を｜、ポーズを<pz>で示すと「フクオカ<pz>ニ｜イキマシタ」のように、アクセント区内部の語境界にポーズが生じた場合に付与する。なお、ポーズが単語の内部に生じた場合は表 3-5 の P を用いることになるので注意。

2+p：アクセント句境界にポーズが後続することによって、アクセント境界における韻律上の不連続性が、ポーズのないものよりも強く感じられる場合に付与する。

2+b：アクセント句境界に BPM が生じると、BI は 3 となるのが普通であるが、ときとしてピッチレンジがリセットされないことがある。その場合に付与する。

2+bp：アクセント句境界に BPM が生じており、ポーズも後続しているが、ピッチレンジはリセットされていない場合に付与する。

3.5.2 非流暢性

表 3-5 も X-JToBI において拡張された BI ラベルである。これらのラベルは表 3-4 のラベルと違ってその強さを単一の次元に配列することが困難であるので、整数値は与えていない。

表 3-5: X-JToBI の BI ラベル（その２：非流暢性など）

ラベル	用法	備考
P	語中のポーズ	
<D	語断片の直前	3.8.2 節も参照
D	語断片の終端	3.8.2 節も参照
<F	フィラー区間の開始	3<F, 2+b<F”のように先行アクセント句の BI とともに示す。
F	フィラー区間の終了	<F と共に利用
F2	対話の応答詞のうち言語情報を保有していないと判断される部分の終端	対話音声にのみ使用
W	短単位内部のアクセント境界	下記の説明参照
PB	句末境界音調の継起	下記の説明参照

P：「フク<pz>オカデ」（福岡で）のように、短単位の内部にポーズが生じた場合に付与している。

<D 及び D：短単位が断片化した発話で、かつ、そこに韻律上の非流暢性が知覚される場合、語断片の直前に＜D を、終端に D を付与している。例えば「先日、ト、名古屋で」という発話には、「名古屋」を「東京」と言い誤りかけたところで発話を中断したために生じた語断片「ト」が含まれている。この発話の BI を( )に入れて示すと「センジツ(<D)ト(D)ナゴヤ(1)デ(2)」のようになる。本ラベルの用法については、3.8.2 節 で別途説明している。CSJ の転記テキストにおける語断片の扱いについては小磯・西川・間淵(2006)の(D)タグに関する記述を参照。

<F 及び F：<F はフィラー区間の始端に付与している。フィラーに先行するアクセント句の BI とともに用いて、後続区間がフィラー区間であることを示す。一方 F はフィラー区間の終端を示す（F は単独で使用し、後続するアクセント句との関係は示さない）。例えば「私がエー学生だったとき」という発話の BI を( )に囲って示すと、「ワタクシガ(2<F)エー(F)ガクセーダッタトキ」のようになる。フィラーについては 3.8.1 節 でも別途説明している。

F2：このラベルは対話音声においてのみ使用される。対話音声には「ウン」「フーン」「ンー」等の応答表現やあいづちが生じる。これらの表現は CSJ の転記テキストでは（F）タグで囲われている。しかし（F）で囲われたテキストのなかには、発音が明瞭でなく、また、言語情報を保有しているかどうかも明らかでないものがある。例えば相手の発話を聞きながら「アー」「ンー」と母音や鼻音を発音しつづけているような場合である。この種の音声にラベリングを施すことは困難であるので、ラベリングの対象から除外した。BI 層にラベル F2 が記入されている短単位区間には、トーン情報は何も記入されていない。ただし、転記テキストにおいて（F）で囲われている表現であっても、以下の条件を満たす発話は通常のアクセント句として認定している。1)は有核アクセント句、2) は無核アクセント句である。

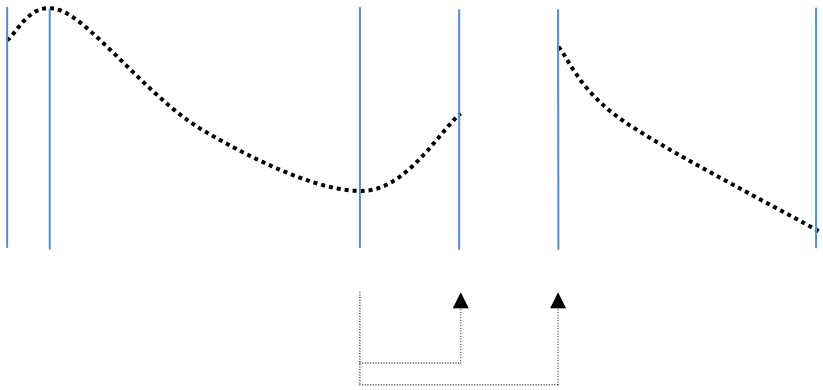
1) 発話中にアクセント類似のピッチ下降が知覚される場合
2) アクセント類似のピッチ下降は知覚されないが、テキストが明確な言語情報を保有している場合（ほぼテキストが「ハイ」の発話に限られる）

W：ひとつの短単位内部において、複数のアクセント核が生じるかアクセント核の後ろに句頭の上昇類似の上昇が生じた場合、その境界に付与している。下の例では‖の位置に W が付与される（┐はアクセント核、┌は句頭の上昇、そして｜は短単位境界）。

ファ┐ッショ‖ナ┐ブル｜ナ｜カタ┐｜ガ｜オ┐ーク｜テ（ファッショナブルな方が多くて）
W 1 2 1 2 1

エ┐キ‖マ┌エ｜ワ｜カワラ｜ナ┐イ｜シ（駅前は変らないし）
W 1 1 1 1

PB：PB は“Parasitic Boundary”の略である。句末境界音調は、原則として表 3-3 上部に示された 5 種類のどれかひとつが選択され、アクセント句末に生じる。しかし自発音声ではふたつの境界音調が隣接して継起したと解釈できる発話が生じることがある。その場合、最初の境界音調末尾に BI ラベル PB を付与し、２番目の境界音調には通常の BI を付与している。PB が生じる典型的環境は何らかの語句が引用された場合である。例えば「『そうなの』と訊いた」という発話は、図 3-3 のようなイントネーションで発音されることがある（<pz>はポーズ）。この場合、被引用句「そうなの」の末尾に L%H%が生じているが、後続する引用の助詞「と」も高いピッチで発音されているので、境界音調 H%を保有しているとみなせる。「そうなの」末尾の L%は、図 3-3 下部の矢印のように、直後のH%と上昇句末境界音調を形成すると同時に、2 番目の H%とも上昇句末境界音調を形成していると解釈できる（なお、ここで助詞「と」の F0 最高値の時点に H%と%L が付与されているのは、上記 3.4.3 節 のケース 5）に該当するからである）。なお、PB は注釈層における AYOR, QQ ラベルと共起することが多いので、3.7 節の記述も参照すること。



トーン　%L A　L%　H%　H%%L　L%

ソ┐　ー　|　ナ　|　ノ　||　ト　|　キー　|　タ　|

BI　1　1　PB　2+b　1　3

（　アクセント句１　）（アクセント句２）

図 3-3: PB の使用例

3.6 プロミネンス層のラベリング

プロミネンス層は X-JToBI において新たに設けられた層であり、表 3-6 に示す 4 種類のラベルが記入される。これらのラベルは、句末境界音調の下位分類に関わる情報と無核アクセント句のプロミネンスに関わる情報を提供することによって、トーン層ないし BI 層のラベルの解釈を補助することを目的としている。

表 3-6: X-JToBI のプロミネンスラベル

ラベル	用法	備 考
PNLP	アクセント句次末モーラに生じた非語彙的な卓立	L%HL%と共起
FR	川上蓁の「浮き上がり調」	L%H%と共起
HR	川上蓁の「つり上げ調」	L%H%と共起
EUAP	強調された無核アクセント句	

PNLP： PNLP は Penultimate Non-Lexical Prominence の略である。アクセント句の最後から 2 番目のモーラ（以下では次末モーラと呼ぶ）に非語彙的な卓立（プロミネンス）がおかれる現象をさす。例えば「ディズニーラ┐ン↑ド↓デ」という発話において「ラ」から「ン」にかけてのピッチ下降（┐）はアクセント核に起因しているが、「ド」における上昇（↑）・下降（↓）は、アクセント核とは無関係であるので PNLP として記述される。X-

JToBI ではこの現象を句末境界音調 L%HL%と同一のトーン列の時間的配置が変更されたものと解釈している。すなわち、通常は最終モーラ（「デ」）内部に生じる上昇下降調の句末境界音調が、発話末の２モーラ（「ドデ」）に広がって実現されたという解釈である。通常の L%HL%と PNLP 化した L%HL%とはトーン層のラベルを見るだけでは区別できない。そこで両者を区別するために PNLP 化した L%HL%については、プロミネンス層にラベル PNLP を付与している。ラベルの付与位置はアクセント句末の BI 位置である。

PNLP に関する注意：CSJ では、音声学的には PNLP と同じイントネーション形状を持つ発話であっても、次末モーラにアクセント核が存在する可能性がある場合は、PNLP として解釈しない。例えば「名古屋まで」が「ナ┐ゴヤ↑マ↓デ」と発音されていても PNLP とはしていない。助詞「マデ」が単独の発話では「マ┐デ」のようにアクセント核をもちうるからである。上記の発話は助詞がアクセント句として独立していると解釈して以下のようにラベリングする。このように助詞が韻律的に独立する現象については第 7 章で詳しく分析する。

	ナ┐ゴヤ ｜	マ┐	デ｜
BI		2 or 3	
トーン	%L A	L%%L A	L%

FR：FR は Floating Rise の略であり L%H%(上昇調１)の変種を示す。通常の L%H%では F0 の上昇は専ら発話の最終音節（ないしモーラ）内部で生じるが、FR に分類される変種では、上昇の始端が、次末ないし次々末モーラ（を含む音節）に置かれる。これは、川上(1963)が「浮き上がり調」と呼んだ句末境界音調現象である。上昇調１と浮き上がり調では、トーン層のラベルの種類はともに L%H%であり、ラベルの位置（時刻）だけが相違する。浮き上がり調の句末境界音調をもつアクセント句にはプロミネンス層にラベル”FR”を付与している。ラベルの位置はアクセント句末の BI ラベル位置である。なお、FR は稀にではあるが上昇下降調(L%HL%)に生じることもある。

HR：Hooked Rise の略。このラベルは、やはり川上(1963)が「つり上げ調」と呼んだ句末境界音調のために用意されたものである。川上は、この音調の特徴を、アクセント句末が「～です」「～ます」で終わる発話に生じ、末尾の２モーラが des, mas と１音節に発音され、その始端から急にピッチが上昇するタイプの句末境界音調と説明している。CSJ でも、ほぼこの環境に限ってつり上げ調が観察されている。つり上げ調音調のトーン層ラベルは“L%H%”のままであるが、プロミネンス層にラベル”HR”を記入している。ラベルの位置は句末 BI の終端である。

EUAP：EUAP は Emphasized Unaccented Accentual Phrase の略である。このラベルは強調を受けた無核アクセント句と後続アクセント句との境界に顕著なピッチレンジの差が観察される場合に付与している。ラベルの位置は強調を受けた無核アクセント句終端の BI にあわせる。EUAP が必要とされる理由は、東京方言ではアクセント句が有核であるか無核であるかによって強調の音声学的実現形が異なってくるからである。東京方言の強調は、基本的にはアクセント句のピッチレンジの拡大によって実現される。有核句の場合アクセントに起因する F0 のピークが顕著に高まることによってピッチレンジの拡大が実現されるが、無核アクセント句は、F0 曲線に顕著なピークが存在しないために強調を実現しにくいことが知られている。無核句における強調を実現するためには、以下の手段のいずれかが用いられる。

1) 強調の置かれた無核句のピッチレンジを拡大すると同時に後続するアクセント句のピッチレンジを狭める。
2) 無核句のピッチレンジはほとんど変化させないが、句末に BPM を置くと同時に後続するアクセント句のピッチレンジを狭める。

いずれの場合も、強調を受けた無核句から後続句にかけて、あたかも有核句におけるダウンステップが生じたかのような F0 形状が観察されることになる。その場合強調を受けたとみなされる無核句終端の BI ラベルの位置に EUAP を付与している。

図 3-4 の模式図中 A)-D)は、上記１）の典型例として、ふたつの無核アクセント句が継起した場合の F0 形状を示している。A)はどこにも強調がない場合（ふたつの句が融合している）、B)は無核句１に強調がおかれているが句境界に EUAP の対象となるダウンステップ類似の F0 形状が観察されない場合、C)と D)は EUAP が付与される場合である（図 3-1 には上記 C に該当する EUAP が生じている）。E)と F)は上記２）の典型例として無核句 1 の末尾に L%H%が生じた場合を示している。E)では無核句 2 のピッチレンジが狭められていないので EUAP の対象とはならない。

EUAP に関する注意：

1) 有核句に強調が置かれていることを示すラベルは存在しない。
2) EUAP の対象とした無核句の BI は 2、2+b、2+bp のいずれかであり 3 とはならない。

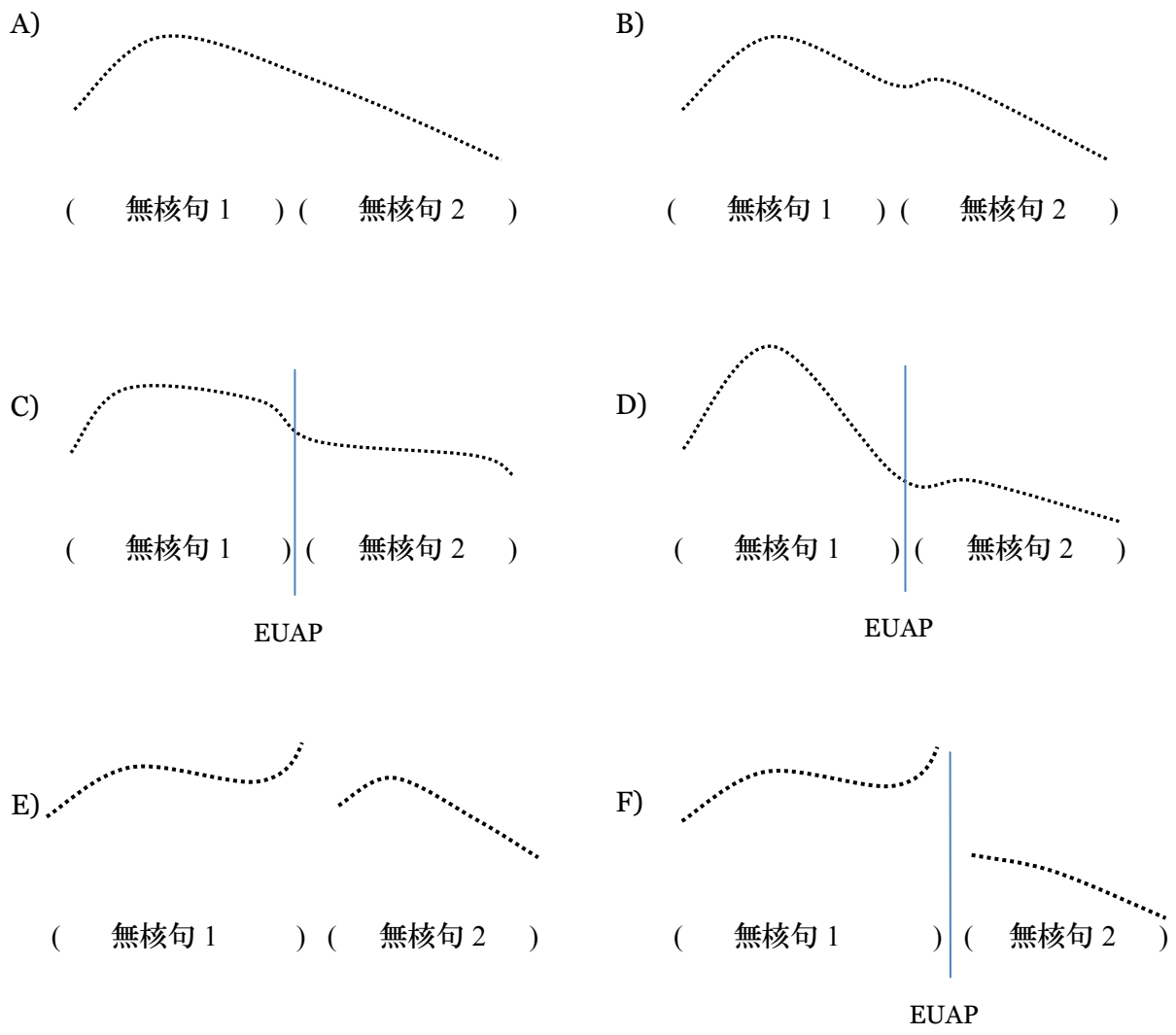


図 3-4: EUAP の適用例

3.7 注釈層のラベリング

注釈層に用いられるラベルは表 3-7 に示した 3 種類である。これらはいずれも、X-JToBI による自発音声ラベリングにおいて十分には解決されていない問題の所在を示している。

表 3-7: X-JToBI の注釈層ラベル

ラベル	用法	備 考
HBP	H で始まるアクセント句	High Beginning Phrase
QQ	いわゆる「半疑問」イントネーション	Quasi-question
AYOR	韻律の解釈に成案を得ていない箇所	At your own risk

HBP：High Beginning Phrase の略。アクセント句の冒頭が例外的に高いトーンで始まっている場合に当該アクセント句冒頭の短単位終端に付与している。「例外的に」とは以下の意味である。音韻論では東京方言のアクセント句は必ず低いトーン(%L)から始まると考えられている。しかし実際にはアクセント句頭で F0 が最高値をとる発話が若干存在しており、その大部分は、下記のいずれかに該当する。

A) アクセントが句頭にあるケース
B) 直前のアクセント句が H%ないし LH% 境界音調で終了しており、その高さが当該アクセント句頭まで持ち越されていると分析できるケース
C) 直前の語が低いトーンで終わっているときに、高いトーンで始まる後続アクセント句冒頭の語が機能語であり、かつ、統語的には先行アクセント句末の語とともにアクセント句を形成すると解釈できるケース（図 3-5 参照）

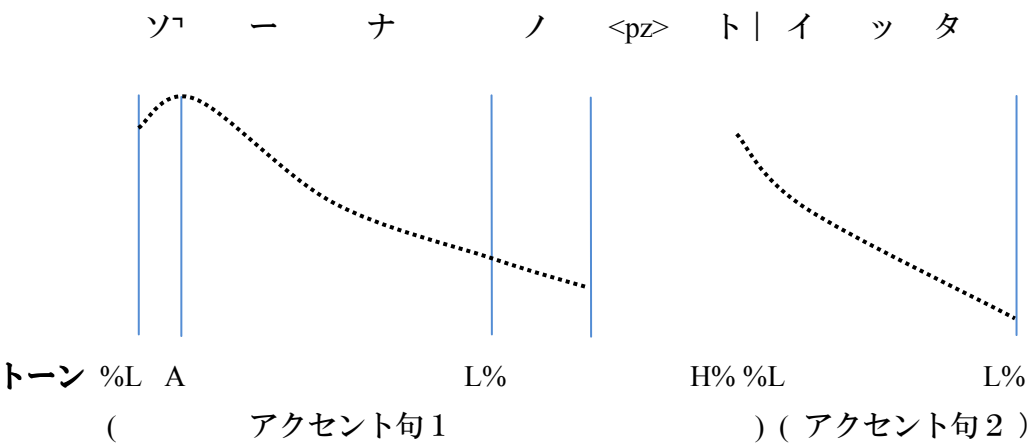


図 3-5: 句頭が高く始まるアクセント句の例（C の場合）

しかし、上記 A)-C)のような合理的説明が不可能なケースも、ごく稀にではあらうが、生じる。例えば講演の冒頭や、長いポーズに続く発話が、いきなり「というわけで」、「ってことで」等で始まり、その際、「ト」や「ッテ」のピッチが顕著に高いことがある。これはちょうど上図の右半分が先行文脈なしに生じるケースである。このような発話は、句頭にアクセントがなく、また、先行するイントネーション上の文脈も存在しないので、発話冒頭に高いピッチとして実現される句頭境界音調が存在するとみなさざるをえない。

CSJ のイントネーションラベリングでは、上例のように、1)講演の冒頭ないしポーズの直後に生じる発話が、2)韻律的には高いピッチで始まっており、3)形態的には機能語（上例では助詞）で始まっている場合に限り、トーン層に境界音調（H%）を付与し、プロミネ

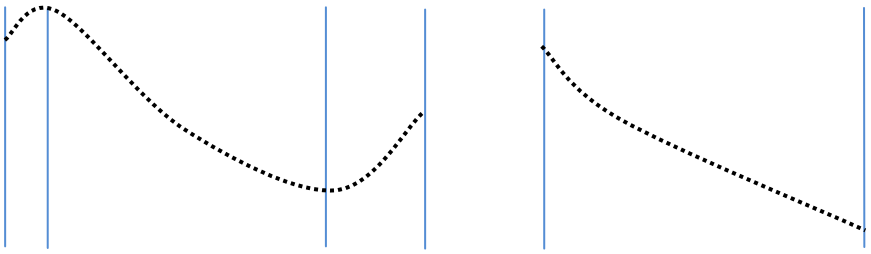
ンス層にラベル HBP を付与している。なお、ラベル（H%）が句頭境界音調でありながら（%H）ではないのは以下の理由による。

上記 1)-3)の条件にかなう発話が生成されるのは、やはり何らかの先行発話が（たとえ表層的には実現されていなくても）話者に意識中に存在しているからであると考えられる。その潜在的発話の影響によって当該発話が高いピッチで始まったという解釈がラベル（H%）によって含意されている。

QQ：Quasi-question の略。俗に「半疑問」「半クエスチョン」と呼ばれるイントネーションが生じたことを示すラベルである。「半疑問」の認定基準は明示化されておらず、ラベリング作業者の直感に依存している。その位置に上昇句末境界音調（H%ないし LH%）が生じていることが必要条件である。ラベル QQ の付与位置は半疑問と判断される句末境界音調が生じたアクセント句の終端である。？で半疑問の音調、｜でアクセント句境界を示せば、「スマート？と｜言うのかな」の場合助詞の「と」の終端である。

QQ はしばしば PB と共に生じる。アクセント区内部に（半疑問の）BPM が生じ、後続する機能語が高いピッチを有する場合である。例として図 3-6 に「農業？<pz>の|衰退が」（<pz>はポーズ）の模式図を示す。この種の発話は下図のようにラベリングするが、PB と共起する場合にかぎり、ラベル QQ はラベル PB と同じ BI 位置に付与している。

ノ┐　ー　ギョ　ー　|　<pz>　ノ|　ス イ タ イ |ガ |



トーン	%L A	L%	H%	H%%L		L%
BI			PB|	2+b|	1| 3|	
注釈			QQ|			
	（	アクセント句１	）	（	アクセント句２	）

図 3-6: QQ の適用例

AYOR：At your own risk の略。いわゆる引用の発話では、X-JToBI によるラベリングに困難が生じることがある。典型的には「～と言う（ことです）」「～って（ことです）」などの発話形式である。

引用表現には、先に説明した PB と QQ が多く生じるが、その他に「という」「ってこと」などの引用句部分が低く（稀には高く）平らなピッチで（あたかもフィラーのように）発音されることもある。これら、引用にかかわるイントネーション特徴の解釈には現在でも十分な成案を得ていないので、イントネーションラベルの利用目的によっては、引用の発話を分析対象から除外する方が賢明な場合もある。そこで、引用とみなされる発話に AYOR ラベルを付与することにした。なお、ここで「引用の発話」と呼んでいるのは、文法的に正確な用法ではなく「典型的には引用の表現に生じる韻律上の特徴を有する発話」という意味である。上に挙げた例のほかに「～と聞いた（んです）」「～みたいな」などの発話にも“AYOR”が付与されていることがある。AYOR ラベルは対象となる発話の最後の BI 位置に付与している。

3.8 非流暢性の BI ラベリングについての補遺

自発音声を朗読音声から区別する顕著な特徴のひとつが非流暢性(disfluency)である。非流暢性については、ここまでの説明で折に触れて言及しているが、もっとも頻繁に生じる非流暢性要素であるフィラーと語断片に関わる BI の処理についてまとめて説明する。

3.8.1 フィラー

自発音声を特徴づける非流暢性(disfluency)現象のうち、もっとも頻度が高いのは、「ア(ー)」「エ(ー)」「アノ(ー)」「アノデスネ」等、言い淀みによって生じるフィラー(filled pauses)である。CSJ の転記作業では、フィラーないし感情表出系感動詞（アー、オー、ホー、ギャ、フーン等）を（F）タグで囲って表記している（対話では応答表現類も（F）で囲われている。（小磯・西川・間淵 2006 参照）。X-JToBI では、転記テキストにおいて(F)タグを付与された発話のうち、以下の基準によって韻律的なフィラーを認定し、通常のアクセント句とは異なる方法でラベリングしている。

1) フィラー候補の韻律的長さが 1 モーラのものは、すべて韻律的フィラーとみなす。

2) 長さが 2 モーラ以上のフィラー候補については、A)句頭の上昇が認められず、かつ、B)発話中にアクセント類似の局所的なピッチ下降が認められないものだけを韻律的フィラーとみなす。

上記の基準にかなう発話は、全体を通してピッチが平坦であり、これが韻律的フィラーの必要条件となっている。なお、アクセント句末で最終モーラの母音が顕著に延長される発音は、機能的にはフィラーの役目を果たしていると考えられるが、韻律的フィラーとは認定していない（ただし転記テキストでは<H>タグが付与されており、X-JToBI ラベルの単語層にもそのまま反映されている）。

韻律的フィラーはその前後を特殊な BI で区切る。フィラーの始端にはラベル<F を、終端にはラベル F を付与している。このうち始端ラベルは 2<F, 3<F, 2+bp<F のように、フィラー区

間の直前・直後に出現する通常のアクセント句間について判定した BI 値と共に記入されている。例えば以下の例は、「サキニ」と「モーシマシタヨーニ」の間の BI 関係が“3”であることを示している。

	サキニ	エー	モーシ	マシ	タ	ヨー	ニ
BI		3<F	F	1	1	2	1

フィラーのピッチは、全体を通して高いか低いかを判定し、それぞれ FH, FL というトーンラベルで表現する。このラベルは対象とするフィラー区間の内部に位置していればよい。複数のフィラーが継起している場合は、下の例のように、各フィラーの終端に BI ラベル F を記入し、各フィラーの高さを FH ないし FL で判定する。この場合も、数字 3 は「サキニ」と「モーシマシタヨーニ」の間の関係を示している。

	サキニ	エー	マー	モーシ	マシ	タ	ヨー	ニ
BI		3<F	F	F	1	1	2	1
トーン		FL	FL					

3.8.2 語断片

語断片(word fragment)とは、「ワ、私が」における「ワ」のように、語（短単位）を発音している途中で発話が中断されることによって生じた語の断片のことである。自発音声には語断片が頻繁に現れ、これに対して消極的な立場をとると、場合によってはほとんどラベルが付与できなくなることがある。そこで、語断片が生じていてもできるだけ多くの情報を記録するよう努めた。

語断片の認定にあたっては、転記テキストにおいて(D)タグが付与されている発話を語断片候補とした。ただし転記テキスト上の語断片と韻律上の語断片とは必ずしも一致しない。統語的・文脈的には断片化している要素であっても、そのイントネーションは滑らかに連続していることがある。その場合、韻律上の語断片を認定する必要はない。

CSJ のラベリングでは、語断片候補があればその前後で BI の判断を実施し、BI が 1 以外の場合（つまり語断片の境界において何らかの韻律上の不連続性が知覚される場合）にかぎり、BI 層に語断片ラベルを付与することにした。語断片の前側境界か後側境界のどちらかの BI が 1 以外ならば（両方でなくてもよいことに注意）、対象とする語断片の前側に<D, 後側に D を付与している（ラベリング例は 3.5.2 参照）。

ちなみに転記テキストでは、（D）タグの他に（D2）タグも用いられている。これは、「話者ごと(D2 の)に」、「言語(D2 に)では」のように、機能語（助詞、助動詞、接頭辞、接尾辞）が機能語で言い直された場合に用いられるタグである(小磯・西川・間淵 2006)。転記

テキストでは上例のように先に出現した機能語に対して(D2)タグを付与するが、韻律的にはむしろ転記テキストで（D2）タグを付与された機能語とそれに後続する機能語との間に不連続性が生じることが多い。そこで、イントネーションラベリングでは、後続要素（上例では「に」と「では」）を語断片候補としている。

語断片に関する注意：

トーン層のラベリングにおいては、語断片に対しても極力トーンラベルを付与している。そのため語断片の周辺では、トーンの文法に従わないラベル系列が現れることがある（例えばアクセント句頭の境界音調が存在しても句末の境界音調が存在しない）。

3.9 分節音ラベル

以上 X-JToBI の韻律ラベルについて解説した。3.2.2 節で触れたように X-JToBI には分節音層があり、分節音ラベルも X-JToBI ラベルの一部として定義されている。CSJ の分節音ラベルは基本的には音素ラベルだが、子音の口蓋化などの音声的な特徴ももっている。

本研究では第 4、 5、6 の各章で分節音ラベルを利用した研究を行なう。関連する分節音ラベルはそのつど詳しく説明するので、ここでは分節音ラベルの一覧表を示すにとどめる。表 3-8 が母音ラベル、表 3-9 が子音ラベル、表 3-10 が補助ラベルである。これらのラベルは ASCII キャラクタの組合せで表現されており，小文字と大文字を区別する。

表 3-8: X-JToBI 分節音層母音ラベル

ラベル	対象	ラベル	対象	ラベル	対象
a	ア	aH	アの長母音*	A	無声化したア
i	イ	iH	イの長母音*	I	無声化したイ
u	ウ	uH	ウの長母音*	U	無声化したウ
e	エ	eH	エの長母音*	E	無声化したエ
o	オ	oH	オの長母音*	O	無声化したオ

* ラベル「H」は長母音の第二要素を表す。

表 3-9: X-JToBI 分節音層子音ラベル

ラベル	対　象
k	カ行子音（軟口蓋無声破裂音）
g	ガ行子音（軟口蓋有声破裂音）
s	サ行のうち「サ，ス，セ，ソ」の子音（歯茎無声摩擦音）
z	ザ行子音（歯茎有声摩擦音ないし破擦音）→　6.4.2 節参照
t	タ行のうち「タ，テ，ト」の子音（歯茎無声破裂音）
c	タ行のうち「ツ」の子音（歯茎無声破擦音）
d	ダ行のうち「ダ，デ，ド」の子音（歯茎有声破裂音）
n	ナ行子音（歯茎鼻音）
h	ハ行のうち「ハ，ヘ，ホ」の子音（声門無声摩擦音）
F	ハ行のうち「フ」の子音（両唇無声摩擦音），IPA ならば[F]
b	バ行子音（両唇有声破裂音）
p	パ行子音（両唇無声破裂音）
m	マ行子音（両唇鼻音）
y	ヤ行（拗音）子音（硬口蓋接近音）
r	ラ行子音（歯茎弾き音）
w	ワ行子音（両唇軟口蓋接近音）
v	ヴァ行子音（唇歯有声摩擦音）
kw	「クヮ」の子音，いわゆる合拗音，IPA ならば[k^w]
gw	「グヮ」の子音，いわゆる合拗音，IPA ならば[g^w]
kj	母音イの前で口蓋化した k：「キ」の子音，IPA ならば[k^j]ないし[c]
gj	母音イの前で口蓋化した g：「ギ」の子音，IPA ならば[g^j]ないし[ɟ]
sj	母音イの前で口蓋化した s：「シ」の子音，IPA ならば[ʃ]
zj	母音イの前で口蓋化した z：「ジ」の子音，IPA ならば[ʒ] ないし[ʤ]
cj	母音イの前で口蓋化した c：「チ」の子音，IPA ならば[ʧ]
nj	母音イの前で口蓋化した n：「ニ」の子音，IPA ならば[ɲ]
hj	母音イの前で口蓋化した h：「ヒ」の子音，IPA ならば[ç]
ky	カ行拗音の子音
gy	ガ行拗音の子音
sy	サ行拗音の子音
zy	ザ行拗音の子音
cy	タ行拗音の子音（「チャ」「チュ」「チョ」の子音）
ty	「テャ」，「テュ」，「テョ」の子音，IPA ならば[t^j]
dy	「デャ」，「デュ」，「デョ」の子音，IPA ならば[d^j]
ny	ナ行拗音の子音
hy	ハ行拗音の子音
Fy	「フャ」，「フュ」，「フョ」の子音，IPA ならば[F^j]
by	バ行拗音の子音
py	パ行拗音の子音
my	マ行拗音の子音
ry	ラ行拗音の子音
N	撥音
Q	促音

表 3-10: X-JToBI 分節音層補助ラベル

ラベル	対　象
<cl>	破裂音ないし破擦音中の閉鎖区間の終端（バースト位置）
<pz>	発話中のポーズ（200ms 未満）
<uv>	声帯振動終了後に母音フォルマントが持続している場合，その終端を示す
<sv>	母音終了後に声帯振動が持続している場合，その終端を示す
<fr>	ボーカルフライが言語音と重畳せずに生じている区間の終端に付与
<fv>	母音であるが，音素としての属性が分類不能のもの（母音不確定音）
<?>	子音であるが，音素としての属性が分類不能のもの（子音不確定音）
<N>	種々の雑音
<b>	呼吸音
#	発話単位の先頭に付与

最後に表 3-11 に現代日本語のモーラ表と CSJ の分節音ラベルの対応関係を示す。この表は現代日本語に生じるモーラを網羅的に示したものであり、そのすべてが CSJ に生じているわけではなく、またすべてのラベルを CSJ のラベリングで利用したわけでもない。特にガ行子音に関する g, G, @の区別を CSJ の分節音ラベリングでは実施していないことは、第 6 章で報告する研究との関係において重要である。

表 3-11: 現代日本語のモーラと CSJ 分節音ラベルの対応表

ア	イ	ウ	エ	オ	ヤ	ユ	ヨ	イェ		
a	i	u	e	o	ya	yu	yo	ye		
カ	キ	ク	ケ	コ	キャ	キュ	キョ	クヮ		
ka	kji	ku	ke	ko	kya	kyu	kyo	kwa		
ガ	ギ	グ	ゲ	ゴ	ギャ	ギュ	ギョ	グヮ		
ga	gji	gu	ge	go	gya	gyu	gyo	gwa	（有声破裂音）	
Ga	Gji	Gu	Ge	Go	Gya	Gyu	Gyo	Gwa	（有声摩擦音）	
カ゜	キ゜	ク゜	ケ゜	コ゜	キ゜ャ	キ゜ュ	キ゜ョ	ク゜ヮ		
@a	@ji	@u	@e	@o	@ya	@yu	@yo	@wa（軟口蓋鼻音=鼻濁音）		
サ	シ	ス	セ	ソ	シャ	シュ	ショ	シェ	スィ	
sa	sji	su	se	so	sya	syu	syo	sye	si	
ザ	ジ	ズ	ゼ	ゾ	ジャ	ジュ	ジョ	ジェ	ズィ	
za	zji	zu	ze	zo	zya	zyu	zyo	zye	zi	
タ	チ	ツ	テ	ト	チャ	チュ	チョ	ティ	トゥ	テュ
ta	cji	cu	te	to	cya	cyu	cyo	ti	tu	tyu
					ツァ	ツィ	ツェ	ツォ	チェ	
					ca	ci	ce	co	cye	
ダ	デ	ド						ディ	ドゥ	デュ
da	de	do						di	du	dyu

表 3-11（続き）

ナ	ニ	ヌ	ネ	ノ	ニャ	ニュ	ニョ	ニェ	
na	nji	nu	ne	no	nya	nyu	nyo	nye	
ハ	ヒ	フ	ヘ	ホ	ヒャ	ヒュ	ヒョ	ヒェ	
ha	hji	Fu	he	ho	hya	hyu	hyo	hye	
					ファ	フィ	フェ	フォ	フュ
					Fa	Fi	Fe	Fo	Fyu
バ	ビ	ブ	ベ	ボ	ビャ	ビュ	ビョ		
ba	bi	bu	be	bo	bya	byu	byo		
パ	ピ	プ	ペ	ボ	ピャ	ピュ	ピョ		
pa	pi	pu	pe	po	pya	pyu	pyo		
マ	ミ	ム	メ	モ	ミャ	ミュ	ミョ	ミェ	
ma	mi	mu	me	mo	mya	my	myo	mye	
ラ	リ	ル	レ	ロ	リャ	リュ	リョ		
ra	ri	ru	re	ro	rya	ryu	ryo		
ワ				ヲ	ウィ	ウェ	ヲ		
wa				o	wi	we	wo		
撥音	促音	長母音（第二要素）							
N	Q	H							

3.10 ラベリング精度

X-JToBI のラベリング精度については菊池他(2003)に報告がある。この研究では、CSJ から選択された自発音声のサンプルを、J_ToBI と X-JToBI とでラベリングし（ラベラーは 3 名）、両者の精度を κ（カッパ）で比較した。κ は、期待される（偶然の）一致率を P(E)、観察された一致率を P(O)とすると、

$$\kappa = \{ P(O) - P(E)\} / \{ 1 - P(E)\}$$

で与えられる。

BI 層ラベルの場合、J_ToBI が κ=0.73(P(O)=0.78, P(E)=0.40)であったのに対し、X-JToBI では κ=0.78(P(O)=0.83, P(E)=0.37)であった。ラベル数が増加しているにも関わらず、ラベリング精度が向上しているところに、X-JToBI の優越性が表れている。

次にトーン層ラベルの代表として、BPM ラベルの一致度を検討したところ、J_ToBI では κ=0.41(P(O)=0.53, P(E)=0.19)、X-JToBI では κ=0.61(P(O)=0.68, P(E)=0.18)であり、この場合も、X-JToBI を利用することによってラベリング精度が大幅に向上することがわかった。

さらに本研究の 6.2 節では有声破裂音の閉鎖調音のラベリングについて 2 名による判定結果が、また 7.2.4 節では 2 名のラベラーが助詞のアクセントを判定した結果がそれぞれ報告されている。

3.11 結論

自発性の高い日本語（東京方言）音声に韻律アノテーションを施すために、J_ToBI を大幅に拡張した X-JToBI を提案した。X-JToBI は 2002 年の発表後、日本語韻律の研究に広く利用されてきている。X-JToBI は J_ToBI の上位仕様なので、X-JToBI をさして Japanese ToBI と呼ぶ研究者もいる(http://web.mac.com/jen.venditti/iWeb/Site/Japanese%20ToBI.html)。

X-JToBI ラベルは、第 4 章以下の研究で重要な役割を果たす。特に第 10 章の研究では、X-JToBI において拡張されたラベル群の有益性が、韻律特徴によるレジスターの判別問題を扱うなかで客観的に示される。

第Ⅱ部

第 4 章. 母音の無声化

4.1 問題

本章と続くふたつの章では、日本語自発音声の分節的特徴をとりあげて分析する。本章ではまず母音の無声化現象をとりあげる。日本語では、前後を無声子音に挟まれた環境において狭母音（/i/と/u/）が無声化することがある。つまり本来有声音である母音の調音時に声帯が振動せず無声音として生成される。

母音の無声化は日本語に生じる副次調音(secondary articulation)のひとつとして、よく知られた現象であり、日本語の音声をあつかった教科書類ではほぼ例外なく言及されている現象である。しかし、この現象には音声学的ないし言語学的にみてよくわからないところが少なくない。例えば以下の無意味語の無声化を考えてみる。アクセントはすべて平板型に統一し、「あの____が」というキャリアセンテンスに入れて発音したとする。

チツ　　チシ　　チキ
シツ　　シシ　　シキ
キツ　　キシ　　キキ

教科書的な説明に従えば、これらの語の第 1 モーラの母音は等しく無声化することになる。しかし実際には、後に CSJ のデータで確認するように（4.3 節）、上のマトリクスの第 2 列中の語は、他の 2 列の語に比べて著しく無声化を起こしにくい。この単純な事実を把握するためには、ある程度のデータをまとめて分析する必要があるが、従来そのような研究は十分には行なわれてきていない。

母音の無声化の研究史をふりかえってみても、佐久間(1929)は内省に基づく調音音声学的な観察とカイモグラフによる観察（外部から喉頭に接触させたカンチレバーによる声帯振動の記録）に依拠しており、どの程度の被験者を観察したかは不明である。Han(1962)はサウンド・スペクトログラフによる声帯振動の観察を行なっているが、やはり被験者数は不明である。前川(1989)の一部に報告されている鳥取方言のデータは、29 名のデータをサウンド・スペクトログラフで分析しているが、分析対象音声は共通語で表記された文章の朗読音声である。Tsuchida(1997), Kondo(2005), Hirayama(2009)なども、理論的関心から実験データ（朗読音声）を扱った研究であり、無声化そのものを幅広く観察しようとしたものではない。

結局、標準語（東京語）の自発音声における母音の無声化の生起実態は、これまで十分精密かつ広範囲に観察されたことがないと言ってよい。言語的／社会的要因によって確率的な変動をともなうことが知られている現象を議論するためのデータとして、これは満足すべき状態ではない。

以下本章では、CSJ に含まれる大量の無声化データを分析することによって、母音無声化現象の確率的性質を再確認し、次いで、典型的な無声化環境においても無声化率が 100%に達しない原因、すなわち無声化が回避されるメカニズムの一端を解明することを試みる。本

章前半（4.3.3 節まで）で紹介する研究成果は、Maekawa and Kikuchi(2005)に報告したものであり、当時構築中であった CSJ-Core のサブセットを利用している。このサブセットは 41 個の学会講演 と 44 個の 模擬講演から構成されており、427973 個の母音を含んでいる。話者は女性 29 名、男性 56 名である。

また以下では、言語学の習慣に従って、音声の音韻表記はスラッシュで囲んで表示し、音声表記は四角括弧で囲んで表示する。「土」「滑車」の音韻表記は/cuci/, /kaQsya/、音声表記は[tsutʃi]、[kaʃ:a]である。以下、本章の音声記号では、母音は簡略表記を用いる。

4.2 母音無声化の概観

最初に、日本語のすべての母音について、無声化がどの範囲に生じているかを概観する。表 4-1 は以下の分析で利用する分節音ラベルの一覧である。表中 plain consonants は口蓋化のない普通の子音（五十音表のイ段以外の直音子音）であり、phonetically palatalized consonants は、音声学的な口蓋化が生じた子音、すなわち母音/i/の直前にあってともに 1 個のモーラを構成する子音（イ段直音モーラの子音）である。また phonologically palatalized consonants は、キャ、チュ、ヒョなどの拗音モーラ中の子音である。/H/, /Q/, /N/はいわゆる特殊モーラに関する記号であり、/H/は長母音の後半要素、/Q/は促音、/N/は撥音に該当する。

表 4-2 に、データに含まれるすべての母音について、生起頻度（N）と有声、無声の内訳、及び無声化率を示した。この無声化率は、CSJ-Core の母音ラベルに記録されている無声化の有無（表 3-8 参照）を集計したものである。無声化率は短母音の/i/と/u/において最も高い率をとり、長母音(/aH/, /eH/, /iH/, /oH/, /uH/)においては、おしなべて非常に低い。

表 4-3 は、表 4-2 に示したデータのうち、前後に子音を伴っている母音の無声化率を、先行子音（C1）及び後続子音（C2）との関係において集計した結果である。ここからは以下の重要な事実を読みとることができる。

第一に、典型的な無声化環境、すなわち狭母音が前後を無声子音に囲まれた/CoVcCo/環境における無声化率は 100％ではなく、90％にも達していない。

第二に、無声子音直後の狭母音の直後に有声子音をともなう/CoVcCv/環境においても、狭母音はかなりの率で無声化している（/i/は 17.4％、/u/は 20.9％）。

第三に、非狭母音も前後に無声子音をともなう/CoVncCo/環境においては、無声化することがある（無声化率は/a/が 2.1%、/e/が 3.3%、/o/が 3.5%）。

最後に、母音の無声化が最も生じにくいと考えられる環境、すなわち/CvVncCv/（前後に有声子音をともなう非狭母音）においてすら、わずかではあるが無声化が生じている。

これらの現象は母音無声化の確率現象としての特性を示すものであり、いずれも検討に値する現象であるが、本章では主に第一の問題を検討する。典型的な環境における無声化こそ、従来の分析が対象としてきた無声化であり、CSJ を利用することによって、そこにどのような新知見をくわえることができるかを示すためには、この環境における無声化を詳しく

検討することが有益と考えられるからである。/CoVcCv/,/CoVncCo/,/CvVncCv/の各環境における無声化の実態に興味のある読者は Maekawa and Kikuchi (2005)を参照されたい。

表 4-1: 本章で利用する分節音ラベルの整理

Vowels:

a, i, u, e, o (有声)
A, I, U, E, O (無声)

Plain consonants:

k, g, G[ɣ], @[ŋ], s, z, t, c [ts], d, n, h, F, b, p, m, r[ɾ], w, y[j]

Phonetically palatalized consonants:

kj, gj, Gj, @j, sj[ʃ], zj[ʒ], cj [tʃ], nj[ɲ], hj[ç]

Phonologically palatalized consonants (*yoo'on*):

ky, gy, Gy, @y, sy, zy, cy, ny, hy, by, py, my, ry

Moraic phonemes:

H (長母音後半要素　long vowel)
Q (促音 geminate)
N (撥音 moraic nasal)

Vowel class:

Vc　(狭母音 closed vowels)
Vnc (非狭母音 non-closed vowels)

Consonant class:

Co (無声子音 voiceless consonants)
Cv (有声子音 voiced consonants)

表 4-2: データにおける各種母音の頻度と無声化率

VOWEL	N	有声	無声	%無声
/a/	109624	108432	1192	1.09
/aH/	3956	3954	2	0.05
/e/	58154	57401	753	1.29
/eH/	12363	12361	2	0.02
/i/	75581	60675	14906	19.72
/iH/	2650	2646	4	0.15
/o/	88412	87282	1130	1.28
/oH/	19445	19437	8	0.04
/u/	49448	33917	15531	31.41
/uH/	8340	8307	33	0.40

表 4-3 : 先行子音(C1)と後続子音(C2)と単母音無声化率の関係*

VOWEL	C1	C2	有声	無声	%無声
/a/	Co	Co	12214	262	2.10
	Co	Cv	18570	92	0.49
	Cv	Co	24943	481	1.89
	Cv	Cv	19867	29	0.15
/e/	Co	Co	5550	190	3.31
	Co	Cv	10890	116	1.05
	Cv	Co	11552	323	2.72
	Cv	Cv	11388	29	0.25
/i/	Co	Co	1475	12124	89.15
	Co	Cv	10556	2219	17.37
	Cv	Co	9200	126	1.35
	Cv	Cv	12072	133	1.09
/o/	Co	Co	12247	437	3.45
	Co	Cv	19752	365	1.81
	Cv	Co	14650	13	0.09
	Cv	Cv	16802	14	0.08
/u/	Co	Co	1732	9267	84.25
	Co	Cv	11851	3133	20.91
	Cv	Co	5562	127	2.23
	Cv	Cv	7748	61	0.78

*Co: 無声子音, Cv: 有声子音

4.3 典型的環境における無声化の問題点

典型的な無声化環境である/CoVcCo/環境において解明すべき問題は、この環境における平均無声化率が 100%よりも低く、90%にも達しない原因の解明である。表 4-4 と表 4-5 は、母音/i/と/u/の無声化状況を前後の子音環境によって詳細に分類した結果である。これらの表においては音韻論的な子音分類を採用しているので、表 4-1 に示されている音声学的な異音の一部は音素に統合されていることに注意してほしい。また、生起頻度が 10 未満の C1-C2 の組み合わせをデータから除外していることにも注意してほしい。

表 4-4: /CoVnCo/環境における/i/の無声化率*

VOWEL	C1	C2	有声	無声	%無声
/i/	/c/	/c/	16	73	82.02
		/h/	35	7	16.67
		/k/	31	358	92.03
		/p/	7	44	86.27
		/Q/	16	16	50.00
		/s/	64	41	39.05
		/t/	32	181	84.98
	/h/	/c/	5	80	94.12
		/h/	22	9	29.03
		/k/	15	342	95.80
		/Q/	21	39	65.00
		/s/	11	3	21.43
		/t/	21	883	97.68
	/k/	/c/	19	62	76.54
		/h/	167	65	28.02
		/k/	73	476	86.70
		/Q/	32	51	61.45
		/s/	144	262	64.53
		/t/	53	791	93.72
	/p/	/Q/	118	9	7.09
	/s/	/c/	7	259	97.37
		/h/	47	14	22.95
		/k/	50	1102	95.66
		/Q/	25	92	78.63
		/s/	259	178	40.73
		/t/	49	6507	99.25
	/t/	/k/	11	0	0.00
		/Q/	13	0	0.00

* C1:先行子音, C2:後続子音

表 4-5: /CoVnCo/環境における/u/の無声化率*

VOW	C1	C2	有声	無声	%無声
u	/c/	/c/	16	57	78.08
		/h/	24	10	29.41
		/k/	44	872	95.20
		/Q/	13	32	71.11
		/s/	137	140	50.54
		/t/	19	207	91.59
	/h/	/c/	4	86	95.56
		/h/	17	16	48.48
		/k/	15	227	93.80
		/Q/	25	7	21.88
		/s/	6	46	88.46
		/t/	10	248	96.12
	/k/	/c/	48	123	71.93
		/h/	132	56	29.79
		/k/	151	246	61.96
		/p/	3	21	87.50
		/Q/	114	26	18.57
		/s/	380	1202	75.98
		/t/	148	1021	87.34
	/p/	/k/	8	7	46.67
		/s/	12	18	60.00
		/t/	6	12	66.67
	/s/	/c/	3	8	72.73
		/h/	4	8	66.67
		/k/	31	2207	98.61
		/p/	2	154	98.72
		/Q/	23	31	57.41
		/s/	60	195	76.47
		/t/	37	1210	97.03

* C1:先行子音, C2:後続子音

4.3.1 子音調音様式の影響

表 4-4, 4-5 は、母音無声化における先行ないし後続子音の調音様式の重要性を示している。これは先行研究でも指摘されてきた要因である（前川 1989, 吉田 2002, 藤本・桐谷 2003）。表 4-4, 4-5 を調音様式の観点から整理しなおすと、表 4-6, 4-7 の結果を得る。

表 4-6: /CoVnCo/環境における/i/の無声化率(%)に及ぼす子音の調音様式の影響

		C2		
		破擦音	摩擦音	破裂音
C1	破擦音	81.1	33.3	89.4
	摩擦音	96.3	38.1	98.4
	破裂音	80.2	51.5	89.3

表 4-7: /CoVnCo/環境における/u/の無声化率(%)に及ぼす子音の調音様式の影響

		C2		
		破擦音	摩擦音	破裂音
C1	破擦音	77.2	48.1	94.5
	摩擦音	95.1	61.2	97.5
	破裂音	80.8	74.0	80.1

表 4-6, 4-7 における無声化率は、いずれの表においても C1 が摩擦音(fricative)で C2 が破裂音(stop)である場合に最高値をとっている。反対に無声化率が最低になるのは、いずれの表においても、C1 が破擦音(affricate)、C2 が摩擦音の場合である。それに次いで無声化率が低い環境が C1, C2 がともに摩擦音の場合である点もまたふたつの表に共通している。

表 4-6, 4-7 は、C1 と C2 の調音様式の間に交互作用が存在していることを示唆している。実際、C1 と C2 の調音様式を主要因とする二元配置分散分析を/i/と/u/をまとめたデータに適用すると、主要因とともに交互作用にも有意差が認められた（C1 の主効果は DF=2, F=44.38, P<0.0001。C2 の主効果は DF=2, F=1959.43, P<0.0001。交互作用項 C1*C2 は DF=4, F=263.24, P<0.0001）。

なお、表 4-6, 4-7 の集計においては、C2 が促音(/Q/)であるサンプルを集計から除外している。/Q/自体は音韻論的には調音様式を指定されないことがひとつの理由であるが、次節で論じるように/Q/が後続する音声環境は特別な扱いを必要とすると思われることがもうひとつの理由である。

4.3.2 促音の影響

表 4-8 に C2 が促音であるかないかによって狭母音全体の無声化率を比較した結果を示す。この表からわかるように、C2 が促音であると、C1 の調音様式によらず、無声化率が顕著に低下する。さらに詳しく検討すると、無声化率は C1 が摩擦音で、後続音声環境によって決

定される促音の調音様式が破裂音である場合に最高値をとり、反対に C2 と促音がともに摩擦音である場合に最低値をとることがわかる。これは先に表 4-6, 4-7 に観察された傾向と一致している。

表 4-8: 後続促音の有無が無声化に及ぼす影響（/i/と/u/をプールして計算）

	C2 が/Q/以外			C2 が /Q/		
C1	有声	無声	%無声	有声	無声	%無声
破擦音	454	2,021	81.7	29	49	62.8
摩擦音	860	14,099	94.3	112	181	61.8
破裂音	1,464	4,954	77.2	282	87	23.6

4.3.3 無声化連鎖の影響

典型的な無声化環境である/CoVcCo/が連鎖することがある。/…CoVcCoVcCoVcCo…/のように、ある環境における C2 が後続環境における C1 として機能する環境である。この環境は日本語に頻繁に生じている。例えば「菊池」/kikuci/、「色紙」/sikisi/、「ススキ」/susuki/、「知識」/cisiki/、「複式」/hukusiki/などの名詞はいずれもこの環境に属しており、最初の２ないし３母音が典型的な無声化環境におかれている。

これらの無声化連鎖環境にあっては、隣接するふたつの狭母音が連続して無声化することもあるが(Maekawa 1990)、無声化を回避する傾向が生じることが古くから指摘されている（佐久間 1929）。この回避傾向が実在するならば、典型的な無声化環境における無声化率を低下させる要因になっているはずである。

無声化連鎖は、上例のような語中だけでなく、無声化環境中に語境界を挟んだ場合（「岸から」「父しか」の類）にも生じるが、ここでは Kondo(1997)が指摘する語境界の影響を避けるために語中の環境だけを対象として検討を進めることにする。今回のデータには語中に無声化連鎖が生じうるサンプルが 318 個含まれていた。

表 4-9 はこれらのサンプルにおける無声化連鎖環境に含まれる最初の 2 個の母音における声の分布を示している。全 318 サンプル中、無声化連鎖が生じていたのは 84 サンプル(26.4%)だけであり、他のサンプルではいずれかひとつの母音の無声化が回避されるか、２母音がともに有声で発音されていた。無声化の連鎖を回避しようとする傾向が存在していることが確認できる。

表 4-9: 無声化連鎖環境中の最初の 2 母音における声の分布

第 1 母音 \ 第 2 母音	有声	無声
有声	17	44
無声	171	84

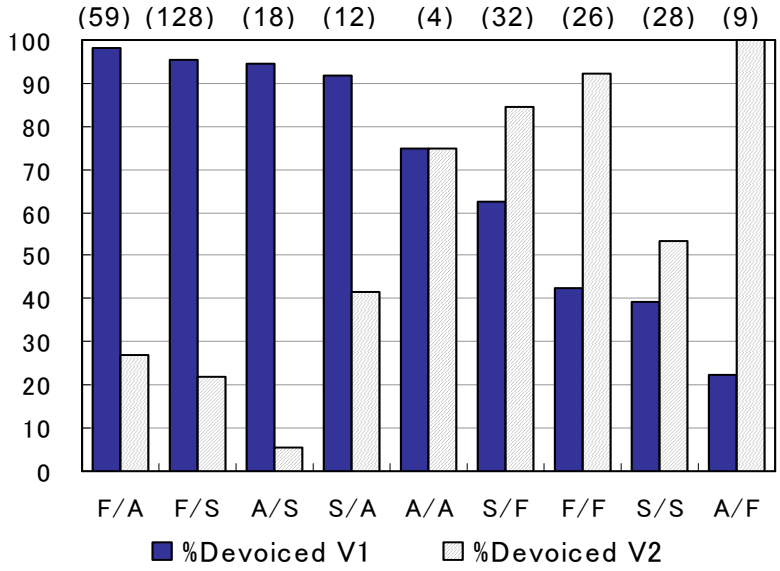


図 4-1: 無声化連鎖環境中の 2 母音の無声化率
括弧中の数字はサンプル数

無声化連鎖の回避傾向を知るために、連鎖環境中の 2 モーラの子音の調音様式に注目して 2 母音の無声化率を表示したのが図 4-1 である。横軸は C1 と C2 の組み合わせを示しており、第 1 母音の無声化率の降順にソートされている。横軸に用いられている記号 A, F, S はそれぞれ「破擦音 Affricates」「摩擦音 Fricatives」「破裂音 Stops」を意味している。例えば「靴下」/kucusita/の最初の 2 母音を例とすれば、最初の母音/u/の C1 は破裂音/k/、C2 は破擦音/c/であるから F/A となり、二番目の母音/u/の C1（すなわち最初の母音の C2）は破擦音/c/、C2 は摩擦音/s/であるから A/S となる。

図 4-1 は、隣接する 2 個の母音の無声化率はおよそのところ反比例の関係にあることを示しており、無声化連鎖の回避傾向が実在することを示している。また、子音調音様式の組み合わせにおける無声化率の高低は、表 4-6、表 4-7 に観察された傾向と一致している。表 4-10 は、図 4-1 における第 1 母音の無声化率と表 4-6、表 4-7 において対応する無声化率とを、/i/と/u/をプールして比較した結果である。この事実は、無声化連鎖環境における無声化も基本的には、無声化全般と同一の原則に則っていることを示唆している。

表 4-10: 無声化連鎖中の第 1 モーラと通常の無声化の無声化率(%)の比較

組み合わせ	図 4-1	表 4-6, 4-7
F/A	98.3	95.9
F/S	95.3	98.1
A/S	94.4	92.6
S/A	91.6	80.7
A/A	75.0	79.3
S/F	62.5	68.6
F/F	42.3	48.9
S/S	39.2	84.4
A/F	22.2	43.1

表 4-11: トーンの影響

トーン	有声	無声	%無声
%L	52	540	91.2
A	82	42	33.9
H%	20	0	0.0
L%	112	281	71.5
No tone	187	3,009	94.1

4.3.4 韻律の影響

典型的無声化環境における無声化率を低下させる要因に発話の韻律特徴の影響がある。表 4-11 は X-JToBI トーンラベルとの関係において無声化率を集計した結果である。/i/と/u/をプールした結果である。

アクセント核(A)をもつモーラでは無声化率が顕著に低下している。またアクセント句末に上昇 BPM（H%、 9 章参照）が生じると無声化は完全に阻止されている。一方、アクセント句の冒頭（%L）及び句末(L%)を示す境界トーン(boundary tone)は、無声化にほとんど影響を及ぼしていない。

4.4 議論

以上、本章では CSJ-Core における母音無声化の実態を分析し、次いで典型的無声化環境における無声化率を低下させていると思われる言語的要因を検討した。本節では、そのうち子音の調音様式の問題に着目して無声化回避のメカニズムを議論する。

4.4.1 先行研究

子音の調音様式が無声化率に影響を及ぼしうることは先行研究においても指摘されている。武田・桑原(1987)は、母音無声化率は全般的に先行子音(C1)が摩擦音のときに上昇するとい

い、また、モーラ「シ」に含まれる母音の無声化率は、後続子音が閉鎖音のときに最高値をとると報告している。吉田・匂坂(1990)も、先行子音に無声子音をもつ狭母音の無声化率は、後続子音が閉鎖音のときに最高値をとると述べている。しかし、これらの研究では先行子音と後続子音の影響が別個に論じられているため、両者の交互作用については多くを知ることができなかった。

近年、吉田(2002)と藤本・桐谷(2003)は、母音に隣接する子音の交互作用を分析して、本章と類似した結論に達している。しかしこれらの研究は、すべての子音環境を検討したものとはなっていない。吉田の実験は/k/と/s/だけが扱われており、藤本と桐谷の実験も/k, t, s, h/を扱うにとどまっている。また以上すべての先行研究に共通する問題として、特定の調音様式の組合せが何故無声化率の低下をもたらすのかについての説明が与えられていない。

本章の分析結果は、摩擦音と摩擦音の組合せ、及び、破擦音と摩擦音の組合せにおいて、無声化率が著しく低下することを明らかに示していた。本章の分析は、先行研究よりも広い範囲の子音環境を、より多くの被験者の発話について検討しており、さらに、何の統制もくわえられていない自発音声の分析であるから、その結果には高い信憑性があると考えられる。

そして、上述の無声化率低下現象は、モーラ境界知覚の容易性という観点から解釈すれば、自然に説明できると考えられる。

4.4.2 モーラ境界知覚の容易性

狭母音と摩擦音もしくは破擦音から構成される CV モーラに無声化が生じるとき、音声学的にみると先行子音（破擦音においては後半部）の延長が生じる。つまり無声化した母音に該当する時間区分には先行子音（破擦音においては摩擦要素）と同一の摩擦ノイズが生じる。

その結果、摩擦音と摩擦音ないし破擦音と摩擦音の組合せからなる無声化環境のモーラに無声化が生じると、第 1 モーラに属する摩擦ノイズと第 2 モーラに属する摩擦ノイズが隣接して[s.s], [s.h], [s.F], [s.ʃ], [ts.s], [ts.h], [ts.F], [ts.ʃ]（ドットはモーラ境界）のような音声連続が生じるが、この種の音声連続においてモーラ境界を知覚することには困難が伴う。表 4-6, 表 4-7 において、摩擦音と摩擦音ないし破擦音と摩擦音の組合せにおける無声化率が低下していたのは、この知覚上の困難を低減するためではないかと想像される。

同様の困難は、先行子音が破裂音、後続子音が摩擦音の場合にも生じる。先行モーラの母音に無声化が生じると、モーラ境界は、破裂音の破裂後に生じる帯気音(aspiration)と後続する摩擦音とによって形成されるからである。しかし、この組合せにおけるモーラ境界の知覚は、摩擦音と摩擦音ないし破擦音と摩擦音の組合せに比べれば、相対的に容易であると考えられる。何故ならば、破裂音の存在自体は、閉鎖の破裂に伴うバースト音の存在によって容易に知覚することができ、また[t^h.s], [p^h. F]などの音声連続においても帯気音（上付きの h で表す）と摩擦音との音質上の相違は、[s.s], [F.F]などよりも大きいと思われるからである。実際、表 4-6、表 4-7 の無声化率は、この予測に合致する値をとっている。

以上のケースに比べると、[s.t], [F.t], [ʃ.k], [ts.t]等、摩擦音ないし破擦音に破裂音が後続する環境における無声化では、モーラ境界が無音区間の存在によって明示されるので、モーラ境界の知覚は著しく容易である。表 4-6、表 4-7 において、摩擦音ないし破擦音に破裂音が後続する環境における無声化率が高いのは、この環境においては、無声化がモーラ境界を曖昧化する可能性が低いので、無声化を回避する必要性が低いためであると考えられる。

4.4.3 促音と無声化連鎖の説明

以上の説明は、先に表 4-8 で確認した促音の無声化率低減効果にも適用できる。よく知られているように、日本語の促音は、「河童」[kap:a]、「買った」[kat:a]、「作家」[sak:a]、「発作」[hos:a]、「滑車」[kaʃ:a]のように、子音の持続時間を延長して一種の長子音を作り出すことによって生成される。

破裂音の促音では、閉鎖調音の延長によって、モーラ境界の知覚は完全に曖昧化されている。つまり延長された閉鎖区間の音声だけを聴取しても、そこにはモーラ境界の存在の手がかりは一切ふくまれておらず、それが促音の重要な音韻的特徴となっている。

典型的無声化環境において、後続子音が促音であるモーラ、例えば「切符」/kiQpu/、「きっと」/kiQto /、「オリンピック」/oriNpiQku/、「失政」/siQseH/等に無声化が生じると、隣接する 2 個のモーラ境界がともに曖昧化することになり、通常よりも一層大きな知覚上の困難が生じる。例えば「ひっそり」/hiQsori/の第 1 モーラに無声化が生じると[çs:ori]となり、聞き手は硬口蓋無声摩擦音[ç]から歯茎無声摩擦音[s]への音質の変化で最初のモーラ境界を知覚し、次に[s]の持続時間が通常よりも長いことによって促音の（従って第 2 の）モーラ境界を知覚する必要がある。

より極端な例が「失笑」/siQsyoH/において見出される。この語に無声化が生じた音声は[ʃ::o:]である。記号[ʃ::]は、無声後部歯茎摩擦音[ʃ]が通常の促音よりも長いことを表している。この場合、音声の聞き手は[ʃ]の持続時間について、第 1 モーラに無声化の生じていない「支所」[ʃiʃo]、第 1 モーラに無声化が生じた「支所」[ʃ:o]、そして第 1 モーラに無声化が生じ、促音が後続する「失笑」[ʃ::o:]の 3 段階の知覚判断を要請されることになる。このような処理を回避しようとする傾向が日本語に存在することは驚くにあたらないであろう。

これと同じ説明が、無声化連鎖環境における無声化の回避についてもある程度まであてはまる。無声化連鎖が生じると、隣接する 2 個のモーラ境界がともに曖昧化するからである。

4.5 コーパスの再分析

前節で提案したモーラ境界知覚の容易性に基づく無声化回避メカニズムの仮説を検証するために、CSJ-Core の再分析を試みた（河津・前川 2009）。この分析には CSJ-Core の全体を利用した。表 4-12 と表 4-13 にデータ全体の無声化率を示す。当然であるが、表 4-6、4-7 と同一の傾向が読みとれる。

表 4-12: 母音/i/の無声化率(%)*

		C2					
		破擦音	摩擦音		破裂音	Total	
C1	破擦音	89.1	33.1	(39.4)	92.7	81.8	(84.9)
	摩擦音	97.4	37.5	(40.6)	98.8	95.5	(96.1)
	破裂音	84.3	46.2	(61.7)	90.7	76.6	(84.1)
	Total	94.2	41.3	(49.3)	97.4	91.5	(93.5)

*( ) 内は/h/, /hy/が後続する場合を除いた値

表 4-13: 母音/u/の無声化率(%)*

		C2					
		破擦音	摩擦音		破裂音	Total	
C1	破擦音	78.3	48.7	(49.9)	95.9	86.5	(87.1)
	摩擦音	97.5	62.3	(74.2)	98.4	94.7	(96.5)
	破裂音	85.1	76.4	(78.5)	84.3	80.1	(81.3)
	Total	88.4	70.9	(74.6)	95.0	88.0	(89.4)

*( ) 内は/h/, /hy/が後続する場合を除いた値

上述の仮説を定量的に検討するためには、無声化によって隣接することになる子音セグメント間の距離を何らかの方法で表現することが必要になる。今回は、ケプストラム距離尺度を利用して、無声化の生じたモーラ/C1V̥/（V は/i/ないし/u/）と直後の無声子音/C2/とのスペクトルパターンの類似性を定量化した。/C1V̥ /, /C2/それぞれの区間において 12 次元までのメルケプストラム係数を求め，セントロイド間のユークリッド距離を求めた。サンプリング条件は 16kHz/16bit。フレーム長 16ms、フレームシフト長 8ms で、Hamming 窓を利用している。計算には、Mike Brooks 氏が公開している Matlab 用の音声処理ライブラリ VOICIBOX 中の mcepstrum.m を利用した。メルケプストラム距離は音声認識のフロントエンドで多用されることからわかるように、ある程度まで人間の知覚を反映した距離になっていると考えられる。

図 4-2 に前後の子音の調音様式の組合せごとに、母音の無声化率とケプストラム距離との関係を示した。左が/i/、右が/u/の結果であり、図中の A, F, S は破擦音、摩擦音、破裂音を表している。メルケプストラム距離は/CoVCo/ 環境で無声化が生じた場合を対象として計算し、その中央値を◆印で示した。

C1 が破擦音ないし破裂音の場合は、破裂のバースト及び閉鎖終了後の摩擦音的ノイズ区間を対象としてケプストラム距離を求めた。また C2 が破擦音ないし破裂音の場合は、閉鎖区間を対象としてケプストラム距離を求めた。いずれの場合も C1（正確には/C1V̥/－V̥は無声化した/i/又は/u/）ないし C2 に属するフレーム群のセントロイドを求めて C1 ないし C2 の代表値とし、セントロイド間の距離を計算している。

C1 と C2 とのケプストラム距離が小さい場合、つまり類似性が大きい場合には無声化率が低く、逆にケプストラム距離が大きい場合には無声化率が高いことがわかる。無声化率とケプストラム距離との相関係数は母音/i/において 0.88、母音/u/において 0.82 であった。

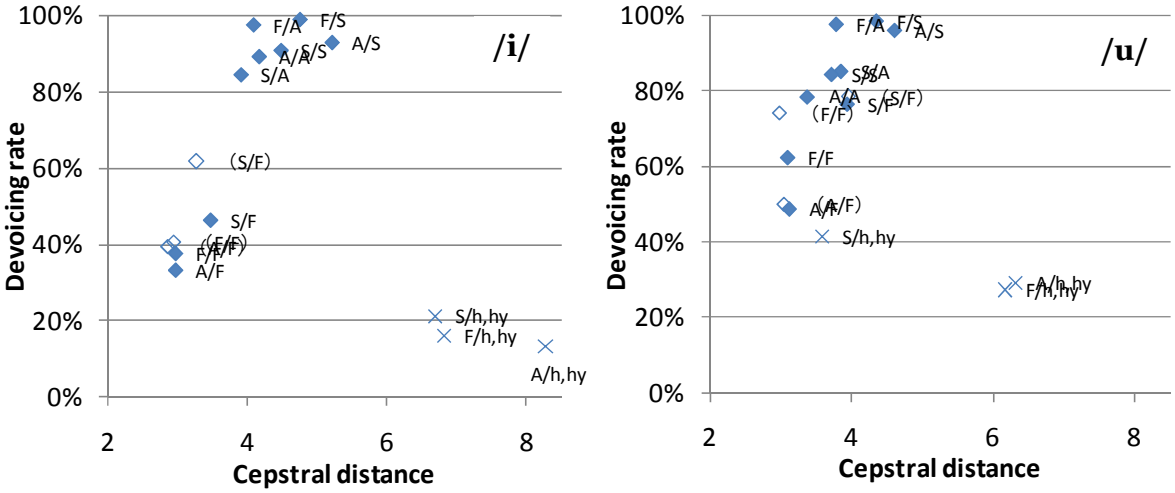


図 4-2: 子音セグメント間の音響距離と無声化率の関係

ただし、C2 が/h/または/hy/[ç]の場合だけは、ケプストラム距離が大きいにも関わらず、無声化率が低い値にとどまっていた（図 4-2 中の×印参照）。一見すると奇異な現象であるが、実はこれらの/h/ないし/hy/の大部分には有声化が生じており（佐久間 1929, 天沼・大坪・水谷 1978）、これが母音の無声化を阻止する原因になっていた。これらの/h/ないし/hy/に関わるサンプルは典型的無声化環境から除外するのが適当と考えられる。

/h/, /hy/が後続する場合を除いて母音の無声化率を再計算した結果が、表 4-12、表 4-13 の( ) 内に示してある。同様に/h/, /hy/が後続する場合を除いて、無声化率とケプストラム距離との関係を計算した結果が、図 4-2 中に◇印で示してある。

以上の分析結果は、C1, C2 間の音響距離が無声化率と密接に関係していることを明らかにした。このような関係の存在は、モーラ境界知覚上の困難が母音無声化回避の要因となっているという上述の仮説を支持する証拠である。

4.6 結論

CSJ-Core に含まれる大量の母音サンプルの分析によって、母音の無声化の実態を検討したところ、従来、典型的な無声化環境とされてきた環境における無声化率は 90%に達しないことが判明した。その原因としては、無声化環境を形成する先行子音（C1）と後続子音（C2）の調音様式の組合せ、無声化連鎖の回避、アクセント核及び上昇イントネーションなどの韻律的特徴などの関与が考えられる。

このうち調音様式の組合せの影響は、無声化によってモーラ境界の知覚が曖昧化する場合に無声化を阻止する機能を果たしている可能性が高い。この仮説の当否を検討するために C1-C2 間の音

響距離と無声化率の相関を検討したところ、両者間に強い相関が存在することが明らかとなり、仮説を支持する結果が得られた。

このような研究成果は、CSJ を利用することによって、先行研究に比べてはるかに広い範囲の子音の組合せを分析できたことによってもたらされたものである。また実験的統制を一切くわえていない自発音声データにおいて首尾一貫した分析結果が得られたことは、分析結果に高い信憑性を与えている。

第 5 章. /z/の調音様式

5.1 問題

日本語の標準語ではザ行モーラの子音音素である/z/の調音様式に変異が観察され、/z/は有声摩擦音ないし有声破擦音のいずれとしても発音される。この事実は良く知られたものであり、日本語音声の教科書にはほぼ例外なく記載されている。しかし、次節で論じるように、この変異現象がどのようなメカニズムに従っているのかについては、よくわからないことが多い。本章では、CSJ-Core の X-JToBI ラベル情報を活用することによって、/z/の変異を律している根本的な要因が何であるのかを検討する。

5.2 先行研究

日本語ないし日本語学の教科書の多くは、/z/の変異を、語中位置との関係で説明している。例えば日本語の入門書である Yamaguchi(2007)は、“/z/ is variably realized as either [dz]or [z], as in [dzak:a](/zaQka/ ‘sundries’) and [kaze] (/kaze/ ‘wind’)” (p.19)と述べて、語頭では破擦音[dz]が、語中では摩擦音/z/が調音されることを示唆している。このような説明は古くから行われており、日本語音声学の教科書として有名な天沼・大坪・水谷(1978)は「「ズ」,「ジ」は，語頭では，ほとんど，いつも[dz], [dʒ]で発音されるが，語中では，環境によって，[z], [ʒ]のように発音されることもある」と述べている(p.71)。/z/の変異を語中位置と関係づける試みは、管見のかぎり、服部(1951)にまで遡ることができる。日本の音声学界に強い影響を及ぼしたこの著作で服部は「日本語の母音に先立たれない「ジ」「ズ」などの子音は、東京方言などにおいて[dʒ][dz]と表すべき破擦音」と述べている (p.184)。

語中位置による説明は日本語研究者全般に広く受け入れられているが、この説明の問題を指摘した研究も少なくともふたつある。川上(1977)は「語頭以外の位置にあり、かつ「ん」の直後にない「ざ、ず、ぜ、ぞ」は[za, z, ze, zo]であることが多い」と述べた直後に「常にそうであるかのように説かれるのが常であるがそれは誤りである」と付け加えている(p.52)。

Vance(1987)は上記の川上の見解を紹介したあとで、”Kawakami’s account is certainly the most accurate for standard speakers that I have observed, although I think [z] may sometime appear word-initially when a pause does not precede”と述べている(p.24)。

上に紹介した研究は、川上と Vance も含めて、語頭と語中では/z/の音声学的な姿に差が生じることを認めている。意見が一致しないのは、/z/の変異がどの程度まで位置によって制御されているかについてである。服部とその追随者は/z/の変異が位置によってほぼ完全に説明できると考えているが、川上や Vance はそう考えていない。

先行研究には、さらにいくつかの問題がある。まず、ひとつの大きな問題は、上に紹介したすべての研究が摩擦音[z]と破擦音[dz]のどちらが音素/z/の本来的な姿であるかという問題について何の考察も加えていないことである。この問題を考察した研究に有坂(1940)がある。有坂は「實際の發音に於ては，[dz]と[z]との間に位する無數の中間音が現れて來るが，それ

らはすべて/dz/を意圖しつつ發音されたものであって、[z]又はそれに近い形で現れてゐるものと雖も，丁寧に發音される場合にはすべて[dz]の形になるのである」と述べている(p.58)。

第二に、有坂まで含めたすべての先行研究は、いわゆる主観的な研究方法を用いており、変異を量的に考察しようとしていない。これは Labov(2001)などに代表される言語変異研究の観点からすれば大きな欠陥である。この観点からすれば、すべての先行研究が観察対象としているのがどのようなレジスターの発話であるかを明らかにしていないことも問題になる。

最後に、やはりすべての先行研究は、「語」とは何かをはっきり説明していない。これは日本語のように語境界の言語学的規定が安定していない言語を研究する場合には、大きな問題を引き起こす可能性のある問題である。

5.3 分析

5.3.1 データ

本章の分析には CSJ-Core の自発モノローグ（学会講演と模擬講演）を利用する。表 5-1 に話者の観点から見たデータの内訳を示す。話者の異なり総数は 123 名であり、年齢、性別とも広い範囲に分布している。また 6 名の話者は学会講演と模擬講演の両方を提供している。

表 5-1: 本章で利用したデータにおける話者の分布

レジスター	N	女性話者*	男声話者*	生年 <1950*	生年 1950s*	生年 1960s*	生年 1970s*	N of /z/
学会講演	56	17(18)	38(38)	2(2)	6(6)	21(22)	26(26)	6627
模擬講演	106	37 (53)	37 (53)	19(19)	11(21)	22(32)	22(34)	7976

*括弧外の数字は話者の異なり人数。括弧内は講演数。

5.3.2 X-JToBI ラベルによる/z/の変異の表現

図 5-1 に CSJ-Core に記録されていた、同一男性話者による 2 種類の「次元」の音声を示す。左パネルは破擦音[dz]として実現された/z/で、「5 次元」の一部として発音されたもの。右パネルは摩擦音[z]として実現された/z/で「12 次元」の一部として発音されたものである。両パネルとも、最上段に音声波形、その下にサウンドスペクトログラムが表示され、その下に X-JToBI の Word ラベル（短単位）と分節音ラベルが表示されている。

この例からわかるように、X-JToBI の分節音ラベリングでは、破擦音[dz]の声道閉鎖区間の末端に<cl>ラベルが付与されており、摩擦音にはこれが付与されていないので、これを利用して破擦音と摩擦音を別個に集計することができる。データ全体での/z/の破擦率は 35%であり、これが以下における比較のベースラインとなる。なお、X-JToBI の分節音ラベル/zj/は音声学的な口蓋化（母音/i/の直前での口蓋化）を意味している（表 3-9 参照）。

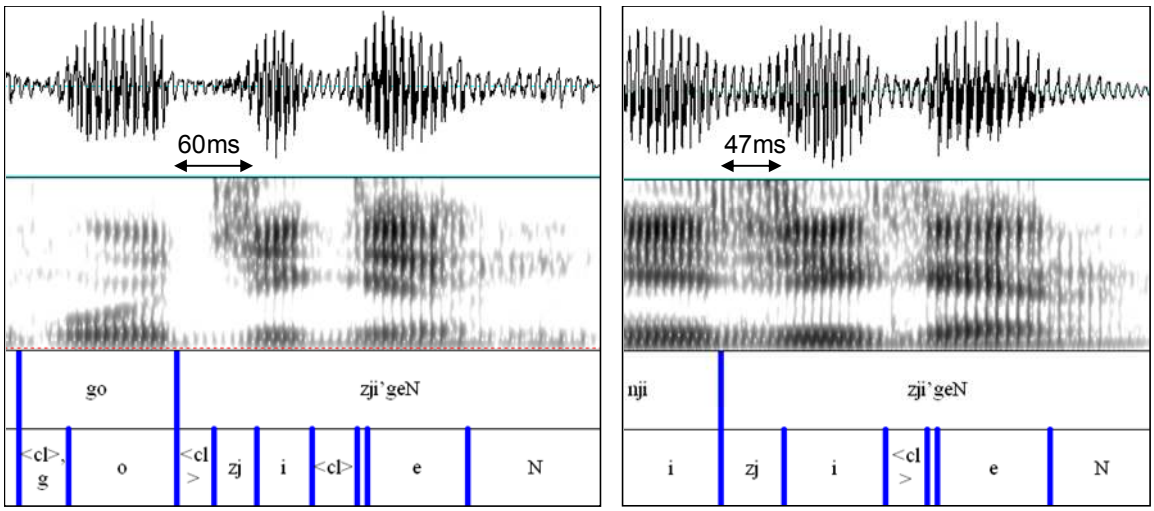


図 5-1：破擦音(左)及び摩擦音(右)として実現された/z/
（図中の矢印は/z/の時間区分を示している）

5.3.3 音素の位置に関する分析

A. 短単位中のモーラ位置の分析

最初に短単位(SUW)中のモーラ位置によって/z/の調音に生じる変異を検討する。例えば「時間」「孤児」「必ず」の各短単位において/z/はそれぞれ第 1、第 2、第 4 モーラに位置している。このような語中位置についての集計結果を表 5-2 に示す。表の列は左から、短単位中のモーラ位置、/z/の摩擦音としての生起数、同じく破擦音としての生起数、破擦率（RAA: Rate of Affricate Articulation 破擦音としての生起数／全生起数×100）である。最後の列（Mean TACA）には 5.5 節で言及する。

表 5-2 ではモーラ位置が 1 の場合、すなわち短単位の語頭において、RAA（破擦率）が最高値をとっている。しかしながら、その絶対値は高々51%にとどまっており、「語頭では破擦音」という範疇的な変異が生じているとは解釈できない。

B. 長単位冒頭の短単位の分析

前節ではすべての短単位を対象に集計をおこなったが、ここでは長単位(LUW)の冒頭に位置する短単位に対象を限って同様の集計をおこなった。短単位「時間」「孤児」は、長単位「時間差」「孤児院」においては長単位冒頭位置にあるので本節の集計の対象となるが、「勤務時間」「戦災孤児」では長単位の語中に位置しているので集計対象からはずれる。

集計結果を表 5-3 に示す。表の構造は表 5-2 と同じである。表 5-3 における破擦率の分布は全般的に表 5-2 に酷似しているが、/z/が長単位冒頭（従って同時に短単位冒頭）に位置するときは破擦率が 58.3%まで上昇する。

C. アクセント句冒頭の短単位の分析

表 5-4 はアクセント句(AP)冒頭の短単位について同様の集計をおこなった結果である。破擦率の分布は全般的に表 5-2, 5-3 に酷似しているが、/z/がアクセント句冒頭（従って同時に短単位及び長単位冒頭）に位置するときは破擦率が 63.7%まで上昇する。

表 5-2: 短単位中のモーラ位置の影響

モーラ位置	摩擦音	破擦音	RAA[%]	Mean TACA[ms]
1	3,333	3,482	51.1	154.2
2	2,869	334	10.4	54.5
3	2,851	1,195	29.5	76.5
4	274	52	16.0	60.3
5	99	34	25.6	67.4
5<	72	8	10.0	60.7

表 5-3: 長単位冒頭の短単位におけるモーラ位置の影響

モーラ位置	摩擦音	破擦音	RAA[%]	平均 TACA[ms]
1	2,047	2,856	58.3	183.1
2	2,505	298	10.6	54.8
3	2,477	1,070	30.2	77.2
4	259	49	15.9	59.9
5	92	28	23.3	67.0
5<	66	8	10.8	60.6

表 5-4: アクセント句冒頭の短単位におけるモーラ位置の影響

モーラ位置	摩擦音	破擦音	RAA[%]	平均 TACA[ms]
1	1,159	2,035	63.7	236.6
2	1,842	241	11.6	54.9
3	1,806	815	31.1	78.2
4	212	42	16.5	60.4
5	58	23	28.4	69.7
5<	49	8	14.0	63.3

5.3.4 近傍分節音に関する分析

A. 後続母音の影響

表 5-5 に/z/直後の分節音、すなわち/z/とともにモーラを構成する母音の種別による破擦率の変動を示す。後続母音が前舌母音(/i/, /e/)であると破擦率が上昇することがわかる。なお表 5-5 の集計からは拗音が除外されている。

表 5-5: 直後の分節音の影響

後続母音	摩擦音	破擦音	RAA[%]	平均 TACA[ms]
/a/	631	168	21.0	68.5
/e/	651	406	38.4	74.9
/i/	2,979	2,252	43.1	78.2
/o/	734	117	13.7	58.9
/u/	1,506	368	19.6	62.7

B. 拗音の影響

表 5-6 は拗音（音韻的口蓋化）の影響を母音の別を問わずに全体として評価した結果である。また後続母音ごとに拗音中の/z/の破擦率を集計した結果を表 5-7 に示す。

拗音化した/z/の破擦率は直音よりも高いが、その差は僅少である。実際この差はデータにおける語彙的な偏りに起因するものであることを 5.5.2 節で論じる。

表 5-6: 拗音化（音韻的口蓋化）の影響

/z/	摩擦音	破擦音	RAA[%]	平均 TACA[ms]
拗音	2,997	1,794	37.4	106.0
非拗音	6,501	3,311	33.7	108.1

表 5-7: 拗音における後続母音の影響

モーラ	摩擦音	破擦音	RAA[%]	平均 TACA[ms]
/zya/	302	289	48.9	81.3
/zyo/	1,543	681	30.6	67.0
/zyu/	1,082	800	42.5	86.7

C. 直前の分節音の影響

表 5-8 に直前の分節音の影響を示す。/z/の直前に促音/Q/ないし撥音/N/があると破擦率が顕著に上昇することは注目に値する。この現象は 5.4.2 節以降で詳しく検討する。

なお、伝統的な日本語では促音は有声音である/z/の直前には生起しないとされているが、外来語はその例外になる。CSJ に記録されている/Qz/は「ロッジ」「ベイブリッジ」「カバレッジ」などの外来語に生じている。促音、撥音と同じくモーラ音素である長母音/H/には破擦化促進効果が認められない。これについては 5.4.2 節参照。

直前の母音が/i/である場合の破擦率が低い。これには語彙的な偏りが関係している。この環境に生じる高頻度語を頻度の降順に示すと、「非常」(N=407)、「以上」(N=201)、「二十」(N=165)、「規準」(N=76)、「自然」(N=56)、「相槌」(N=53)等である。これらの語では/z/が語中に位置している（正確に言えば後述する TACA の値が小さい）ために破擦率が低く（9.5%）、それが全体の平均値を低下させている。

表 5-8: 直前の分節音の影響

直前分節音	摩擦音	破擦音	RAA[%]	平均 TACA[ms]
/a/	1,847	983	34.7	67.7
/e/	562	387	40.8	76.4
/i/	2,279	404	15.1	57.9
/o/	1,292	881	40.5	68.0
/u/	980	280	22.2	58.7
/H/	1,655	822	33.2	67.4
/N/	877	1,329	60.2	120.6
/Q/	5	14	73.7	106.1

5.3.5 韻律的特徴に関する分析

A. アクセント

/z/を含むモーラにアクセント（アクセント核）が置かれているかどうかによる集計結果を表 5-9 に示す。「火事」/ka'zi/のように、語としてはアクセント核(')をもつが、/z/が属するモーラ以外の位置にアクセント核が位置しているケースは無に分類されている。アクセントが存在すると破擦率が上昇している。この問題は 5.5.3 節で検討する。

表 5-9: アクセントの影響

アクセント	摩擦音	破擦音	RAA[%]	平均 TACA[ms]
有	1400	1026	42.3	76.6
無	8098	4079	33.5	73.2

B. 直前韻律境界の BI

対象とする/z/の直前に位置する韻律境界に付与された BI ラベル（表 3-4 参照）による集計結果を表 5-10 に示す。この集計からは頻度が 10 以下の BI ラベルは除外してある。

1 とポーズが後続する 1+p を比較すると後者の破擦率が高く、2 ないし 2+b とポーズが後続する 2+p ないし 2+bp を比較するとやはり後者の破擦率が高い。さらに語中のポーズを表す P の場合も破擦率は相対的に高い値をとる。このようにポーズが破擦率を顕著に上昇させる傾向が認められる。

表 5-10: 直前の韻律境界の影響

BI	およその特徴	摩擦音	破擦音	RAA[%]	平均 TACA[ms]
1	SUW (短単位) 境界*	2,584	882	25.4	68.3
1+p	SUW 境界にポーズ有	13	23	63.9	109.7
2	AP （アクセント句）境界*	1,862	828	30.8	69.3
2+b	AP 境界に BPM 有	371	149	28.7	66.9
2+p	AP 境界にポーズ有	212	212	50.0	91.1
2+bp	AP 境界にポーズと BPM 有	124	86	42.0	85.3
3	IP （中間句）境界*	2,773	1,926	41.0	78.6
F	フィラー	1,545	978	38.8	73.8
P	SUW 内部のポーズ	10	9	47.4	82.8

*SUW, AP, IP はそれぞれ短単位、アクセント句、中間句

C. 直前のポーズ

ポーズの効果をより直截に検討するために、/z/の直前にポーズが位置することの効果を検討した結果を表 5-11 に示す。/z/直前にポーズが位置することによって破擦率は顕著に上昇し、80% に達する。これは、ここまでに観察されたなかで最も高い破擦率である。

表 5-11: 直前のポーズの影響

ポーズ	摩擦音	破擦音	RAA[%]	平均 TACA[ms]
有	193	778	80.1	69.6
無	9,305	4,327	31.7	131.3

5.3.6 言語外的特徴に関する分析

ここまでに検討してきたのは言語自体の特徴であった。ここでは言語そのものではなく、発話が行なわれた状況や話者の属性などの言語外的特徴を検討する。

A. 講演の種別

最初に CSJ のレジスターの影響を検討する。模擬講演の平均破擦率は 39.6%(N=7976)で29.4%の学会講演(N=6627)よりも高い。

B. 話者の性別

女性話者の平均破擦率は 41.3%(N=6172)で、男性話者の 30.3%(N=8431)よりも高い。

C. 話者の年齢

表 5-12 は話者を 10 年幅の生年代に分類して、その影響を示したものである。話者の生年代と破擦率の間には明瞭な相関があり、話者が若くなるほど破擦率は低下している。

表 5-12: 話者の生年代の影響

生年代	摩擦音	破擦音	RAA[%]	平均 TACA[ms]
<1950	745	799	51.7	82.3
1950s	1,220	794	39.4	74.9
1960s	3,470	1,765	33.7	72.7
1970s	4,063	1,747	30.1	72.0

D. 発話速度

表 5-13 に発話速度の影響を示す。発話速度はアクセント句を領域として 1 秒あたりの生成モーラ数で表している。発話速度と破擦率の間には明瞭は負の相関があり、発話速度の上昇につれて破擦率は低下する。

表 5-13: 発話速度の影響

SpkRate [mora/sec]	摩擦音	破擦音	RAA[%]	平均 TACA[ms]
=<3	66	54	45.0	98.3
3-6	959	718	42.8	90.9
6-9	3,709	2,203	37.3	76.8
9-12	3,612	1,699	32.0	68
12<	1,152	431	27.2	61.3

5.4 議論

5.4.1 分析のまとめ

前節のコーパス分析により、/z/の調音様式の変異が、先行研究の多くが想定していた「語頭では破擦音、それ以外では摩擦音」という範疇的な要因によって条件付けられた異音現象

ではないことが明らかになった。短単位、長単位、アクセント句の各「語」頭位置においては、たしかに破擦音の比率が上昇するが、その効果は限定的であり、破擦率が最も上昇するアクセント句頭においても高々6 割程度が破擦音化するにとどまる。またアクセント句のように複雑な単位を「語」とみなすことにはそもそも無理がある。

/z/の破擦化を促す単独の要因として最も影響力が大きかったのは、/z/直前に位置するポーズであり、また/z/を含む短単位の直前にポーズが位置する場合にも破擦化の促進効果が認められた。さらに/z/の直前に促音ないし撥音が位置している場合にも破擦音化が促進されていた。

5.4.2 仮説

/z/の変異を正しく説明する仮説を得るためには、破擦化の促進要因である、語頭、ポーズ、撥音、促音の背後に存在する要素を見出す必要がある。そのような要素として考えられるのは、子音/z/の調音運動に利用することのできる時間長である。すなわち、/z/の調音に利用可能な時間が十分にあれば/z/は破擦音として調音され、時間が不足するにつれ、摩擦音として実現されることが多くなるという仮説である。

この仮説はポーズの破擦促進効果を直截的に説明するほか、各種の語頭位置は、その直前にポーズが生起する確率の高い位置であることを考えれば、語頭効果の一部もこの仮説によって説明することができそうである。

最後に促音と撥音の破擦化促進効果もまた、この仮説で説明可能である。よく知られているように、日本語の促音と撥音はそれ自体としては調音点の指定をうけていない。実際の調音においては、直後に位置する分節音（子音）の調音点の調音点が促音ないし撥音の調音点情報として用いられる。

例えば/kaQpa/「河童」, /kaQta/「買った」, /kaQka/「閣下」, /kaQsya/「滑車」における促音とその後続子音は、音声学的には[kap:a], [kat:a], [kak:a], [kaʃ:a]のように長い子音として実現される。これは撥音に鼻音が後続する場合も同様であり、/saNma/「秋刀魚」 /kaNna/「鉋」には、[sam:a], [kan:a]のように長い鼻音が生じる。

撥音に後続する子音が鼻音でない場合、/kaNpai/「乾杯」, /kaNtai/「歓待」, /kaNkai/「官界」, /saNba/「産婆」, /kaNda/「神田」, /kaNgo/「漢語」はそれぞれ[kampai], [kantai], [kaŋkai], [samba], [kanda], [kaŋgo]（ないし[kaŋɣo]ないし[kaŋ:o]。この変異については第 6 章参照）となる。ここでは厳密な意味での長子音は生じていないが、[mp], [nt], [ŋk],[mb], [nd], [ŋg], [ŋɣ]はいずれも同じ調音点を共有する鼻音と非鼻音の連鎖であるから、唇や舌の調音運動は促音などと同様、一個の長い子音の調音に該当するものと考えられる。

撥音ないし促音に/z/が後続する場合、/roQzi/「ロッジ」は[rod:ʒi]（もしくは[roʒ:i]）、/kaNzi/「漢字」は[kandʒi]（もしくは[kanʒi]）になるから、舌の調音運動に関する限り、そこには一種の長子音が生じているとみなせる。

さて、長子音においては通常の短子音よりも多くの時間を調音に利用することができる。従って直前に促音ないし撥音が位置する場合の/z/の破擦化促進効果もまた、調音に利用可能な時間という観点から統一的に説明することができる。

ちなみに日本語の音韻分析において、撥音や促音と同じくモーラ音素として位置づけられる長母音（の後半要素）/H/は、/z/の直前に位置しても/z/の破擦化を促進する効果は認められないはずである。何故ならば/H/もまたそれ自体としては調音点情報をもたない音素であるが、/H/の場合、調音位置情報は後続する子音からではなく、先行する母音音素のものが利用されるからである。実際、表 5-4 における/H/直後の/z/の破擦率は 33%であり、コーパス全体の破擦率と一致していることがわかる。

5.4.3 Time allotted for consonant articulation

前節で述べた仮説を定量的に検討するためには、子音調音に利用可能な時間に係る何らかの定量的測度が必要である。その測度のことを以下で Time Allotted for Consonant Articulation の頭文字をとって TACA と呼ぶことにする。

TACA の定義には様々なものが考えられる。本来 TACA は調音運動の観測情報から導かれるべきであるが、ここでは CSJ において利用可能な音響情報を利用して TACA を定義することとし、3 種類の定義を比較検討する。

最も単純な定義は、音響的に規定される/z/の持続時間そのものを何の加工もせずに TACA とする定義である。これを TACA0 と呼ぶ。

第二の定義は、/z/の直前に促音/Q/ないし撥音/N/が存在していたら/Qz/ないし/Nz/全体の時間を TACA とみなすものである。これを TACA1 と呼ぶ。

第三の定義は/z/の直前ないし/Qz/, /Nz/の直前にポーズが位置していれば、その持続時間も TACA に組み入れるものである。これを TACA2 と呼ぶ。/z/の直前に短いポーズがあり、その直前に促音ないし撥音が位置している場合も、促音ないし撥音とポーズと/z/の全体を TACA2 とみなす。TACA2 の計算ではさらにポーズの持続時間に一定に制約を加えることが必要だが、これについては後述する。

TACA0, TACA1, TACA2 の値は、/z/の直前にポーズがなく、促音も撥音も存在していなければ同一である。/kore wa zasiki desu/「これは座敷です」（音素表記中の空白は短単位境界）の/z/がその例である。しかし/kore wa <pz> zasiki desu/（記号<pz>でポーズを示す）のように/z/直前にポーズがあれば、TACA0 と TACA1 は等しいが、TACA2 は相対的に大きな値をとる。

一方/kazi/「火事」と/kaNzi/「漢字」を比較すると（発話速度その他の条件が同一であれば）、TACA1 及び TACA2 は/kaNzi/において/kazi/よりも大きな値をとる。しかし TACA0 には変化は生じない。また非流暢な発話において/ka<pz>zi/のように語中にポーズが出現すれば、TACA2 は TACA0 及び TACA1 よりも大きな値をとる。

以上の説明からわかるように TACA は、発話速度と発話の局所的な音韻構造によって決定される測度であり、発話の形態論的、統語的、ないし韻律的構造には一切関係がない。

図 5-2 に上述の 3 種類の TACA と/z/の破擦率の関係を示した。横軸は TACA をミリ秒単位で、縦軸は破擦率を百分率で示している。TACA のいずれの定義においても、TACA が増大するにつれ破擦率が上昇するという全体的傾向は共通して観察される。しかし TACA0 と TACA1 では、横軸の両端において、単調増加関係に乱れが生じているのに対して、TACA2 では TACA が 20ms から 240ms まで変化するにつれ、破擦率が 5%から 95%までほぼ単純に増加しており、この点で TACA2 は 3 者中最も優れた測度になっていると言える。

ここで TACA と破擦率の間に単調増加関係が存在すると考える根拠は以下のとおりである。TACA には破擦調音が可能となる閾値が存在し、その閾値は正規分布していると仮定すれば、図 5-2 はその正規分布密度関数を積分して累積確率を計算した結果とみることができる。この仮定にたてば TACA と破擦率の間には単調増加の関係が存在していなければならなくなる。

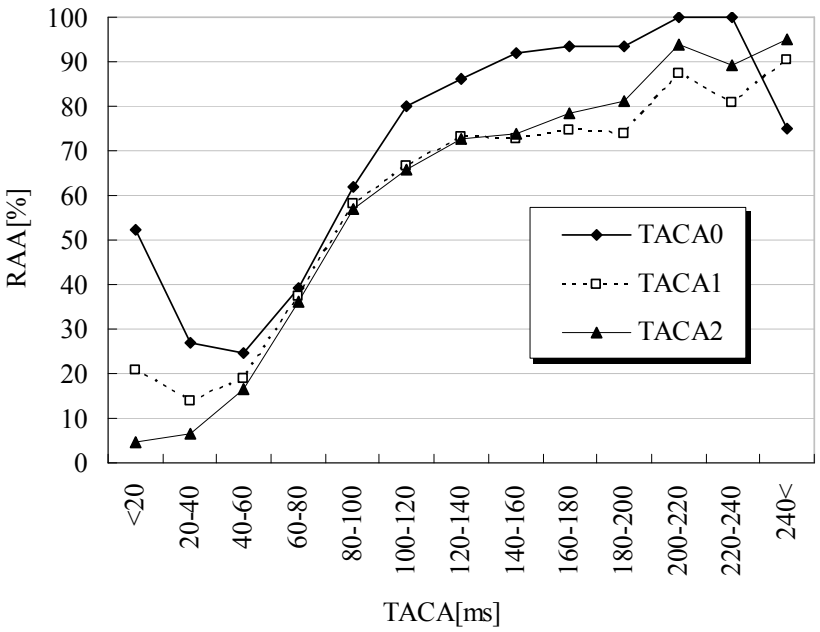


図 5-2: 3 種類の TACA と破擦率の関係

表 5-14 は、z/の調音を破擦ならば 1、摩擦ならば 0 という 2 項変数とみなして、TACA を独立変数とするロジスティック回帰分析（単回帰分析）で予測した場合の平均予測精度（closed data）と AIC 値である。TACA2 による予測精度が最も高く、AIC も最小値をとることがわかる。

表 5-14: TACA による破擦率の予測

TACA の種類	予測精度 [%]	AIC
TACA0	70.5	17,565
TACA1	72.4	16,440
TACA2	74.2	15,539

最後に TACA2 におけるポーズのとりあつかいを説明する。/z/に先行するポーズの長さは、それが発話境界に合致したような場合には、極端に大きな値をとることがある。今回のデータでは、/z/に先立つ 971 個のポーズのうち、1000ms 以上のものが 207 個、2000ms 以上が 80 個、そして 3000ms 以上が 40 個あり、なかには 10 秒以上に達するポーズもあった。このようなポーズの時間長の全体が/z/の調音に利用されるとは考えられないので、過大なポーズ長をそのまま TACA2 の計算に用いることは、TACA の測度としての価値を低下させることになる。この問題を解消するため、図 5-2 及び表 5-14 に示した TACA2 の計算には、ポーズ長は最大でも/z/そのものの持続時間長を超えないという制約を導入した。

このほかに、コーパスから計算されるポーズの平均長(72ms)を利用して TACA2 を計算することも考えられる。表 5-14 と同じ方法で評価すると、この方式の方が良い結果をもたらすのだが（平均予測精度 75.0%、AIC は 15312）、本研究ではこの方式は採用しなかった。その理由は、ポーズの平均値は全面的にコーパスに依存しているため、コーパスを利用しない研究では利用できないからである。

5.5 TACA による再分析

図 5-2 に示された TACA2 と破擦率の間の単調増加関係と表 5-14 に示された TACA による破擦率の高い予測精度は、5.4.2 節に述べた仮説の妥当性を強く示唆するものである。本節では、/z/の破擦率の予測変数としての TACA2 の有益性をより詳しく証明するために、5.3.2 から 5.3.6 の各節で検討した変数群と TACA の関係を検討することにする。以下では TACA2 のことをただ TACA と呼ぶことにする。以下の議論で参照する TACA(TACA2)の値は既に表 5-2 以下に掲載されている。

5.5.1 位置

図 5-3 は、短単位の語頭及び語中（非語頭）位置における TACA と/z/の破擦率の関係を示している。この図からは３つの重要な事実を読みとることができる。第一に TACA と破擦率の間には、短単位中の位置を問わずに高い相関が存在している。ピアソン積率相関係数値は語頭で 0.849、語中で 0.933 である。

第二に TACA の全域にわたって、短単位語頭位置における平均破擦率は、対応する語中位置における平均値よりも高い。これは TACA の効果とは別に語頭に位置することの効果が存在することを示唆している。

第三に、しかしながら、語頭の効果は、TACA の効果に比較すると限られたものにとどまっている。TACA の値にしたがって破擦率が最低 5%から 95%まで変化するのに対して、語頭の効果（すなわち図中の 2 本のカーブの間の差）は最大でも 30%に達していない。

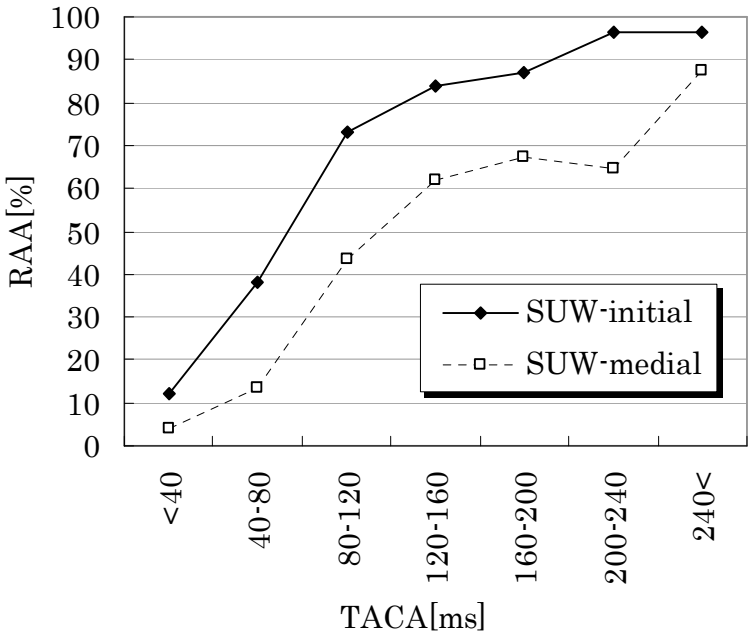


図 5-3 短単位の語頭及び語中における TACA と破擦率の関係

図 5-4 は長単位について図 5-3 と同じ分析を施した結果である。短単位と同じ結果が得られており、相関係数は長単位語頭において 0.879、語中において 0.948 である。

最後に図 5-5 はアクセント句を領域として、句頭と句中における TACA と破擦率の関係を分析した結果である。結果は短単位及び長単位の場合と同一である。TACA と破擦率の相関係数はアクセント句頭において 0.850、句中において 0.965 である。

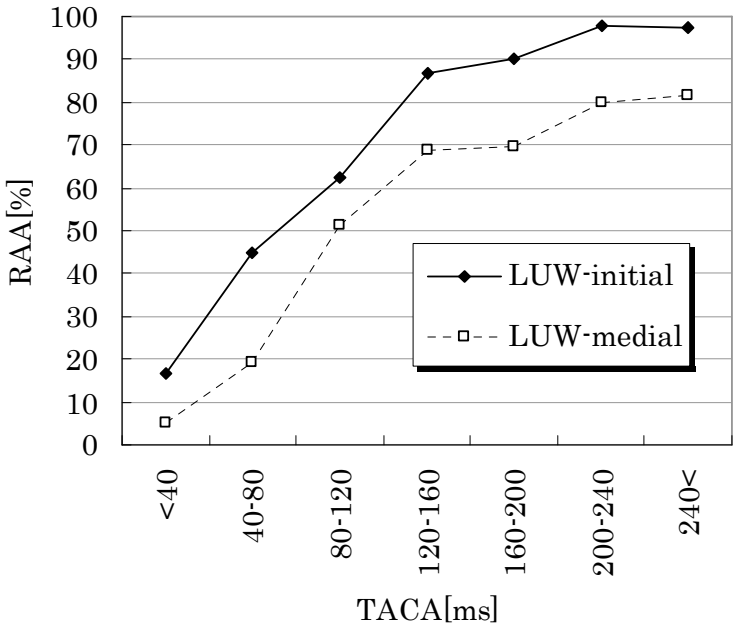


図 5-4: 長単位の語頭及び語中における TACA と破擦率の関係

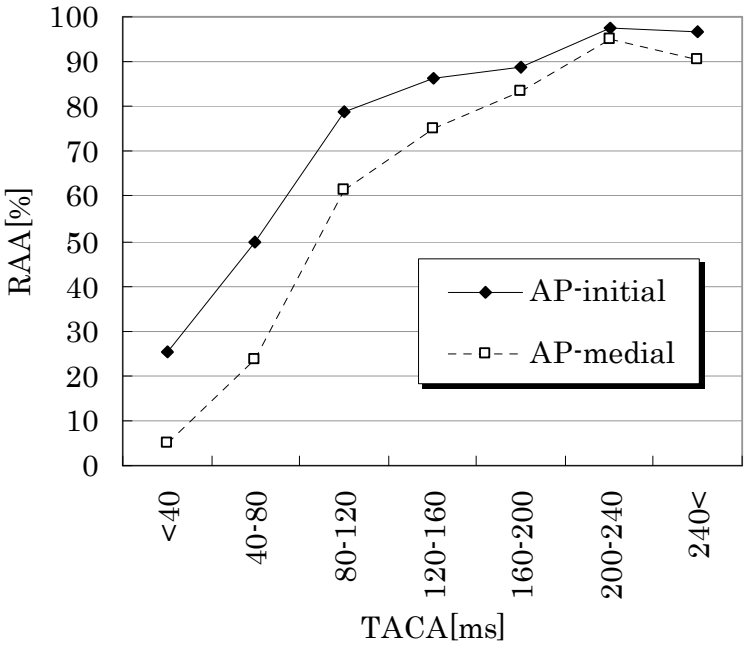


図 5-5: アクセント句の句頭及び句中における TACA と破擦率の関係

5.5.2 分節音

5.3.4 節でとりあげた分節音の特徴のうち直前の分節音に関する効果の一部、すなわち直前の促音、撥音、ポーズの効果は、TACA の定義そのものに反映されているので、ここでは、それ以外の要素を検討する。

A. 後続母音

この問題を検討するために必要な TACA の情報は、表 5-5 に掲載されている。TACA と破擦率の相関係数（表 5-5 の第 4 列と第 5 列の相関、以下同様）は 0.966 に達する。

B. 先行母音

表 5-8 に掲載されている TACA と破擦率の相関係数は、0.901 である。表から促音と撥音を除外して計算すると 0.910 である。

C. 拗音

表 5-6 では、直音の TACA と破擦率が 108.1ms と 33.7%、拗音が 106.0ms と 37.4%である。破擦率の上昇が TACA の減少に対応しており、仮説に反する関係になっている。また、そもそも観察された TACA の差は破擦率の差を説明するには僅少すぎると思われる。

また同じく拗音といっても、表 5-7 に示されているように、破擦率は/z/とモーラを構成する母音の種類によって変動しており、/zya/の場合に最高値(48%)をとっている。これらの事実は、表 5-6, 5-7 における破擦率の変動がデータの語彙的なバイアスによって生じている可能性を疑わせる。

実際にデータを仔細に検討すると、/zya/においては 591 個のサンプルのうち 559 個が短単位の語頭に位置しており、また 221 個は撥音の直後に生じていた。このような偏りは接続詞の「じゃあ」(119 個)と判定詞（断定の助動詞）「だ」に由来する「じゃ」(310 個)によってもたらされていた。後者は「～のじゃない」の「の」が撥音化することによってしばしば「～んじゃない」の形で用いられていた。これら 2 語を除外して再集計すると、/zja/の破擦率は 30.8%となる。

次に/zyu/の破擦率を上昇させているのは、の 3 個の数詞「十」「三十」「四十」であった。「三十」と「四十」では撥音の効果によって破擦率が向上しており、「十」は、原因はよくわからないが、大きな言語単位の冒頭に多く生じていた。すなわち「十」の生起総数 358 回のうち、321 回は長単位の語頭であり、268 回はアクセント句の句頭であった。後に 5.6 節で再確認するように、これらの位置には、従来考えられていたものよりは限定されてはいるが、それでも一定の破擦化促進効果が存在する。これらの数詞を除外して再集計すると、/zju/の破擦率は 32.8%になる。

5.5.3 アクセント

アクセントについては、アクセントの存在と破擦率の上昇の間に音声学的な因果関係を想定することが困難である。日本語のアクセントには、英語のストレスなどとは異なって、アクセントが生じたモーラの持続時間を延長する効果は認められないからである(Beckman 1986)。また表 5-9 に生じている 8.8%の破擦率の差を説明する要因としては、観察された TACA の差(3.4ms)はあまりに僅少である。

ここでもまた一見アクセントの効果とみえるものは、語彙の偏りに起因している。アクセントをもつモーラに生じている/z/は 2426 個あるが、そのうち 1745 個(75%)は短単位の語頭に生じているのである。ちなみにアクセントをもたない/z/の場合、短単位語頭に生じているのは 42%（12177 個中の 5070 個）にとどまっている。この偏りは、そもそも頭高型アクセントが日本語（東京語）において有力なアクセント型であることによってもたらされる偏りである。

5.5.4 言語外的特徴

A. 話者の性別とレジスター

話者の性別を TACA の観点から検討するとき、やや複雑な議論が必要になる。先に 5.3.6 節で見たように、女性話者の破擦率は男性話者よりも高い。表 5-15 は性別による破擦率と TACA 及び発話速度の関係を示しているが、この表を見ると、女性話者の方が男性話者よりも発話速度が低いことがわかる。発話速度の低さは TACA の上昇に直結するから、破擦率の高さを説明するためには都合が良いのだが、ここに問題がある。

先に 2.4.1 節で指摘したように CSJ の学会講演には男性話者が多く、また学会講演は講演時間に強い制約が課されているために模擬講演よりも発話速度が大きくなる傾向がある。そのため、発話速度と性別の問題を考えるときには、レジスターによる偏りに配慮する必要がある。

そこで表 5-16 においてレジスターと話者の性別の関係を検討した。この表を見ると学会講演においても模擬講演においても女性の発話速度が小さいので、発話速度の性差はレジスターの影響ではないと判断できる。

それでは破擦率の性差をもたらしているのは発話速度の差による TACA の差なのかと言えば、そうとは考えにくい。表 5-16 における破擦率の変動幅に対して TACA の変動が小さすぎるからである。

結局のところ、女性話者の破擦率の高さは、女性は一般に男性よりも発音が明瞭であるためとして説明するのが良いと思われる。図 5-6 は TACA の値に対する破擦率の変化を男女別に示しているが、ほぼすべての TACA の値に対して、女性話者は男性話者よりも高い破擦率を示しており、上述の解釈を支持するデータとなっている。

表 5-15: 話者の性別による変動

性別	N	RAA[%]	TACA[ms]	平均発話速度 [mora/sec]
女性	6,172	41.3	76.8	8.4
男声	8,431	30.3	71.5	9.2

表 5-16: 話者の性別とレジスターによる変動

レジスター	性別	N	RAA [%]	TACA [ms]	平均発話速度 [mora/sec]
学会講演	女性	2,460	36.5	73.8	8.7
	男声	4,167	25.2	65.5	9.6
模擬講演	女性	3,712	44.4	78.8	8.2
	男声	4,264	35.3	77.3	8.9

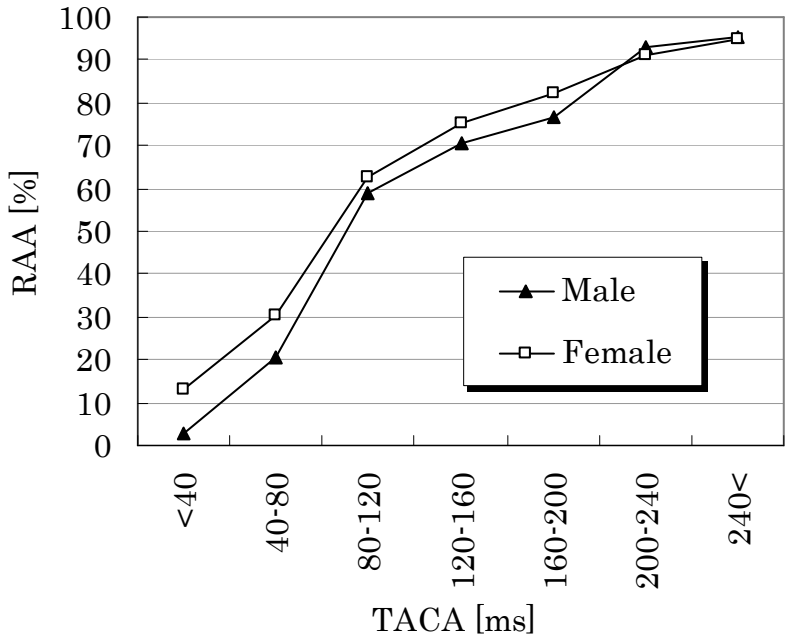


図 5-6: TACA と破擦率の関係における性差

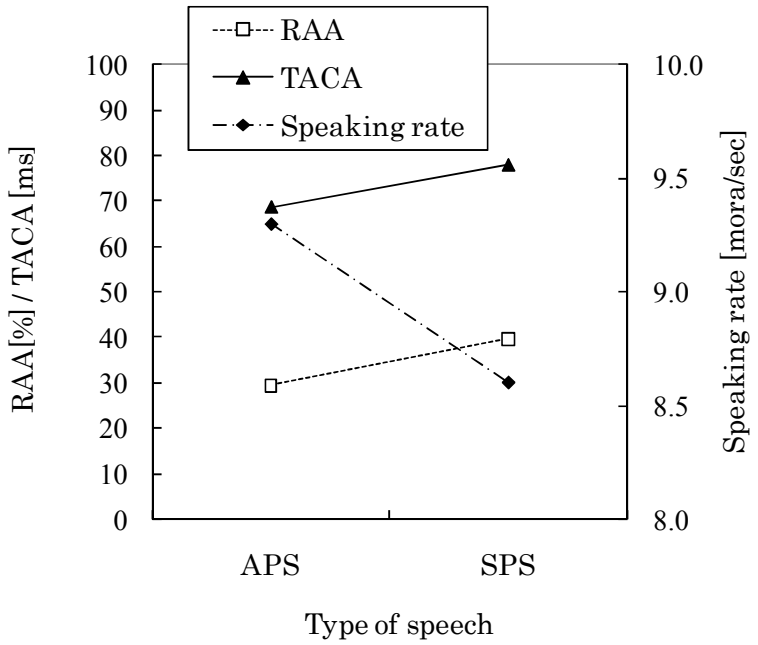


図 5-7: 破擦率, TACA, 発話速度とレジスターの関係

B. 話者の年齢

先に表 5-12 に示したように、TACA と破擦率は生年代に関して強く相関しており、両者の相関係数は 0.99 に達する。後に第 6 章（6.4.4 節、図 6-5）で議論するように、この相関は加齢にともなう発話速度の低下によって生み出されたものである可能性が高い。

C. レジスター

図 5-7 は、破擦率・TACA・発話速度の三者をレジスターによって比較した結果である。この図からは、発話速度の大小が TACA に反映し、それがさらに破擦率に反映していることが読みとれる。

レジスターによる分節音の分布の差も破擦率の差の要因となっている可能性がある。表 5-17 は、促音ないし撥音の頻度とポーズの頻度がいずれも学会講演よりも模擬講演において高いことを示している。

表 5-17: レジスターによる分節的特徴の分布の差

レジスター	N	直前の/N,Q/	直前のポーズ	/N,Q/の割合 [%]	ポーズの割合[%]
学会講演	6,627	739	397	11.2	6.0
模擬講演	7,976	1,486	573	18.6	7.2

D. 発話速度

表 5-13 に示されているように、破擦率と TACA は発話速度を介して強く相関しており、両者の相関係数は 0.99 に達する。しかし、これをもって TACA が発話速度そのものであると解釈してはならない。図 5-8 は、アクセント句を領域として測定された発話速度の高低によってデータを三分した上で、TACA と破擦率の関係を示したグラフである。いずれの発話速度帯においても、TACA は横軸の全域にわたって分布し、さらに破擦率と強く相関している。換言すれば、TACA と破擦率の関係は発話速度によって変化していないことがわかる。

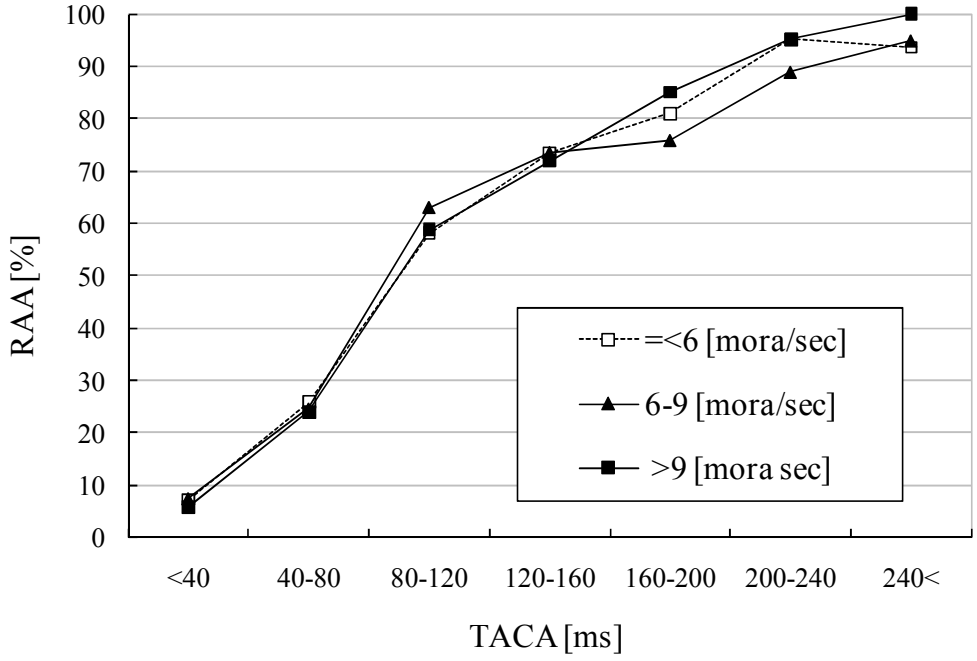


図 5-8: TACA と破擦率の関係に対する発話速度の影響

5.5.5 話者の個人差

ここまで話者の個人差の問題を無視してきたが、実際には、今回のデータにも著しい個人差が存在している。図 5-9 は平均破擦率の個人差を示している。話者(speaker)と講演(speech)を区別しているのは、今回のデータでは 6 名の話者が学会講演と模擬講演の両方を提供しているためである。いずれの場合も、平均破擦率は 20%以下から 80%以上まで幅広く分布していることがわかる。5.6 節で論じる統計的なモデリングでは、一部で個人差を積極的に利用することを試みている。

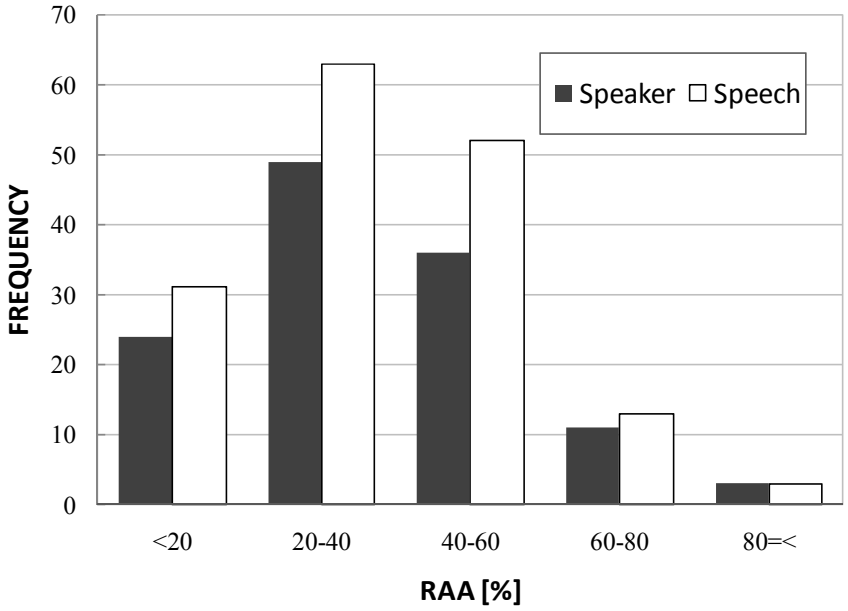


図 5-9: 平均破擦率の個人差

5.6 統計的予測

本節では/z/の破擦率に影響を及ぼす種々の要因の相対的な重要性を検討するために、統計的なモデリングをおこなう。そのために二分木による決定木を構築する。

5.6.1 決定木

予測すべき変数は/z/の調音様式であり、二値変数である。分類のために利用する独立変数の候補としては、これまでに報告したすべての言語的ないし言語外的な特徴を利用する。さらに個々の話者もまた独立変数に加えることにした。5.5.5 節で論じたように、今回のデータには著しい個人差が認められるからである。独立変数のうち、短単位、長単位、アクセント句の位置情報は、語頭かそれ以外（語中）かの二項変数として扱っている。

R 言語(Ver. 2.8.1)の rpart 関数をデフォルト設定で用いて、CART アルゴリズムによる二分決定木を計算した。図 5-10 に計算された決定木を示す。

決定木のルートノードでの分岐、従って最も説明力の高い分岐は TACA の値が約 70ms より小さいかどうかによっている。ルートノードの左右の枝（図中のノード A, B）では、ともに話者のグループ化による分岐がおこなわれている。図中の speaker group 1 と 3、及び 2 と 4 に属する話者は全く同一ではないが、図 5-11 に示されているように、平均破擦率の低い話者が speaker group1,3 に属し、平均破擦率の高い話者が speaker group2.4 に属している。図 5-10 のノード A,B が類似した規準による分岐であることがわかる（5.6.2 節の分析も参照）。

決定木における最終の分岐は図中の C, D のノードでともに長単位の語頭であるかどうかを規準にしておこなわれている。

決定木において同一の高さに属するノード A,B 及びノード C,D において、実際上同一と考えられる規準による分岐がおこなわれていたことは、決定木に採択された独立変数（TACA、個人差、長単位中の位置）の間には統計的な交互作用が存在しないことを示唆している。この問題は次節で検討する。

図 5-10 の各ターミナルノードの下に分数のように示されている数字は、決定木による分類結果であり、分子が摩擦音、分母が破擦音に分類されたサンプル数を示している。各ノードにおける平均破擦率は(分母÷(分子＋分母))×100 で計算できる。図 5-10 では平均破擦率が 50%以上のノードは affricate に、50%未満のノードは fricative に分類されている。

決定木全体による分類の精度は 79.8%であり、破擦音は 5105 個中 3531 個が、摩擦音は 9498 個中 8123 個が正しく分類されている。またルートノードのみによる分類の精度を計算すると約 74%である。

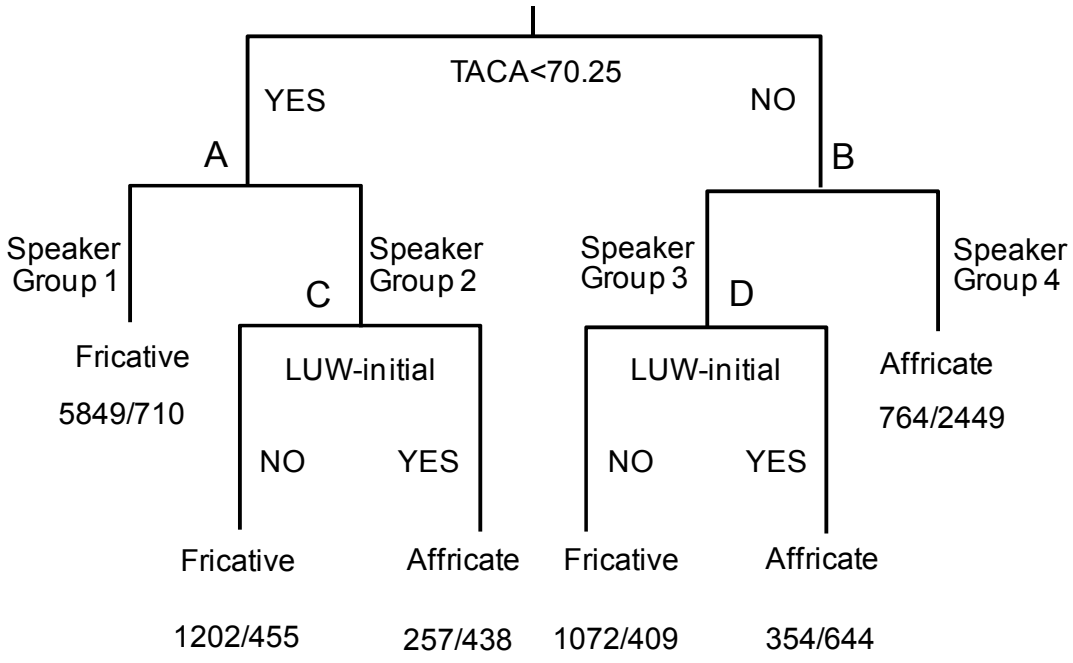


図 5-10: CART アルゴリズムによる決定木

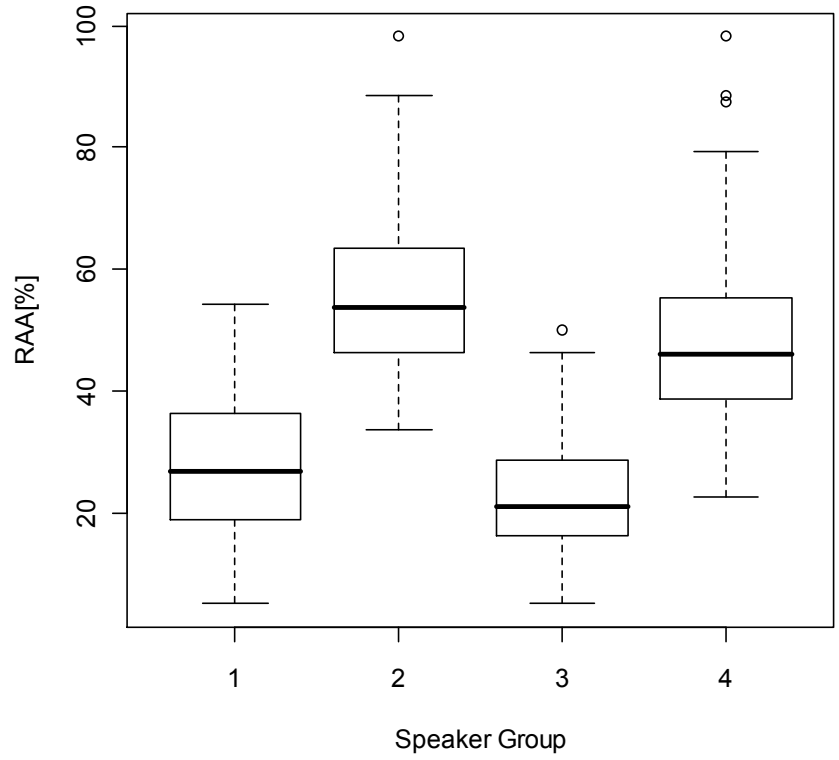


図 5-11: 話者グループごとの平均破擦率

5.6.2 交互作用の分析

先に指摘したように、図 5-10 の決定木に含まれる独立変数間には交互作用がないと予想される。これを実証するために、図 5-10 と同じ独立変数を用いたロジスティック回帰分析を実施した。結果を表 5-18 に示す。独立変数のうち LUW は長単位中の位置を表す二値変数、GroupLow は平均破擦率が 40%未満の講演であることを示す二値変数である。

表 5-18: 交互作用項を含むロジスティック回帰分析の結果

	Estimate	Std. Error	Z value	P(>|z|)
Intercept	-2.9045	0.1096	-26.499	<0.0001
LUW	1.5265	0.2101	7.265	<0.0001
GroupLow	-1.4315	0.1537	-9.315	<0.0001
TACA	34.9653	1.4629	23.901	<0.0001
LUW*GroupLow	-0.0977	0.2720	-0.359	0.719
LUW*TACA	-0.9715	2.9281	-0.332	0.740
GroupLow*TACA	-3.1615	1.8764	-1.685	0.092
LUW*GroupLow*TACA	5.5230	3.6287	1.522	0.128

表 5-18 の Estimate はモデルによって推定された偏回帰係数、Std.Error は標準誤差、Z value は偏回帰係数を標準誤差で割って標準化した値で、その確率が Pr(>/z/)である。表 5-18 の解釈はきわめて容易である。3 個の主効果にはすべて p<0.0001 で有意差が認められるが、考えうるすべての交互作用項には有意差が認められない。これによって先述の予想の正しさが証明される。また表中の偏回帰係数はその独立変数の従属変数(/z/の破擦化)への影響の大きさを表しているが、TACA とダミー変数とでは単位が異なるので直接比較できない。そこで 3 個の独立変数の標準化偏回帰係数（偏回帰係数と当該独立変数の標準偏差の積を従属変数の標準偏差で除した値）を計算すると、TACA が 3.767、LUW が 1.608、GroupLow が-1.357 となり、TACA の貢献の大きさが明白になる。

5.7 結論

日本語のザ行子音/z/の調音様式の破擦音と摩擦音にまたがる変異を CSJ-Core に付与された X-JToBI ラベルを利用して分析した結果、この変異を音素が語中で占める位置による条件変異とみる従来の分析は妥当性を欠くものであることが明らかになった。

この変異を説明するために最も有効な観点は、/z/の調音にどれだけの時間を利用することができるかであり、本研究ではその量的指標として TACA を提案した。TACA は単独でデータ全体の約 74%を正しく分類することができる。

/z/が語中で占める位置は説明変数として全く意味をなさないわけではなく、TACA とは独立に副次的な要因として機能している。しかし標準化偏回帰係数で評価すると、語中位置の貢献は TACA に比べると格段に小さく、話者の個人差と同程度にとどまっている。

TACA による分析はまた/z/の本来的な姿が破擦音であろうという推論を可能にする。図 5-2 等において横軸の右端では RAA が 100%に接近しているが、これは理想的な状況下における/z/の調音状態（金田一 1965 の言う「丁寧な発音」参照）を近似していると考えられるからである。

最後に従来の分析の欠点を指摘すると、まず従来の研究では「語」が何を意味するかがはっきりと規定されていなかった。孤立発声された短単位の語頭は、短単位語頭であると同時に長単位の語頭でもあり、アクセント句の句頭、さらに発話の冒頭でもある。従来の研究が語中位置の効果を過大評価してしまった原因はここに求められる。従来の研究のもうひとつの欠点は、発話速度に注意を払っていなかったことである。TACA は発話速度だけで決まる指標ではないことは上述のとおりであるが、促音、撥音とポーズの効果にだけ注目したのでは、TACA の効果を正しく評価することは困難だと思われる。

CSJ には種々の言語単位のアノテーションが施されているので、種々の言語単位の効果を比較検討することができた。また、発話速度も大幅に変動しているので、発話速度の影響も正しく把握することができた。自発音声の多様性が問題解決の端緒となったと言える。

第 6 章. 有声破裂音における閉鎖調音の弱化

6.1 問題

本章では、前章の成果を応用して、日本語有声破裂音素/b/, /d/, /g/の音声変異を分析する。/z/における有声破擦音と有声摩擦音との変異は声道に明瞭な閉鎖とその急激な開放が認められるかどうかに係るものであるが、音声学的にこれとよく似た変異は有声破裂音にも認められる。例えば有声両唇破裂音/b/に典型的な[b]に加えて [β]で表記すべき有声摩擦音が生じることは川上(1977)、天沼・大坪・水谷(1978)、Vance(2008)などに報告されている。有声軟口蓋摩擦音/g/においては[g]と[ŋ]が生じることは周知のとおりだが、さらに[ɣ]が生じることもよく知られている（川上 1977, Vance 2008）。最後に有声歯茎（歯裏）破裂音/d/にも閉鎖が明瞭でない有声摩擦音に類した音声が観察される。本稿ではこれを Vance(2008)に倣って[ð]の記号で表すことにする。これらはすべて有声破裂音における閉鎖及び開放調音が弱化することによって生じる現象である。以下本稿ではこれを閉鎖調音の弱化と呼ぶ。

この変異の要因については、どの文献もあまり言及していないが、川上(1997)は/b/について「語頭では確かに破裂音だが語頭以外では摩擦音[β]である場合がある」(p.32)と述べて語に係る位置の関与を指摘している。一方 Vance(2008)は “This weakening of a voiced stop to a fricative normally happens only if the affected segment is both preceded and followed by a vowel or a semivowel” (p.76)と述べて音声環境の関与を示唆している。

以下では 6.2 節で弱化の判定方法について説明し、6.3 節以下で日本語有声破裂音における閉鎖調音の弱化を/z/とほぼ同一の手法で分析する。

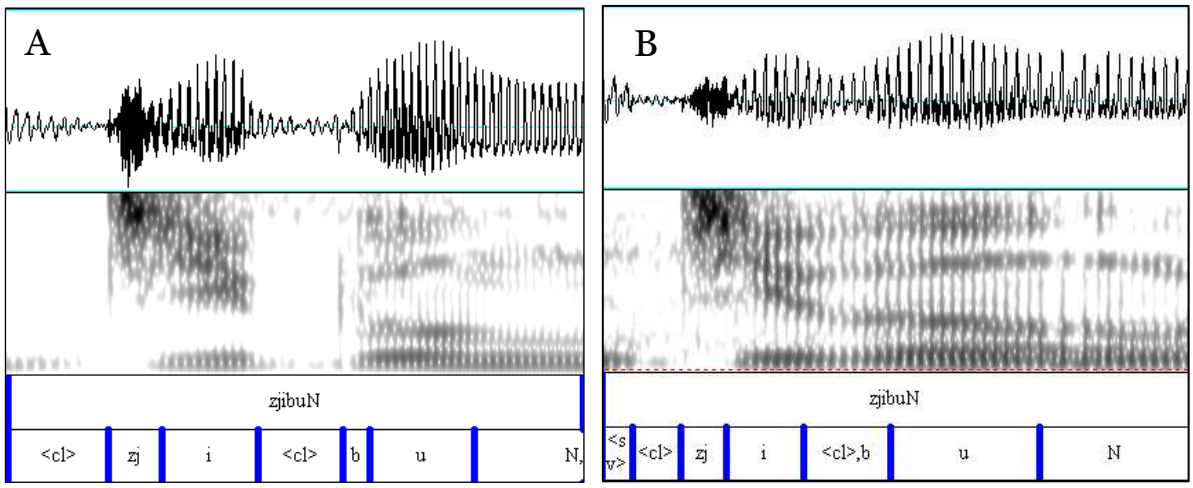
6.2 破裂音弱化の判定

本節では 6.3 節以下で分析の対象とするデータと分析において利用する変数について説明する。紙幅の制約上できるだけ簡潔な説明を試み、詳細は参考文献に譲ることとする。

図 6-1 に同じ女性話者が発音した「自分」の例をふたつ示す。音声波形とサウンドスペクトログラムの下に X-JToBI の短単位ラベルと分節音ラベルが表示されている（X-JToBI では口蓋化した/z/を「zj」で示す）。

X-JToBI の分節音ラベルでは、閉鎖区間とバーストが明瞭に観察される典型的な破裂音は 2 個のラベルによって表現される。閉鎖区間の終端（すなわちバースト音の直前）にラベル「<cl>」が付与され、バースト音に続く後続母音の始端に破裂音の種別を表すラベル「b, d, g」のいずれかが付与される（図 6-1A 参照）。本稿で問題とする閉鎖が弱化した有声破裂音においては、音声波形ないしサウンドスペクトログラムに明瞭な閉鎖区間が認められないか、明瞭なバースト音が観察されないか、あるいはその両方である。このような場合、X-JToBI の一般規約に従って、上述の両ラベルを融合させた「<cl>,b」「<cl>,d」「<cl>,g」が子音区

間の終端(後続母音の始端)に付与されている（図 6-1B 参照）。従って融合ラベルが用いられているかどうかによって、有声破裂音の弱化を検索できる。



閉鎖区間が明瞭でバースト音も明瞭　　閉鎖区間が不明瞭、バースト音無し

図 6-1: 音素/b/の変異と X-JToBI ラベル

ただし CSJ-Core における有声破裂音のラベリング規準にはひとつ問題がある。X-JToBI の分節音ラベリングを解説した藤本・菊池・前川(2006)は、撥音/N/に有声破裂音が後続する環境において「破裂音のバーストは観察されるが子音の閉鎖区間と/N/との境界が特定できない場合」には破裂の時刻にラベル「N」を、後続母音の始端に融合ラベル「<cl>,d」等を付与したと述べている。この規準に従ったラベリングでは明瞭なバーストが観察される有声破裂音にも融合ラベルが付与されてしまうので、本研究のデータとしては不都合である。

そこで上記の音声環境におかれた有声破裂音はすべて再ラベリングすることにした。該当するサンプルは/Nb/が 826 個、/Nd/が 7733 個、/Ng/が 3336 個である。ラベリングは筆者が単独で実施した。それ以外のサンプルについては CSJ-Core のラベリング結果をそのまま利用する。

最後にラベリングの精度を簡単に評価しておく。弱化といってもその程度にはかなり大きな幅があり、それがラベルのゆれになってコーパスに反映していることが予想される。そのゆれが大きすぎればデータとしての価値がなくなるので、この問題を予め検証しておく必要がある。

CSJ-Core の学会講演と模擬講演には/b/が 9279 個、/d/が 34289 個、そして /g/が 21953 個含まれている。このなかから各音素のサンプルを無作為に約 200 個ずつ抽出して、筆者が弱化の有無について判定をおこなった（ちなみに筆者は X-JToBI の仕様を策定したが、CSJ-Core のラベリング作業には参加していない）。

その結果を CSJ-Core における融合ラベルの有無と比較して κ 統計量を計算したところ、/b/では 0.70、/d/では 0.62、/g/では 0.71 の値を得た。Kappa が 0.6 以上であれば、経験的にふたつのラベリング結果は「よく一致している」と判断してよいとされているので、今回利用するデータは信頼に足る精度をもっていると判断する。

なお/g/の異音には、特に/N/に後続する環境において[ŋ]が生じている。鼻音の典型的調音では声道に閉鎖もしくは強い狭窄とその開放が生じるが、本稿では鼻音を破裂音の一種とはみなしていないことに注意。

6.3 分析

CSJ-Core の学会講演と模擬講演全体において、有声破裂音が典型的な破裂音として実現された百分率（以下これを Rate of Stop Articulation --RSA-- と称する）を計算すると、/b/は 52.9%、/d/は 76.1%、/g/は 29.3%である。以下の分析ではこれらの数字が RSA の高低判断のベースラインとなる。

6.3.1 位置との関係

短単位(SUW)、長単位(LUW)、アクセント句(AP)、発話転記単位(IPU)の各単位において、その冒頭（語頭）とそれ以外（語中）における RSA の平均値を比較したのが表 6-1 である。IPU は CSJ の転記ファイルで採用された転記基本単位（2.3.1 節参照）で前後を 200ms 以上のポーズで区切られた音声区間（小磯・西川・間淵 2006）である。これは言語学上の単位とは言えないが、ここでは仮に発話に該当する単位として扱うことにした。表 6-1 から表 6-7 には RSA の他にサンプル数（N）と TACA が示されている。TACA については 6.4 節以下で言及する。表中の R についてもそこで言及する。表 6-2 中の記号~/Ng/については 6.4.2 節参照。

表 6-1 で RSA はいずれの単位においても、冒頭においてそれ以外よりも顕著に高い値を示している。しかし SUW ないし LUW 冒頭における RSA は、最も低い/g/においては 30%強、最も高い/d/においても 80％弱であり、これを語中位置に係る条件異音とみる解釈は成立しがたいことがわかる。

表 6-1 におけるもうひとつの顕著な傾向は、冒頭位置において、SUW<LUW<AP<IPU の順に RSA が上昇していることである。以上ふたつの傾向が/z/に観察されることは第 5 章で報告した（ただし第５章では IPU は検討対象としていない）。

表 6-1: 各種言語単位中の位置による RSA（閉鎖率）と TACA の変動

音素	言語単位	単位頭			単位頭以外		
		N	RSA [%]	TACA [ms]	N	RSA [%]	TACA [ms]
/b/	SUW	4341	66.5	64.8	4938	41.0	52.5
	LUW	3286	70.7	65.3	5993	43.1	54.4
	AP	1969	76.3	69.6	7310	46.6	55.2
	IPU	436	92.9	94.8	8843	50.9	56.5
R				0.976			0.959
/d/	SUW	25911	78.5	62.7	8378	68.6	47.3
	LUW	19498	78.3	58.1	14791	73.1	59.9
	AP	7736	90.3	66.3	26533	71.9	56.7
	IPU	3899	98.2	74.6	30390	73.2	56.9
R				0.948			0.962
/g/	SUW	14524	33.1	51.3	7429	22.0	52.1
	LUW	9431	33.0	50.6	9431	24.5	52.8
	AP	1957	76.7	66.6	19969	24.7	50.1
	IPU	558	93.4	103.7	21395	27.7	50.2
R				0.908			-0.606
~/Ng/	SUW	12994	32.6	44.3	5809	20.6	38.5
	LUW	11394	32.8	44.7	7409	22.9	39.2
	AP	1897	77.0	64.8	20056	24.8	40.0
	IPU	522	95.0	92.9	18281	27.0	41.1
R				0.958			0.979

6.3.2 先行分節音との関係

表 6-2 に有声破裂音の直前に位置する分節音による RSA の変動を示す。/b/, /d/の場合、直前が撥音/N/もしくは促音/Q/であると RSA が顕著に上昇している。これと同じ現象は/z/についても観察されている。他方/g/の RSA は/Q/の直後では顕著に上昇しているが、/N/の直後ではベースラインからほとんど変動していない。この現象は 6.4.2 節で論じる。

表 6-2: 先行分節音による RSA と TACA の変動

音素	/b/			/d/			/g/		
	N	RSA [%]	TACA [ms]	N	RSA [%]	TACA [ms]	N	RSA [%]	TACA [ms]
/a/	1,659	47.4	52.7	4,957	74.2	48.9	3,067	29.9	44.9
/e/	1,099	50.9	53.7	5,241	67.8	46.7	1,180	33.6	45.2
/i/	1,636	45.5	51.7	4,211	67.4	43.0	3,712	21.2	38.8
/o/	1,882	54.8	52.8	5,897	77.0	46.8	3,523	30.1	42.6
/u/	1,109	43.2	48.2	2,935	81.9	52.8	4,090	30.7	42.2
/H/	1,102	57.5	60.3	3,545	72.9	46.6	3,176	31.1	43.6
/N/	770	84.4	114.3	7,448	86.4	100.8	3,150	32.0	105.4
/Q/	21	95.2	93.2	32	100.0	93.7	16	93.7	90.3
R			0.917			0.845			0.564 (~/Ng/ 0.997)

表 6-3: 先行短単位境界 BI による RSA と TACA の変動

BI	/b/			/d/			/g/		
	N	RSA [%]	TACA [ms]	N	RSA [%]	TACA [ms]	N	RSA [%]	TACA [ms]
1	2,400	48.0	55.0	20,467	71.0	57.3	13,081	22.1	47.0
1+p	25	72.0	74.3	329	97.3	73.9	214	80.8	116.8
2	1,840	51.3	55.3	3,876	72.5	52.5	2,294	33.5	53.7
2+b	292	56.6	60.9	408	77.2	52.4	449	32.5	57.3
2+bp	138	65.2	70.0	146	89.7	63.6	174	41.4	58.9
2+p	273	66.3	65.5	486	85.8	69.2	399	42.6	62.0
3	2,802	54.7	60.0	6,728	89.8	66.3	3,363	42.2	58.4
F	1,490	54.8	60.8	1,749	82.8	59.2	1,919	39.9	56.2
P	11	45.5	52.5	53	81.1	70.3	32	43.7	45.2
R			0.975			0.759			0.954

6.3.3 先行 BI 及びポーズとの関係

表 6-3 は、直前の短単位境界に付与された BI 値による RSA の変動を示している。生起頻度が 10 以下のラベルは集計から除外した。ポーズに関係する BI（1+p, 2+bp, 2+p）の直後では RSA が上昇する傾向が認められる。これも/z/と同じである。

ポーズの影響を直接に分析すると表 6-4 の結果を得る。いずれの有声破裂音においても、その直前にポーズが存在すると RSA が 90%以上に達する。

表 6-4: 先行ポーズの有無による RSA と TACA の変動

音素	先行ポーズ有			先行ポーズ無		
	N	RSA [%]	TACA [ms]	N	RSA [%]	TACA [ms]
/b/	461	92.4	97.0	8,818	50.8	56.2
/d/	3,979	97.8	75.3	30,311	73.2	56.7
/g/	594	91.8	104.4	21,359	27.6	50.1

6.3.4 外的要因

言語外的な要因にも簡単に触れておく。RSA を学会講演と模擬講演とで比較すると、/b/において 51.6%(学会講演):53.9%(模擬講演)、/d/において 76.5%:75.8%、/g/において 31.2%:27.8%であり、一貫した傾向は認められない。

表 6-5 は 5 段階に区分した発話速度と RSA の関係を示している。発話速度は AP ごとに計算しており単位は mora/sec である。いずれの音素に関しても、発話速度の上昇につれて RSA は例外なく単調に低下している。これも/z/と同一の結果である。

表 6-6 は話者の性別による変動を示している。いずれの音素においても女性の方が男性よりも RSA が高い。/z/においてもこれと同じ結果が観察されている。最後に表 6-7 は話者の生年代（10 年間隔）による RSA の変動である。「30s」は話者が 1930 年代生まれであることを意味する。/b/と/g/については話者の年齢が低下するほど RSA が低下する緩やかな傾向が観察される（この傾向は前稿ではより明瞭に観察された）。しかし/g/はこの傾向に従っておらず、RSA はむしろ上昇する傾向をみせている。この問題は 6.4.4 節と 6.6 節で検討する。

表 6-5: 発話速度による RSA と TACA の変動

発話速度	/b/			/d/			/g/		
	N	RSA [%]	TACA [ms]	N	RSA [%]	TACA [ms]	N	RSA [%]	TACA [ms]
=<3	59	69.5	86.1	487	89.3	84.3	163	49.1	71.9
3~6	988	61.3	70.2	3,936	86.5	71.0	2,343	40.5	65.5
6~9	3,814	55.0	60.8	13,616	79.1	60.0	9,003	32.5	53.1
9~12	3,440	51.4	54.6	12,709	72.6	55.2	8,196	25.5	47.6
12<	978	40.4	47.5	3,541	63.8	50.8	2,248	17.6	43.9
R			0.972			0.926			0.983

表 6-6: 話者の性別による RSA と TACA の変動

話者性別	/b/			/d/			/g/		
	N	RSA [%]	TACA [ms]	N	RSA [%]	TACA [ms]	N	RSA [%]	TACA [ms]
女性	3,826	57.8	60.5	14,520	82.8	59.4	9,475	35.3	52.9
男声	5,435	49.5	56.7	19,769	71.1	58.5	12,478	24.8	50.5

表 6-7: 話者の生年代による RSA と TACA の変動

生年代	/b/			/d/			/g/		
	N	RSA [%]	TACA [ms]	N	RSA [%]	TACA [ms]	N	RSA [%]	TACA [ms]
30s	350	66.6	70.2	1,735	69.8	82.6	922	21.4	61.9
40s	468	53.0	57.4	1,876	61.9	83.7	1,163	27.4	54.0
50s	1,294	61.7	60.7	5,814	58.3	75.7	3,286	31.9	51.4
60s	3,264	51.7	56.8	11,787	57.5	76.8	7,883	29.3	51.7
70s	3,903	49.7	57.7	13,077	58.6	73.6	8,699	29.5	50.1
R			0.903			0.733			-0.948

6.4 議論

6.4.1 TACA の導入

ここまでに示した分析結果は、6.3.2 節における/N/直後での/g/の RSA の低さと 6.3.4 節における/g/と生年代との相関を例外として、/z/についての分析結果と一致していた。短単位においても長単位においても語頭位置における/b/, /d, /g/の RSA は高々80%程度であり、閉鎖の弱化を語中位置に係る条件変異とみることは困難である。撥音と促音とポーズの直後で RSA が顕著に上昇することからは、むしろ子音の調音に利用可能な時間の増大に応じて RSA が上昇するという関係が示唆される。撥音と促音における声道閉鎖は後続子音の調音位置において後続子音と一体化して実現されるから、撥音ないし促音に先立たれた/b/, /d/, /g/においては撥音ないし促音の持続時間も閉鎖調音のための調音運動に利用することができる。

この時間を第 5 章と同じく TACA(Time Allotted for Consonant Articulation)と呼ぶことにする。TACA には幾通りもの定義が可能だが、ここでは第 5 章と同様、当該子音の直前に撥音、促音、ポーズがあれば、その持続時間を子音自体の持続時間に加算したものを TACA と呼ぶことにする。撥音、促音、ポーズのうちふたつ以上が子音の直前に生起している場合は、すべての持続時間を加算の対象とする。ただしポーズのなかには持続時間が 10 秒を越えるようなものもあるので、ポーズ長は最長で子音自体の持続時間を越えないという制約を設けることにする。

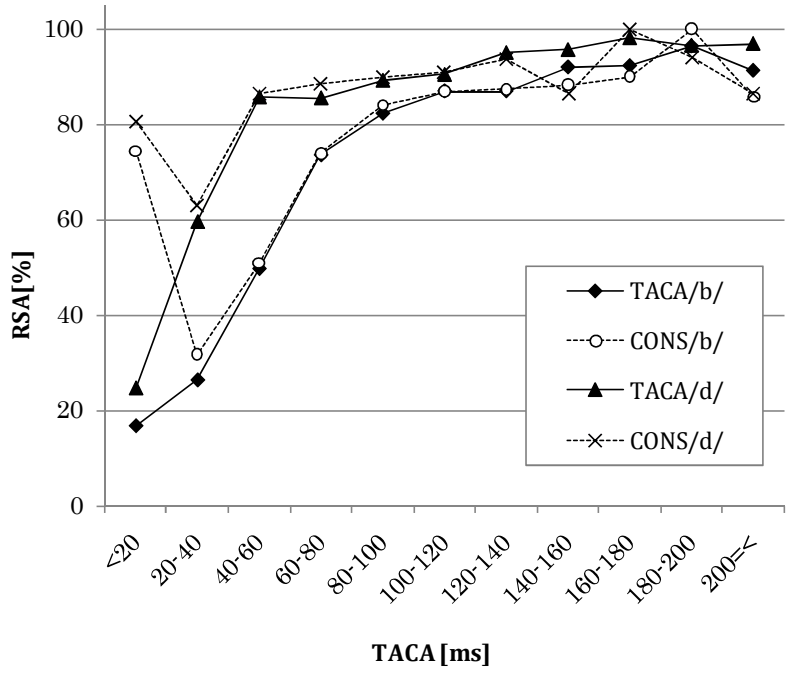


図 6-2: TACA 及び子音持続時間長と RSA

図 6-2 はこのようにして計算された TACA（横軸、単位は ms）と RSA（縦軸、単位は%）の関係を/b/と/d/について示している。比較のために TACA の代わりに子音持続時間単体と RSA の関係(CONS/b/と CONS/d/)も示してある。/b/, /d/のいずれにあっても、TACA と RSA の間にはほぼ単調増加とみなせる関係が成立しているのに対し、子音持続時間と RSA の関係では横軸の両端（殊に左端）において単調性が破綻している。これによって RSA を予測するための変数としては、子音持続時間よりも TACA の方が優れていることがわかる。

ここで有声破裂音の持続時間の測定法を説明しておく。子音/b/, /d/, /g/の持続時間とは閉鎖ないし強い狭窄の持続時間と VOT(バースト音から後続母音の開始時刻までの時間長）の合計である。図 6-1A においては「<cl>」ラベルと「b」ラベルの時間長を合計したものがこれにあたり、図 6-1B においては融合ラベル「<cl>,b」の時間長がこれにあたる。

しかし、/b/, /d/, /g/が IPU 頭（CSJ ではほぼすべての場合に 200ms 以上のポーズを伴う）に位置しているときには閉鎖の持続時間を正確に知ることができないので、何らかの便法を用いる必要がある。CSJ-Core では典型的な有声破裂音ないし閉鎖が弱化した異音の多くにおいて、声帯振動がバースト音ないし摩擦ノイズに先行して始まることに注目して、IPU 冒頭での声帯振動の開始時点から当該子音に後続する母音の始端までの時間をもって子音の持続時間としている。本稿でもこれに従って子音持続時間を決定している。IPU 内部に 200ms 未満のポーズがあり、その直後に/bdg/が生じている場合もこれと同じ方法で対処している。ただし稀に IPU 冒頭にバースト音が観察されるものの、それに先行する声帯振動が観察されない

発話もあった。その場合は、閉鎖の時間長を知る手段が存在しないので、閉鎖の時間長はゼロとして処理している。

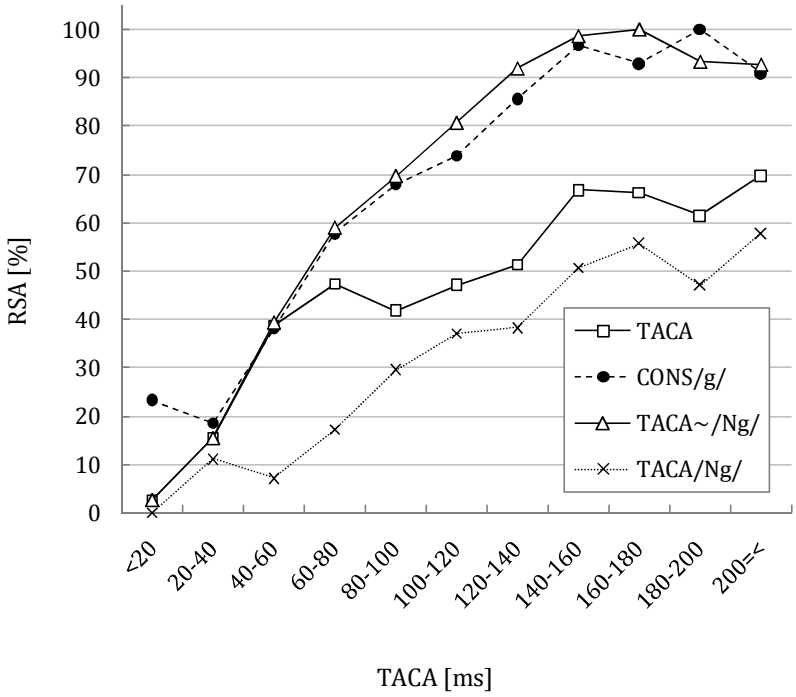


図 6-3: /g/における TACA-RSA 関係

6.4.2 /g/の特異性

図 6-2 には示していなかった/g/のデータを図 6-3 に示す。破線と●の組合せが/g/における子音持続時間(CONS/g/)と RSA の関係、実線と□の組合せが TACA と RSA の関係である。グラフの左端で CONS/g/が単調性に乱れを見せているのに対し、TACA はそうなっていない点は/b/, /d/と同じである。しかし TACA の値が小さいうちに RSA が頭打ちになり、最終的にも 70%に達しない点が/b/, /d/との顕著な相違である。

その原因として考えられるのは/g/の異音には軟口蓋鼻音[ŋ]が含まれることである。CSJ-Core の分節音ラベルでは[ŋ]と[ɣ]を区別していないが、実際にラベリングを行うと撥音直後の環境では/g/は多くの場合に[ŋ]として実現されていることがわかる（以下この環境を/Ng/環境と呼び、直前に撥音が存在しない環境を~/Ng/環境と呼ぶことにする）。/Ng/環境における TACA は相対的に大きな値をとるが、2.4 節に述べたように今回の分析では[ŋ]は破裂音に認定されないのでグラフが頭打ちになるのだと考えられる。

この仮説の正しさの傍証として、図 6-3 には~/Ng/環境のグラフも示してある（実線と△の組合せ）。そこには/b/,/d/と同様の単調増加関係が認められることから仮説の正しさが示唆される。最後に/Ng/環境のグラフ（細点線と×の組合せ）は、他のいずれのグラフよりも下

に位置しており RSA は最高でも 60%に達しない。ただしこの環境においても TACA と RSA の間には緩やかな単調増加関係が認められることは注目に値する。

6.4.3 TACA による分析

有声破裂音の弱化要因としての TACA の有効性を示すために 6.3 節で検討した諸要因を TACA との関係という観点から再分析する。TACA の測定値は表 6-1 から表 6-7 に掲載済である。位置の影響を分析した表 6-1 において、TACA と RSA は共変関係にあり、いずれの音素に関しても高い正の相関が認められる（表中の R がピアソン積率相関係数）。また単位冒頭位置だけでなく非冒頭位置でも高い相関が認められる。唯一の例外は/g/における非冒頭位置であるが、この例外は非冒頭位置では冒頭位置よりも/g/に[ŋ]が生じやすいことに起因していると考えられる。実際、撥音に後続する/g/を除外して再計算すると高い正の相関が得られる（表最下列の~/Ng/参照）。

直前の分節音の影響を分析した表 6-2 においても/b/, /d/に関しては TACA と RSA の間に高い相関が認められる。/g/の場合データ全体を分析すると相関が低いが、ここでも/Ng/環境のサンプルを除外すると高い相関が得られる（相関係数の列中の~/Ng/参照）。

直前短単位境界における BI の影響を検討した表 6-3 では/b/, /g/については非常に高い相関が、/d/についても高い相関が認められる。

表 6-4 においては、直前にポーズが存在する場合に TACA は顕著に大きな値をとる。これは TACA の定義上、当然である。

表 6-5 において発話速度と TACA が相関するのは TACA の定義から当然予測される結果であり、RSA との相関も音素によらず高い。ただしこのことから TACA と発話速度を同一視してはならない。図 6-4 は発話速度を 6 mora/sec 以下、6~9mora/sec、9mora/sec 以上に区分化した上で、TACA と RSA の関係を検討した結果である。どの群においても TACA は広い範囲に分布しており、RSA との間にはほぼ単調増加の関係が成立している。TACA は直前の撥音・促音・ポーズによっても変動するために、どの発話速度群においても幅広く分布しているのである。図 6-4 では RSA の単調増加のあり方が音素ごとにかなり異なっていることが注目されるが、この現象については 6.6 節で論じる。

以上のように TACA と RSA との間には広い範囲に高い相関が認められる。言語単位中の位置、先行分節音の種別、先行韻律境界の種別など表面的には大きく異なる言語的要因が TACA を利用することによって統一的に解釈できることは、TACA が有声破裂音弱化の本質的要因であることを強く示唆している。

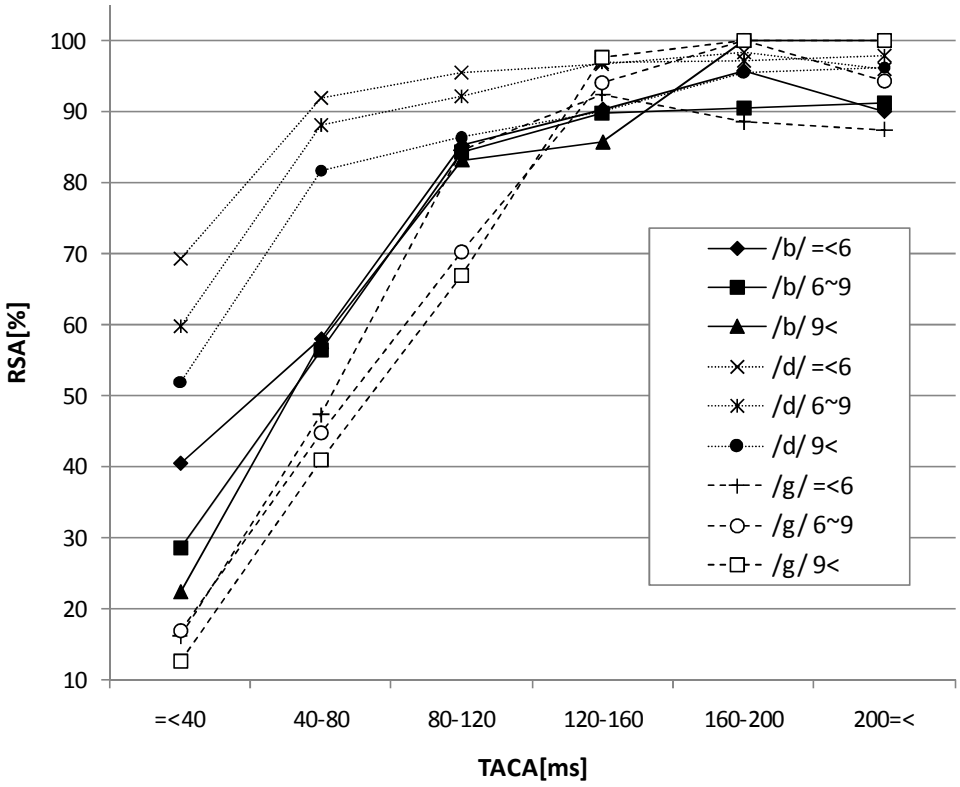


図 6-4: 発話速度と TACA

6.4.4 例外の検討

しかし説明を要する例外がふたつある。表 6-6 における性別の影響と表 6-7 のうち/g/にかかわる負の相関である。まず表 6-6 においても TACA と RSA は共変関係にある。ただしそこで生じている TACA の差は RSA の差の原因とするには僅少すぎる。TACA が RSA の差を生み出しているのではなく、むしろ所与の TACA の水準において、女性音声では男性音声よりも高い RSA 値が実現される（つまり女性の方が男性よりも調音が丁寧である）と考えるべきであろう（図 5-6 も参照）。

次に表 6-7 の問題。この表では/b/, /d/, /g/のすべてにおいて TACA の平均値は若い世代ほど減少する傾向をみせている。そこから予想されるのは若い世代ほど RSA が低下するという傾向である。実際、/b/と/d/にはその傾向が観察され、TACA-RSA 間に高い正の相関が生じているのだが、/g/だけは反対に若い世代の RSA が上昇している。第 5 章で分析した/z/の破擦率も若い世代ほど下降していたことを考えあわせると、一般的なのは/b/, /d/のパタンであり、特殊なのは/g/のパタンであると言える。

ここで図 6-5 は CSJ-Core 全体を使って生年代ごとの平均発話速度を計算した結果である。この図からは若い世代ほど発話速度が速くなる傾向が読みとれる。TACA の定義上、平均発話速度の変化はそのまま TACA に反映されるので、表 6-7 において若い世代ほど TACA が低

下しているのはその結果と考えられる。そして本稿で想定している TACA と RSA の因果関係からすれば、この傾向は若い世代における RSA の低下を引き起こすことが予想される。

実際、表 6-7 の/b/, /d/においては予想どおりの変化が生じている。しかし何らかの理由で/g/においては RSA が低下せず、逆に上昇している。その原因は残念ながら不明であるが（ただし 6.6 節の議論参照）、/g/の生成には発話速度に帰着させることのできない世代差が存在していることは確実である。

なお、上述の/g/の特異性は先行する撥音の影響によるものではない。~/Ng/環境における RSA と TACA の相関係数は-0.830 であり、表 8 と同様、負の相関を示すからである。

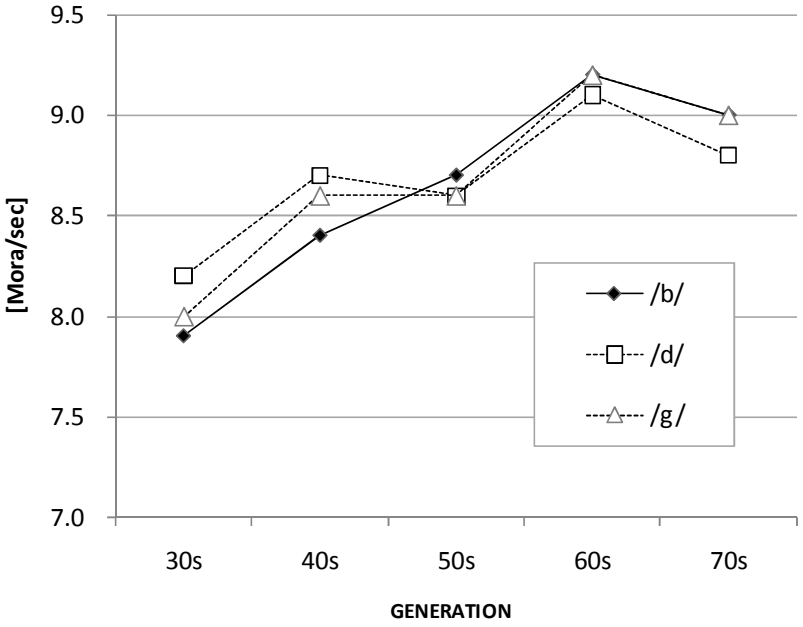


図 6-5: 話者の生年代による平均発話速度の変化

6.5 統計的予測

前節までの検討で TACA が有声破裂音弱化の有力な要因であることが明らかになった。本節では、統計的にはどの程度まで有声破裂音の弱化を予測できるかを検討する。予測のための統計手法としてロジスティック回帰分析を用いる。従属変数は有声破裂音における弱化の有無という 2 値変数である。独立変数としては 6.3 節でとりあげた各種変数を利用する。分析には R 言語の glm 関数を用い (Baayen 2008;青木 2009)、３種の有声破裂音それぞれを独立に分析した。

表 6-8 に分析結果をまとめた。表の各行が独立変数に対応している。Talktype は学会講演か模擬講演かの別、spksex は話者の性別、generation は世代の別、precphoneme は直前分節音素、precsuwbi は直前短単位境界 BI、pause は直前ポーズの有無、suwinitial から ipuinitial ま

では、短単位、長単位、アクセント句、発話転記単位の冒頭に位置しているかどうかの別である。

TACA 以外のすべての変数はカテゴリー変数なので、独立変数にカテゴリー変数を含む回帰分析（一般化線形モデル）の定石として、各変数に含まれる各水準が２値のダミー変数として展開されている。例えば spksexM は話者の性別が M であることを示す２値変数であり、pause1 はポーズの存在を示す２値変数である。一般に N 個の水準をもつ変数は N-1 個のダミー変数に展開されるので、性別やポーズのような２値変数では 1 個のダミー変数が、generation のように 5 水準の変数では 4 個のダミー変数が用いられている。

表 6-8: ロジスティック回帰分析の結果*

	/b/				/d/				/g/			
	Estimate	z value	Pr(>|z|)	SPRC	Estimate	z value	Pr(>|z|)	SPRC	Estimate	z value	Pr(>|z|)	SPRC
(Intercept)	9.22	0.05	0.963		-0.77	-1.13	0.257		-3.22	-3.63	0.000	
talktypeS	0.22	4.08	0.000	0.21	-0.19	-6.11	0.000	-0.22	-0.16	-4.47	0.000	-0.18
spksexM	-0.27	-5.44	0.000	-0.27	-0.69	-22.77	0.000	-0.80	-0.61	-17.58	0.000	-0.67
generation40s	-0.44	-2.64	0.008		0.17	1.75	0.081		0.69	5.53	0.000	0.34
generation50s	-0.10	-0.70	0.485		-0.29	-3.68	0.000	-0.25	1.03	9.58	0.000	0.81
generation60s	-0.43	-3.17	0.002		-0.20	-2.62	0.009		0.87	8.37	0.000	0.91
generation70s	-0.62	-4.58	0.000	-0.61	-0.36	-4.89	0.000	-0.41	1.01	9.75	0.000	1.08
precphonemea	-11.39	-0.06	0.954		13.48	0.03	0.974		-0.20	-0.20	0.844	
precphonemee	-11.42	-0.06	0.954		13.35	0.03	0.975		-0.24	-0.23	0.819	
precphonemeH	-11.07	-0.06	0.955		13.67	0.03	0.974		-0.12	-0.12	0.906	
precphonemei	-11.31	-0.06	0.954		13.52	0.03	0.974		-0.38	-0.37	0.709	
precphonemeN	-11.03	-0.06	0.955		12.27	0.03	0.977		-1.66	-1.61	0.107	
precphonemeo	-11.16	-0.06	0.955		13.87	0.03	0.974		-0.26	-0.25	0.804	
precphonemeQ	-8.88	-0.05	0.964		25.76	0.06	0.953		2.54	1.71	0.087	
precphonemeu	-11.32	-0.06	0.954		13.76	0.03	0.974		0.13	0.12	0.901	
precsuwbi1	0.00	0.00	0.998		-13.39	-0.03	0.975		-0.52	-0.58	0.563	
precsuwbi1+	-11.16	-0.08	0.936		-13.52	-0.03	0.974		-9.09	-0.11	0.910	
precsuwbi1+p	0.99	0.77	0.439		-10.88	-0.03	0.979		0.65	0.70	0.487	
precsuwbi2	0.19	0.16	0.875		-13.61	-0.03	0.974		0.05	0.06	0.955	
precsuwbi2+b	0.31	0.26	0.793		-13.34	-0.03	0.975		0.06	0.06	0.951	
precsuwbi2+bp	0.55	0.46	0.648		-12.69	-0.03	0.976		0.47	0.51	0.609	
precsuwbi2+p	0.61	0.51	0.609		-13.26	-0.03	0.975		0.44	0.48	0.630	
precsuwbi3	0.35	0.30	0.767		-13.14	-0.03	0.975		0.39	0.43	0.666	
precsuwbiF	0.38	0.33	0.745		-12.79	-0.03	0.976		0.78	0.86	0.389	
precsuwbiP	-0.27	-0.20	0.842		-13.32	-0.03	0.975		0.47	0.48	0.632	
precsuwbiPB	--	--	--		-1.27	0.00	0.998		1.29	1.04	0.299	
precsuwbiW	-0.93	-0.50	0.618		-13.39	-0.03	0.975		--	--	--	
pause1	-0.86	-1.60	0.110		-2.00	-5.07	0.000	-1.50	-1.46	-3.37	0.001	
suwinitial1	0.51	4.70	0.000	0.51	-0.15	-2.62	0.009		1.00	11.07	0.000	1.04
luwinitial1	0.54	5.65	0.000	0.52	0.06	1.39	0.165		-0.13	-2.12	0.034	
apinitial1	0.19	1.63	0.102		0.69	9.27	0.000	0.67	1.26	12.44	0.000	0.79
ipuinitial1	2.27	4.00	0.000	0.96	4.08	9.91	0.000	3.04	2.77	5.94	0.000	0.96
taca	0.04	22.56	0.000	2.27	0.05	48.70	0.000	4.64	0.03	32.95	0.000	2.34

*表中の--は変数が生起しなかったことを示す。

Estimate はモデルによって推定された偏回帰係数である。z value は偏回帰係数を標準誤差で割って標準化した値で、その確率が Pr(>/z/)である。偏回帰係数はその独立変数の従属変数（閉鎖の弱化）への影響の強さを表しているが、TACA とダミー変数とでは単位が異なるので直接比較できない。そこで Pr(>/z/)が 0.001 未満の変数について、標準化偏回帰係数（表中の SPRC。偏回帰係数と当該独立変数の標準偏差の積を従属変数の標準偏差で除した値）を計算した。SPRC を用いれば各独立変数の従属変数への影響の強さを直接に比較することができる。いずれの音素についても TACA の SPRZC の絶対値が最も大きな値を示しているのは、前節までの分析結果と合致する結果である。

表 6-9 にはロジスティック回帰式による予測精度（closed data での正答率）を示した。表 6-8 に示した全独立変数を用いた分析に加えて、言語的変数（suwinitial, luwinitial, apinitial, ipuinitial）と TACA を独立変数に用いた分析と TACA だけを独立変数とした単回帰分析の結果も示した。言語的変数＋TACA による予測精度は全変数よりも 2%弱低下し、TACA のみによる予測精度はさらに最大で 5%弱低下するが、それでも 69~77%の精度を保っている。

表 6-9: 有声破裂音弱化の予測精度

	/b/	/d/	/g/
表 6-8 の全変数	72.0%	81.1%	78.1%
言語的変数と TACA	71.5%	79.2%	76.3%
TACA のみ	68.8%	76.7%	71.3%

6.6 音韻体系との関係

本節では上に報告した分析結果について音韻体系の観点から検討を加える。日本語（東京語）の有声破裂音と調音位置を共有する有声摩擦音と鼻音は図 6-6 の音素記号のように分布しており、/b/, /d/, /g/の 3 音素はすべて異なる対立関係におかれている。対立が最も複雑なのは歯裏（歯茎）音であり、破裂音、摩擦音、鼻音のすべてが対立している。両唇音には破裂音と鼻音の対立があるが、軟口蓋音には破裂音しか存在しないので、三者の調音様式間に積極的な対立は存在しない。

このような場合、各音素の異音の実現範囲も音韻対立の複雑さを反映して変化すると考えられる。図 6-6 中の楕円は音韻対立が単純な場合には異音の実現範囲が変化し拡大する様を概念的に表したものである。このような関係が先に図 6-4 に示した TACA と RSA のグラフにも反映されている。図 6-7 は図 6-4 と同じ TACA-RSA 関係のグラフであるが、発話速度による区分を行っておらず、代わりに/g/を/Ng/環境と~/Ng/環境とに二分し、また第 5 章に報告した/z/のデータを追加表示したものである。

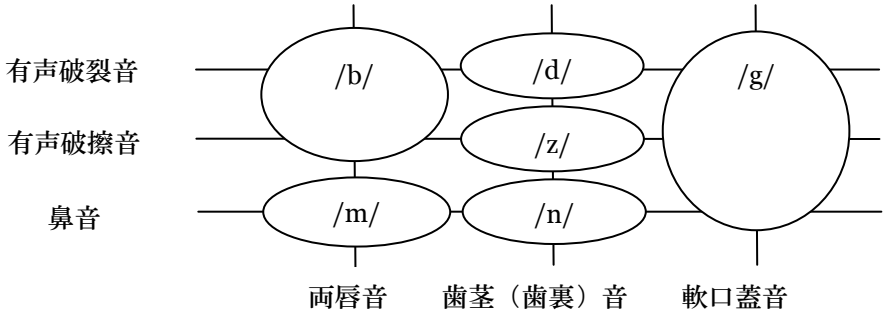


図 6-6: 日本語有声破裂音・有声破擦音・鼻音の分布

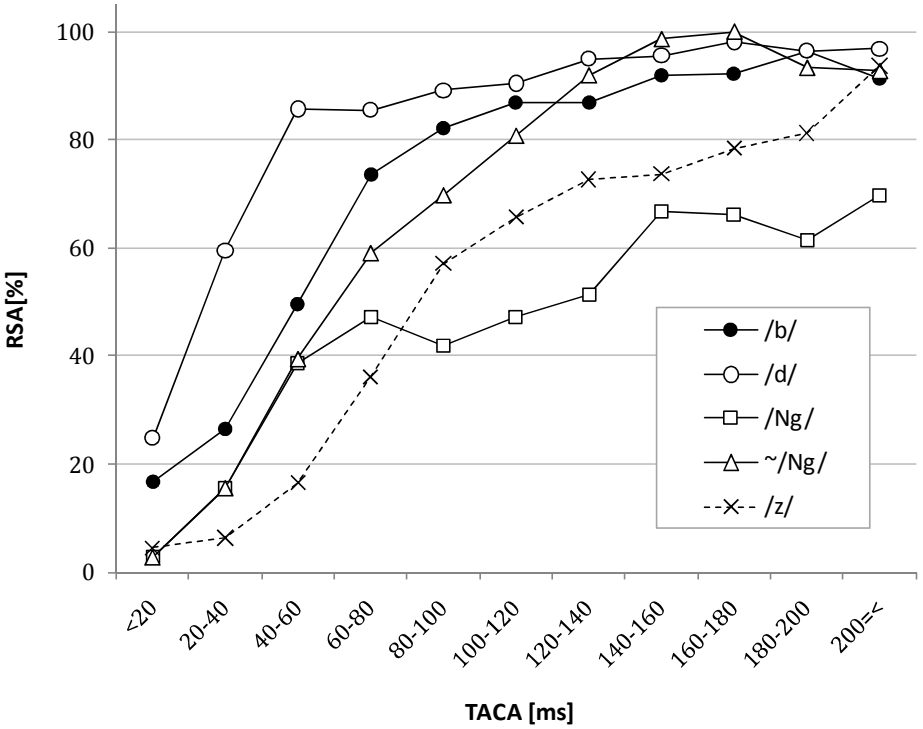


図 6-7: /b/, /d/, /g/及び/z/における TACA-RSA 関係

興味深いのは RSA が 90%以上に達するグラフの右端を除けば、TACA の広い範囲において、所与の RSA に対応する TACA の値について/d/ < /b/ < /g/の関係が成立していることである。例えばグラフから RSA が 80%に達する TACA の値を読みとると/d/では 50ms 前後、/b/は 90ms 前後、/g/(~/Ng/)では 110ms 前後であることがわかる。これは図 6-6 との関係でいえば音韻対立が複雑な破裂音であれば相対的に短い TACA で破裂調音が実現され、音韻対立が複雑でない場合にはより多くの TACA が必要とされることを示唆している。

さらに興味深いのは/z/と/Ng/の比較である。/z/における声道の閉鎖（つまり破擦調音）は音韻の対立には関与していない。そのため図 6-7 における/z/のグラフは/b/, /d/よりも緩やかに上昇する。しかし/z/のグラフは TACA とともにほぼ単調に増加し、最終的な破擦音の比率は 95%以上に達する。第 5 章の結論ではこの事実をもって/z/は理想状態においては破擦音となるべく調音されていると推定した。

これに対して/g/のうち/Ng/環境のサンプルはグラフの横軸の終端に達しても高々70%の水準に達するに過ぎない。グラフの概形から判断しても TACA 値 160ms 前後で既に高原状態が生じており、グラフの右端を越えて TACA 値を増大させても RSA が 100%に達することはないと思われる。この点で/Ng/環境における/g/は/b/, /d/, ~/Ng/はもとより/z/とも異なったふるまいを示している。今回のデータに記録された音声に関する限り、/Ng/環境における/g/が理想状態においては有声破裂音となるべく調音されているとみなせる証拠は見つからなかった。一方~/Ng/環境の/g/は図 6-3 のデータから有声破裂音となるべく調音されているものとみなすことができる。

ひとつの可能性として指摘するならば、現代東京語の/Ng/環境では、鼻音の直後という自然な同化の環境において、かつて東京語においてひとつの音素として存在していたと思われるガ行鼻濁音/ŋ/の調音運動が化石的に保存されているのかもしれない。そうであるとするならば、/g/において/b/,/d/とは反対に若い世代ほど RSA が上昇していたこと（6.4 節）は、化石化した/ŋ/の調音が最終的に消滅する方向への変化であるとみなすことができるかもしれない。

6.7 結論

本章では、現代日本語（標準語）の有声破裂音/b/, /d/, /g/における閉鎖調音弱化現象の大部分が第 5 章で扱った/z/の変異と同様、TACA によって説明できることを示した。先行研究において変異の要因として指摘されることがあった語ないし言語単位中の位置は、副次的要因として限定的に機能しているが、その効果が発揮されるのはアクセント句や発話といった上位単位の冒頭に限られていた。この点においても、本章の分析結果は/z/の分析結果と良く一致している。第 5 章と本章における分析結果の一致は、それらが同一のコーパスの分析結果であることによって一層高い信頼性を獲得している。

ただし、/g/については一部で/b/, /d/, /z/と異なるふるまいが観察された。その大部分は撥音直後の/g/が [ŋ]として実現されやすいことに起因していたが、/g/では若い世代の話者ほど破裂率が上昇する傾向については撥音との関係が認められなかった。この問題については今後の研究が必要である。

本章ではもうひとつ新しい知見の発見があった。TACA と閉鎖調音の実現率の関係が音素によって異なっており、そこには各音素の調音位置における調音様式に係る音韻対立の複雑さが反映されていた。音声的なパラメータである TACA と本質的に記号の問題である音韻対立との間にこのように明白な関係が成立していることは Lindblom(1998)の唱える音声変異の

Hyper- and Hypo-articulation 理論とも合致する事実であり、大変に興味深い。この事実は一見すると調音運動の生理学的限界によって生じているようにみえる/z/や/b/, /d/, /g/の音声変異が、実は広い意味では言語の体系によっても制御された変異であることを示唆している。

第Ⅲ部

第 7 章. 助詞のアクセントの独立性

7.1 問題

朗読音声と自発音声とでは、そもそも研究対象とする現象のあり方自体が異なっていることがある。本章では、そのような韻律現象の代表として、2 モーラ有核助詞のアクセントをとりあげる。

東京語の長さ 2 モーラの助詞には、「マ┐デ」「ヨ┐リ」「ノ┐ミ」のように、有核語（アクセントをもつ語）として辞書に記載されているものがある。ただし、これらの助詞は、先行する自立語のアクセント属性によって実現形を変化させると考えられている。例えば、東京語の包括的なアクセント規則を示した秋永(2001)の規則 71 は、助詞が名詞に後続する場合を扱っているが、「ほとんど名詞類のアクセントの型は変わらない。おもに型が変わるのは助詞のほうである。規則的。」と述べている。助詞が動詞、形容詞に後続する場合についても同様の記述がなされている。

同書から例をひけば、「ヨ┐リ」「サ┐エ「デ┐モ」などの助詞は、無核語につけば「トモダチヨ┐リ」のようにアクセント核を保存するが、有核語につくと「オトート┐ヨリ」「アサ┐ガオヨリ」「コ┐ーモリヨリ」のように核を喪失すると記述されている。

ところで、このような助詞のアクセント変化が常に生じるかというと、そうとはいえない。個人によって、あるいは発話の状況によって、「朝顔より」が「アサ┐ガオヨ┐リ」と発音されることがある。以下、本稿では、このような場合、助詞が韻律上の独立性を保持していると言うことにする。反対にアクセント核が消去される発話においては、助詞は韻律上自立語に従属していると言うことにする。助詞が韻律上の独立性を保持する発話がありうることを明言した論文に川上(1957b)や上野(1977)がある。

「アサ┐ガオヨリ」と「アサ┐ガオヨ┐リ」ないし「コ┐ーモリヨリ」と「コ┐ーモリヨ┐リ」の間には、主知的な意味の違いは存在しないから、これは東京語における韻律上の変異とみることができる。しかし通常の変異現象とはちがって語の音韻的構成要素ではなく、語と語の連接のありかたに係る変異である点に特徴がある。

東京語韻律の基本単位としてアクセント句(accentual phrase)を認める立場から解釈すれば（Poser 1984; Pierrehumbert and Beckman 1988）、この現象は先行する自立語と 2 モーラ有核助詞がひとつのアクセント句を形成するか、個々にふたつの句を形成するかにかかわる変異である。また Hattori (1961)、服部(1973)などのアクセント素(prosodeme)の理論にしたがえば、助詞のアクセントが通常のアクセント素であるか副次アクセント素であるかの変異ということになるであろう。

音韻論上どのような立場をとるにせよ、この現象は東京語の韻律構造の基本にかかわる重要な変異現象である。例えばテキストからの音声合成（TTS）の規則は必ずこの現象を処理しなければならない。しかしこの種の変異現象についての調査研究は、従来十分にはおこなわれてきていない。

以下本稿では現代の東京語における有核２モーラ助詞の発音実態に対する定量的な理解を深めるために、自発音声コーパスを解析した結果を報告する。

7.2 データ

7.2.1 『日本語話し言葉コーパス』

データベースとして『日本語話し言葉コーパス』（以下 CSJ）を利用する。CSJ は現代日本語の自発音声を多量に格納した音声データベースである。話者の出生地は日本全国にわたるが、首都圏出身者が多数ふくまれているので本研究の対象として好適である。また発話のレジスターについても、ある程度幅のある発話が収録されているので、それを分析に利用することができる（7.3.5 節参照）。CSJ の設計及び実装技術については第 2 章参照。

7.2.2 助詞の選択

２モーラ有核助詞のうち表 7-1 に示した 10 語「スラ、サエ、ノミ、マデ、ヨリ、コソ、シカ、ナド、デワ、デモ」を調査対象とした。このうち「デワ、デモ」の 2 語は、1 モーラ助詞の連鎖であるが、本来の２モーラ助詞と比較するために分析対象にくわえてある。

２モーラ有核助詞には、この他に「シモ、トテ、ヤラ、ユエ」などがあるが、これらは CSJ における出現頻度が非常に低いか、「必ずシモ」のように固定された文脈にしか生起していないので、分析対象にはくわえなかった。

表 7-1: 分析対象データのサンプル数

助詞	CSJ	首都圏	聞き取り	先行有核	先行尾高	N
スラ	90	47	47	21	7	14
サエ	284	118	118	51	8	43
ノミ	1,031	368	368	210	14	196
マデ	10,599	4,979	350	137	13	124
ヨリ	4,175	1,898	350	191	16	175
コソ	564	265	265	96	6	90
シカ	2,188	1,021	350	227	36	191
ナド	5,163	2,496	350	188	16	172
デワ	14,541	6,051	350	189	17	172
デモ	11,233	6,069	350	154	12	142

表 7-1 の第 2 列は分析対象とした 10 助詞の CSJ における生起数である 。次に話者の出生地を東京、神奈川、埼玉、千葉に限定すると第 3 列のサンプルが残る。第 3 列のサンプル数は助詞によって大きく異なっており、100 倍以上のひらきがある。そこで生起数が 400 以上

の助詞については、等間隔抽出法によってランダムに 350 サンプルを抽出した。この処置を施した後のサンプル数が表の第 4 列である。

7.2.3 音声環境の選択

次に表 7-1 第 4 列に該当する全サンプルを聴取して、以下の 3 条件に該当するか否かを判定した。このうちひとつにでも該当するサンプルは分析対象から除外することにした。

(1) 対象とする助詞に先行するアクセント句が無核であるもの。
(2) 助詞直前の語のアクセント型が尾高のもの。
(3) その他不適当と判断されるもの

条件(1)によって助詞に先行する語を有核語(有核アクセント単位)に限定する。ただし先行語が尾高型であると聞き取りが難しいので、条件（2）によって除外する。

条件(3)に該当するのは、「デハ」「デモ」のうち、「デハ、始めましょう」「デモ、これで良かったんです」のように接続詞として機能しているもの、助詞が言い直されたり繰り返されたりしているもの（「高校の先生よヨリも」「給付金しかシカ支払われない」など）、また、わずかではあるが形態素解析が誤っているもの（「スラって喋ってる」）などである。

表 7-1 の第 5 列は、条件(1)で除外されるサンプルを除外したのちに残るサンプルの数である。また、第 6 列は条件(2)ないし(3)によって除外されるサンプルの数を示している。第 7 列は第 5 列から第 6 列を引いた値であり、これが最終的に分析対象としたサンプルの数である。「スラ」はサンプルが少ない。

7.2.4 助詞アクセントの聞きとり

表 7-1 の第 7 列に該当する全サンプルについて、アクセント核が音声的に実現されているかどうかの聞きとりを実施した。判定者 2 名（KM と YI。YI のみ東京出身者）が、助詞部分にアクセントが知覚されるかを判定した。両名とも、F0 を視認するなどの補助手段は一切利用せず、聴覚印象のみに基づいて判断をくだした。また聞きとりに先立って判定者間で聞きとりの規準について相談したり調整したりすることもおこなっていない。

実際の判定では X-JToBI の BI 判定基準(五十嵐・菊池・前川 2006)に従って 3 段階の判定をくだしたが、本稿では助詞にアクセント核が知覚されたか否かに二値化した結果を利用する。判定結果は核が知覚されれば 1 、そうでなければ 0 である。

表 7-2: 判定者２名の判定結果

KM \ YI	無核	有核	合　計
無核	676	229	905
有核	48	370	418
合　計	724	599	1323

表 7-2 に判定者 2 名の聞き取り結果のクロス集計を示す。これはすべての助詞をプールした場合の結果である。表 7-3 には助詞ごとに細分して判定の一致率を示した。全体としては約 8 割の一致率が得られている。また、判定の不一致には方向性があり、KM が無核と判定するサンプルを YI が有核と判定することが、その逆よりも圧倒的に多い。つまり不一致の大部分は、YI が KM よりも核の存在に敏感であることによってもたらされている。

表 7-3: 判定結果の一致度

助詞	スラ	サエ	ノミ	マデ	ヨリ	コソ	シカ	ナド	デワ	デモ
一致率	92.9	82.2	81.1	76.6	87.4	81.1	85.9	58.7	79.7	78.9
N	14	43	196	124	175	90	191	172	172	142

7.3 分析

以下の分析では、助詞の独立性に関する KM と YI の判定結果を単純に平均した値を分析の対象とする。そのため独立性の判定値は 0.0, 0.5, 1.0 のいずれかの値をとる。0.5 は２名の判定が一致していない場合である。

7.3.1 助詞の意味特性の影響

図 7-1 に判定結果の比率を百分率で示す。判定 0、すなわち 2 名の判定者が一致してアクセント核が知覚されないと判定したサンプルの比率が高いのは、「マデ」「ヨリ」「シカ」であるが、そのような助詞においても、全体の 5%から 10%程度のサンプルは１と判定されている。反対に「スラ」「サエ」「ノミ」「コソ」では判定１の率が 50%以上に達している。図 7-1 は、助詞による程度の差はあれ、韻律上、先行自立語に完全に従属してしまう助詞はひとつもないことを示している。

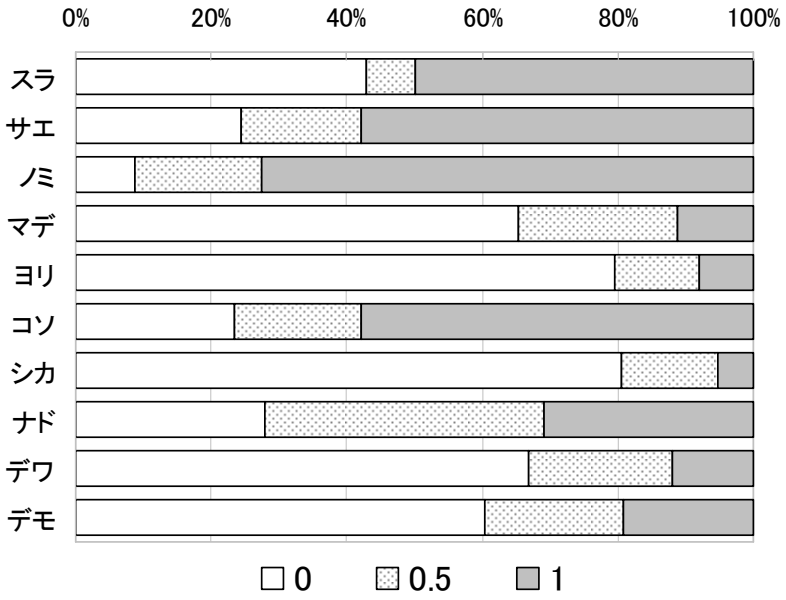


図 7-1: 判定結果の助詞ごとの比率（KM と YI の平均値）

次に助詞の意味的特性と判定結果の間の相関を検討する。独立することの多い助詞―「スラ」「サエ」「ノミ」「コソ」―はおしなべて副助詞であり、さらにとりたての意を含んだ助詞である。

反対に格助詞「ヨリ」や格助詞に係助詞が後続した「デワ」「デモ」は独立することが最も少ない。「ナド」「マデ」は文法的には副助詞に分類されるが、「スラ」「サエ」等に認められる限定／対比の意味は認められない。念のため注記すれば「マデ」には「お前マデ裏切るのか！」のようにとりたての意味での用法がありうるが、今回のデータにはその用例は含まれていなかった 。

上の観察からは、とりたての助詞は独立しやすいという仮説が導かれる。ところがその例外となる助詞がある。「シカ」である。この副助詞は明らかにとりたての意味を表すにもかかわらず、独立することが非常に少ない。

しかし、これは例外でないと考える根拠がある。「シカ」のアクセント核は音声的な要因によって実現を阻害されているからである。「シカ」のアクセント核を担う第 1 モーラの母音/i/は、前後を無声子音に挟まれており、しかも後続モーラの母音は非狭母音である。これは母音に最も無声化が生じやすい環境である(前川 1989、Maekawa and Kikuchi 2005)。母音が無声化したモーラにアクセントを置くことは、不可能ではないにせよかなり困難であるから、この助詞が他の比況の助詞に比べて独立することが少ないのは当然である。ちなみに今回分析対象とした助詞で母音無声化の典型的音声環境を含んでいるものは「シカ」以外にないことに注意。

7.3.2 核間距離の影響

上野(1977)は東京語の助詞「マデ」の核について、先行する自立語が「有核の自立語の場合でも、核がより語頭に近くて、マデの核から離れていればいるほど（中略）実現しやすくなる傾向がある」と述べている(p.287)。この傾向が今回のデータにも認められるかどうかを検討した。最初にすべての助詞をプールして分析した結果を図 7-2 に示す。

この図の横軸は助詞に先行する要素中のアクセント核位置を要素の末尾からのモーラ数で表している。数値が大きくなると助詞中のアクセント核と先行要素中のアクセント核との距離（核間距離）がひらいていることになる。ただし尾高型はもともと除外しているので距離が 1 となることはない。バー上端の数字は各バーに含まれるサンプルの総数である。

助詞が韻律的に独立したサンプル（判定 1）の比率は、距離が 2 から 5 までの範囲では、距離に比例して増大する。また、同じ範囲では、助詞が韻律的に独立していない（判定 0）サンプルの比率は距離に反比例して減少している。

距離が 6 以上になるとこの傾向がくずれる。これには原因がある。この条件に該当する 16 サンプルのうち 12 までは「本研究では」という発話（ﾎﾝｹﾝｷｭｰﾃﾞﾜ；距離＝6）である。前節で述べたように「デワ」は独立性が弱い助詞であるので、それがバイアスとなって上記の傾向が崩れている。試しに「デワ」を除いた 4 サンプルだけで計算してみると判断 0 の率が 25%、同 1 の率が 50%となって、期待される傾向に近づく。ちなみに上記 4 サンプルに含まれる助詞は「スラ」「コソ」「ナド」（2 例）であった。

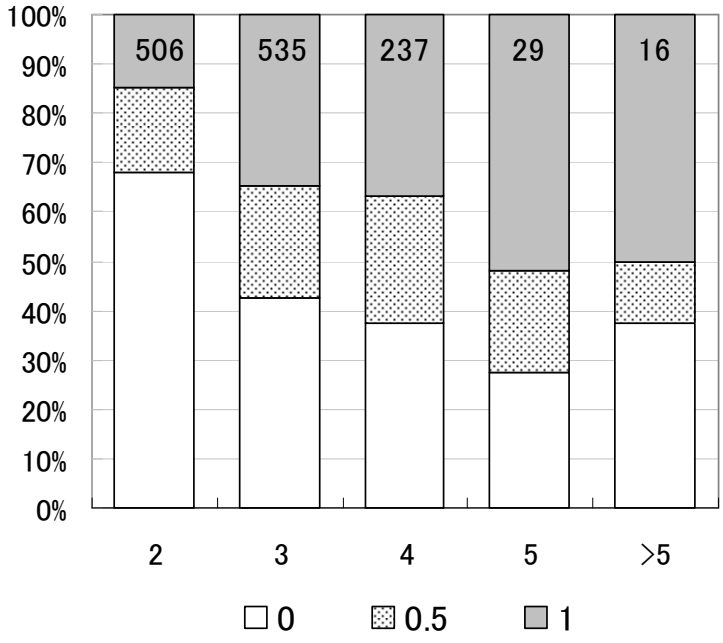


図 7-2: 核間距離との関係

次に助詞を個別に分析する。図 7-3 は、「スラ」を除いた 9 助詞について、核間の距離による判定 1 の変動を示したグラフである。距離に反比例して判定 1 の百分率が増加するのは「ナド」と「コソ」であり、「マデ」もこれに近い。それ以外の助詞には距離と独立性の相関を認め難い。図 7-2 に観察された距離との相関は、「ナド」「コソ」「マデ」における相関が反映されたものと考えられる。

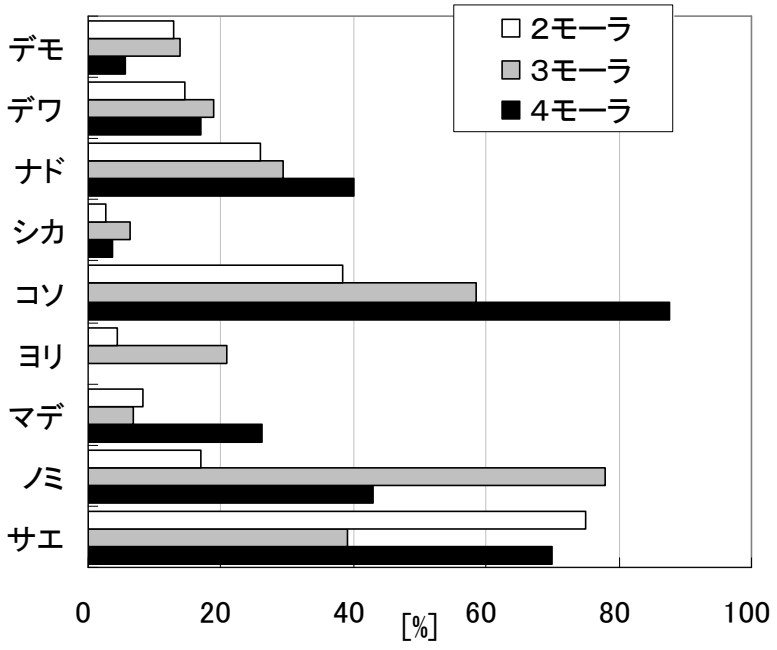


図 7-3: 助詞ごとにみた核間距離の影響（判定 1 の百分率）

7.3.3 句末音調の影響

問題となる助詞に句末音調が生じていることがある。その影響を評価した結果を表 7-4 に示す。句末音調は『日本語話し言葉コーパス』のために開発した X-JToBI の BPM カテゴリー（第 3 章参照）に従って分類している。

表 7-4: 句末音調との関係*

判定 / 句末音調	0	0.5	1
L%H%（含 FR）	121 (82.9)	18 (12.3)	7(4.8)
L%HL%	25 (89.3)	1(3.7)	1 (3.7)
L%HL%(PNLP)	17 (100.0)	0 (0.0)	0 (0.0)
L%（句末音調無し）	518 (45.9)	248 (22.0)	362 (32.1)

*()内は百分率

L%H%は句末の１モーラ内部で上昇が生じる単純な上昇音調である。この上昇調には上昇音調の開始時刻が次末モーラないしは次々末モーラに移動しているものがある。川上(1963)が「浮き上がり調」と呼んでいるものであり、X-JToBI ではこれを FR(floating rise)と呼んでいるが、表 7-4 ではこの二種を区別せず、ひとつにまとめている。

L%HL%は上昇下降調である。上昇下降が句末モーラ内にほぼ収まるものは単純な L%HL%だが、上昇が早く始まって上昇・下降のピークが句末からふたつ目のモーラに位置するものを、X-JToBI では PNLP(Penultimate Non-Lexical Prominence)と呼んでいる。表 7-4 で L%HL%と PNLP とを別個に扱っているが、これについては注意が必要である（下記参照）。

最後に L%は句末に特徴的な音調が生じずに下降しているケースである。この他に L%LH%―川上(1963)が「反問の上昇」と呼ぶもの―が 4 例生じており、判定はすべて 0 であったがサンプルが少なすぎるので表からは除外した。

表 7-4 の各セルの左側にはそのセルに該当するサンプルの実数を示し、その右側の括弧中には百分率を示した。何らかの特徴的音調が生じている場合とそうでない場合（L%）を比較すると、句末音調が助詞の独立を妨げていることがわかる。

ここで PNLP についての注意を述べる。記号↑と↓でピッチの上昇と下降を表すならば、PNLP は「デ┐ー↑タ↓ガ」「ディズニーラ┐ン↑ド↓デ」のように句末からふたつ目のモーラにピッチの上昇下降が観察される音調である。しかし、X-JToBI では、句末からふたつ目のモーラに（語彙的な）アクセント核が存在すると解釈できる発話には PNLP を認定しないことにしている。そのため、「道路より」「彼すら」「今でこそ」のように本稿が対象としている 2 モーラ有核助詞が句末に位置した発話には定義上 PNLP が生じない。「ド┐ーロ↑ヨ↓リ」のような発話は助詞が韻律的に独立した発話として（すなわち「ド┐ーロヨ┐リ」として）認定されるだけである。

表 7-4 で L%HL%（PNLP）として認定されているのは、本稿の分析対象である助詞群にさらに別の助詞が後続し、そこが句末となっている発話である。具体的には「了解度よりも」「トラックなどが」などの発話である。これらが「リョーカ┐イドヨ↑リ↓モ」「トラ┐ックナ↑ド↓モ」のように発音された場合に PNLP を認定している。

7.3.4 先行語の品詞の影響

助詞の直前、直後の語の品詞の影響も検討した。直後の語の品詞については興味ある結果は得られなかった。直前の語については表 7-5 の結果を得た。表の表示方法は表 7-4 と同一である。また表 7-6 は判定対象の助詞と先行語品詞によって、サンプル数の分布をクロス集計した結果である。

表 7-5: 先行語の品詞との関係

判定 / 品詞	0	0.5	1
助詞	106 (45.1)	38 (16.2)	91 (38.7)
代名詞	40 (88.9)	4 (8.9)	1 (2.2)
動詞	29 (51.1)	6 (14.3)	7 (16.7)
名詞	495 (51.9)	212 (20.1)	262 (28.0)

表 7-6: 助詞と先行語品詞のクロス集計

先行品詞 / 助詞	助詞	代名詞	動詞	名詞
コソ	65	0	0	24
サエ	20	0	0	25
シカ	72	1	15	98
スラ	7	0	0	7
ノミ	28	0	4	161
デワ	7	0	1	164
デモ	12	7	0	117
ナド	7	0	3	153
マデ	14	27	9	73
ヨリ	3	10	10	147

表 7-5 をみると、先行語が助詞である場合に判定 1 が増加していることがわかる。しかし、ここに言語学上の因果関係を看取するのは早計である。表を詳しく検討すると、「コソ」「サエ」「シカ」「スラ」が助詞に後続しやすいのに対して、「デハ」「デモ」「ナド」「マデ」「ヨリ」は、大部分が名詞に後続している。つまり、7.3.1 節で指摘したとりたての意味機能を有する助詞、すなわち韻律的に独立しやすい助詞がほぼそのまま助詞連鎖を生じやすい助詞と重なっているのである。

この問題をさらに検討するために、助詞を「コソ」「サエ」「スラ」「ノミ」（すなわちとりたての助詞）と「デハ」「デモ」「ナド」「マデ」「ヨリ」（その他の助詞）に二分し、先行語の品詞を助詞とそれ以外に二分して集計した（表 7-7）。各セルの数字は判定の平均値であり、括弧中はサンプル数である。助詞を二分するにあたって「シカ」を除外したのは、この助詞の判定には音声上の特殊要因が関係しているからである(7.3.1 参照)。

表 7-7 は、助詞を意味特性のクラス別に検討すると、先行品詞による判定値の差は無視できることを示している。したがって、助詞の独立性に影響する真の要因は助詞の意味的特性であると考えられる。

表 7-7: 助詞の意味特性と助詞連鎖の関係*

先行品詞 / 助詞	助詞	その他	全体
コソ、サエなど	0.73 (120)	0.76 (225)	0.75 (345)
その他	0.28 (43)	0.28 (744)	0.28 (787)
全体	0.61 (163)	0.39 (969)	0.43 (1132)

*()内はサンプル数

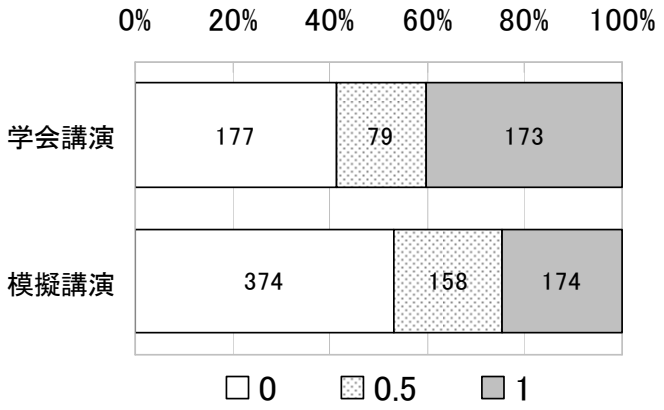


図 7-4: 学会講演と模擬講演の差

7.3.5 レジスターの影響

以上で言語的要因の検討を終え、続いて社会的要因を検討する。話者の性差、年齢差、レジスター（発話タイプ）の影響を検討した。このうち顕著な影響が認められたのはレジスターであった。

CSJ には「学会講演」「模擬講演」「一般講演」「朗読」「対話」などの発話が記録されている。学会講演は各種学会での研究発表の音声であり、相対的にあらたまった口調の発話が用いられやすい。模擬講演は人材派遣会社から派遣された話者による日常的話題のスピーチであり、全般的に学会講演よりもくだけている。一般講演は多数の一般聴衆を相手にした研究者による講演であるが、あらたまり度は学会講演よりも模擬講演に近いことが多い。対話にはインタビュー、課題指向対話、自由対話が含まれており、いずれも明らかにくだけた発話になっている。最後に朗読には自然科学の啓蒙書の文章を朗読したものと、自分が既に

実施した学会講演ないし模擬講演の転記テキストを再朗読したものとがあり、今回のデータにも両者がともに含まれていた。

今回のデータは、これらのレジスターをすべてカバーしているが、対話（N=31）と一般講演(N=33)はサンプル数が少ないので分析対象から除外することにした。また朗読については後述する理由で別途検討する。

図 7-4 に学会講演と模擬講演における判定値の分布を百分率で示す。バー内部の数字は判定値ごとのサンプル数である。学会講演音声では助詞の独立性が高まっていることがわかる。あらたまった発話では助詞が独立しやすくなる可能性を示唆する結果である（ただし 7.4.4 の議論参照）。

次に朗読音声を検討する。図 7-4 からは除外したが、今回のデータには朗読音声のサンプルも含まれており、図 7-4 と同様に集計すると、判定 0 が 79.8%、判定 0.5 が 8.9%、判定 1 が 11.3%となり、助詞の独立性が低い水準にとどまっていることがわかる。

ところで、今回分析した朗読音声中のサンプルは 124 個とさほど多くはない。また朗読音声にどの助詞がどれだけ生じるかはテキストによって完全に決まってしまうので、テキストに依存した偏りが生じる。実際、今回分析対象とした朗読音声に生じていた助詞は、「シカ」3 回、「デワ」4 回、「デモ」1 回、「ナド」6 回、「ノミ」11 回、「マデ」35 回、「ヨリ」64 回であり、「マデ」と「ヨリ」の頻度が高い。上に指摘した独立性の低さは、この偏りに起因している可能性がある。

表 7-8 は、助詞の意味特性とレジスターの関係を分析した結果である。セルの数字は判定値の平均、括弧内がサンプル数である。意味特性の分類は、「シカ」を除外している点も含めて、表 7-7 と同一である。このように助詞の意味特性を考慮すると朗読音声中の助詞が独立しにくいとはいえず、むしろ学会講演よりも独立性が高い可能性が示唆されている。

表 7-8: 助詞の意味特性とレジスターの関係

レジスター / 助詞	学会講演	模擬講演	朗読
とりたての助詞	0.82 (168)	0.67 (160)	0.86 (11)
その他の助詞	0.30 (226)	0.32 (417)	0.09 (110)

7.4 考察

前節における分析の結果、2 モーラ有核助詞の韻律上の独立性を促進する要因として、1）とりたての意味機能を有する助詞であること、2）先行する核との距離が大きいこと、3）助詞に句末音調が用いられないこと、4）発話のあらたまり度が高いことの 4 点を指摘した。これらの一々について簡単な考察を試みる。

7.4.1 とりたての助詞

図 7-1 に示された韻律上の独立性の助詞による変動は、他の 3 要因よりも大きな変動を生じている。また他の 3 要因のなかに助詞の意味特性統計的な交互作用を生じさせるものはない。むしろ、他の要因単独では結果が明らかでない場合に、助詞のクラスを区別することによって明快な結論が得られることがあった（7.3.4 節及び 7.3.5 節参照）。

これらのことから、助詞がとりたての意味機能を有するか否かが、前節で指摘した 4 要因のうち、もっとも強い要因であると考えられる。

とりたての助詞が韻律的に独立しやすいのは、とりたての助詞はそもそも意味的に強調をうけやすい助詞であるために、韻律的にも助詞を強調した（プロミネンスをおいた）発音がおこなわれやすく、助詞を強調する音声学的手段のひとつとして助詞を独立させることがおこなわれるのだと解釈できる。

川上(1953)は声の上昇をプロミネンスの特徴量とみなしているが、助詞が韻律上独立した発話においては、助詞の第 1 モーラに核を実現するための上昇が生じるから、助詞の独立は、程度の差こそあれ、自動的にプロミネンスを付加することになる。

ただし、2 モーラ有核助詞にプロミネンスを付与する音声学的手段はこれに限られるわけではない。他に、助詞の第 2 モーラだけにプロミネンスを付与する（高く発音する）方法がある。これは、郡(1997)が 1 モーラ助詞について「フォーカス語につく助詞やフォーカス語末を高める言い方」として指摘している方法である。この方法と本稿で考察した助詞を独立させる方法とでは、強調のあり方に何らかの語用論的な差異があるのかどうかは句末音調との関係（7.4.3 節参照）も含めて今後検討すべき課題である。

7.4.2 核間距離

アクセント核間の距離の効果については助詞による著しいばらつきがあった。以下に述べる理由によって、この要因は本稿が対象としたサンプルの範囲では本来あまり強力な影響を及ぼすものではないと考えられる。

核間距離が短い場合に韻律上の独立性が低下しやすい原因は、短い時間区間において複数の核を実現するために急激なピッチ制御をおこなうことが困難であるためと考えることができる。しかし本稿で分析の対象としているのは距離が最短で 2 モーラまでのサンプルにすぎないから、核の実現が全く不可能になるとは考えられない。そして実際のデータもそのとおりの結果を示しているのである。

ちなみに核間距離が 1 モーラになった場合、すなわち助詞に先行する語のアクセント型が尾高型である場合はどうなるかを予想してみると、この場合にも助詞が独立した発音は不可能ではないと思われる。その根拠としては、7.3.3 節に示したように同一の助詞内に核と句末音調とが共存しうることと、東京語では尾高型に頭高型が接続した場合（例えば「花高し」）にあっても双方のアクセント核が実現可能であることが指摘されていること（川上, 1953）の 2 点を指摘できる。

7.4.3 句末音調

句末音調が用いられるとそれによって助詞の独立が抑止された。そしてその効果は上述の核間距離が近接している場合の抑止効果よりも強力であった。この抑止効果は核間距離の場合と同じく基本的には複雑なピッチ制御の忌避に起因するものと考えられるが、句末音調を有する発話においては、核を担うモーラと句末音調を担うモーラとが同一助詞内で隣接している。すなわち核と句末音調間の距離が 1 モーラである。句末音調の抑止効果が相対的に強いのはそのためであろうと考えられる。

なお、句末音調のうち、L%HL%（上昇下降調）には稀にではあるが助詞が独立したサンプルが存在したのに対して、PNLP タイプ（第 9 章参照）の場合には助詞の独立が一切観察されなかった（表 7-4 参照）。サンプル数が少ないので明確な結論を下すことはできないが、表 7-4 で PNLP に分類した「トラ˥ックナ↑ド↓モ」の発話は、助詞の独立とは両立しえない可能性が高い。

この種の発話で助詞のアクセント核を実現し、なおかつ PNLP を実現するのは、7.4.2 節末で言及した「花高し」と同じではないかと思われるかもしれないが、PNLP におけるピッチのピークはプロミネンス実現のために一層高く実現されることが多いので（9.6 節参照）、核による下降との両立は困難になると考えられる。

7.4.4 あらたまり度

助詞の独立しやすさをレジスターとの関係で分析すると、「模擬講演」＜「学会講演」という序列が得られた。また「模擬講演」＜「学会講演」＜「朗読」であろうことが示唆された。

この序例を発話のあらたまり度の序列とみるのは自然な解釈である。しかし CSJ では、平均的にみると、学会講演では予め準備した原稿に頼る傾向が模擬講演よりも高いし（前川 2004）、朗読では完全に原稿に依存して発話がおこなわれている。そうすると、上の序列を原稿に対する依存度の序列ないしは自発性の低さの序列とみることも可能である。

どちらの解釈がより妥当であるかは今後の検討課題であるが、いずれの解釈をとるにせよ、助詞の韻律的独立性には言語的要因だけでなく社会的（非言語的）要因が関与していることは明らかである。

7.5 結論

本章では 2 モーラ有核助詞が有核語に後続した場合にも助詞のアクセント核が頻繁に実現していることを示した。このような現象は、朗読音声の研究ではあくまで例外的現象とみなされてきたが、現実の自発音声では、助詞の種類によっては高頻度で生じていることがわかった。助詞の種類としては、とりたての意味をもつ助詞において、とりわけ助詞の独立性が

高い。このような事実は、アクセントの音韻的記述に反映されるべきであるし、日本語学習者に対するアクセントの教育においては、とりわけ真剣に考慮される必要がある。

第 8 章. Final Lowering

8.1 問題

分節音の分析のうち、/z/の変異や/b/, /d/, /g/の変異の分析では、朗読音声の研究において未解決のまま残されていたか、あるいは誤った結論に導かれていた問題が、自発音声の分析によって解決された。本章では韻律現象の研究においても同様の事例があることを示す。本章で分析するのは final lowering と呼ばれるイントネーションの現象である。

Final lowering （以下 FL と省略）とは、平叙文の末尾で F0 が局所的に下降して、発話の終了を示す現象である。FL をイントネーション研究のなかに積極的に位置づけた最初の研究は英語の FL を扱った Liberman and Pierrehumbert (1984)である。このような発話末の下降現象は比較的多くの言語に観察されると思われるが、それについての実証的な研究は必ずしも多くない。FL の効果を端的に図示すると図 8-1 のようである。

図上部に示されているのは日本語(東京方言)における発話の韻律構造(Pierrehumbert and Beckman 1988)であり、図下部には F0 の下降傾向が模式的に示されている。F0 declination は韻律構造とは無関係に発話に要する時間の関数として単調に F0 が低下する傾向であり、ここでは直線をもって表されている。これに対して FL は発話の終端部にだけ生じる構造依存の F0 下降現象である。

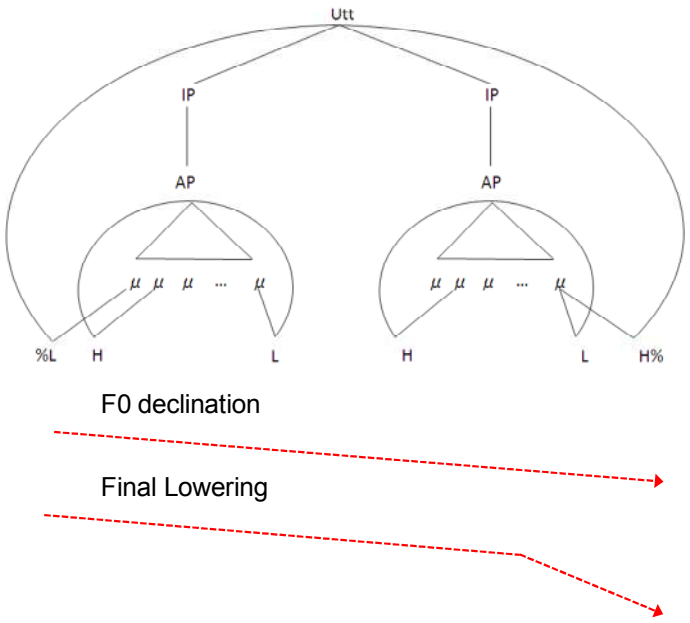


図 8-1: Final lowering の模式図

日本語の FL を扱った研究には Poser(1982)と Pierrehumbert and Beckman (1988)がある。これらの先行研究を以下では Poser、P&B と略記する。Poser 及び P&B は、日本語東京方言イントネーションの F0 下降メカニズムの一要因として FL を認める点では一致しているが、

FL の効果が及ぶ範囲については異なる結論を述べている。Poser が FL の効果が発話の最終モーラにのみ及ぶと考えているのに対し、P&B はより広い範囲に FL の効果が及ぶと考えている。

両者の相違点を図式化して示せば図 8-2 のようになる。この図は 3 個のアクセント句（AP1, AP2, AP3）から構成される名詞句の発話「彼が食べたラーメン」と、2 個のアクセント句から構成される動詞述語文の発話「彼が食べた」を比較している。実線は名詞句のイントネーションを模式的に示しており、点線と破線は動詞文末尾に FL が生じたことによるイントネーションの変化を示している。

Liberman and Pierrehumbert (1984)に始まる FL の理論は、平叙文発話の末尾で F0 はその話者のピッチレンジの下限(図中の base line)まで下降すると考えるので、点線と破線はいずれも名詞句発話の末尾（ここにも FL が生じる）と同一レベルまで低下している。

ここで点線は Poser による予測、破線は P&B による予測を示している。Poser は動詞文発話のイントネーションは「タ┐ベタ」の第 2 モーラ「ベ」までは名詞句発話と同一であり、最終モーラ「ダ」だけが FL によって低下しているとみる。これに対して P&B は FL の効果は最終モーラにとどまらず、より遠くにまで及ぶと述べているが、その範囲がどのようなものであるかについては明言を避けている。図 8-2 では仮に FL の効果が AP2 の冒頭まで遡るとみなして作図してある。

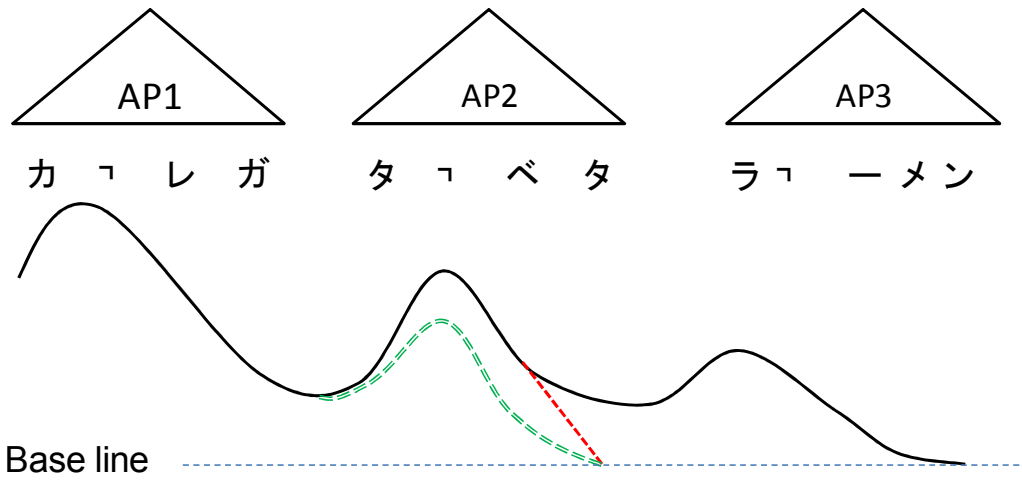


図 8-2: Poser と P&B における FL の扱い

以下本章では、まず、そもそも自発音声には FL が認められるかという問題を検討する。Declination や FL については、Umeda(1982)のように、自発音声には必ずしも存在しないとする立場があり、また藤崎博也氏の F0 重畳モデル(藤崎・須藤 1971, Fujisaki and Hirose 1984)や後述する NTT の音声合成システムのように、FL を積極的に表現しないイントネーションモ

デルも存在するからである。次いで、Poser と P&B の予測はどちらがデータに合致するか、また仮に P&B の予測が正しい場合、FL の効果はどのような範囲に及んでいるかを検討する。

8.2 データ

データとして、X-JToBI ラベリングが施された CSJ-Core の学会講演（APS）, 模擬講演（SPS）, 再朗読音声を利用する。本章では F0 値のデータと節(clause)末の形態論的特徴の関係に注目して、FL の効果を検討する。これらのデータの性質について簡単に説明を加える。

8.2.1 F0 値

F0 のデータは、X-JToBI 情報を記録した CSJ の XML ファイルに記録されている F0 値を利用する。これは X-JToBI の tone ラベルの時刻の F0 値を記録したものであるが、F0 値は母音の無声化などによって測定不能となることがあるので、必ずしもすべての tone ラベルに F0 値が付与されているわけではない。

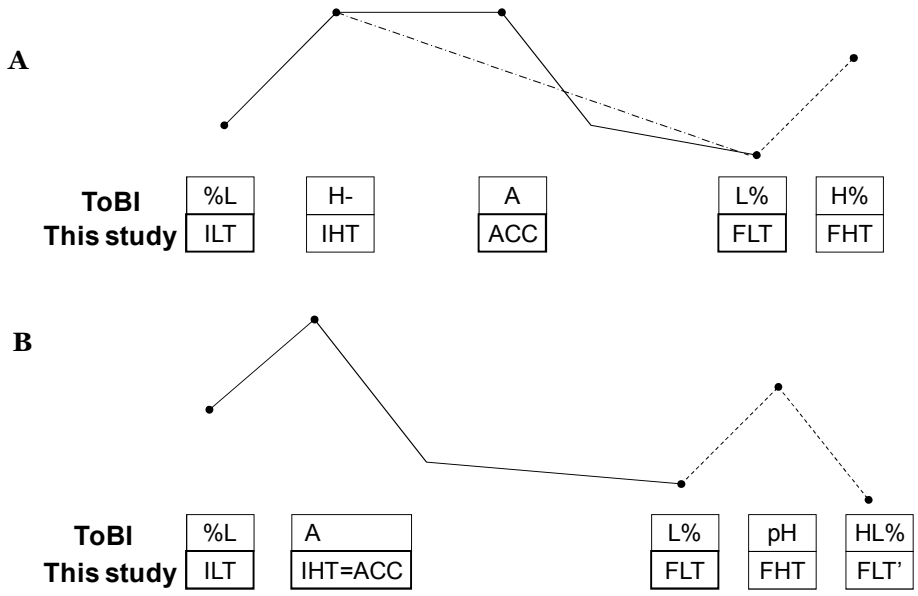


図 8-3: F0 測定点の名称

図 8-3 に本章で用いる F0 測定点の名称を示す。図中のパネルは有核（実線）及び無核（点線）のアクセント句の F0 形状を模式化して表現している。点は F0 測定点を表し、各点下部に X-JToBI での Tone の名称と、本研究での名称を上下に分けて示している。

アクセント句の冒頭を特徴づける F0 上昇は、基点（ILT: Initial Low Tone）と頂点(IHT: Initial High Tone)によって表される。有核アクセント句の場合はアクセントに対応する F0 の

ピーク値が（ACC:Accent）によって表現される。アクセント句末における F0 の最低値が FLT(Final Low Tone)である。

パネル A のように、アクセント句末に上昇 BPM（第 9 章参照）が生じていれば、上昇の頂点の F0 値が FHT(Final High Tone)によって表され、パネル B のように上昇下降 BPM が生じていれば、上昇の頂点が FHT で、下降後の最低値が FLT’で表される。

ちなみにパネル B はアクセント句冒頭にアクセントが位置する頭高型のアクセント句の例である。その場合、句頭の上昇の頂点である IHT とアクセント由来の頂点である ACC を区別することが不可能なので、本研究ではこれを一律に IHT として扱っている。本章では、これらの測定値のうち主に FLT に注目するが、一部で ILT と IHT にも言及する。

表 8-1: 節単位情報

クラス	節境界の種類
絶対境界(ABSOLUTE)	[文末]，[文末候補]，[と文末]
強境界(STRONG)	/並列節ガ/，/並列節ケレドモ/，/並列節ケレド/， /並列節ケドモ/，/並列節ケド/，/並列節シ/，/ヨウニ節/
弱境界(WEAK)	<条件節タラ>，<条件節タラバ>，<条件節ト>， <条件節ナラ>，<条件節ナラバ>，<条件節レバ>， <理由節カラ>，<理由節カラニハ>，<理由節カラ-助詞>， <理由節ノデ>，<タリ節>，<タリ節-助詞>， <テ節>，<テハ節>，<テモ節>，<テカラ節>， <テカラ節-助詞>，<テ節-助詞>，<トカ節>， <トカ節-助詞>，<ノニ節>，<連用節>，<引用節>， <引用節-助詞>，<引用節トノ>，<トイウ節>， <間接疑問節>，<間接疑問節-助詞>，<連体節テノ>， <並列節ダノ>，<並列節デ>，<並列節ナリ>， <フィラー文>，<感動詞>，<接続詞>， <接続詞 C>，<接続詞 L>，<接続詞 CL>，<接続詞 M>

8.2.2 節単位情報(CBL)

自発的なモノローグや対話のデータに「文」や「発話」を認定することには困難がつきまとうが、「節」(clause)であれば、比較的に安定した認定が可能であることが知られている（丸山・高梨・内元 2006）。CSJ には節を一旦 49 種類に分類した上で、3 種類のクラスに再分類した節単位情報(Clause Boundary Label)が付与されている。表 8-1 にその内訳を示す。絶対境界がもっとも強い文末であり、この文末の前後では統語的な修飾関係の大部分がリセットされる。以下、強境界、弱境界の順に境界の強さが低下する。

節単位情報の各クラスを代表する例を挙げる。いずれも丸山・高梨・内元(2006）に示されている例文であり、CSJ からの実例である（本研究第 3 章も参照）。

A) 絶対境界

- 簡単に最初に復習をしておきたいと<引用節>思います**[文末]**
- あたしニワトリ歩いてるのって見たことが今までなかったんですね**[文末]**
- まず左側の方を御覧ください**[文末]**
- 一番左が最初千文書を取得した段階での初期検索の精度ですね**[文末候補]**
- それにしても<テモ節>私が本で読んだあの情報は一体何だったのでしょうか**[文末候補]**
- で<接続詞>総論では賛成なんだけれども/並列節ケレドモ/ 少し煮詰める必要があるんではないだろうかと**[と文末]**

B) 強境界

- で<接続詞>結果ですが/**並列節ガ**/　まず絶対音感群の結果から見ていきたいと<引用節>思います[文末]
- まずその教室の様子ですけれども/**並列節ケレドモ**/ 教室の名前は俳句文法教室と言います[文末]
- で<接続詞>その豪華な船に乗れたっていう<トイウ節>体験もできたし/**並列節シ**/　楽しかったと<引用節>思っています[文末]
- 先程申し上げましたように/**ヨウニ節**/ 条件付き確率の式はこのように異なっています[文末]

C) 弱境界

- で<接続詞>その時にちょうど先生がもしかしたら**<条件節タラ>**私は東京都指定の特殊な難病かもしれないと<引用節>言われました[文末]
- 翌日朝食を済ませると**<条件節ト>**私達は旅行代理店へ向かいました[文末]
- もし形式的にこの会話の中で起きた他の相づちに習うならば**<条件節ナラバ>**ここではそうかなそうだよといったものになるはずなるのではないでしょうか[文末候補]
- で**<接続詞>**ついでにこれもいい日本語に替えてくれれば**<条件節レバ>**いいなと<引用節>そんな気がしてます[文末]
- どうしてもこの遊びの方が先に入ってたから**<理由節カラ>**そっちを優先しちゃったっていう<トイウ節>ことでした[文末]
- 縮むことがないので**<理由節ノデ>**安心して<テ節>何度でも洗います[文末]

- ある時は失恋して<**テ節**>泣いている私に寄り添ってもくれました[文末]
- もう初日から彼らを探しては<**テハ節**>きゃあきゃあ騒ぎ過ぎて<**テ節**>もう声枯れしちゃって<テ節>大変でした[文末]
- 寝たきりになってからは<**テカラ節-助詞**>どんどんと衰弱していきました[文末]
- 洗濯機の音ですとか<**トカ節**>テレビの音ですとか<**トカ節**>色々あります[文末]
- 花火が上がり<**連用節**>紙吹雪が舞い<**連用節**>みんなもう酔っ払ったアメリカ人が踊りながらハッピーニューイヤーと叫び合います[文末]
- 何となく雰囲気は分かったと<**引用節**>思います[文末]
- 何か汗がすぐ塩になっちゃう感じで<**並列節デ**>あんまりだらだらと汗をかいたっていう<**トイウ節**>イメージがないです[文末]
- とにかく夢のような八日間で<**並列節デ**>今でも本当にそこに行ったのか<**間接疑問節**>時々分からなくなるくらいです[文末]
- 横軸が時間で<**並列節デ**>縦軸が周波数になっています[文末]
- だから<接続詞>そういう意味では歴史のある町なんですね[文末]

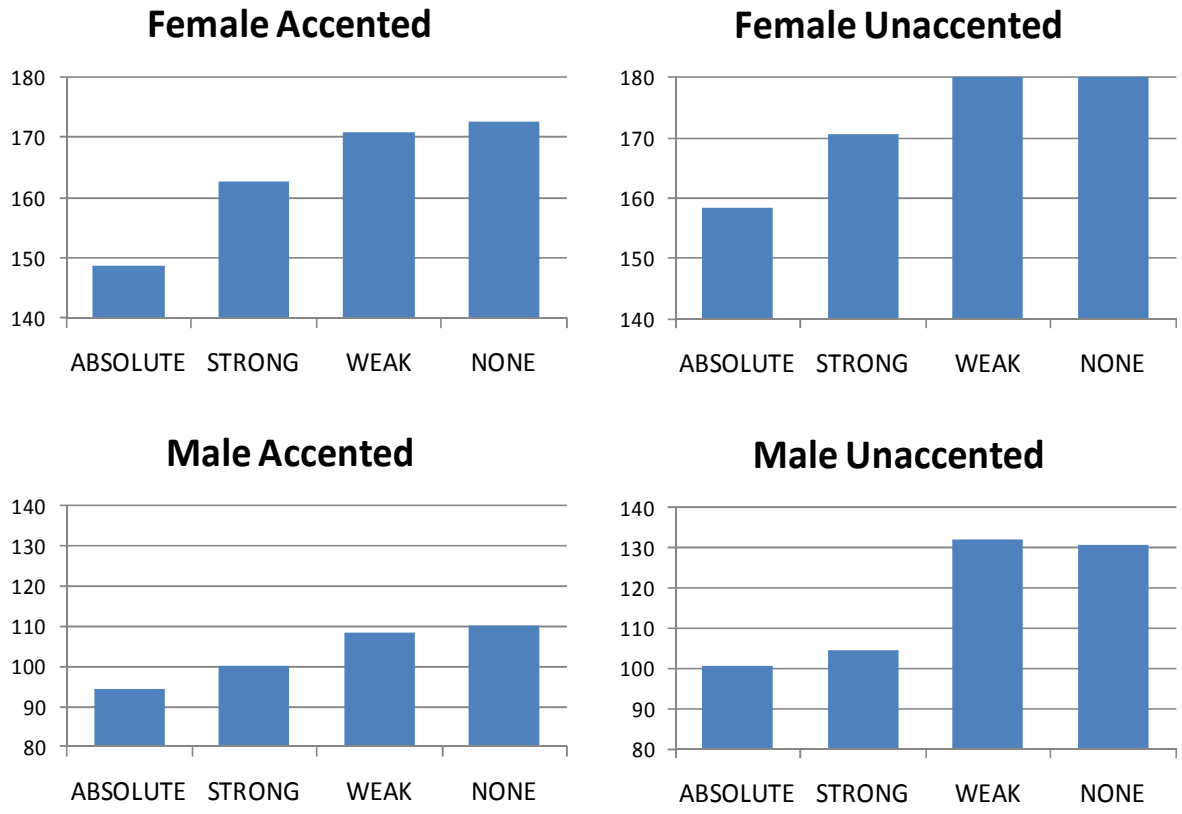


図 8-4: 節単位情報とアクセント核の有無による平均 FLT 値の変動

8.3 分析

8.3.1 節単位情報とレジスター

最初に自発音声に FL が存在するかどうかを検討するために、アクセント句末の FLT の平均値とアクセント句末位置に付与された節単位情報の関係を分析する。図 8-4 は話者の性別とアクセント句中のアクセントの有無（有核：Accented、無核：Unaccented）による分類の結果である。自発音声における FLT を検討するために、再朗読データは除外して学会講演と模擬講演のみを分析した。

縦軸が平均 FLT 値（単位 Hz）、横軸が節単位情報のクラスである。節単位情報が NONE に分類されているのはアクセント句末に節単位ラベルが付与されていない場合である（名詞句などの場合であり、データ数は NONE が最も多くなる）。

男声、女声を問わず、また核の有無を問わず、平均 FLT 値は ABSOLUTE（絶対文末）において最低値をとることがわかる。また全体を通して、ABSOLUTE < STRONG < WEAK < NONE という序列を見てとることができる。

図 8-5: レジスターと節単位情報による平均 FLT の変動

次にレジスターの影響を検討する。今度は、再朗読データを対象に加え、有核と無核を区別せずに分析する。図 8-5 に結果を示す。縦軸は平均 FLT 値（単位 Hz）で、横軸はレジスター（A:学会講演、S:模擬講演、R:再朗読）と節単位情報の組み合わせである。

ここでもやはり全体を通して、ABSOLUTE < STRONG < WEAK < NONE という序列を見てとることができる。唯一の例外は、男声の朗読音声において ABSOLUTE が STRONG よりも高い値をとっていることであるが、その差は 2.5Hz（ABSOLUTE が 81.0Hz、STRONG が 78.5Hz）に過ぎない。

8.3.2 アクセント句の位置

図 8-4, 8-5 に見られた節単位情報と FLT との間の相関には FL だけでなく、当該アクセント句が発話中で占める位置も関与している可能性がある。絶対文末は発話末尾に位置することが多いはずであるから、発話冒頭から絶対文末にいたるまでの時間は他の節単位情報クラスよりも長くなると予想される。その場合、絶対文末の FLT は declination（図 8-1 参照）の効果によって一層低下すると考えられるからである。

図 8-6 に発話冒頭からの距離（DIST、単位はアクセント句数）毎に集計した節単位情報クラスと平均 FLT 値の関係を示す。集計においてはアクセント核の有無にも配慮した。4 から 7 アクセント句までのいずれの距離においても、またこの図には示していない他の距離においても、FLT 値は、ABSOLUTE < STRONG < WEAK < NONE の序列を示している。

次に回帰分析を行なう。FLT 値を従属変数とし、節単位情報のクラス、発話冒頭からの距離を独立変数として用い、男女別に R 言語(Ver. 2.10.1)の lm()関数を利用して分析を実施した結果を表 8-2, 8-3 に示す。この分析では、N 個の値をとる質的変数は N-1 個のダミー2 値変数に展開されている。actdU は、アクセント核が無核か有核かを示す質的変数のうち無核の状態を表すダミー変数、dist は発話冒頭からの距離をあらわす連続変数（整数値）、cblcNONE から cblcWEAK までは節単位情報のクラスをあらわすダミー変数である。

女声データを例として偏回帰係数（Estimate）の値を検討する。actdU は actdA（変数化されておらず基準点となる）に比べて、FLT を 22Hz ほど上昇させる。また dist は距離が 1 アクセント句増えるごとに FLT を 0.7Hz ほど下降させる。節単位情報クラスについては、ABSOLUTE（基準点）に比べて、NONE は FLT を 16.4Hz 上昇させる。同じく WEAK は 14.7Hz、STRONG は 3.8Hz、FLT を上昇させる。

ただし独立変数の測定単位は同一でないので上の比較は dist に対して不利になっている。ダミー変数が 0 か 1 かの値しかとらないのに対して、連続変数である dist の値は 20 以上にまで分布しているからである。この問題を解決するために、標準化偏回帰係数(SPRC)を計算して表中に示した。SPRC の絶対値で dist と節単位情報関係の変数を比較すると前者の貢献は後者の変数よりも小さいことがわかる。男声データの場合も同一の結論を得る。

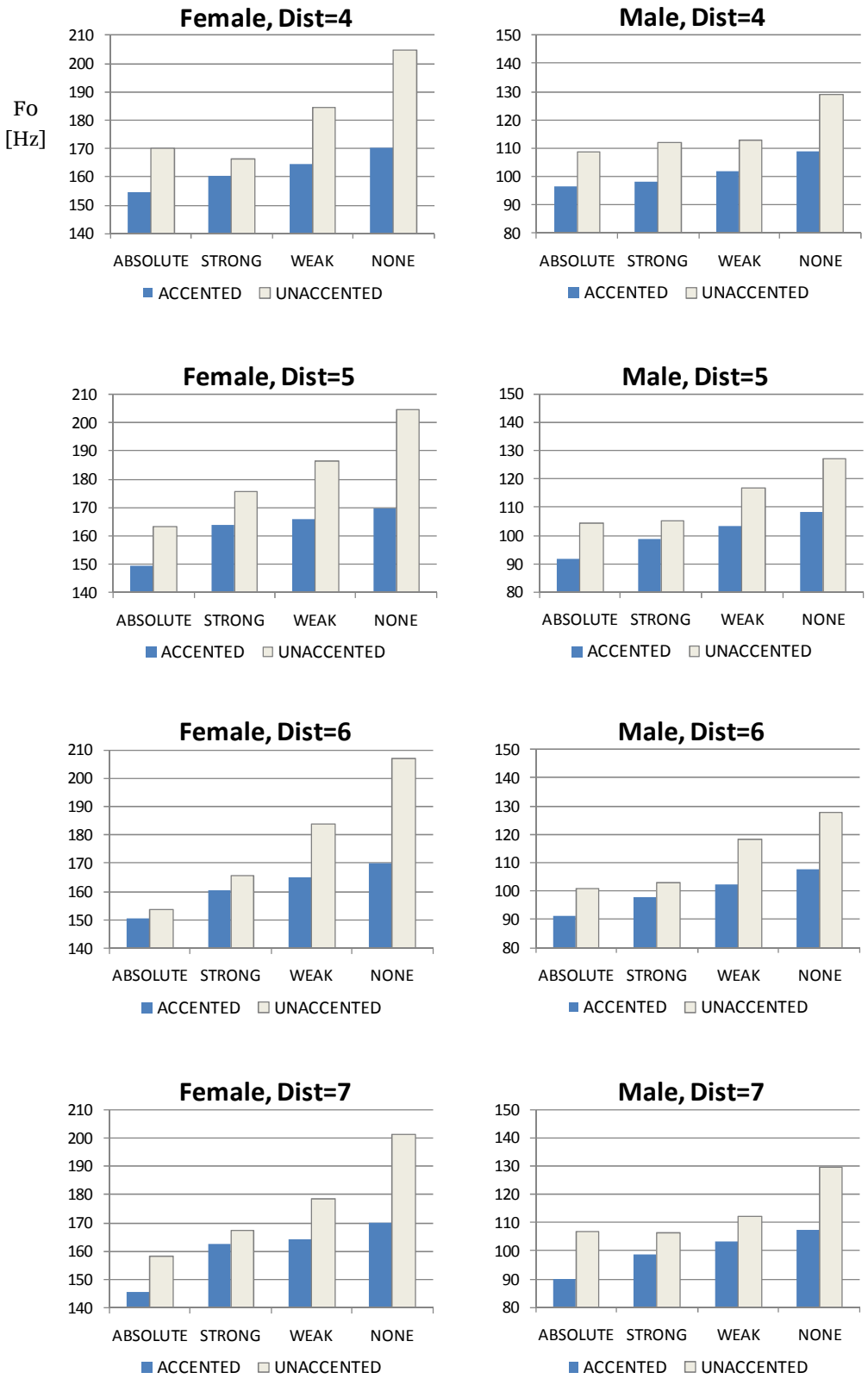


図 8-6: 発話冒頭からの距離による FLT と節単位情報の関係の再分析

表 8-2: FLT の回帰分析（女声データ）

	Estimate	Std.	Error	t value	SPRC
(Intercept)	97.89757	0.98714	99.173	2.00E-16	
actdU	22.09148	0.27284	80.968	2.00E-16	0.254
dist	-0.7027	0.02281	-30.813	2.00E-16	-0.105
cblcNONE	16.44416	0.98126	16.758	2.00E-16	0.174
cblcSTRONG	3.8375	1.18103	3.249	0.00116	0.018
cblcWEAK	14.6585	1.02013	14.369	2.00E-16	0.138

表 8-3: FLT の回帰分析（男声データ）

	Estimate	Std.	Error	t value	SPRC
(Intercept)	97.89757	0.98714	99.173	2.00E-16	
actdU	22.09148	0.27284	80.968	2.00E-16	0.317
dist	-0.7027	0.02281	-30.813	2.00E-16	-0.130
cblcNONE	16.44416	0.98126	16.758	2.00E-16	0.210
cblcSTRONG	3.8375	1.18103	3.249	0.00116	0.022
cblcWEAK	14.6585	1.02013	14.369	2.00E-16	0.168

8.3.3 Final Lowering の領域

前節の分析によって自発音声にも FL が存在すること、またその効果は発話冒頭からの距離の効果を割り引いても存在することが確認できた。本節ではもうひとつの問題、すなわち FL の効果がおよぶ範囲を検討する。この問題を検討するために、有核アクセント句のみから構成される発話をとりあげて、そこに含まれる F0 測定値の変動を分析する。

ところで CSJ には発話単位のラベルは存在しない。ここで発話と呼んでいるのは以下の条件を満たすアクセント句の連鎖である。

1) 節単位情報が絶対境界もしくは強境界であるアクセント句の直後に始まり、
2) 節単位情報が絶対境界もしくは強境界であるアクセント句に終わり、
3) アクセント句連鎖の内部には絶対境界も強境界も含んでいない

このような有核アクセント句の連鎖においては、catathesis の効果によってピッチレンジの上限が一定の割合で下降してゆくことが予想される。その下降傾向からのずれを評価することによって FL のおよぶ範囲を決定することを試みる

表 8-4 は、アクセント句数 2 個から 5 個までの範囲で、有核句のみからなら発話の数を男女別に示したものである。データ数は発話長の増加につれて急速に減少することがわかる。最低でも 10 個以上のデータが確保できる範囲として 5 アクセント句の発話までを対象とすることにした。

表 8-4: 有核アクセント句のみからなる発話の数

発話長 [AP]	女声	男声
2	101	109
3	53	75
4	38	39
5	26	14

この表に示された発話にふくまれるすべてのアクセント句について、ILT、ACC、FLT の F0 平均値を計算すると図 8-7 の結果を得る。図中、ILT1、ILT2 などの添字は発話の先頭から数えて何番目のアクセント句にその測定点が位置しているかを示しており、F0 を表す縦軸の単位は Hz である。

図 8-7 の女声データのうち、第 2 アクセント句に属する測定点（ILT2 、ACC2、FLT2）に注目すると、5AP（アクセント句 5 個からなる発話）、4AP、3AP に属する測定点の値はほぼ等しいのに対して、2AP に属する測定点だけが著しく低下していることがわかる。同様に、第 3 アクセント句においては、5AP と 4AP に属する測定点がほぼ等しく、3AP の測定点だけが低下しており、第 4 アクセント句においては、4AP の測定点が 5AP よりも低下している。男声データについても全く同一の傾向を観察することができる。

この傾向は、FL の効果が発話の最終アクセント句の全体には及ぶが、それ以前には及ばないことによって生じていると考えることができる。ここで、無核アクセント句の連鎖に対して同様の分析を施してみることが考えられるが、無核句のみからなる発話は非常に少なく、2 個のアクセント句からなる発話においても十分な数のデータを確保することができないことが判明したので、分析は断念した。

図 8-7 にはもうひとつ指摘すべき事実がある。それは発話の長さによらず、最終アクセント句のピッチレンジがほぼ同一範囲に納まっていることである。最終アクセント句の FLT は、女声データにあっては 160Hz 前後、男声データにあっては 90Hz 前後に分布している。

これは発話の F0 制御において、発話長を考慮に入れた制御(pre planning)が行なわれていることを示唆する事実である。発話のイントネーションの計算にあたって、発話冒頭と末端における F0 値を固定して declination を求めるモデルは従来から提案されてきており（Maeda 1976; ‘t Hart, Collier and Cohen 1990）、日本語については NTT 研究所の TTS システム（佐藤 1987）が知られている。しかしこのシステムにおいて declination を表現する「話調」成分には FL に該当する F0 下降メカニズムが含まれておらず、発話の冒頭から末尾にかけて F0 はほぼ直線的に低下するものと想定されている。そのため、図 8-7 に観察されたパタンを予測するものとはなっていない。

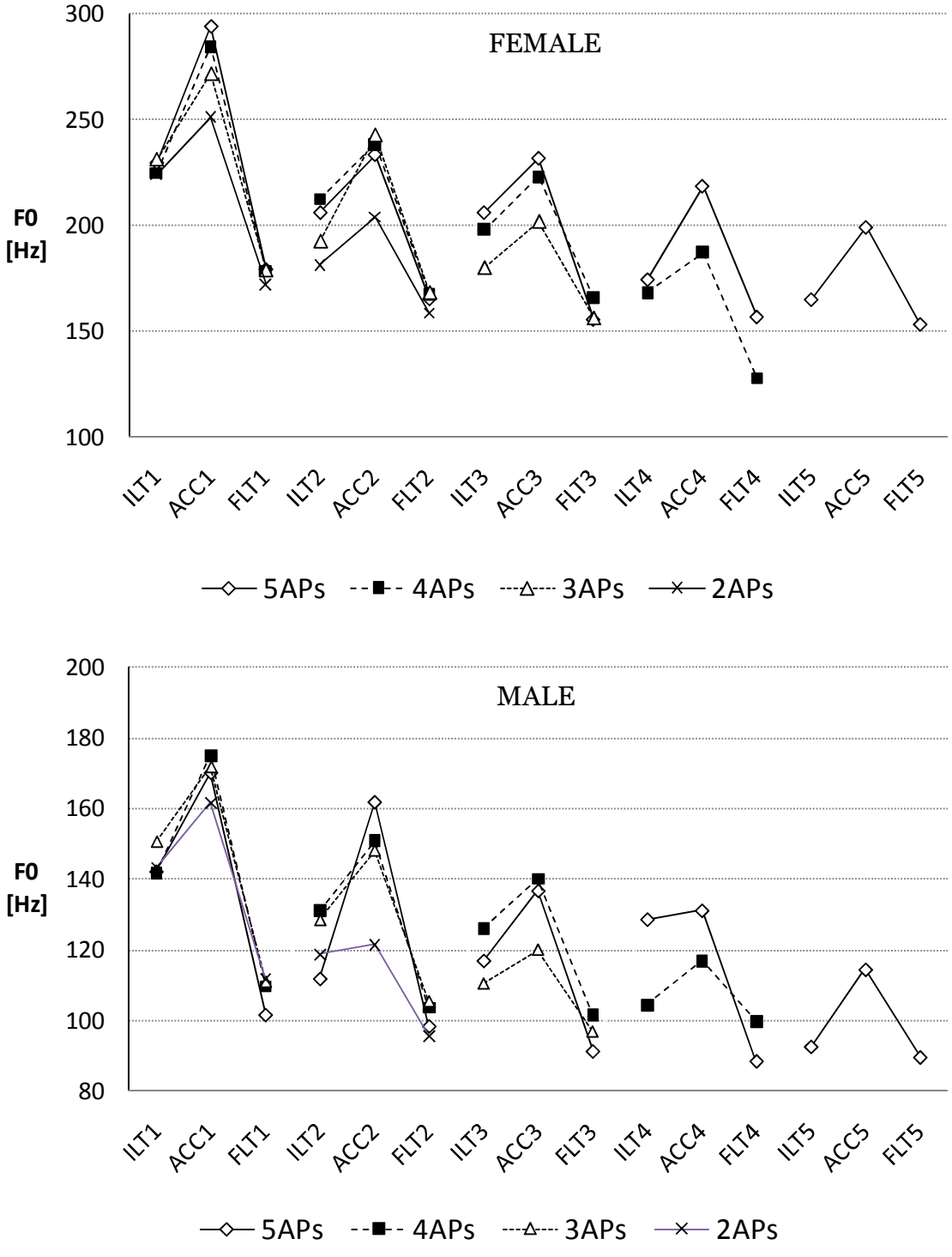


図 8-7: 有核アクセント句連鎖中の F0 測定点の平均値

8.4 結論

本章ではまず FLT と節単位情報の相関を分析することによって、自発音声にも FL が存在することを明らかにした。図 8-5 を見るかぎり、自発音声における節単位情報と FLT の相関は、再朗読音声よりもむしろ自発音声において一層明瞭である。また FL の効果が発話の長さとは独立して存在していることも回帰分析によって明らかになった。

次に、有核アクセント句連鎖を分析することによって、FL は発話の最終モーラだけに働くものではないことを明らかにした。図 8-7 の結果から見るかぎり、FL は発話の最終モーラにのみ働くとした Poser の主張には明らかに問題があり、P&B の主張の方に合理性が認められる。P&B は FL の及ぶ範囲を明言していなかったが、本章の分析によって、少なくとも有核アクセント句連鎖については、その範囲は発話の最終アクセント句に限られることが明らかになった。

第 9 章. Boundary Pitch Movements

9.1 問題

第 7 章と第 8 章でとりあげたのは、従来から朗読音声を対象に研究が行われてきた韻律現象であった。本章と次章では朗読音声の研究では従来ほとんど研究されてきていない現象の分析を試みる。本章ではアクセント句末に生じる局所的なイントネーションをとりあげる。

日本語ではアクセント句末に多用なピッチの変動が生じて、発話に様々なニュアンスを添えることが知られている。伝統的な国語学では、この局所的特徴のことをイントネーションと呼んできたが、近年の音声研究ではイントネーションという術語を発話全体の F0 形状の意味で用いることが多く、本研究もこれに従っているので、アクセント句末のイントネーションには別の名称を用意する必要がある。以下では Venditti, Maeda and Santen (1998)などに従って BPM (Boundary Pitch Movements)という名称を用いることにする。

日本語の BPM が語彙的な意味の対立に関係しないことについては研究者の意見は一致している。それでは BPM はどのような非語彙的情報を伝達しているのだろうか。

BPM についての先行研究の大部分は、「よ」「ね」のような終助詞に各種 BPM が加わった場合の語用論的意味を論じている（小山 1997, 片桐 1997, 轟 2008 等）。この種の研究の対象は当然、対話音声である。BPM が対話音声において、終助詞と結合して重要な情報を伝達していることは事実であるが、それは BPM の機能の一部にすぎない。CSJ のモノローグにも BPM は頻発しており、その大部分は終助詞以外の語に生じている。

この種の BPM が伝達する情報についての実証的な考察はこれまでほとんど行われてきていないと思われる。以下、本章では CSJ-Core1 に記録された BPM の機能をふたつの観点から考察する。発話の主観的印象の観点と談話の境界表示機能の観点である。

9.2 日本語の BPM

9.2.1 日本語 BPM の概観

日本語（東京方言）の主要な BPM の概形と X-JToBI におけるラベル表現を図 9-1 に示す。BPM の生起する言語単位はアクセント句だと考えられるので(Venditti, Maekawa and Beckman 2008, Maekawa 2010b)、ここでは有核アクセント句の末尾に生じる各種 BPM を模式的に示している。

図最上部の N は BPM が生じていない有核アクセント句であり、その句末は L%トーンで終了している。A は上昇 BPM である。アクセント句末に H% トーンが追加されることによって上昇 BPM（L%H%）が形成される。

B は上昇下降 BPM（L%HL%）である。この BPM のラベル表現はやや複雑であり、アクセント句の終端に HL%ラベルが付与されて、BPM のタイプを示すと同時に下降の終端位置を示す一方、上昇の頂点を示すために pH ラベルが利用されている。

C は上昇 BPM の変種である L%LH%である。これは川上(1963)が「反問の上昇」と呼んでいる BPM であり、A との相違点は、上昇に先だって低いピッチが一定時間延長される点にある。ラベルは上昇の頂点に LH%が付与され、低ピッチ延長区間の終端に pL が付与される。

D は上昇下降上昇 BPM である。ラベルは全体の末尾に HLH%が付与され、最初の上昇の頂点と下降の終端に pH と pL が付与される。後述するようにこの BPM は CSJ-Core にはほとんど出現しないが、女性同士のくだけた対話音声などには頻出している。

E と F は extender と呼ばれる記号 > の用法を示している。Extender は発話の冒頭ないし末端で%L ないし%H トーン（及びそのトーンが結合しているモーラ）が引き延ばされて発音されている場合に用いられる。E は上昇 BPM(L%H%)末の H%が延長された場合である。上昇の頂点に H%が付与され、延長区間の末端に>が付与される。一方、F はアクセント句末の L%に延長が生じた場合である。

最後に G はアクセント句冒頭の%L に延長が生じた場合である。この場合、低ピッチ区間の始端に%L が付与され、終端に%L が付与される。G は BPM にアクセント句頭の音調であり、狭義の BPM ではないが、CSJ には実際に出現する音調である。また前川・北川(2002)が示しているように、「疑い」や「感心」などのパラ言語情報の表出に利用される音声特徴のひとつである。

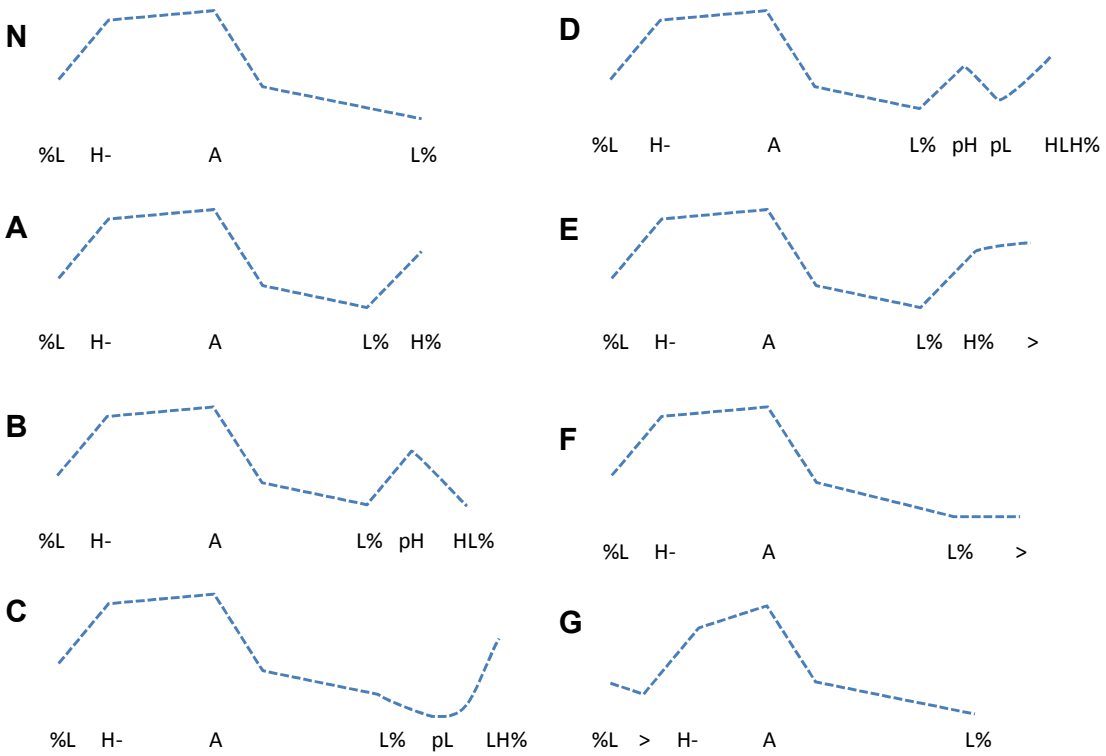


図 9-1: 日本語（東京方言）の BPM

表 9-1: CSJ-Core における主要な BPM の生起頻度*

レジスター	L%	H%	HL%	HLH%	LH%	Total	PNLP
学会講演	53,213	18,691	2,953	1	115	74,973	617
模擬講演	58,766	9,939	6,677	7	220	75,609	376
対話	6,746	1,721	564	4	34	9,069	16
再朗読	4,772	1,536	31	0	1	6,340	19
Total	123,497	31,887	10,225	12	370	165,991	1,028

*HL%の欄は PNLP を含めた数字

図 9-1 の BPM はすべて CSJ-Core に生じているが、その生起頻度には著しい偏りがある。表 9-1 は CSJ-Core のアクセント句全体を BPM の観点から分類した結果である。この表を含めて以下では BPM を構成するトーンのうち L%を除外して表記することにする。H%, HL%はそれぞれ L%H%, L%HL%の意味であり、L%は BPM なしのアクセント句末を表す。Extender は下位区分の基準であるから表 9-1 の集計規準としては用いていない。

全アクセント句のうち約 25%に BPM が生じているが、そのうち 75%が H%、24%が HL%であり、HLH%と LH% はあわせて 1%に達さない。以下本章では CSJ における BPM の 99%を占める H%と HL%に対象を限って検討を進める。

9.2.2 上昇及び上昇下降 BPM の変異形

本節では H%及び HL%の変異形を検討して分析の対象を明確化する。日本語の句末上昇音調について画期的な分類を施した川上(1963)は、上昇調に 5 種の変異形を認めている。「普通の上昇」「反問の上昇」「浮き上がり調」「吊り上げ調」「強調の上昇」である。

このうち「反問の上昇」には X-JToBI の LH%が該当することは先述したとおりである。残る 4 種の上昇調にはすべて L%H%のトーンラベルが付与されるのでトーンラベルだけでは区別できない。しかし「浮き上がり調」と「吊り上げ調」には X-JToBI のプロミネンス層において FR (floating rise)と HR(hooked rise)ラベルがそれぞれ付与されているので、これによって識別が可能である。残る「強調の上昇」は X-JToBI のどの層でも特別のラベルを付与していないので、識別不可能である。

以下本章で行なう上昇 BPM(H%)の分析対象とするのは、川上の「普通の上昇」である。表 9-1 の H%には FR ラベルが 2310 回、HR ラベルが 202 回生じていたが、これらは分析対象から除外する。菊池・前川(2007)は、浮き上がり調の H%ないし HL% は「～けれども」「～ですが」「～ですけど」のような逆節表現の節末に生じやすいことを報告している。分析対象から浮き上がり調の H%を除外するのは、このような意味的な偏りを排除しておくためである。

川上は HL%の変異形については何も述べていない。しかし HL%にはすくなくともひとつの変異系が存在する。本研究ではこの変異形を PNLP(penultimate non-lexical prominence)と呼ぶ（Maekawa et al. 2002, Venditti, Maekawa and Beckman 2008）。通常の HL%と PNLP の主要

な、そしておそらくは唯一の差異は、BPM を構成する 3 個のトーン(HLH)の時間的な配置である（この問題については 9.6 節参照）。

図 9-2 は通常の HL%と PNLP を有核アクセント句「キーワードを」の句末において模式的に比較したものである。図中の音素記号/H/は長母音の後半要素である。図中の記号 FLT, FHT については前章の図 8-3 とその解説参照。

通常の HL%においては、HL%による F0 の上昇と下がともに句の最終モーラである格助詞「を」の内部に納まっており、結果として F0 のピークもまた格助詞の内部に納まっている。一方、PNLP では F0 上昇の起点は句末から 2 番目(penultimate)のモーラである「ド」か更に左側に位置しており、F0 のピークは penultimate モーラと最終モーラの境界付近か、そこから最終モーラ内部にやや入り込んだところに位置している。ただし聴覚印象としてはピークは penultimate モーラに位置しているように聞こえる。

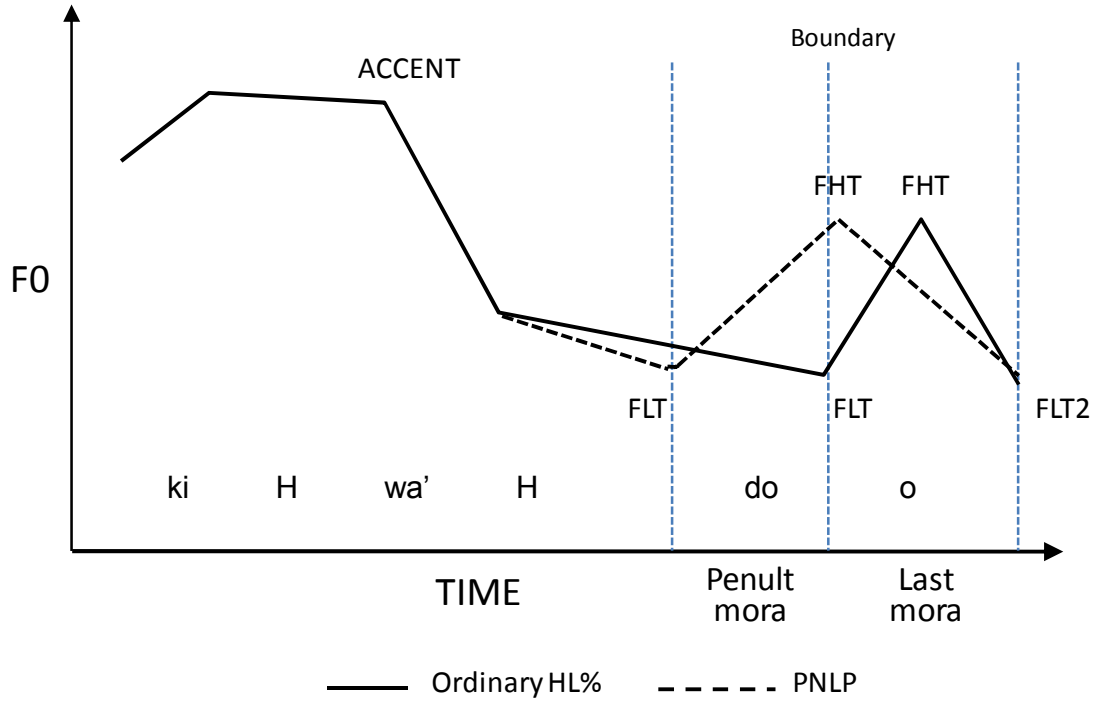


図 9-2: 通常の HL%と PNLP の比較

PNLP の認定には、もうひとつ重要な言語的条件がある。BPM によって引き起こされる F0 の下降はアクセント核として解釈できてはならないという条件である。この条件を課す必要があるのは、PNLP は（その名が示しているように）非語彙的な句レベルの韻律現象として規定しているからである。

日本語のアクセント句には、最後から 2 番目のモーラにアクセント核が位置しているものが少なくない。例えば「弟が」は、尾高型有核名詞である「弟」に無核助詞が後続している

ので、「オトート┐ガ」となる。また前章でとりあげた韻律的に独立した 2 モーラ有核助詞も「キョ┐ートマ┐デ」となって、PNLP とよく似た F0 形状をとるが、これらは PNLP とは認定しない。

真の PNLP は、PNLP による上昇と下降を↑と↓で表すならば、「ギ┐ジュ↑ツ↓デ」（技術で）、「ジッカ↑カ↓ラ」（実家から）のような例である。PNLP は「技術」のような有核語にも「実家」のような無核語にも生じる。

9.3 データ

本節の分析では、CSJ-Core に付与された X-JToBI の BPM 関係トーンラベル、トーンラベルに対応する時刻の F0 データ、CSJ-Core 固有のアノテーションである集合評定印象評定データ、そして節単位情報（CBL）の相互関係を分析する。このうち BPM 関係トーンラベルは 9.2 節及び第 3 章で、また節単位情報は 8.2 節及び第 2 章で説明済みである。ここでは、F0 情報と印象標的データについて説明する。

9.3.1 F0 データ

X-JToBI のラベル情報を保存した XML ファイルには、トーンラベルの時刻に対応する F0 値が記録されている。以下の分析ではこの値を XSLT で抽出して分析に供した。しかしラベルによっては、母音の無声化その他の原因によって F0 値が抽出できていないこともある。そのようなラベルは分析対象から除外している。

9.3.2 印象評定データ

X-JToBI と並ぶ CSJ 独自のアノテーションが印象評定データ(IRS: Impression Rating Score)である。これは CSJ に記録されたモノローグが日本語の母語話者に与える種々の主観的印象を尺度化して表現した情報である。

印象評定データのための尺度は、筆者らが CSJ のために独自に開発したものであり（山住ほか 2005, 2007, 籠宮ほか 2007）、「好悪」「上手さ」「速さ感」「活動性」「スタイル」と命名された 5 種の尺度が用いられる。各尺度はそれぞれ 4 個の評定語対から構成されている。例えば「スタイル」の場合、評定語対は、「礼儀正しい」対「無礼な」、「きまじめな」対「不真面目な」、「丁寧な」対「ぞんざいな」、「上品な」「下品な」の 4 対である。

これら 20 対（4 対×５尺度）に加えて、発話スタイルにかかわる 6 対（「あらたまった」対「くだけた」、「生真面目な」対「奔放な」、「きちんとした」対「くつろいだ」等）も評定に利用した。これらの尺度と評価語対は、大規模な予備実験を経て構築されたものである。

印象評定データの評定実験のために、対象とするすべてのモノローグの冒頭、中央、末尾から、それぞれ約 1 分の連続した音声区間をとりだし、ランダムな順番で評点者に提示した。評定者には、個々の音声サンプルからうけた主観的印象を 26 個の評価後対のそれぞれにつ

いて 7 段階評価で評定することを要求した。実験は、評定者が GUI を操作して各自のペースで評定を進めることができる環境で実施された。

このようにして収集された印象評定データデータを以下では BPM の分析に利用する。

9.4 発話スタイルに関する分析

本節と次節では、H%、HL%、及び PNLP の相違点をさまざまな観点から分析する。これまでと同様、特にまぎれる可能性がないかぎり、上昇 BPM を H%、上昇下降 BPM を HL%と呼ぶ。

9.4.1 レジスターによる BPM の分布の相違

表 9-1 を見ると、BPM の分布はレジスターによって変動していることがわかる。相対的にみて、H%の頻度は学会講演と再朗読音声において高く、模擬講演と対話音声において低い。これと反対に、HL%の頻度は模擬講演と対話音声において高く、学会講演と再朗読音声において低い。この事実は、BPM の選択に発話スタイル（ここでは発話のあらたまりの程度と解釈する。9.7.1 節の議論も参照。）が関与していることを示唆している。CSJ の語彙や音声のレジスターによる変異を調査した先行研究は、学会講演と再朗読音声は模擬講演と対話よりも発話のあらたまり度が高いことを報告している（Maekawa et al., 2003, Maekawa 2009）

9.4.2 印象評定データとの関係における H%と HL%の相違

BPM と発話スタイルの関係を客観的に検討するために、印象評定データと H%及び HL%の生起頻度の関係を分析した。この目的のために、評定者 10 名の評定データの平均値を計算した。1 個の講演の始端、中央、末端の 3 箇所についてそれぞれ 10 個の評定データがあるから、欠損値がないかぎり 1 個のファイルには評定語対ごとに 30 個の評定データがある。これらを平均した値を以下では利用する。

図 9-3 は、H%と HL%を対象として、4 個の印象評定項目（表 9-2 の A17, 18, 21, 25）について印象評定データの平均値と BPM の生起率の関係を示したものである。図は左右にふたつの独立した縦軸をもっており、左軸は H% 、右軸は HL% の生起率（全 AP 数に対する比率）を百分率で示している。また図の横軸は平均値の順位を表している。最低値である「１」は当該サンプルの平均値が、全印象評定データの分布の下位 20%に属していることを、「5」は平均値が上位 20%に属していることを示している。

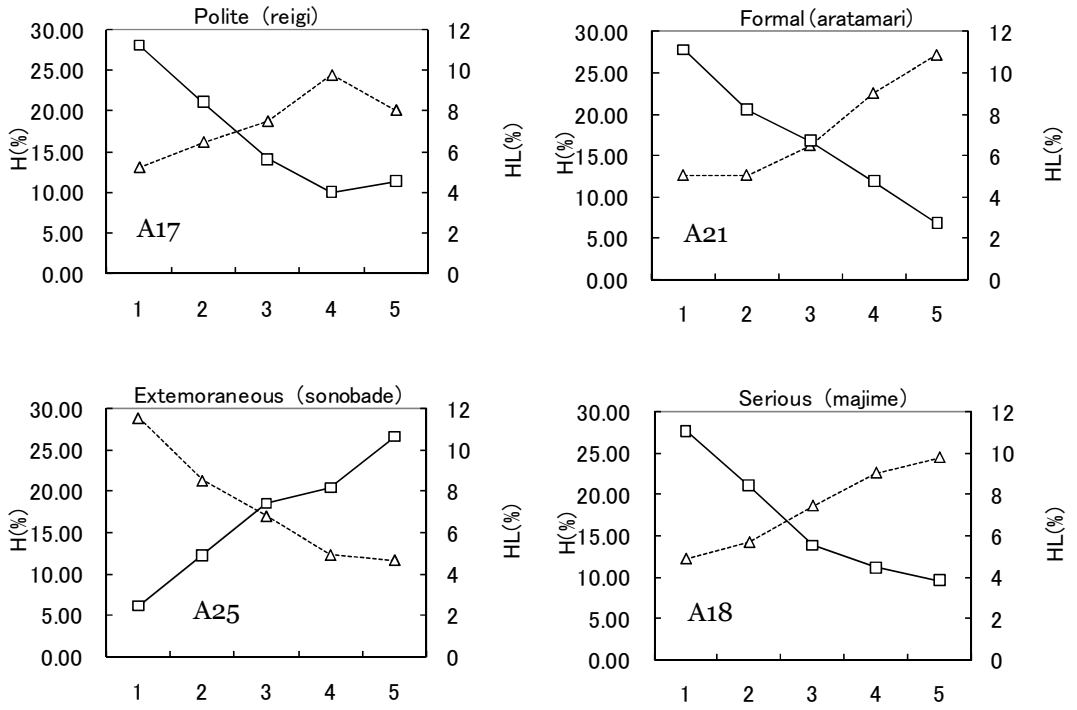


図 9-3: 印象評定データと H%及び HL%生起率。三角が H%(左軸)、四角が HL%(右軸)。

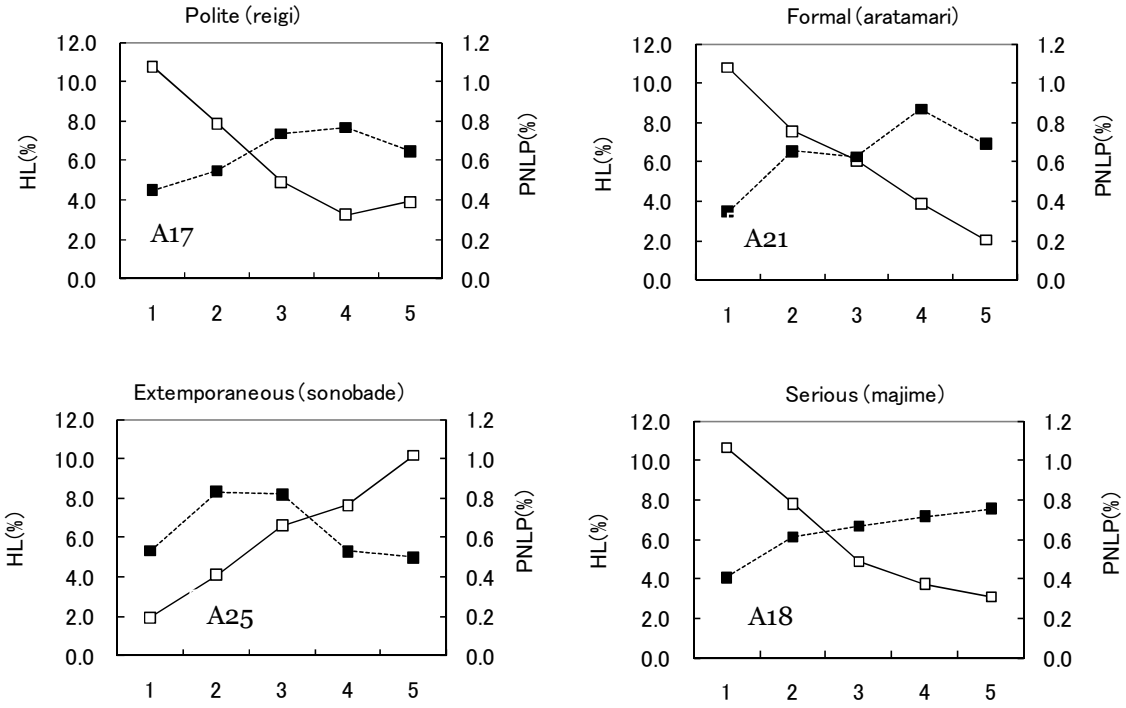


図 9-4: 印象評定データと HL%及び PNLP 生起率。白が HL% (左軸)、黒が PNLP(右軸)

表 9-2: 印象評定データ全評価語対の H% ないし HL% 生起率との相関*

尺度	項目ID	評価語対	IRS~H%	IRS~HL%	H%~HL%
好悪	A01	好きな～嫌いな	**-0.72**	**0.79**	-0.63
	A02	心地よい～不快な	-0.76	0.34	-0.78
	A03	感じの良い～悪い	-0.68	0.53	-0.91
	A04	親しみやすい～親しみにくい	**-0.97**	**0.95**	-0.91
上手さ	A05	流暢な～たどたどしい	0.04	-0.69	0.29
	A06	話し慣れた～話し慣れていない	-0.47	0.37	-0.60
	A07	なめらかな～しどろもどろな	-0.07	-0.71	0.53
	A08	上手い～下手な	-0.80	0.29	-0.64
速さ感	A09	速い～遅い	**0.77**	**-0.85**	-0.93
	A10	スピード感のある～ゆったりした	**0.92**	**-0.85**	-0.96
	A11	せわしげな～のんきな	**0.89**	**-0.95**	-0.75
	A12	落ち着きのない～落ち着きのある	**0.92**	**-0.80**	-0.88
活動性	A13	声の大きい～声の小さい	0.23	0.12	-0.91
	A14	力強い～弱弱しい	-0.12	0.14	0.07
	A15	元気のある～元気のない	-0.31	0.87	-0.58
	A16	積極的な～消極的な	-0.31	0.16	-0.95
スタイル	A17	礼儀正しい～無礼な	**0.83**	**-0.93**	-0.94
	A18	まじめな～ふまじめな	**0.99**	**-0.96**	-0.96
	A19	丁寧な～そんざいな	0.64	-0.88	-0.66
	A20	上品な～下品な	0.61	-0.74	-0.78
話し方	A21	あらたまった～くだけた	**0.96**	**-0.99**	-0.93
	A22	きまじめな～奔放な	**0.97**	**-0.99**	-0.98
	A23	きちんとした～くつろいだ	**0.98**	**-1.00**	-0.98
	A24	甘えた～そっけない	**-0.92**	**0.94**	-0.82
	A25	その場で考えて話している～原稿を読み上げている	**-0.96**	**0.99**	-0.96
	A26	聞き取りやすい～聞き取りにくい	-0.94	0.51	-0.32

*IRS は印象評定項目値（impression rating score）の意

図 9-3 では、印象評定データ平均値に対するふるまいが H%と HL%とで 180 度異なっている。印象評定データが「A17 礼儀正しい」（図左上のパネル）、「A21 あらたまった」（右上）、「A18 まじめな」（同右下）の場合、評定値のランクが上昇するにつれ、H%の生起率が上昇し、反対に HL% の生起率は下降する。「A25 その場で考えて話している」(左下)ではパタンが逆転しているが、H% と HL% が正反対にふるまう点では変わりはない。

このような H%と HL%の対照的なふるまいは、図 9-3 に示した 4 個の項目にだけかぎって観察されるものではない。表 9-2 は、全 26 個の評価語対（3 列目が評価語対）について、図 9-3 と同様の集計を施したうえで、各評価語対と H%生起率及び HL% 生起率、そして H%生起率と HL% 生起率のピアソン積率相関係数を計算した結果を示している。ほとんどすべての評価語対において、H% と HL%の相関係数は符号が逆転していることが確認できる。その結果、H%と HL% の生起率間には強い負の相関が生じている。表 9-2 では評価語対と BPM

生起率の相関係数値が 0.7 を超えるケースを太字で示してあるが、H% と HL%がともに太字になっている評価語対（13 件）においては例外なく両 BPM 間に強い負の相関が生じている。

9.4.3 印象評定データとの関係における HL%と PNLP の相違

続いて HL%とその変異形である PNLP の関係を分析する。図 9-4 は HL%と PNLP を対象として、図 9-3 と同じ 4 個の評価語対について印象評定データと BPM の生起率の関係を示したものである。図の縦軸は左側が HL%、右側が PNLP の生起率であり、横軸は図 9-3 と同じである。なお、本節で HL%と呼んでいるのは、HL%の全データから PNLP のデータを除外したものである（前節で HL%と呼んだデータには PNLP も含まれていた）。非常に興味深いことに、HL%と PNLP もまた印象評定データに関して正反対のふるまいを示している。

表 9-3: 印象評定データ全評価語対の H% ないし PNLP 生起率との相関*

尺度	項目ID	評価語対	IRS~HL%	IRS~PNLP	HL%~PNLP	H%~PNLP
好悪	A01	好きな～嫌いな	0.79	-0.29	-0.46	0.70
	A02	心地よい～不快な	0.35	-0.19	-0.50	0.41
	A03	感じの良い～悪い	0.53	-0.06	-0.26	0.43
	A04	親しみやすい～親しみにくい	0.97	-0.65	-0.54	0.66
上手さ	A05	流暢な～たどたどしい	**-0.74**	**0.87**	-0.82	0.08
	A06	話し慣れた～話し慣れていない	0.04	0.73	-0.59	-0.26
	A07	なめらかな～しどろもどろな	**-0.78**	**0.87**	-0.52	-0.01
	A08	上手い～下手な	0.22	0.89	0.47	-0.77
速さ感	A09	速い～遅い	-0.87	0.60	-0.41	0.57
	A10	スピード感のある～ゆったりした	**-0.87**	**0.73**	-0.53	0.68
	A11	せわしげな～のんきな	**-0.95**	**0.97**	-0.89	0.63
	A12	落ち着きのない～落ち着きのある	-0.79	0.51	-0.76	0.63
活動性	A13	声の大きい～声の小さい	0.04	0.55	0.05	0.12
	A14	力強い～弱弱しい	0.03	0.49	0.35	-0.28
	A15	元気のある～元気のない	0.86	0.40	0.29	-0.74
	A16	積極的な～消極的な	0.08	0.37	-0.11	-0.19
スタイル	A17	礼儀正しい～無礼な	**-0.93**	**0.74**	-0.93	0.91
	A18	まじめな～ふまじめな	**-0.96**	**0.93**	-0.97	0.90
	A19	丁寧な～そんざいな	-0.88	0.59	-0.85	0.68
	A20	上品な～下品な	-0.74	0.37	-0.52	0.45
話し方	A21	あらたまった～くだけた	**-0.99**	**0.76**	-0.82	0.63
	A22	きまじめな～奔放な	-0.99	0.53	-0.52	0.39
	A23	きちんとした～くつろいだ	**-1.00**	**0.94**	-0.95	0.96
	A24	甘えた～そっけない	**0.94**	**-0.91**	-0.99	0.81
	A25	その場で考えて話している～原稿を読み上げている	0.99	-0.35	-0.28	0.17
	A26	聞き取りやすい～聞き取りにくい	0.54	-0.46	-0.31	0.18

*IRS は印象評定項目値（impression rating score）の意

表 9-3 は、すべての評価語対と HL% 及び PNLP との相関係数値を示している。ここでもまた、多くの評価語対において、HL%と PNLP の間には負の相関が認められる。そして、表 9-2 の場合と同様、印象評定データと HL%及び PNLP のいずれもが絶対値で 0.7 以上の相関を示している場合には、HL%と PNLP の間には例外なく負の相関が認められる。表 9-3 には HL%と PNLP、また H% と PNLP の相関係数値も示されているが、これらには 9.7.1 節で言及する。

9.5 談話機能に関する分析

続いて HL%と PNLP の機能的な相違を検討する。韻律特徴の機能分析には様々な可能性が考えられるが、ここでは特定の発話単位において HL%と PNLP の分布（生起頻度と生起位置）を比較することによって、両者の機能差を推測する方法を試みる。

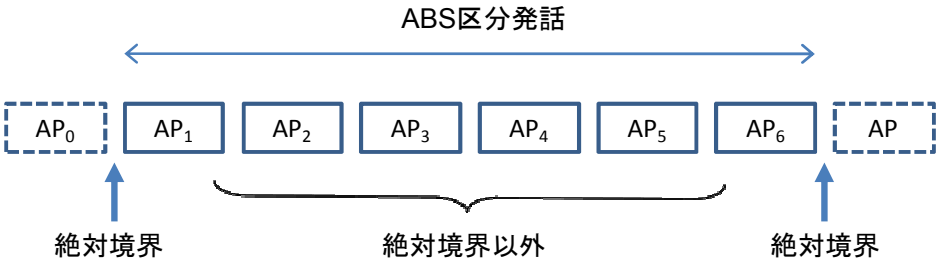


図 9-5: ABS 区分発話の概念図

9.5.1 ABS 区分発話

両者の分布を比較する領域を限定するために、CSJ のアノテーションのひとつである節単位情報を利用する（丸山・高梨・内元 2006）。節単位ラベリングは節(clause)の形態論的・統語的特徴を客観的に処理することによって、自発音声に言語処理のための単位分割を施すことを目的としたアノテーションであり、CSJ コアに含まれるすべての節は、「絶対境界」「強境界」「弱境界」の 3 種に分類されている（8.2.2 節参照）。以下では、このうち文法的に最も深い境界である「絶対境界」(absolute boundary)によって区切られた発話単位を分析の単位として採用し「ABS 区分発話」と呼ぶことにする。

図 9-5 に ABS 区分発話の概念図を示す。図中の AP はアクセント句である。実線で囲われた AP1 から AP6 までの区間は、前後を絶対境界で区切られており、内部に絶対境界を含まないので、1 個の ABS 区分発話を構成する。この図では 6 個のアクセント句が ABS 区分発話を構成しているが、次節に示すように CSJ の ABS 区分発話にはこれよりもはるかに長いものが含まれる。

9.5.2 生起頻度の計算

今回の分析では、AP 数にして 5 個から 15 個までの長さの ABS 区分発話を対象として、PNLP と HL%、そして比較のために H% （上昇調）の生起頻度を調査した。その方法は以下のとおりである。

最初に ABS 区分発話の長さごと、句末イントネーションごとに、発話内に 1 個以上の当該イントネーションを含む発話数を調査する。その際、1 個以上であれば何個ふくまれていてもよい。AP10 個からなる発話の場合、PNLP を 1 個以上含む発話は 47 個、HL%は 140 個、H%は 450 個であった。次に発話長ごとに、各句末イントネーションの総生起数を計算する。AP10 個からなる発話の場合、PNLP は 51 個、HL %は 196 個、H% は 1159 個生じていた。この二つの数字があれば、ABS 区分発話に各イントネーションが生起する場合（頻度 1 以上）の平均生起数を計算できる。本例の場合、PNLP は 1.09 個、HL%は 1.40 個、H%は 2.58 個である。同様の計算を 5AP から 15AP までのすべての AP 区分発話に対して実行した結果を図 9-6 に示す。

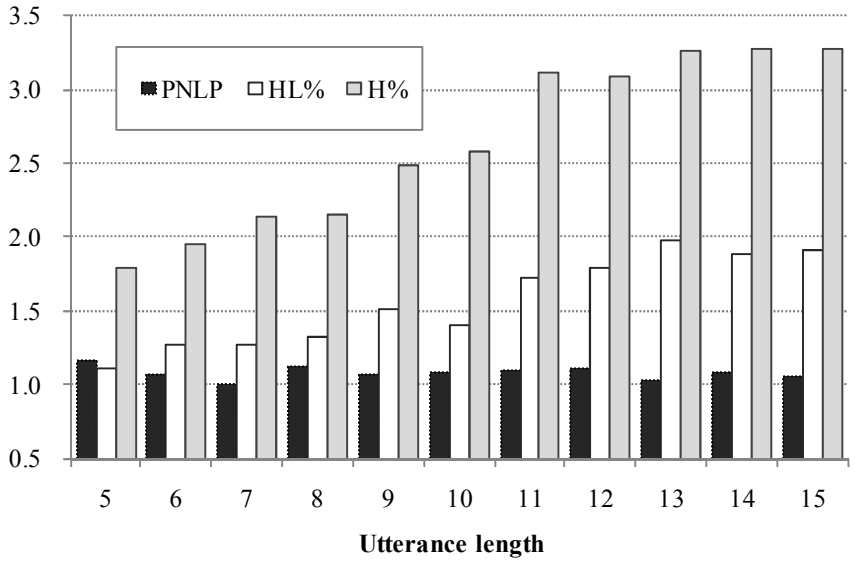


図 9-6: 発話長と句末イントネーションの平均生起数(頻度 1 以上)の関係

発話長が増大するにつれ HL%及び H% の平均生起数はほぼ単調に増大しているのに対し、PNLP の平均生起数は発話長の変動にかかわらず 1.1 前後で一定している。発話長（AP 数）と平均生起率の積率相関係数を計算すると HL% が 0.96、H%が 0.97 であるのに対し、PNLP は-0.25 である。図 9-6 に示された分布の著しい差異は、PNLP とそれ以外の句末イントネーションとの間には機能面で大きな差異が存在していることを示唆している。

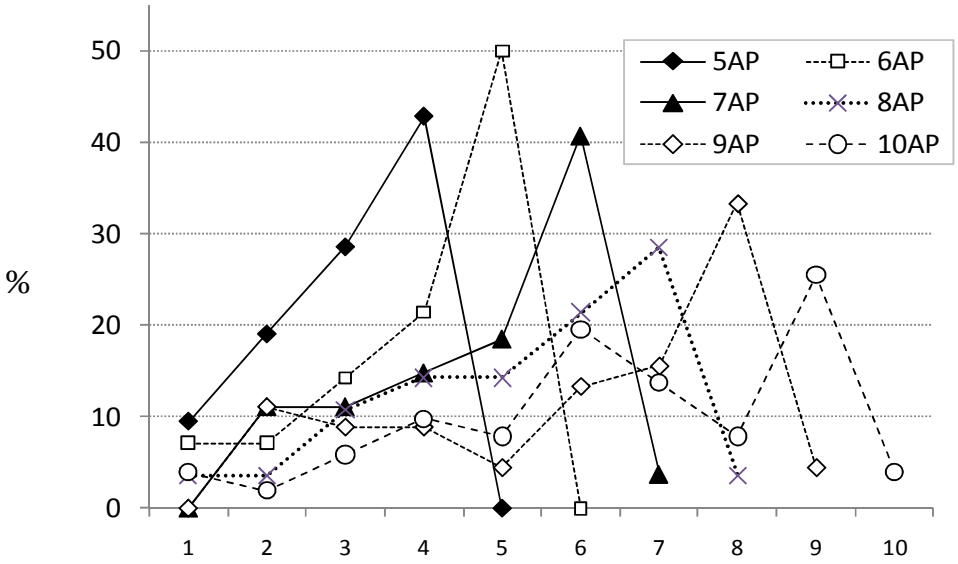


図 9-7: ABS 区分発話における PNLP の位置別生起率
(5AP から 10AP まで)

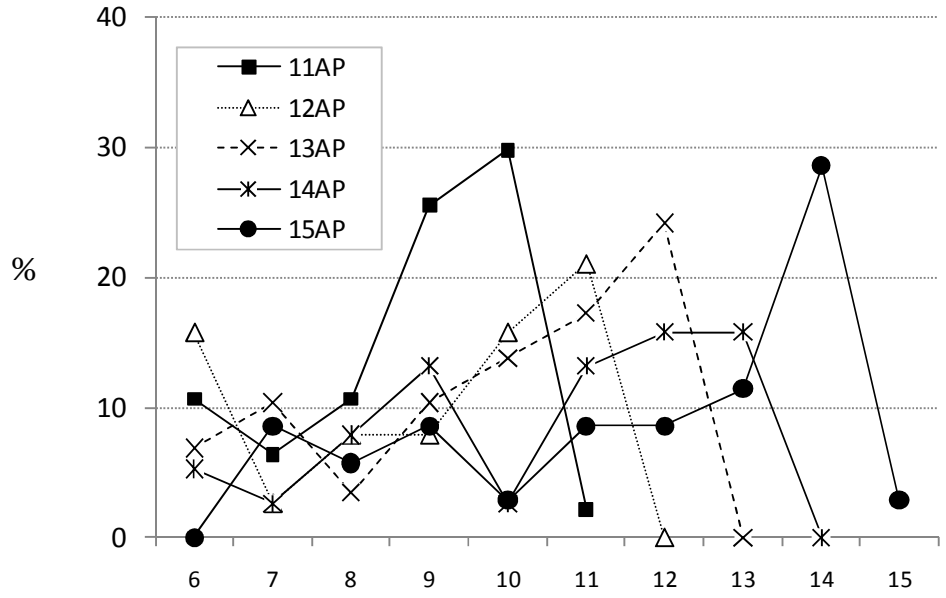


図 9-8: ABS 区分発話における PNLP の位置別生起率
(11AP から 15AP まで)

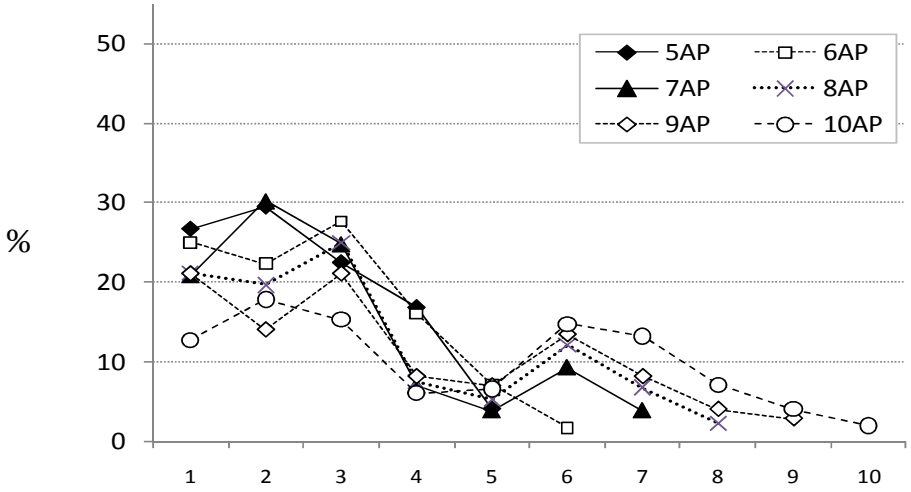


図 9-9: ABS 区分発話における HL%の位置別生起率

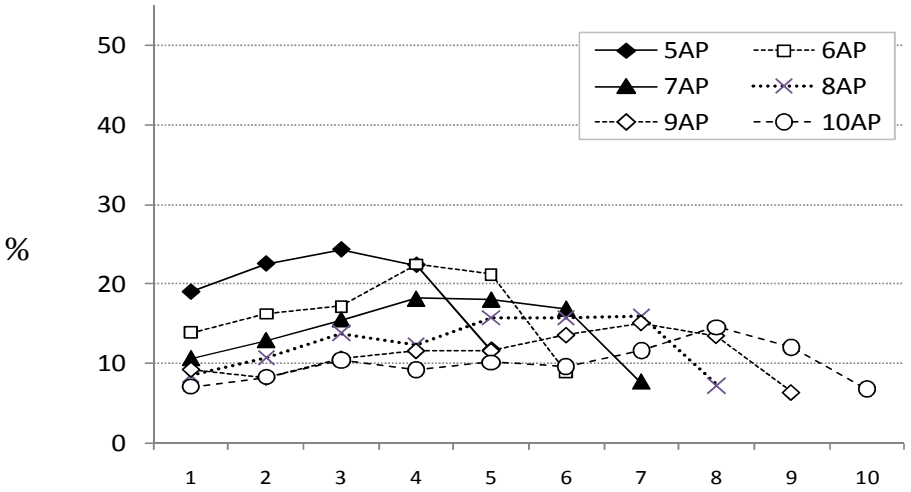


図 9-10: ABS 区分発話における H%の位置別生起率

ここで考えられる仮説は、PNLP の生起率が発話長の長短と無関係に一定しているのは、PNLP には ABS 区分発話（もしくはそれが近似している何らかの言語単位）の頂点ないし境界を表示する機能があり、HL%と H% にはそのような機能がないというものである。

この仮説が正しいとすれば、PNLP の生起位置と ABS 区分発話境界の間には何らかの関係が見出されることが期待される。また、HL%及び H% の生起位置にはそのような関係が認め

られないであろう。この仮説を検証するために、所与の長さの ABS 区分発話に含まれるすべての AP 位置について、PNLP の平均生起率（頻度 1 以上）を計算した。

例えば 5 個の AP から構成される ABS 区分発話で PNLP を 1 個以上含む発話は 18 個あり、生起した PNLP の総数は 21 個である。これらの PNLP は図 2 の記法に従えば、AP1 には 2 回、AP2 には 4 回、AP3 には 6 回、AP4 には 9 回、そして AP5 には 0 回生じていた。これを生起率(百分率)に変換すると 9.5%, 19.0%, 28.6%, 42.9%, 0.0%である。この計算を AP5 個から 15 個までのすべての長さの発話のすべての AP 位置について実行した。

分析結果を図 9-7 と図 9-8 に示す。図 9-7 は AP 長が 5 から 10 までの発話、図 9-8 は AP 長が 11 から 15 までの発話の分析結果である。図 9-8 では見やすさのために AP1 から AP5 までの生起率を省略している。両図とも横軸が AP 位置、縦軸が PNLP の生起率(%)である。すべての発話長において PNLP の生起率は発話冒頭では低く、発話の進行につれて漸増し、次末 AP（N 個の AP から構成される発話における APN-1）において最高値に達し、最終 AP では再度低い値（しばしばゼロ）をとる。図 9-8 において省略されている AP1 から AP5 までの区間でもこの傾向は例外なく成立している。

次に HL% と H% に PNLP と同じ分析を施した結果を示す。図 9-9 が HL% の結果、図 9-10 が H% の分析結果である。これらの図では発話長は 5AP から 10AP までだけを扱っている。これらの句末イントネーションの ABS 区分発話中での位置の分布は、PNLP とは全く異なっている。HL%の分布には AP2 ないし AP3 と AP6 とに二つのピークが存在しており、頂点ないし境界を表示する機能にとっては不都合な分布となっている。一方、H% は発話の全体に万遍なく分布しているが、これも頂点ないし境界表示のためには不都合な分布である。

9.6 音声的形態の分析

CSJ-Core に付与された X-JToBI ラベルを利用して、HL%と PNLP の音声的な特徴を比較する。この比較は、9.2.2 節で模式的かつ主観的に説明した PNLP の音声特徴が実際の自発音声データにおいても成立しているかどうかを確認するために必要である。また、それによって CSJ の韻律ラベリングにおいて通常の上昇下降調と PNLP とが適切に区別されているかどうかも確認できる。

9.6.1 上昇下降タイミングの分析

まず FLT ラベルと FHT ラベルのタイミングを検討する（図 9-2 参照）。FLT ラベルのタイミングは(1)式で計算する。t(FLT)は FLT ラベルの生起時刻、t(Boundary)は最終モーラと次末モーラの境界の時刻である。変数 FLT_loc の符号が負であれば FLT ラベルはモーラ境界に先行して次末モーラ内に位置しており、正であればモーラ境界に遅れて最終モーラ内に位置している。同様に FHT ラベルのタイミングは、t(FHT)を FHT ラベルの生起時刻として(2)式で計算する。

$$FLT\_loc = t(FLT) - t(Boundary) \quad \cdots\cdots\cdots\cdots \quad (1)$$

$$FHT\_loc = t(FHT) - t(Boundary) \quad \cdots\cdots\cdots\cdots \quad (2)$$

計算にあたっては、FLT と FHT についてともに F0 が測定可能であったサンプルだけを利用することとし、母音の無声化によって F0 が存在しなかったり、喉頭化等によって F0 値に信頼性がなかったりするデータは除外した。分析に用いたサンプルは 7232 個であった。

表 9-4 に学会講演と模擬講演に分けて分析結果を示す。FLT_loc の場合、学会講演においても模擬講演においても、HL%と PNLP の間に約 100ms の差が認められ、PNLP においては次末モーラの内部に深く入り込んでいることが分かる。Welch の方法による t 検定（両側検定）も講演のタイプを問わず $p< 0.0001$ で有意である。

FHT_loc についても、HL%と PNLP では約 100ms の差があり、HL%においては最終モーラの内部に位置しているのに対し、PNLP ではほぼ正確にモーラ境界上に位置していることがわかる。t 検定の結果もすべて有意である。

表 9-4: FLT と FHT のタイミング

変数	レジスター	HL%	PNLP	Welch t-test（両側検定）
FLT_loc	学会講演	-20.6	-127.5	t = 28.7, df = 487.6, p < 0.0001
	模擬講演	-18.9	-129.3	t = 30.7, df = 309.1, p < 0.0001
FHT_loc	学会講演	94.3	-4.0	t = 36.5, df = 534.2, p < 0.0001
	模擬講演	111.4	4.2	t = 39.3, df = 317.0, p < 0.0001

以上は CSJ コア全体を 1 個の母集団とみた場合の結果である。これとは別に個々の話者ごとの検定も行う。Maekawa(2010a)が/z/の異音の持続時間について報告しているように、コーパスを全体として検定すると高い有意差が認められても、個々の話者については多くの場合に有意差が認められない音声現象があるからである。このように二重の検定を行った結果、個人レベルでは有意性が成立しにくい現象と個人レベルでも大部分に有意差が認められる現象があれば、後者の方が安定した音声特徴と考えられる。

HL%と PNLP のどちらにも 4 個以上のサンプルのある 50 個の講演（学会講演 23 個と模擬講演 27 個）を Welch の方法で t 検定（片側検定）したところ、FLT については 43 個、FHT については 46 個の講演に $p<0.05$ で有意差が認められた。個人レベルでも比較的に安定した音声特徴であることがわかる。

9.6.2 ピーク F0 の分析

次に上昇下降調のピークをなす FHT の平均 F0 値を分析する。男性、女性ごとに平均値の差を Welch の方法で t 検定（両側検定）した結果を表 9-5 に示す。HL%よりも PNLP の方が有意に平均値が高い。

表 9-5: FHT の平均 F0 値

	話者性別	HL%	PNLP	Welch t-test（両側検定）
FHT_f0	女性	226.6	242.8	t = -3.7, df = 152.4, p = 0.0003
	男声	137.4	164.3	t = -9.7, df = 354.9, p < 0.0001

前節同様、個人レベルでの検定も実施した。HL%と PNLP において各 4 個以上のサンプルをもつ 35 個の講演（学会講演 14 個と模擬講演 21 個）を Welch の方法で t 検定（片側検定）したところ、p<0.05 で有意差が認められた講演は 12 個だけであった。個人レベルでは安定性のない音声特徴であることがわかる。

以上の分析結果から、HL%と PNLP を区別する音声特徴として安定しているものは、FLT と FHT のタイミングであり、FHT の高さは副次的な特徴であると考えられる。結局、図 9-2 は模式図として正確なものであったと言える。

9.7 議論

9.7.1 スタイル依存性

本章における第一の発見は、BPM に強いスタイル差が存在することである。表 9-2, 9-3 を吟味すると、あらたまった発話にはくだけた発話よりも H%が多く、反対にくだけた発話にはあらたまった発話よりも HL% が多いことがわかる。日本語の母語話者は、H% と通常の HL% の頻度を調整することによって、発話の丁寧さを調整していると思われる。

従来の BPM 研究がこの問題にほとんどまったく着目していないのは不思議であるが、それはおそらく、従来の研究が対話音声における終助詞と BPM の交互作用に代表される BPM の意味的側面に注意を奪われすぎていたためであろうと思われる。そのため、従来の研究手法は BPM の語用論的意味を研究者の内省に頼って記述することにとどまっており（9.1 節に挙げた先行文献参照）、発話スタイルの変動をともなう幅広いデータを客観的に観察し、定量的に分析しようとしていない。

本章における第二の発見は、発話スタイルに関して通常の HL% と PNLP とのふるまいが大きく異なっていることである。この発見によって惹起される興味深い問題は、H% と PNLP の関係である。これらの BPM はいずれも HL% と 180 度異なったふるまいを示すことから、これらの BPM のふるまいは類似したものとなっている。実際、表 9-3 の最終列が示しているように、多くの印象評定データ項目において H% と PNLP の間には正の相関が生じており、特にスタイル関係の印象評定データ項目(A17-A26) では例外なく正の相関が生じている。このような事実は、H% と PNLP とが何らかの点で同じ機能をもった BPM であることを示唆していると解釈すべきだろうか。

筆者はこの解釈には賛成できない。9.5 節で論じたように、PNLP の最も大切な機能は境界表示機能であると考えられるからである。PNLP が一見 H% と類似したふるまいを示すのは、

H% と PNLP がともに、しかし別個の理由によって、模擬講演よりも学会講演において用いられやすいからである。

H%が模擬講演よりも学会講演において多く用いられるのは、学会講演があらたまり度の高い発話だからである。一方、PNLP が学会講演において用いられやすいのは、学会での研究発表である学会講演は模擬講演よりも構造化されており、学会講演の話者はその構造の重要な境界を PNLP を含む様々な手段で講演内に示そうとするからである。

このように H% と PNLP の表面的な類似性は、本来独立した要因が学会講演において共存していることによって偶々引き起こされた現象であると考える。

なお、本章で報告した BPM のスタイル依存性は、モノローグの範囲において成立するものであることを指摘しておく必要がある。先行研究が考察の対象としてきたような日常の対話音声における BPM の選択には、モノローグとは別種の原理が働いている可能性がある。

9.7.2 境界表示機能

本章における第三の（そして最後の）重要な発見は PNLP が ABS 区分発話にほぼ 1 回しか生じないという事実である。このことから PNLP には一種の頂点機能(culminative function)が備わっていると考えられる。PNLP のピーク F0 値(FHT_f0)が HL%よりも平均して若干高いという事実もこの解釈を裏付けている(9.6.2 節参照)。

さらに、図 9-7, 9-8 に示されているように、PNLP の平均生起頻度が ABS 区分発話末にむけて単調に上昇することを重視するならば、PNLP に境界機能(delimitative function)を認めることもできる。ただし、これはかなり緩やかな機能である。統計的には PNLP の次の AP が発話末にあたることが多いが、必ずそうなるわけではない。図 9-7, 図 9-8 のデータを使って PNLP と発話末の平均距離を計算すると約 3AP になる。PNLP は発話の終わりそのものを予告するのではなく、終わりが近いことを予告していると言うべきであろう。前川・菊池(2007)が PNLP に「述語の生起を予告するという機能」が備わっているのかもしれないと述べたのは、この境界機能の一端をとらえたものであった。

ところで、音韻論で頂点機能や境界機能を論じるとき、多くは語や分節に相当する小さな句の頂点や境界が問題にされている。しかし PNLP によって頂点や境界が表示されている言語単位は語よりもはるかに大きい。語よりも大きな言語単位の境界が韻律特徴によって表示されること自体はめずらしくない。その一例は英語の nuclear accent である。英語のイントネーション句に複数のピッチアクセントが含まれるとき、最後の、そしてしばしば最もプロミネンスの高いピッチアクセントが nuclear accent である（この説明と用語は Pierrehumbert 1980 に従う）。

しかし nuclear accent と PNLP を比較すると両者には顕著な相違点もある。最大の相違点は nuclear accent が義務的であるのに対して、PNLP は随意的である点である。Nuclear accent をもたない英語の発話はありえないが、PNLP は全 AP の 1%以下にしか生じておらず。適当な仮定のもとに発話単位に換算してもおそらく 5%以下にしか生じていない。

もうひとつの相違点は、nuclear accent 比べても PNLP は一層長い言語単位を表示している点である。PNLP がどのような単位に生じているかの実例を検討しよう。まず学会講演の例をふたつ示す。最初の例は音声知覚について、二番目の例は音声認識についての発表である。縦線はアクセント句境界であり、下線を施したアクセント句に PNLP が生じている。転記テキスト中のタグ(M)はメタ言語的表現（ここでは母音の音価）を、タグ（F）はフィラー(filled pause)を示している。

次に|実験|一では|実験|ゼロよりも|より|単純な刺激音として|
母音のみの|(M ああ)を|用い| (M ああ) | (M あああー)系列の|
長短母音の|同定実験を|行ないました
(学会講演 A01F0122 の 259 秒から 273 秒)

(F えーと) これは|言ってみれば|先程の|図で言うと|(F ま)こちら方向の|
分散が|非常に|小さいので|(F えー) これを|(F えー)ゼロと|強制的に|
置いてしまう|といったことに|相当します
（学会講演 A01M0083 の 392 秒から 403 秒）

次に模擬講演の例をふたつ示す。最初の例は自分の住んでいる町について、二番目は趣味の俳句についてのスピーチである。

そこで|(F えー)大和市の|中に|ある|中央林間|という|ところに|実は|
私は|住んでおる|訳ですが|ちょうど|一年前に|川崎市から|引っ越しました
（模擬講演 S03M1133 の 158 秒から 169 秒）

近頃|感じるんですが|やっぱり|男の方|には|男の方の|視点ていうものが|
あって|やっぱり|男性|女性いて|いいんだな|という|感じは|受けました
（模擬講演 S04F1495 の 163 秒から 175 秒）

いずれの例も単一のトピックについてのまとまった談話である。ABS 区分発話の大部分はこのような談話であり、PNLP が頂点ないし境界を表示している単位は談話上の単位であると考えられる。この単位の言語学的性格を特定することは今後の重要な研究課題である。

今後検討すべきもうひとつの問題は、PNLP の機能は頂点表示ないし境界表示だけかという問題である。大石(1959)や丸山・谷口（2002）は PNLP の機能を専ら強調（プロミネンス）の観点から論じている。本稿に報告した PNLP の特異な分布を強調の観点から説明することは困難であるが、だからといって PNLP に強調の機能がないことにはならない。図 9-6 において PNLP の平均生起率が 1.0 ちょうどではなく 1.1 前後であったことは、1 個の ABS

区分発話に２個（以上）の PNLP が生じる発話があることを意味している。そのような場合、2 個の PNLP がそれぞれ異なる機能を果たしている可能性が否定できないし、1 個の PNLP が境界表示と強調を兼ねて機能することも考えられないわけではない。

9.8 結論

自発音声コーパスとそれに付与された印象評定情報に依拠した定量的分析によって、日本語 BPM の従来知られていなかった特性を報告した。

モノローグにおける BPM の選択には発話スタイルの強い影響が認められ、H%と HL%の生起頻度は発話スタイルに関してほぼ正反対のふるまいを示す。また HL%と PNLP の生起頻度も発話スタイルに関してほぼ反対にふるまう。そして、PNLP には談話レベルの頂点機能と境界機能とが認められる。日本語話者は、数秒から十数秒におよぶ談話単位のまとまりを示す、あるいはその終了が間近であることを予告するために、必要に応じて PNLP を生成するのだと考えられる。

第 10 章. 韻律的レジスター

10.1 問題

第 8 章、第 9 章では日本語イントネーションの構成要素である final lowering と BPM を個別に検討した。そのような分析的研究とは別に、談話全体におけるイントネーションの特徴を総合的に把握しようとする研究もまた必要である。そのことを念頭において、本章では話し言葉のレジスターが発話の韻律的特徴に及ぼす影響を組織的に検討することを試みる。

「レジスター」(registger)は、元来、社会言語学ないしコーパス言語学の術語であり、話題、聴衆、言語行動の目的等によって定まる言語の変種を意味している。レジスターによく似た術語に「発話スタイル」と「テキスト・タイプ」がある。前者は米国流の社会言語学ないし音声科学で、また後者はテキスト言語学や談話分析でよく用いられるが、これらは完全な同義語とはいえない。レジスターと発話スタイルの関係は後に 10.6 節で明確化する。

人々が発話の状況に応じてレジスターを使い分けていることは、専門家にかぎらず広く認知されている。レジスターの選択に影響する要因としては、談話の話題、対話かモノローグか、聴衆の性質と規模、話し手に与えられた時間、発話の自発性などが考えられる。

例えば、CSJ の学会講演 (APS)と模擬講演(SPS)とでは、話題（研究発表と日常的スピーチ）が異なるほか、聴衆の規模と性質（学会講演の聴衆は数が多く、反対意見を表明することがありうる）、発話の自発性などに明白な相違があり、それが両者のレジスターに明確な差異をもたらしていると思われる。

言語学には、発話の状況依存性の研究を中心的な研究課題としている研究領域がある。社会言語学(Labov 1972, Eckert and Rickford 2001 など)、コーパス言語学(Biber 1988, Stubbs 1996, Biber and Conrad 2009 など)、言語学的文体論 (Crystal and Davy 1969)、そして言語教育学(Halliday, McIntosh and Strevens 1964)などの領域である。

こうした領域の研究では、少なくとも部分的には、レジスターの変化によって惹起される音声的ないし言語的な特徴の変化が研究されてきている。さらに近年では音声自動認識研究においても発話スタイルが認識のエラー率に影響を及ぼす重要な要因であることが認識され、その影響についての研究が進められている(Weintraub et al. 1996, Nakamura, Iwano and Furui 2007, Shinozaki, Ostendorf and Atlas 2009)。

ここで指摘しておかねばならないことがある。上記の諸領域で研究されてきたのは、レジスターの影響による音声の分節的特徴の変化であって、韻律的特徴ではなかった。しかしこのことは、韻律特徴の状況依存性が、分節的特徴に比べて小規模であったり、不明瞭であったりすることを意味してはいない。韻律的特徴の状況依存性の研究が不活発であったのは、発話の韻律的特徴の構造に関する我々の理解が、分節的特徴に比べて十分に進歩していなかったこと、またその結果、十分な研究用資源が蓄積されていなかったかことの帰結に過ぎない。

実際、研究史を仔細に検討すると、韻律的特徴のレジスターによる変化を把握しようとした先駆的先行研究を少数ではあるが見つけることができる。Blaauw(1995)は、自発音声と朗読音声の研究を行い、ポーズ長、発話速度、F0 declination、韻律境界の種類、そして声質などの観点から分析を施している。Yaeger-Dror (1996)は、英語の文法的な否定に関連する F0 形状とピーク値の差に対するレジスターの影響を調査しているが、2002 年には英語とフランス語を比較して言語差の存在を指摘している Yaeger-Dror (2002)。

以上は音声学ないし言語学的な領域での研究であったが、Abe(1997)のように音声工学の領域に属する研究も行なわれている。阿部は TTS(text-to-speech)音声合成の制御規則を改良するために、３種類の発話スタイルを演じ分けた朗読音声を用いて、母音のフォルマント周波数とスペクトル傾斜、発話の F0 形状と持続時間長の各特徴の変動を分析し、発話スタイルの特徴づけを試みている。このような acted speech に基づく発話スタイル研究は世界各地で実施されており、Llisterri (1992) や Eskénazi (1993)の文献リストに多数紹介されている。

以上のような一連の実験的研究とは独立に、自発音声コーパスの分析に依拠した一連の研究がある。Crystal and Davy(1969)は、英語の文体を言語学的な観点から包括的に解説した著作であり、韻律とレジスターの関係に関わる研究の嚆矢となった研究である。その髄所に韻律アノテーション付の例文が紹介されている点で当時として画期的な研究であった。

Nevalainen(1992)は、London-Lund Corpus (Svartvik and Quirk 1980)に施された Crystal and Quirk (1964)のラベリング方式による nuclear tone （前章で言及した nuclear accent とほぼ同一の概念の英国流の呼称）の分布を検討して、nuclear tone 全体を上昇クラスと下降クラスに二分すると、レジスターとの相関が明らかになることを報告している。

最後に Hirschberg(2000)は自発音声と朗読音声の韻律的相違点を見出すために、ToBI 方式(Silverman et al. 1992)でラベリングされた複数のコーパスを分析した研究であり、文のタイプ（平叙文、WH 疑問文、Yes-No 疑問文）と境界音調(boundary tone)の分布の関係が朗読音声と自発音声とでは並行的でないことを報告している。

レジスターと韻律の関係を正しく理解するためには、自発音声コーパスの分析が非常に重要である。その理由はふたつある。第一に、我々が現在音声の韻律特徴について保有している知識の大部分は朗読音声の分析によって得られたものであり、韻律的特徴の言語情報（主知的な意味の伝達にかかわる側面）に偏っている可能性が高い。ところで、レジスターの選択は狭義の言語学の問題ではなく、むしろ語用論やパラ言語情報に強く依存して行なわれると考えられるから、この種の情報を実際に豊富に含んだ自発音声の分析が必要である。

第二に、従来の実験的研究では、連続的に変化する音声パラメータ、すなわちピッチ・レンジ、分節音持続時間長、母音フォルマント周波数などが主要な分析対象となっているが、言語学的な観点から真に問題となるのは、物理的なパラメータそのものではなく、その背後にあって、レジスターの影響下に話者が選択する離散的な韻律パラメータである。そのような韻律パラメータの選択は、何らかの形で自発音声の韻律アノテーションに反映されていると期待できる。

以下本章では、CSJ に記録された 4 種のレジスターがどのような韻律的特徴によって特徴付けられているかを検討する。従来、日本語についてはこの種の研究が行なわれておらず、本稿が最初の試みとなる。また本章のコーパス分析は、X-JToBI に含まれるほぼすべてのラベルを検討している点で、特定のラベルの分析しか行っていない従来のコーパス分析（Nevalainen は nuclear tone のみ、Hirschberg は boundary tone のみ）よりも体系的な研究となっている。

10.2 データ

本章の分析対象は CSJ-Core の全体である。CSJ-Core を構成する講演の分布を表 10-1 に示す。講演の大部分は学会講演もしくは模擬講演であり、対話と再朗読はサンプル数が少ない。

表 10-1: CSJ-Core における講演と話者の分布

レジスター	講演数*	話者数*	短単位数	発話総時間数†
学会講演	24 / 46	23 / 45	218 161	14.2
模擬講演	54 / 53	38 / 37	225 572	15.0
対話	9 / 9	3 / 3	41 964	3.0
再朗読	3 / 3	3 / 3	18 977	1.4

*スラッシュの左が女性話者数、右が男声話者数
† 発話間のポーズを除去した実施的な発話時間

10.2.1 X-JToBI データ

本章の議論は X-JToBI ラベルの細部についての知識を前提としている。読者は必要に応じて第 3 章の記述を参照してほしい。X-JToBI のラベルは 6 層（単語層、分節音層、トーン層、BI 層、プロミネンス層、注釈層）に分類されているが、そのうち、本章の分析において重要なのは、トーン層、BI 層、プロミネンス層、注釈層のラベルである。

表 10-2 にこれら 4 層のラベルの簡単な説明を CSJ-Core における生起頻度とともに示しておく。表の第 2 列（「拡張」）に＊印のあるラベルは。オリジナルの J_ToBI には存在せず、X-JToBI への拡張時に追加されたラベルである。説明中に「川上」とあるのは、川上(1963)のことである。

表 10-2: 主要な X-JToBI ラベルとその頻度

層	拡張	ラベル	N	説　明
トーン層		L%	97,556	AP 末境界音調
		H%	24,621	AP 末境界音調(上昇 BPM を構成)†
		HL%	8,863	AP 末境界音調 (上昇下降 BPM を構成)†
	*	HLH%	8	AP 末境界音調(上昇下降上昇 BPM を構成)†
	*	LH%	308	AP 末境界音調(川上の「反問の上昇」を構成) †
	*	L%>	480	延長を伴う L% (ないし %L)
	*	H%>	2,275	延長を伴う H%
BI 層	*	1+p	3,872	短いポーズを伴う AP 内部の韻律境界
		2	42,568	通常の AP 境界
	*	2+p	7,155	短いポーズを伴う AP 境界
	*	2+b	7,098	BPM を伴うがポーズは伴わない AP 境界
	*	2+bp	3,456	BPM とポーズを共に伴う AP 境界
		3	71,383	通常の IP 境界。ピッチレンジリセットを伴う。
	*	W	35	語中（短単位中）の AP 境界
	*	P	263	語中（短単位中）のポーズ
	*	PB	1,033	寄生境界（BPM が 2 個継起）†
プロミネンス層	*	PNLP	856	Penultimate Non-Lexical Prominence（第 9 章参照）
	*	FR	2,535	川上の「浮き上がり調」BPM †
	*	HR	207	川上の「つり上げ調」BPM †
		EUAP	1,667	強調をうけた無核 AP
注釈層	*	QQ	220	疑似疑問

† BPM については 9 章参照

10.2.2 追加データ

X-JToBI ラベルのもたらす情報に加えて、平均発話速度と発話の非流暢性(disfluency)の情報を作成した。以下ではこれらを追加データと呼ぶ。平均発話速度（以下 SR と呼ぶ）は、アクセント句(AP)を単位として、AP に含まれる発話のモーラ数を AP の持続時間（秒）で除して計算した。単位は[mora/sec]である。

発話の非流暢性の情報は、CSJ の転記ファイルから、(F)と（D）のタグ数を抽出した。前者はフィラー(filled pause)のタグ、後者は語断片のタグである。CSJ の転記ファイルにはこの他に（W）タグ（発音の訛り、怠け）もあるが、このタグの対象には語彙的な変異も含まれており、非流暢性の範囲を超えているので、今回のデータには含めなかった。

CSJ-Core の転記ファイルから抽出された（F）は 26994 個、（D）は 4794 個であった。CSJ におけるフィラーの特性については Watanabe(2009)が詳しく検討している。

10.2.3 データの正規化

CSJ-Core に含まれる 201 個の講演は、講演によって講演時間が 429 秒から 1808 秒まで大幅に異なっているので、前節で集計したデータは時間に対して正規化する必要がある。すべての X-JToBI ラベル情報は、集計結果をその講演に含まれる総 AP 数で除することによって AP 当たりの生起数に正規化した。また非流暢性タグの集計結果は、その講演に含まれる総短単位数で除することによって 1 語あたりの生起数に正規化した。いずれの場合も正規化後の値は 0~1 の範囲に分布する。

正規化後のデータは、正規分布への適合性をヒストグラム及び正規累積分布との比較によって検討した。変数（ラベル）のなかには、対数変換した方が正規分布への適合性が高まるものがあったので、以下の分析では対数変換後の値を利用することにした。HLH%, LH%, L%>, H%>, W, P, 2+b, 2+bp, 3, PB, EUAP, FR, HR, PNLP, QQ, (D)が対数変換された変数である。

最後に SR（発話速度）も含めたすべての変数を、平均０、標準偏差 1 となるように Z 変換した。以下の分析ではこの値を用いる。

10.3 データ全体の分析

変換後のデータを線形判別分析(LDA)にかけた。この分析には二つ目的がある。ひとつは韻律特徴の情報だけを利用することによって、CSJ-Core の 4 種のレジスターをどの程度まで自動判別できるかを知ることであり、もうひとつは、LDA の独立変数をとりかえて分析を繰り返すことによって、各種変数（ラベル）の相対的な重要性を評価することである。

10.3.1 X-JToBI ラベルデータを用いた判別分析

最初に X-JToBI ラベルだけを独立変数に用いた判別分析を実施した。R 言語(Version 2.10.1)の MASS ライブラリに含まれている lda()関数をデフォルトの設定で利用した。表 10-3 に分析結果（判別結果）を示す。行が正解、列が LDA によって判別された結果である。以下本章では、この表のように、CSJ-Core の 4 種のレジスターを A(学会講演)、S(模擬講演)、D(対話)、R(再朗読)と略記することがある。

表 10-3: X-JToBI ラベルデータだけを利用した判別分析の結果

	A 学会講演	D 対話	R 再朗読	S 模擬講演
A 学会講演	56	0	0	14
D 対話	0	12	0	6
R 再朗読	2	0	3	1
S 模擬講演	8	0	1	98

表 10-3 の正判別率（対角線要素の合計を 201 で除した値）は 85.1%である。これはいわゆる closed-data の正判別率であるので、leave-one-out 交差評価（cross validation, 201 個の講演データから 1 個を除外し、残る 200 個で分析を実施して得られた判別関数で予め除外しておいたデータのレジスターを予測することを 201 回繰り返す）も実施したところ、78.1%の平均正判別率を得た。以下では Leave-one-out 交差評価のことを単に CV と呼ぶ。

次に X-JToBI ラベルを層ごとに分解して種々に組み合わせて判別分析を実施した。正判別率と CV の結果を表 10-4 に示す。表の上部 3 行には、トーン層、BI 層、プロミネンス層のラベルのみを独立変数に採用した場合の結果であり、4 行目以下は、2 層のラベルをあわせて独立変数とした場合の結果である。

各層を単独で比較した場合に、正判別率が最高となっているのは BI 層のラベルであるが、トーン層ラベルやプロミネンス層ラベルによる判別結果も、BI 層の結果に著しく劣るわけではないことが注目される。２層の組み合わせでは、トーン層と BI 層の組み合わせ、及び、BI 層とプロミネンス層の組み合わせが、ほぼ同程度の成績を挙げており、特に CV においては、X-JToBI の全ラベルを利用した場合（表の最下段参照）と同等以上の成績を挙げていることが注目される。

表 10-4: X-JToBI ラベルの単独及び組み合わせによる判別分析の結果

独立変数	CLOSED-DATA[%]	CV[%]
Tone	70.1	68.2
BI	75.6	74.6
Prominence	71.6	68.2
Tone+BI	80.6	78.1
Tone+Prominence	76.1	72.6
BI+Prominence	79.1	78.6
All tiers (Table 3)	85.1	78.1

表 10-5: 追加データの効果

独立変数	CLOSED-DATA[%]	CV[%]
X-JToBI+SR	86.6	80.6
X-JToBI+(D)+(F)	85.1	78.6
X-JToBI+(D)+(F)+SR	87.6	80.1
(D)+(F)+SR	64.2	63.2

10.3.2 X-JToBI データと追加データによる判別分析

表 10-5 は X-JToBI ラベルデータと追加データを独立変数に用いた判別分析の結果である。SR（発話速度）を追加することによって closed data で 1.5%、CV で 2%程度の成績向上が認められる。一方、非流暢性は、単独で追加しても CV が 0.5%向上するにとどまる。表の最後の行は X-JToBI ラベル情報を利用せずに、SR と非流暢性情報だけで判別分析を実施した場合であり、判別率には著しい低下が認められる。

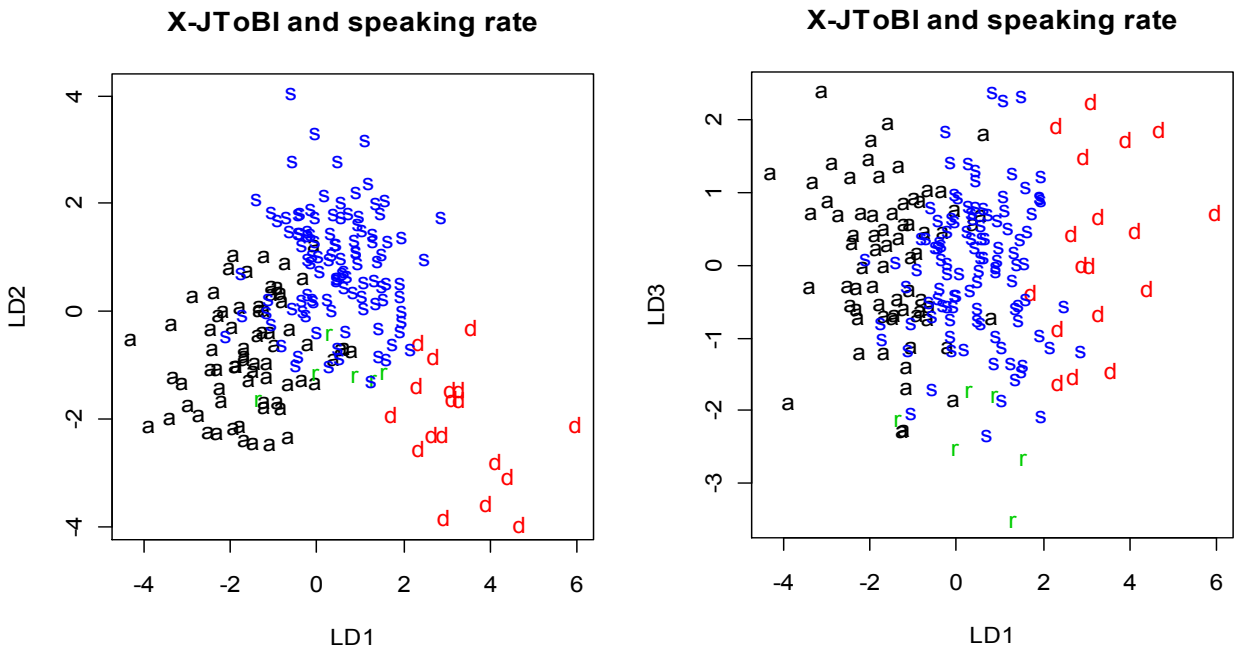


図 10-1: CSJ-Core の全 201 講演の判別空間上の分布

図 10-1 は X-JToBI ラベルデータと SR（平均発話速度）データを用いて構成した判別関数によって構成される判別空間における CSJ-Core の全 201 サンプルの散布状態を示したものである。アルファベットは CSJ のレジスターを示しているが、視認性を高めるためにアルファベットの小文字を利用している。図 10-1 の左パネルは、第 1 判別関数（LD1）と第 2 判別関数（LD2）によって構成される直行空間（平面）を、また右パネルは第 1 判別関数と第 3 判別関数（LD3）によって構成される直行空間を示している。

LD1 は、D とそれ以外のレジスターを効率よく分離しており、さらに S と A もかなりの程度まで分離することに成功している。LD2 は主に S と A の分離に貢献しており、LD3 は R をそれ以外から分離することに貢献している。

10.3.3 データ量と発話中の位置の効果

前節で利用した独立変数の計算では、講演の長短にかかわらず、講演の冒頭から末尾までのすべてのラベル情報を利用していた。しかし人間が講演のレジスターを韻律特徴から判断する課題を考えると、例えば 10 分間の講演の全体を聴取する必要はないものと思われる。

判別分析の独立変数として利用されるラベル頻度情報の計算に利用する音声データ量を最小 60 秒から最大 600 秒まで 60 秒単位で累積的に増加させつつ、判別分析を繰り返し実施した結果を図 10-2 に示す。当然のことながら、正判別率は closed-data も CV もデータ量の増加につれて、ほぼ単調に上昇しつづける。CV に注目すると、冒頭から 600 秒までのデータを利用したあたりで、図の横軸右端に示されている、すべてのラベルを使った場合の結果と同じ水準に達しているが、CSJ の講演の大部分を占める学会講演と模擬講演が平均すると 15 分前後の講演であることを考えると、図 10-2 の結果は、データは多ければ多いほど良いことを示していると解釈できる。

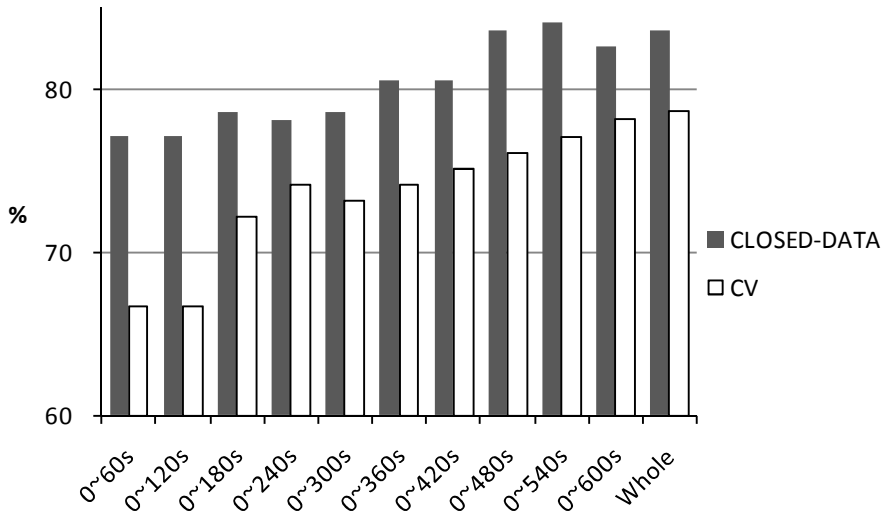


図 10-2: 累積的な判別分析の結果

しかし、図 10-2 は、もうひとつ興味深い問題を提起している。それは最初の 60 秒のデータだけを利用した場合（すなわち横軸の左端）にも、それなり(closed-data で 77%、CV で 67%程度)の判別率が得られていることである。CSJ-Core データにおけるベースラインの正判別率は、すべての講演を模擬講演と判定した場合に得られる 53%であるから、冒頭 60 秒の情報を利用しただけで、正判別率は 20%ほど上昇するわけである。

この事実は発話の冒頭が特別な位置であることを意味しているのだろうか。あるいは発話の冒頭以外の位置からやはり 60 秒幅の音声を抽出して、そこに含まれているラベル情報だ

けを利用した場合にも発話冒頭と同程度の正判別率が得られるのだろうか。この問題は、後に 10.6.3 節でとりあげる韻律特徴の遍在性(pervasiveness)の問題と直接に関係した問題である。

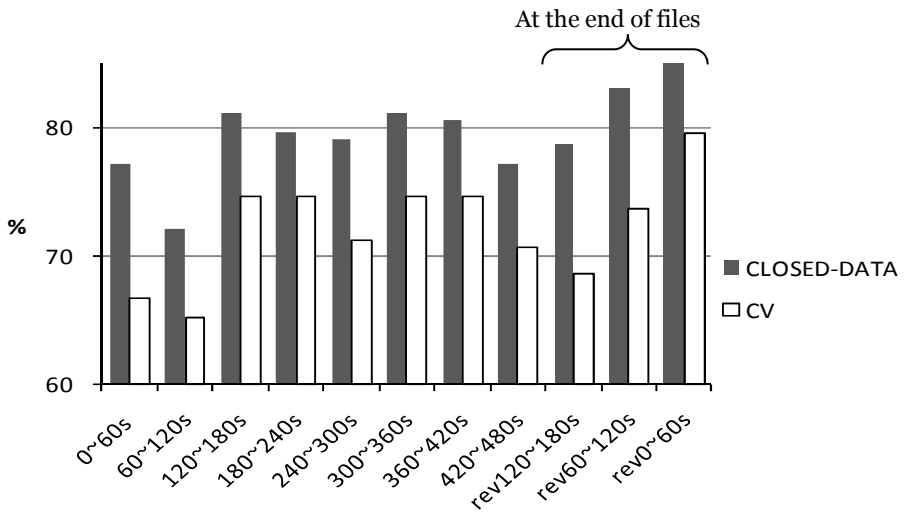


図 10-3: 60 秒幅のデータ窓を利用した判別分析の結果

図 10-3 は、この問題を検討するために、CSJ-Core の全発話の冒頭から 60 秒間隔でデータ窓を移動させつつ、判別分析を繰り替えし実施した結果である（隣接するデータ窓の間に重複がないことに注意）。横軸における「120~180s」のような表記は、講演の冒頭から数えて 120 秒の位置から 180 秒の位置までの 60 秒間に生じたラベルを用いた分析であることを示している。横軸の右端にある「rev60~120s」のような表記（3 個）は、講演の末端から冒頭方向に測って 60 秒の位置から 120 秒の位置までに含まれるラベルを利用した分析であることを示している。なおトーン層ラベルのうち LHL%と H%>は、図 10-3 の分析から除外している。これらのラベルは頻度が低いため、データ窓の設定次第では頻度がゼロになることがあり、その場合、データの Z 変換が行えないからである。

図 10-3 は、正判別率は（closed-data も CV もともに）、発話の冒頭近くにデータ窓が位置する場合にはむしろ相対的に低いことを示している。判別率は、発話冒頭から 120 秒程度が経過した後に上昇し、その後は安定した状態に入り、closed-data で 80%、CV で 70% 前後の正判別率を保持しつづける。最後に、発話全体の末尾(rev~60s)においては、安定状態よりも若干高い正判別率が得られている。

以上の結果をまとめると、CSJ の 4 種のレジスターを判別するための韻律情報は、講演の全体にわかって分布しているが、冒頭部分においてやや希薄であり、末尾部分においてやや濃厚である。また closed-data で 80%、CV で 70%程度の正判別率を得ることが目的であれば、

講演全体を分析する必要はなく、講演冒頭以外のデータを 60 秒程度分析すればよいことがわかった。

10.4 個々の変数の分析

前節では多数のラベルを同時に独立変数として利用した。本節では個々のラベルの分布状態を詳しく検討する。

10.4.1 統計的検定

前節でとりあげた独立変数のすべてに対して Kruskal-Wallis 検定を実施した。主効果はレジスターの差である。表 10-6 に結果をまとめた。表の第 2 列が主効果の検定結果であり、3 列以降は下位検定(post-hoc test)の結果である。A/D, R/S などの表記は、それぞれ A と D、R と S の間の下位検定であることを示している。下位検定の結果は有意水準のみで示している。

主効果の検定結果は、24 個の変数中 20 個に何らかの水準で有意差が認められたことを示している。下位検定の結果は、6 個の変数（2, 2+p, W, P, LH%, HR）では、いずれかのレジスターの組み合わせにおいても有意な差が認められなかった。残る 18 変数では少なくともひとつの組み合わせで有意差が認められた。一方、レジスターの組み合わせ（表の列）に注目すると、いずれの組み合わせにおいても少なくとも 2 個以上の変数に有意差が生じている。

表 10-6: 個々の変数の Kruskal-Wallis 検定結果

変数	Kurskal-Wallis 検定†	A/D‡	A/R	A/S	D/R	D/S	R/S
1+p	χ^2=13.7849, p=0.0032	n.s.	n.s.	**	n.s.	n.s.	n.s.
2	χ^2=1.986, p=0.5753	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
2+p	χ^2=5.0293, p=0.1697	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
2+b	χ^2=85.2834, p<0.0001	***	n.s.	***	n.s.	n.s	n.s
2+bp	χ^2= 24.4407, p<0.0001	**	n.s.	***	n.s.	n.s.	n.s.
3	χ^2= 67.2773, p<0.0001	***	n.s.	***	n.s.	n.s.	n.s.
W	χ^2= 1.2633, p= 0.7379	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
P	χ^2= 1.2633, p= 0.7379	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
PB	χ^2= 53.1543, p<0.0001	*	n.s.	***	n.s.	***	**
L%	χ^2= 19.6915, p= 0.0002	n.s.	n.s.	***	n.s.	n.s.	n.s.
H%	χ^2= 53.1138, p<0.0001	n.s.	n.s.	***	n.s.	n.s.	n.s.
HL%	χ^2= 48.2128, p<0.0001	n.s.	n.s.	***	*	n.s.	***
HLH%	χ^2= 10.3559, p=0.0158	*	n.s.	n.s.	n.s.	*	n.s.
LH%	χ^2= 7.9543, p= 0.0470	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
L%>	χ^2= 36.7312, p<0.0001	n.s.	n.s.	***	**	n.s.	***
H%>	χ^2= 11.9496, p= 0.0076	n.s.	**	n.s.	*	n.s.	*
EUAP	χ^2= 16.7378, p= 0.0008	**	n.s.	n.s.	n.s.	n.s.	n.s.
FR	χ^2= 21.3271, p<0.0001	n.s.	n.s.	**	n.s.	*	n.s.
HR	χ^2= 8.1249, p= 0.0435	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
PNLP	χ^2= 32.9479, <0.0001	***	n.s.	**	n.s.	n.s.	n.s.
QQ	χ^2= 19.943, p= 0.0002	n.s.	n.s.	**	n.s.	n.s.	n.s.
(D)	χ^2= 28.3144, p<0.0001	***	n.s.	n.s.	n.s.	***	n.s.
(F)	χ^2= 27.1941, p<0.0001	**	n.s.	n.s.	*	***	n.s.
SR	χ^2= 62.5086, p<0.0001	***	*	***	n.s.	***	n.s.

† χ^2 統計量の自由度はすべて 3。

‡ 有意差水準のコード: *** 0.001, ** 0.01, * 0.05, n/s Not Significant

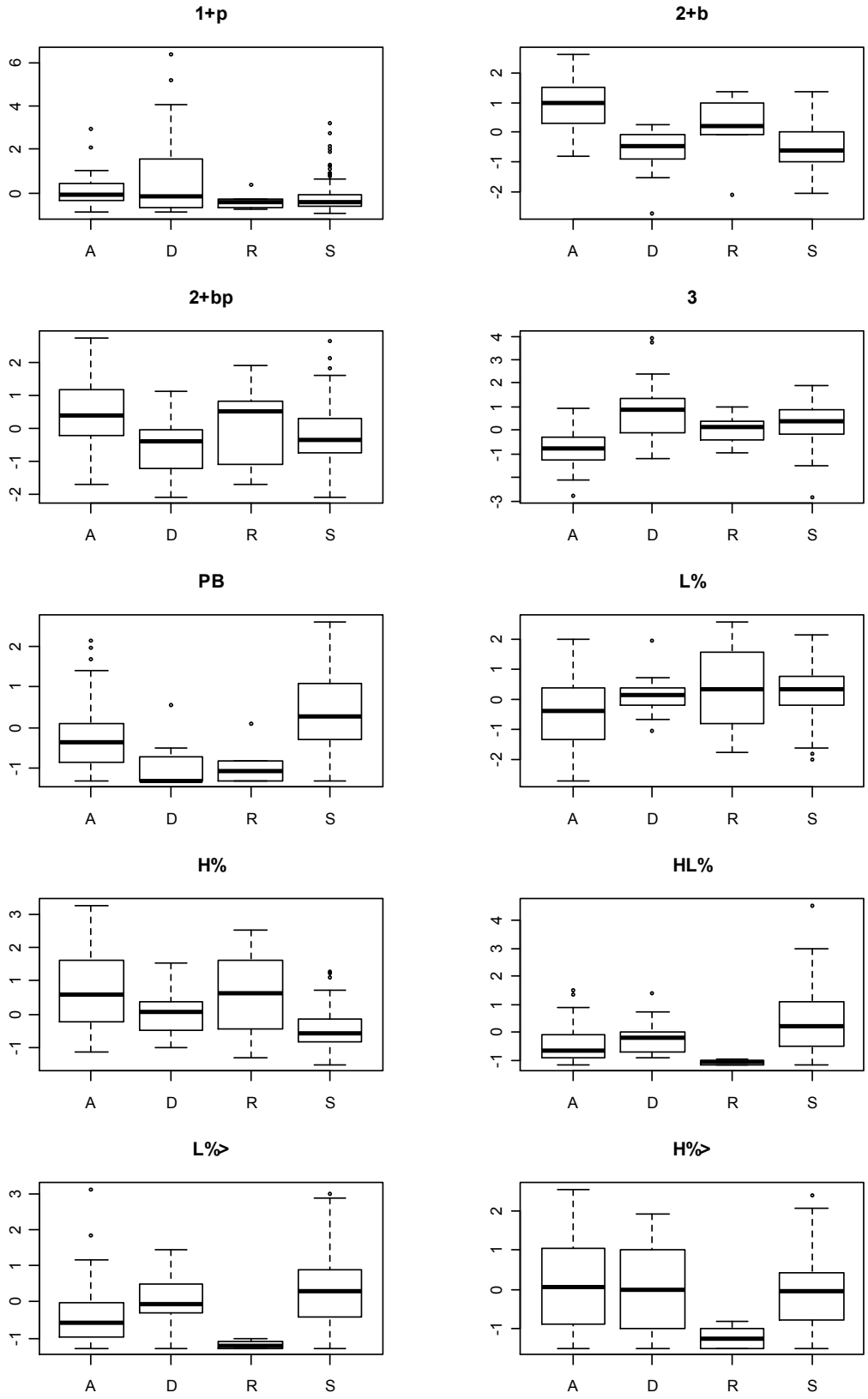


図 10-4: 主要な独立変数のレジスターによる変動

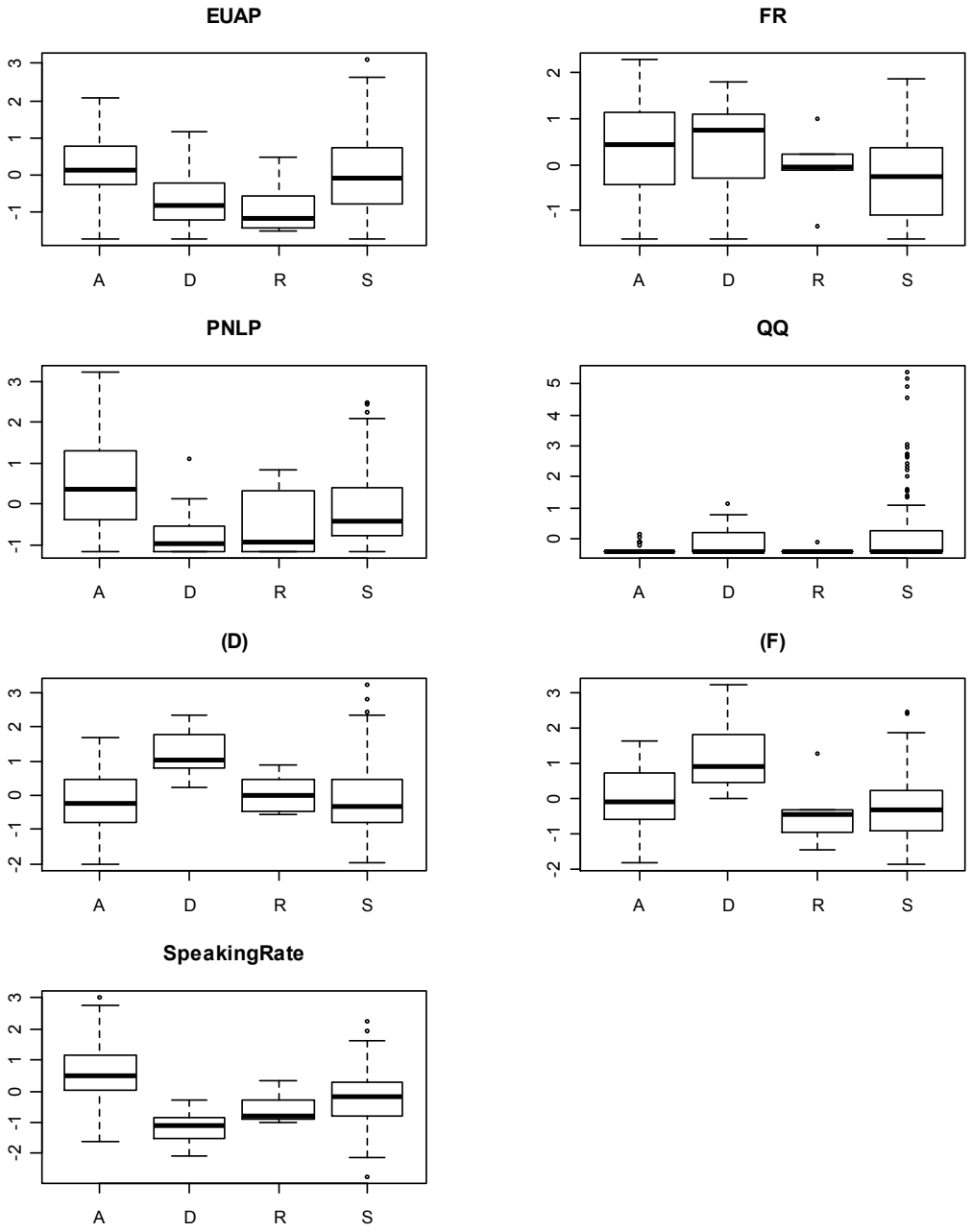


図 10-4: 主要な独立変数のレジスターによる変動（続き）

図 10-4 の各パネルは表 10-6 において $p<0.001$ の有意差を示した独立変数のレジスターによる変動を箱ひげ図として示したものである。これらのパネルの縦軸は平均 0、標準偏差 1 に Z 変換された変数値である。またいくつかの変数は Z 変換の前に対数変換されている（10.2.3 節参照）。図 10-4 に示された結果については 10.6.1 節で詳しく検討する。

10.4.2 変数選択付判別分析

表 10-6 に示されたすべての変数がレジスターの判別に不可欠であるとは考え難い。例えば、2+b と 2+bp の二つの変数の間には変数の定義から推して、ある程度の相関があると予想されるし、実際、表 10-4 においても、これら 2 変数の分布には強い類似性が認められる。予測に最低限必要な独立変数を選択するために、変数選択付の判別分析を実施した。ここまで利用してきた R 言語の lda()関数には変数選択のオプションがないので、青木繁伸氏が開発した sdis()関数を利用して変数選択付の線形判別分析を実施した（この関数の R 言語用のスクリプトは http://aoki2.si.gunma-u.ac.jp/R/sdis.html で公開されている）。変数選択の手法としてはステップワイズ選択法を利用している。

21 個の X-JToBI 変数を独立変数として分析すると、以下の変数が選択された。2+b, PB, P, HL%, 1+p, 3, L%>, 2+p, PNLP の 9 個である。興味深いことに、HL%と 3 以外は X-JToBI において拡張されたラベルである。この事実は後に 10.6.2 節でもとりあげる。

変数選択の妥当性を検討するために、上記の 9 変数を独立変数に用いて R 言語の lda()関数による判別分析を実施した。正判別率は closed-data で 83.6%、CV で 80.6%であり、表 10-4 中の最良の成績と同水準の結果が得られた。変数選択の妥当性を裏付ける結果である。

10.5 決定木

前節で選択された独立変数間の相互関係を検討するために、決定木分析を実施した。R 言語(Ver. 2.10.0)の mvpart ライブラリの rpart()関数の method オプションを class（質的変数）に指定して分析を実施し、cost-complexity pruning で木の剪定を行った。構築された決定木を図 10-5 に示す。ルート・ノードでの分岐は 2+b の値に依存している。これは CSJ-Core のレジスター構成が学会講演(70 ファイル)と模擬講演(107 ファイル)に偏っているため、この両者を効率よく分離できる変数が選択されたものである。

ルート・ノードの右枝における分岐は PB の値によるもので、PB の値が相対的に大きいサンプルを模擬講演に分類している。このノードにおいて模擬講演以外に分類されたサンプルは、最後に HL% の値によって対話と再朗読に分類される。HL% が相対的大きいものが対話に、そうでないものが再朗読に分類されている。

図 10-5 の決定木による分類結果をまとめると表 10-7 のようであり、わずか 3 個の変数のみを利用した分類ながら、正判別率は 79.1%に達している。

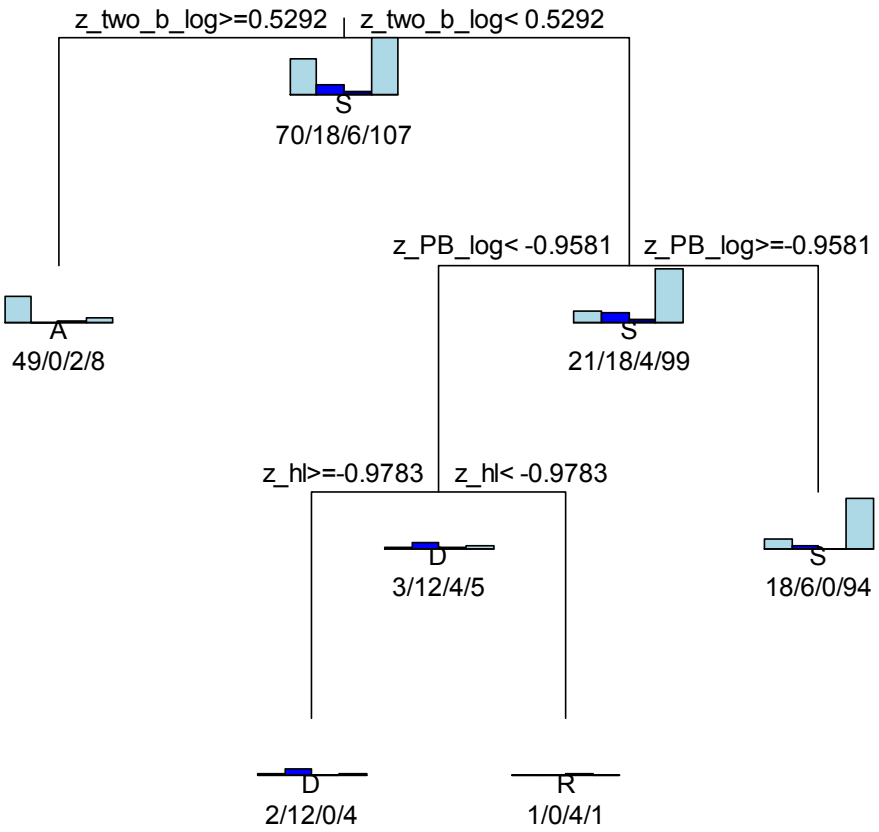


図 10-5: 変数選択後の 9 変数を用いて構成された決定木

表 10-7: 決定木による分類の結果

	学会講演	対	再朗読	模擬講演
学会講演	49	2	1	18
対話	0	12	0	6
再朗読	2	0	4	0
模擬講演	8	4	1	94

10.6 議論

10.6.1 レジスターを特徴付ける韻律特徴

最初に CSJ に記録された 4 種のレジスターがどのような韻律特徴によって特徴づけられているかを検討する。

A. 学会講演（APS）

学会講演の韻律は第一に 2+b の頻度が高いことによって特徴づけられている。図 10-5 の決定木のルート・ノードの分岐規準にこの変数が選ばれているように、学会講演と模擬講演を最もよく区別するのが 2+b である。この BI ラベルは、図 10-4 においてよく似た分布を示す 2+bp と同様、多数の AP から構成され、途中に IP 境界（図 8-1 参照）を含まない、長い発話を特徴付けるラベルである。

つまり、AP 末に BPM が存在していながら、ピッチ・レンジがリセットされない（従って IP 境界とはならずに）発話であり、有核 AP の連鎖であれば、ダウンステップが延々と続いていく類の発話である。主観的な印象としては、そのような発話の典型例は、工学系大学院の大学院生の発表にしばしば見出されるように思える。そのような発表は、文科系の大学院生とは違って講演のために書き下ろした原稿を読み上げているわけではないが、事前に周到に練習を重ねていることが多い。そのため、発話すべき内容（テキスト）は大部分暗記されており、結果として、通常よりも顕著に長いスパンにわたる発話のプラニングが行なわれている状態だと想像される。

先にも指摘したように 2+bp のレジスター間の分布は 2+b に類似しているので、これによって学会講演を模擬講演から区別することが可能であると考えられるが、2+bp の生起数(3456 回)は、2+b(7098 回)に比較すると少ない点でレジスターの識別力は 2+b に劣る。

これらの他に学会講演を特徴付ける変数としては、3 の頻度が低いこと、H% , PNLP, EUAP の頻度が高いこと、そして、SR(発話速度)の値が大きいことを指摘できる。

B. 対話

このレジスターの韻律は、3, (D), (F)の 3 変数の頻度が高いこと、及び、SR の値が小さいことである。 3 の頻度が高いことは、対話における発話長がモノローグに比べて短いことを意味している。非流暢性に関する変数の頻度が高いことも自然な結果であると言える。対話の発話速度は、CSJ に関する限り、学会講演ないし模擬講演よりも顕著に低い。

C. 模擬講演（SPS）

模擬講演の韻律は、2+b と H%の頻度が低いことと、その一方で、PB と HL% の頻度が高いことによって特徴付けられている。模擬講演において 2+b の頻度が低いことは、学会講演において 2+b の頻度が高いことの逆の理由で説明できると思われる。つまり、全般的にみて模擬講演の話者は学会講演に比べると講演前にあまり練習を行わずに講演にのぞんでいるため（2.2.4 節参照）、発話の即興性が高く、長大な区間におよぶプラニングを行いにくいのだと考えられる。

その傍証とするために両レジスターにおける発話長を比較すると、ABS 区分発話を構成する AP 数で比較した場合、学会講演が 9.37 個であるのに対して、模擬講演は 8.99 個である。また、200ms 以上のポーズもしくは文法的な文末形式によって両端を区切られた転記基本単

位（2.3.1 節参照）の時間長で比較すると、学会講演が 2077ms であるのに対して模擬講演は 1841ms であり、いずれも模擬講演の方が短い。

模擬講演では学会講演よりも HL% の頻度が高いこと（そして反対に学会講演では模擬講演よりも H% の頻度が高いこと）は、Maekawa(2009)などに示したとおりであるが、その原因は本論文の 9 章で示唆したように、両レジスターにおける発話スタイル（発話のあらたまり度）の差によるものと考えられる。全般的傾向として HL% はくだけた発話で多用される（そして H% はあらたまった発話に頻出する）傾向がある。

最後に、PB の頻度の高さが模擬講演を特徴付けるのは、模擬講演には第三者の発話を直接引用する発話が多数含まれているのに対して、学会講演にはその種の引用が稀であることによる。PB が引用と関係することについては 3.5.2 節参照。

D. 再朗読

再朗読音声の韻律は 3 個の BPM、すなわち HL%, H%>, L%>の頻度が極端に低いことで、これが再朗読音声の「単調な」（あるいは「棒読み」の）印象を形作っていると思われる。

ところで読者は、再朗読音声には非流暢性要素が少ないはずであり、それが再朗読音声を特徴付けるはずだと考えるかもしれない。しかし実際にはそうなっていない。これは、CSJ の音声転記ファイルには、 (F)や(D)の要素も可能な限り正確に転記されおり、再朗読音声の話者は、これらの非流暢性要素も再現するよう、録音時に指示を与えられているためである。

10.6.2 X-JToBI ラベルの有効性

本章の分析によって、X-JToBI ラベルによって記号的に表現された発話の韻律的特徴を利用するだけで、CSJ に記録された発話のレジスターを 8 割程度の正解率で判別できることが明らかになった。また、本章で検討した 26 個の X-JToBI ラベルのうち、レジスターの判別に大きく貢献しているのは、J_ToBI から X-JToBI への拡張作業において追加されたラベルであったことを強調しておく必要がある。

表 10-6 において $p<0.0001$ 水準の有意差が認められた 9 ラベルのうち 6 個は X-JToBI 拡張ラベルであった（2+b, 2+bp, PB, L%>, FR, PNLP）。また 10.4.2 節のステップワイズ変数選択判別分析では、選択された 9 変数のうち 7 個が X-JToBI 拡張ラベルであった（2+b, PB, P, 1+p, L%>, 2+p, PNLP）。最後に図 10-5 の決定木に含まれる 3 変数のうち 2 個（2+b と PB）が X-JToBI 拡張ラベルであった。2+b と PB は、いずれの分析においても選択されている。

これらの事実は、X-JToBI への拡張が成功であったことを明白に物語っている。実際、従来の J_ToBI ラベル（すなわち H%, HL%, L%, 2, 3 の 5 個）だけを利用した判別分析を実施すると、正判別率は closed data で 74.6%、CV で 72.6%であるから、表 10-4 と比較すると、X-JToBI への拡張によって closed data で 10.5%、CV で 5.5%の向上が認められる。

10.6.3 韻律的レジスターと発話スタイル

レジスターによって音声の韻律的特徴が変動することは広く信じられてきているが、それを客観的に証明するための研究は、従来ほとんど試みられてこなかった。本章の研究は上記の信念について客観的な証拠を提供したものである。

ここで、発話のレジスターのうち韻律的特徴の面にかぎって言及するための用語として「韻律的レジスター」(prosodic register)という新術語を提案したい。ここでも「レジスター」の語は、コーパス言語学における用法と同一の用法で用いられている。つまり、韻律的レジスターの区別は、特定のレジスターにしか出現しない特殊な韻律的特徴の有無によってなされるのではなく、どのレジスターにも遍在している(pervasive)が、その相対頻度がレジスター間で大きく変動する韻律特徴によってなされる (Biber and Conrad 2009)。

2+b や PB に代表される X-JToBI ラベルにはそのような変数としての資格が十分に備わっている。X-JToBI ラベルはすべてのレジスターにわたって分布しており（図 10-4 参照）、また講演中の特定位置に偏って存在しているとも考えられない(図 10-2, 10-3 参照)。

ところで、言うまでもないことだが、韻律的レジスターの概念は、発話スタイル(speaking style)の概念と密接に関係している。しかし両者は同じ概念ではない。両者の間には、自明とは言い難い差異が少なくとも二つ存在している。

まず、Eskénazi (1993)などの批判にも関わらず、発話スタイルはしばしば、発話のあらたまり度に関わる一次元の尺度であると理解される。そのような解釈が典型的に通用しているのは米国流の社会言語学であり(Labov 1972 など参照)、発話スタイルは、casual style から formal style にかけて変化する 1 個の尺度として、分析に利用されている。

このように種々の発話を何らかの尺度上で序列化することは、韻律的レジスター分析の目指すところではない。韻律的レジスターは、「研究発表」「一般的なスピーチ」「大学での講義」「ニュースの報道」「くだけた会話」などのように、発話が行なわれる社会的状況の特性に注目した発話タイプの分類である。当然、どの言語においても「韻律的レジスター」は相当数に及ぶことが予想される。郡(2006)は、インターネット上で「口調」を表現するために利用されている言語表現を検索し、分類することによって、「宝塚口調、講義口調、アナウンサー口調、歌舞伎口調、ナレーター口調、バスガイド口調」など 15 個の「場面限定的な」口調の名称をとりだしているが、これらは日本語の韻律的レジスターの候補の一部であると考えられる。

次に、発話スタイルに関する従来の音声学的研究は、専ら物理的な、従って連続的な音声特徴に注目している。F0、発話速度、発話ないし分節音の持続時間長、母音フォルマント空間の縮約などや(Llisterri 1992; Abe 1997)、その背景に存在すると考えられるモデルパラメータ(Higuchi et al. 1997)である。これに対して、韻律的レジスターの研究は、音声の連続的特徴量の研究を排除するものではないが、韻律的特徴の記号的側面により強く注目する。本研究の例で言えば、PB や EUAP などのラベルによって表現されている韻律特徴は、単一の

物理的特徴で表現されるものではなく、形態素境界などの情報まで含めた総合的な言語的分析によって導き出されたものである。

そして、この点において、韻律的レジスターの研究は、コーパスの韻律アノテーションに利用されたラベリング体系の設計に全面的に依存した関係にある。仮に X-JToBI ラベルに 2+b や 2+bp が含まれていなかったならば、学会講演の韻律的レジスターは 3 の頻度が低いことによって特徴付けられることになる可能性が高いが、音声学的な詳細性の観点からすれば、これが不十分な特徴づけであることは言うまでもない。

10.7 結論

本章の目的は発話のレジスターが韻律特徴に及ぼす影響を組織的に検討することであった。本章の分析結果は、X-JToBI ラベル情報だけがあれば、CSJ-Core の 4 種のレジスターが 80% 以上の精度で判別可能なことを示しており、さらに各レジスターを特徴付ける韻律ラベル、及びラベル間の関係も明らかにすることができた。このような成果は総体として J_ToBI から X-JToBI への拡張が有意義であったことを示している。

本章で提案した韻律的レジスターの分析は、音声科学における重要な未開拓領域のひとつである。今後より広範なレジスターをとりあげて、その韻律的レジスターを客観的な手法で解明していけば、それによって、我々の音声行動に対する理解が深まるのは当然であるが、同時に音声処理技術にも貢献することができる。例えばより多様性に富んだ合成音声の実現などの可能性が考えられる。

逆にいえば、本章の研究の限界は、研究対象が CSJ-Core に記録された 4 種のレジスターに限定されてしまっている点にある。CSJ のレジスターの幅は、従来のコーパスに比べれば、決して狭いものではないが、しかし日本語のレジスター全体からすれば、明らかに限られた一部分しか取り扱っていない。

今後より広い範囲のレジスターの韻律を分析するためには、当然ながら、CSJ-Core のようなコーパスが必要とされる。この点で、10.3.4 節で得た知見は、韻律的レジスター研究用コーパスの設計の重要な指針を与えるものになっており、コーパス構築のコスト軽減に貢献すると思われる

第Ⅳ部

第 11 章. 結語

11.1 本研究のまとめ

本研究全体の目標は、大規模なアノテーション付自発音声コーパスを分析することによって自発音声の科学的研究が可能となり、それによって音声学に新しい可能性が拓けることを示すことにあった。この目標に沿って、本研究では三種類の研究を実施した。

11.1.1 コーパスの設計と実装

CSJ の設計にあたっては、音声認識研究での利用を念頭においてモノローグを中心にすえたが、同時に発話スタイルの異なるふたつのレジスター（学会講演と模擬講演）をほぼ同量収録することで、スタイルにかかわる比較研究を可能にした。また、CSJ-Core には、韻律ラベルを含め、可能な限りのアノテーションを実施した。完成した CSJ の予備的分析結果は、音声、語彙、文法のいずれの面においても、CSJ には自発音声ならではの好ましい特性が存在することを示していた。

CSJ-Core の韻律アノテーションのために、日本語（東京方言）自発音声の韻律アノテーション体系である X-JToBI を考案した。X-JToBI は、従来の J_ToBI よりも多くの面で複雑化しているが、実際に自発音声に適用すると、J_ToBI よりも高いラベリング精度が得られる。本研究後段の研究は X-JToBI ラベルがあって、はじめて可能になったものである（X-JToBI の有効性については 11.1.3 節参照）。

11.1.2 分節的特徴の分析

日本語音声の教科書的記述では、狭母音は前後を無声子音に挟まれた環境で無声化するとされているが、実際にはそのとおりに無声化が生じるわけではない。CSJ-Core の場合、無声化率は 90%程度である。本研究では、無声化によってモーラ境界の知覚が曖昧化することを回避することが無声化率低下の主要な原因だと考えて、この仮説をコーパス全体の音響分析によって立証した。

やはり教科書では、日本語の音素/z/は、語等では破擦音、語中（母音間）では摩擦音で発音されると記述されている。CSJ-Core のデータはこの定説が成立しないことを示している。/z/の破擦率の説明要因としては、本研究が提案する TACA の貢献が大きい。TACA は、/z/自体の持続時間に、直前に位置するポーズ、促音、撥音の持続時間を加算した量であり、話者が/z/の調音に利用可能な時間を表現している。TACA の値が大きければ、/z/は語中位置でも破擦音で発音される。

TACA による説明は、有声破裂音/b, d, g/の閉鎖調音の弱化現象の分析にも有効であった。また TACA と声道閉鎖調音の実現率の関係を 4 種の音素(/z, b, d, g/)について比較すると、日本語の子音体系の特性がそこに反映されていることがわかった。

以上の研究では、X-JToBI に含まれる分節音ラベル情報と CSJ-Core に施された韻律構造、形態論情報等との相関を検討することで新しい知見が得られた。また、CSJ には発話速度の面で非常に幅広いレンジの音声が収録されていることが問題の解決を容易にした。

11.1.3 韻律的特徴の分析

東京方言アクセントの教科書的記述では、「など」「さえ」等の２モーラ有核名詞は有核名詞と結合するとアクセント核を失うとされている。CSJ の分析結果は、この環境にあっても助詞のアクセントが保持されることがあり、助詞によってはアクセントが保持されることの方が多いことを示していた。この傾向は、強調の意味をもつ助詞において特に顕著である。

日本語のイントネーションに final lowering が存在するかどうかについては研究者によって意見がわかれており、また、その作用域の広さについても意見が分かれている。CSJ-Core の分析結果は、自発音声にも final lowering が存在すること、また、その作用域は発話の最終 AP に限られていることを明確に示していた。この研究では、節境界情報と F0 値の相関分析が中心的な役割を果たした。

アクセント句末のイントネーションである BPM については、従来主に対話音声の観点から研究が行なわれてきており、モノローグのデータについての知見が不足していた。CSJ-Core のモノローグに生じた BPM と印象判定データとの関係を検討すると、BPM には強いスタイル依存性が存在していることが判明した。

最後に、韻律特徴の総合的分析として、話し言葉のレジスターと種々の X-JToBI ラベルの生起頻度がどのように関係しているかを分析した。その結果、発話内容（テキスト）の情報を全く利用することなく、韻律情報だけで、CSJ-Core の 201 ファイルのレジスターを 85%の精度で判定可能であることがわかった。

判定に必要な情報は、講演の全体に分布しており、3 分間程度の音声韻律ラベルデータがあれば、80％程度の精度での判定が可能になることがわかった。

レジスターの判別に貢献しているラベルの大部分は J_ToBI から X-JToBI への拡張の過程で追加されたものであり、X-JToBI の設計の有効性も示された。

11.2 得られた成果の分類

本研究第 1 章で自発音声研究の必要性を論じた際に、自発音声の研究が要請される研究に３種の類型を示した。その類型に沿って本研究後半（第 4 章から第 10 章）の成果を評価する。

第一の類型は自発音声には自然に生じるが、朗読音声で実現することが非常に困難な現象である。BPM の分類（第 9 章）と韻律レジスター（第 10 章）がこのタイプに属する。これらの章の研究では、コーパスのデータマイニングによって、従来の日本語音声学では報告されたことのない新事実が発見された。具体的にはモノローグにおける BPM のスタイル依存

性、PNLP の頂点ないし境界機能、そしてレジスターと相関した韻律特徴の組織的変化である。

第二の類型は、朗読音声と自発音声とで生じ方に差がある現象であり、母音の無声化（第 4 章）と２モーラ有核助詞の独立性（第 7 章）とがこのタイプに属する研究であった。いずれの研究についても、コーパスの分析によって、朗読音声の研究では知られていなかった新しい規則性（無声化においてはモーラ境界知覚の重要性、助詞のアクセントにおいては助詞の意味特性）を発見することができた。

第三の類型は、朗読音声にも生じるが、朗読音声を研究しているだけでは正しく理解することの困難な現象であり、子音調音の変異（第 5 章、第 6 章）と final lowering（第 8 章）がこのタイプに属する研究であった。これらの研究では、音声学上の新事実が発見されたわけではないが、従来の定説に根本的な修正をもたらしたり（/z/の変異）、従来対立していた解釈の優劣を決したりすること（final lowering の作用域）ができた。

また子音調音の研究では 10 種以上の言語的、社会的変数を回帰分析、決定木等の統計手法で検討したうえで、最終的に TACA の重要性を確認している。従来の実験研究よりもはるかに広い範囲の変数を検討できたのは、やはり大規模なコーパスが利用できたからである。

以上をまとめると、第 1 章で設定した 4 種の類型のいずれにおいても、大規模なアノテーション付コーパスを利用することによって自発音声の研究が可能となることが示された。

11.3 CSJ の評価

このような成果が得られたのは、本研究が依拠した CSJ に以下のような特性が備わっていたからだと考えられる。

① たとえ十分ではなくとも今回検討した範囲の問題を解決できる程度には規模が大きいこと。
② X-JToBI、節境界ラベル、印象評定データなどのアノテーションが利用できること。
③ たとえ十分ではなくとも比較的幅広いレジスターをカバーしており、話者の年齢、性別などの社会的属性にも広がりがあること。
④ 発話速度が大幅に変動していること。

このうち①は、自発音声を研究するためのデータとして最重要の特性である。統制を施すことのできない自発音声を対象とした本研究の全体を通して、CSJ ならびに CSJ-Core のサイズが問題の解決に寄与したと考えられる事例は枚挙にいとまがない。

②も①とならんで重要な特性である。本研究では、第 4 章、第 5 章、第 6 章の研究が X-JToBI の分節音アノテーションを利用しており、第 8 章、 第 9 章、 第 10 章の研究が全面的に X-JToBI の韻律アノテーションに依存していた。また第 8 章、第 9 章では節単位境界ラベルを利用して ABS 区分発話単位を設定することで、また第 9 章では印象評定ラベルを利用することによって、研究の新しい展開が可能になった。さらにこれらすべての研究において

形態論情報を有効に利用している。このように複数のアノテーションを有機的に関連付けて利用できることが、アノテーション付コーパスを用いた研究の特徴であり、最大のメリットでもある。

③は言語変異の研究において欠くことのできないスタイル分析にとって重要であり、本研究でもいくつかの章で音声変異とレジスターないし話者の社会的属性の関係を検討した。本研究においてスタイルの問題を正面からとりあげたのは 9 章の BPM 分析と 11 章の韻律レジスターの分析であるが、このうちレジスターの相違を積極的に分析に活用したのは 11 章においてであった。9 章の分析では、通常ならばレジスター間の比較を行なう局面において、より強力な印象評定データを利用することができたのでレジスターには言及していない。

最後に④の発話速度は、従来の実験的研究では不当に軽視されてきた要因である。第 5 章と第 6 章の分析結果は、子音持続時間の決定要因としての発話速度の重要性を疑う余地なく示していた。ちなみにこれらの章の分析では、子音調音の変異と話者の生年代（年齢）の間に明白な相関が存在していたが、最終的にその相関は、発話速度と年齢との相関によって生じた偏相関にしかすぎないことが判明した。従来、社会言語学の研究において年齢との相関が指摘されてきた音声現象（したがって進行中の言語変化の候補とされてきた現象）のなかに、このような再分析を許すものがかなり存在するのではないかと思われる。

以上を要約すれば、CSJ―特に CSJ-Core―には自発音声の音声学的研究の資料として絶大な価値が認められ、その価値の源泉は主にデータサイズとアノテーションの豊富さに求められる。

11.4 今後の課題

言うまでもないことだが、日本語自発音声の音声学的研究が本研究をもって完結したわけではない。今後実施すべき研究は数多い。そのなかで、本研究の継承発展という観点から特に重要と考えられるのは、自発音声の韻律構造モデルの再構築である。本研究の第 9 章で分析した final lowering は、本来、F0 declination, catathesis (downstep)とともに日本語イントネーションの生成モデルの一部を構成しており、final lowering だけを扱った本研究は、他のコンポーネントの研究によって補完されなければならない。今後第一に実施すべき研究である。Final lowering そのものに関しても、それが日本語の述部に生じやすいことから、述語と補語との修飾関係など、発話の文法的側面を考慮に入れた再分析が必要であろう。

次に、本研究全体を補完する意味で重要と考えられるのが、非流暢性要素の音声学的研究である。第 1 章末尾で指摘したように、本研究では意図的に非流暢性要素の研究を行なわなかったが、それは本研究の目標が従来の朗読音声に基礎をおいた音声研究と自発音声分析の成果とを対比することにあったからである。非流暢性要素の音声学には未解明の問題が多く、特に韻律特徴については理論的にも興味深い問題が多い。第 3 章で紹介した X-JToBI ではフィラーのトーンを通常のトーンとは別扱いしているが、フィラーや語断片は韻律的な観点からはあきらかに通常の音声とは異なったふるまいを示している。これをイントネーション生

成モデルのなかにどのように位置づけられるかを解明する必要があり、その結果、韻律構造モデルに一定の修正を加えることが必要になる可能性もある。

分節的特徴についても多くの課題が残されているが、本研究と関係する範囲でひとつだけ指摘すれば、表 2-7, 2-8 などに示された語形のゆれの音声学的・言語学的背景を解明する研究がある。この現象については、第 1 章で紹介したように音声認識システム用発音辞書という形ですでにかなりの研究が行なわれているが、統計的な研究ではぬけおちがちな個々の現象に着目した、音声学的な観点からの研究が有益であろうと考える。

また、分節的特徴、韻律的特徴を問わず、音声現象間の相関関係を探ることも今後に残された重要な課題である。これらの研究は総じて、近年 socio-phonetics と呼ばれるようになった音声行動の個人差、社会差の研究に大きな貢献を果たすものと想像される。

最後に自発音声コーパスの拡張の必要性に触れる。CSJ にはまだ多くの研究の可能性が残されているが、なかには本研究以上の知見を得ることが困難な研究もある。第 10 章で論じた韻律レジスターはその代表例である。CSJ に記録された韻律レジスターが日本語全体からみれば一部にすぎないことは明白であるが、この問題をさらに研究するためには、新しいレジスターを含むコーパスの構築が不可欠である。第 10 章の研究ではレジスターの推定に必要なデータ量を推測したが、そこには、将来のコーパス設計の基礎資料を得るという意図もあった。近い将来、CSJ を補うための自発音声コーパスが構築されることになるであろうが、そのようなコーパスを正しく設計するためにも、自発音声の研究を通じて、次世代コーパスに期待される特性を明らかにしておくことが必要である。

このように、日本語音声学には自発音声の定量的分析を通して再検討されるべき問題が多数残されている。日本語音声学は今後自発音声の実証的研究を通して再構築される必要があり、本研究はその端緒として位置づけられるべきものである。

謝辞

私が大学院（上智大学大学院外国学研究科言語学専攻）に入学したのは 1980 年であった。この年から現在まで 31 年が経過している。修士課程入学から博士号取得までの標準期間を 5 年とするならば、ひとの 6 倍の時間をかけて学位論文を準備したことになる。中身も 6 倍になっていると素晴らしいのだが、残念ながらそのような自信はない。ただ 31 年間に研究面でお世話になった方々の数は時間に比例して大変多い。そのため、この謝辞も長いものにならざるをえない。

最初に、ご多忙中にもかかわらず審査にあたってくださった主査の古井貞熙先生をはじめとする審査委員の先生方に心よりお礼を申しあげる。徳永健伸、仁科喜久子、篠田浩一、藤井敦、佐藤大和の各先生である。

古井先生からのお誘いがなければ、私は引退まで学位と無縁の研究者人生を送ったのではないかと思う。CSJ の開発がほぼ完了したころ、学位論文を提出するよう、最初のお誘いをうけた。たしか京都から東京への新幹線車中であった。頑張ってみますと答えたが、生来の怠け癖と環境の変化（2006 年から文科省科学研究費特定領域研究「日本語コーパス」の領域代表者になったことと 2008 年末に始まった独立行政法人国立国語研究所の廃止と大学共同利用機関法人への移管にまつわる一連の空騒ぎ）とで、2009 年に至っても準備はあまり進んでいなかった。内心もう無理だと考えていたが、その年の秋に郡山市で開催された日本音響学会で久しぶりにお目にかかった古井先生は「前川さんに学位をださないと私は引退できません」とおっしゃった。この言葉を聞いて奮起しない音声研究者はいない。ようやくエンジンがかかり、なんとか先生の退官にまにあわせることができた。

以下、時間の流れにそって研究面でお世話になった方々のお名前を記し、感謝の意を表したい。所属はすべて当時のものである。

大学院在学中、菅原勉、小島慶一の両先生に実験音声学の手ほどきをうけた。金田一春彦先生からは日本語アクセント論をはじめとする日本語学全般の教えをうけた。文学部国語国文学科の森岡健二先生からは研究者としての身の処し方を教えていただいた。やはり文学部心理学科の西川泰夫先生からは数理心理学の基礎を教えていただいた。

大学院と並行して、統計数理研究所の統計技術者養成講座では、清水良一先生ほかの先生方に統計学と確率論と線形代数の手ほどきをうけた。また 1983 年には東京言語研究所理論言語学講座で土田滋先生から調音音声学の手ほどきを受けた。

当時、科学研究費特定研究「言語の標準化に関する総合的研究」(1982-84)が進められていた。私は鳥取大学の今石元久先生の研究班の協力者となって、方言音声の分析に参加した。

1984 年に鳥取大学の助手に応募して首尾よく採用されたのは、今石先生のお勧めによるものであり、その後も何かと親身なご指導をいただいた。

鳥取大学在職中には、大阪樟蔭女子大学の杉藤美代子先生と大阪外国語大学の郡史郎氏が主催しておられた近畿音声言語研究会に研究発表の場を求めることがしばしばであった。ATR（国際電気通信基礎技術研究所）の匂坂芳典博士や南山大学の窪薗晴夫博士はこの研究会の常連であり、発表のたびに有益なコメントをいただいた。

1989 年に国立国語研究所言語行動研究部に移ってからは、東京大学医学部音声言語医学研究施設の廣瀬肇、桐谷滋の両先生と筋電図やファイバースコープを用いた音声生理学の共同研究を行う機会を得た。今川博氏には実験のたびに技術面でお世話になった。同施設ではときどきインフォーマルな講義をする機会も与えていただき、自分の知識を整理するのにおおいに役立った。

この時期には杉藤先生を代表者とする科学研究費重点領域研究「日本語音声における韻律的特徴の実態とその教育に関する総合的研究」(1989-92)が実施された。私は日本語教育班に属して日本語学習者音声データベースの構築に携わると同時に、熊本無アクセント方言のイントネーション構造の研究にも従事した。

重点領域研究終了後、1993 年から翌年にかけては、文科省長期在外研究員として米国オハイオ州立大学（OSU）言語学科の Mary Beckman 教授の研究室に滞在する機会を得た。そこでは英語の ToBI システムが整備されつつあった。大学院生だった Jennifer Venditti 氏が来日して ATR で J_ToBI の原型を考案したのが 1995 年。私が提案した X-JToBI はその拡張版である。

OSU の音声聴覚学科には藤村靖先生がおられた。先生には音声の生理学と物理学のほか言語学の方法論についてもご指導いただいた。

帰国後は米国で学んだ実験音韻論的な方法によるイントネーション研究を進めていたが、イントネーション現象を根本的に理解するためには、パラ言語情報の問題を考察する必要があることを痛感するようになり、藤崎博也先生の先駆的な研究の後を追う形で研究を進めた。

パラ言語情報が調音運動に及ぼす影響についての実験では NTT コミュニケーション科学基礎研究所の誉田雅彰博士と ATR の本多清志博士のご協力をいただいた。また大阪大学の土岐哲先生からはしばしば集中講義の機会を与えていただいた。当時の学生のうち江崎哲也君は CSJ コアの音声ラベラーとして構築作業に参加した。また岡田祥平君はのちに CSJ の分析で学位をとった。この頃、東京都立大学で 3 年にわたって講義する機会があったが、CSJ の構築では、私のクラスに出席していた学生を中心に、都立大学出身者が活躍してくれることになる。この講義の機会を与えていただいたのは篠崎晃一氏であった。

パラ言語情報の研究は大変面白かった。しかし或る程度より先に研究を進めるためには、どうしても朗読音声を脱して自発音声を分析しなければならないことが明らかであった。その可能性を探っていた 1998 年に科学技術庁（当時）が科学技術振興調整費補助金開放的融合研究制度を新設した。国立試験研究機関を対象とするこの補助金を獲得するために、通信総合研究所の井佐原均博士の自然言語処理研究グループと私を代表とする国立国語研究所の研究グループとを融合させて古井先生の統括下に自発音声の音声認識研究を進めるというアイデアは ATR 音声翻訳通信研究所の社長であった山本誠一博士が考案されたものである。

当時の国語研内部では科技庁からの外部資金をうけいれることに若干の慎重論があり、実際に開放的融合研究制度に応募したのは翌 1999 年の 2 月になった。審査委員会に対するプレゼンテーションが古井先生の海外出張と重なり、代役を務めたことを思いだす。プレゼンがよかったせいか申請はただちに採択され、同年 4 月から 5 年間にわたる CSJ の構築作業がスタートした。

開発の現場をささえてくれたのは、小磯花絵さん、小椋秀樹さん、菊池英明さん、籠宮隆之さん、丸山岳彦さん、五十嵐陽介さん、米山聖子さん、藤本雅子さん、間淵洋子さん、西川賢也さん、土屋菜穂子さん、斉藤美紀さん、塚原渉さん、山住賢司さん、槇洋一さん、北川智利さんら新進の研究者たちだった。また阿左美厚子さんは当時から現在まで一貫して研究室の事務作業を担当し、完璧にこなしてくれた。本研究の多くの部分は彼女が生み出してくれる時間の余裕を使って実施したものである。

CSJ の開発が始まったとき国語研は国立試験研究機関だったが、開発終了時には独立行政法人になっていた。そして現在は大学共同利用機関法人である。この行政の混乱期にあって研究を継続できたのは上司に言語資源研究についての理解があったからである。甲斐睦朗、杉戸清樹、影山太郎の歴代所長に感謝したい。相澤正夫氏は過去 10 年間直属の上司として私の研究活動を陰に陽に側面支援してくださった。また国語研入所当時の直属上司であった神部尚武氏は私の気儘な研究活動を黙認してくださった。ともに深く感謝する。

大学共同利用機関法人国立国語研究所の言語資源研究系、同コーパス開発センターの同僚各位には、この 1 年ほどの間、学位論文の準備にかまけて何かと迷惑をかけたと思う。この場を借りてお詫びとお礼を申しあげる。

最後に、いつも家庭にあって私の活動をささえてくれる妻、伸子に感謝する（これからもみすてないでください）。

2011 年 2 月 4 日 論文最終審査の日に記す

参考文献

Abe, M. (1997). “Speaking styles: Statistical analysis and synthesis by a text-to-speech system.” In vanSanten, J.P.H., R. Sproat, J. Olive, and J. Hirschberg (eds.). *Progress in speech synthesis*. NY: Springer-Verlag, 417-428.

Akita, Yuya and Tatsuya Kawahara (2010). “Statistical transformation of language and pronunciation models for spontaneous speech recognition.” *IEEE Trans. Audio, Speech & Language Processing*, 18 (6), 1539-1549.

Aston, G. and L. Burnard (1998). *The BNC Handbook: Exploring the British National Corpus with Sara*. Edinburgh University Press.

Baayen, R.H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R.* Cambridge: Cambridge University Press.

Beckman, Mary E. (1986). *Stress and non-stress accent*. Dordrecht: Foris.

Beckman, M. E. and Janet. B. Pierrehumbert (1986). “Intonational structure in Japanese and English.” *Phonology Yearbook*, 3, 255-310.

Biber, D. (1988). *Variation across speech and writing*. Cambridge: Cambridge University Press.

Biber, D. and S. Conrad (2009). *Register, genre, and style*. Cambridge: Cambridge University Press.

Blaauw, Eleonora (1995). *On the perceptual classification of spontaneous and read speech.* Research Institute for Language and Speech, Utrecht University.

Crystal, D. and R. Quirk (1964). *Systems of prosodic and paralinguistic features in English.* The Hague: Mouton.

Crystal, D. and D. Davy (1969). *Investigating English style*. London: Longman.

Eckert, P. and J. R. Rickford (2001). *Style and sociolinguistic variation*. Cambridge: Cambridge University Press.

Eskénazi, M. (1993). “Trends in speaking styles research”. *Proceedings of Eurospeech 93*, Berlin, 501-509.

Fujisaki, Hiroya (1996). “Prosody, Models, and Spontaneous Speech.” in Y. Sagisaka et al. (eds.) *Computing Prosody*, 27-42, Springer.

Fujisaki, Hiroya and Keikichi Hirose (1984). “Analysis of voice fundamental frequency contours for declarative sentences of Japanese.” *Journal of the Acoustical Society of Japan (E)*, 5 (4), 233-242.

Furui, Sadaoki (2005). “Recent progress in corpus-based spontaneous speech recognition.” *IEICE Transactions on Information and Systems*, Vol.E88-D, no.3, 366-375.

Furui, Sadaoki, Masanobu Nakamura, Tomohisa Ichiba and Koji Iwano (2005). “Analysis and recognition of spontaneous speech using Corpus of Spontaneous Japanese.” *Speech Communication*, 47, 208-219.

Halliday, M. A. K., A. MaCintosh, and P. D. Strevens (1964). *The Linguistic sciences and language teaching*. London: Longman.

Hattori, Shiro (1961). "Prosodeme, syllable structure and laryngeal phonemes." *Bulletin of the Summer Institute of Linguistics*, Vol. 1, 1-27, Tokyo: International Christian University.

Han, Mieko, S. (1962). "Unvoicing of vowels in Japanese." *Study of Sounds*, 10, 81-100.

Higuchi, N., T. Hirai, and Y. Sagisaka (1997). "Effect of speaking style on parameters of fundamental frequency contour." In vanSanten, J.P.H., R. Sproat, J. Olive, and J. Hirschberg (eds.). *Progress in speech synthesis*. NY: Springer-Verlag, 417-428.

Hirayama, Manami (2009). *Postlexical prosodic structure and vowel devoicing in Japanese*. Ph D thesis. University of Toronto.

Hirschberg, J. (2000). "A Corpus-based approach to the study of speaking style." In M. Horne (ed.) *Prosody, theory and experiment: Studies presented to Güsta Bruce.* Dordrecht: Kluwer Academic Publisher.

Kondo, Mariko (1997). *Mechanism of vowel devoicing in Japanese*. PhD Thesis. Faculty of Arts, the University of Edinburgh.

Kondo, Mariko (2005). "Syllable structure and its acoustic effects on vowels in devoicing environments." In In J. van de Weijer, K. Nanjo, and T. Nishihara,(eds.), *Voicing in Japanese*, Mouton de Gruyter, 229-245.

Labov, William. (1972). *Sociolinguistic patterns*. University of Pennsylvania Press.

Labov, William (2001). *Principles of linguistic change: Social factors*. Massachusetts: Blackwell.

Liberman, Mark and Janet Pierrehumbert (1984). "Intonational invariance under change in pitch range and length." In M. Aronoff and R. Oerhle (eds.) *Language sound structure*. Cambrige, MA: MIT Press.

Lindblom, Björn (1998). "Explaining phonetic variation: A sketch of H&H theory." W.J. Hardcastle and A. Marchal (eds.) *Speech production and speech modeling*. Dordrecht: Kluwer Academic Publishers.

Llisterri, J. (1992). "Speaking styles in speech research." *Proc. ELSNET/ESCA/SALT Workshop on integrating speech and natural language*, Dublin.

Maeda, Shinji (1976). *A characterization of American English intonation*. PhD thesis, MIT.

Maekawa, Kikuo (1990). "Production and perception of the accent in the consecutively devoiced syllables in Tokyo Japanese." *Proceedings of International Conference on Spoken Language Processing (ICSLP)*, Kobe, 517-520.

Maekawa, Kikuo (2003). "Corpus of Spontaneous Japanese: Its Design and Evaluation." *Proceedings of ISCA and IEEE Workshop on Spontaneous Speech Processing and Recognition (SSPR2003)*, Tokyo, 7-12.

Maekawa, Kikuo (2004a). “Production and Perception of 'Paralinguistic' Information”, *Proceedings of International Conference: Speech Prosody 2004*, Nara, 367-374.

Maekawa, Kikuo (2004b). “Design, Compilation, and Some Preliminary Analyses of the Corpus of Spontaneous Japanese.” In K. Yoneyama and K. Maekawa (eds.) *Spontaneous Speech: Data and Analysis,* Tokyo: The National Institute for Japanese Language, 87-108.

Maekawa, Kikuo (2005a).”Quantitative anslsysis of word-form variation using a spontaneous speech corpus.” *Proceedings of the Corpus Linguistics 2005*, Birmingham.

Maekawa, Kikuo (2005b). “Toward a Pronunciation Dictionary of Japanese: Analysis of CSJ.” *Proceedings of Symposium on Large-Scale Knowledge Resources (LKR2005)*, Tokyo Institute of Technology 21st Century COE Program, 43-48.

Maekawa, Kikuo (2009). “Analysis of Language Variation Using a Large-Scale Corpus of Spontaneous Speech.” In Shu-Chuan Tseng (ed.) *Linguistic Patterns in Spontaneous Speech*, Institute of Linguistics, Academia Sinica, Taipei, 27-50.

Maekawa, Kikuo (2010a). “Coarticulatoryreinterpretationofallophonicvariation:Corpus-based analysis of/z/inspontaneousJapanese.” *Journal of Phonetics*, 38 (3), 160-174.

Maekawa, Kikuo (2010b). “Final lowering and boundary pitch movements in spontaneous Japanese." *Proceedings of DiSS-LPSS Joint Workshop 2010*, Tokyo, 47-50.

Maekawa, Kikuo (Submitted). “Prosodic registers in spontaneous speech.”

Maekawa, Kikuo, Hanae Koiso, Sadaoki Furui and Hitoshi Isahara (2000). “Spontaneous speech corpus of Japanese.” *Proceedings of the Second International Conference of Language Resources and Evaluation (LREC2000)*, Athens, 947-952.

Maekawa, Kikuo, Hanae Koiso, Hideaki Kikuchi and Kiyoko Yoneyama (2003). “Use of a large-scale spontaneous speech corpus in the study of linguistic variation.” *Proceedings of the 15th International Congress of Phonetic Sciences (ICPhS 2003)*, Barcelona, 643-646.

Maekawa, Kikuo, Hideaki Kikuchi, Yosuke Igarashi, and Jennifer Venditti (2002). “X-JToBI: An extended J_ToBI for spontaneous speech.” *Proceedings of 7th International Conference on Spoken Language Processing (ICSLP2002)*, Denver, 1545-1548.

Maekawa, Kikuo and Hideaki Kikuchi (2005). “Corpus-based analysis of vowel devoicing in spontaneous Japanese: An interim report.” In J. van de Weijer, K. Nanjo, and T. Nishihara (eds.), *Voicing in Japanese*, Mouton de Gruyter, 205-228.

Nakamura, Masanobu, Koji Iwano and Sadaoki Furui. (2007). “Differences between acoustic characteristics of spontaneous and read speech and their effects on speech recognition performance.” *Computer Speech and Language*, 22 (2), 171-184.

Nevalainen, T. (1992). “Intonation and discourse type.” *Text*, 12 (3), 397-427.

Pierrehumbert, J. B. (1980). *The phonology and phonetics of English intonation*. PhD thesis, MIT.

Pierrehumbert, J. B (1981). "Synthesizing Intonation." *Journal of the Acoustical Society of America*, 70, 985-995.

Pierrehumbert, Janet and Mary E. Beckman (1988). *Japanese Tone Structure*. Cambridge: MIT Press.

Poser, William J. (1984). *The phonetics and pholology of tone and intonation in Japanese*, PhD Thesis, MIT.

Sagisaka, Y., K. Takeda, M. Abe, S. Katagiri, T. Umeda and H. Kuwabara (1990). "A large-scale Japanese speech database." *Proceedings of the International Congress on Spoken Language Processing (ICSLP)*, Kobe, 1089-1092.

Savino, Michelina, Martine Grice, Barbara Gili Fivelac and Giovanna Marottad (2006). "Intonational cues to discourse structure in Bari and Pisa Italian: perceptual evidence." *Proc. Speech Prosody 2006*, PS8-9-87 (Dresden).

Shinozaki, T., M. Ostendorf, and L.Atlas. (2009). Characteristics of speaking style and implications for speech recognition. *Journal of Acoustical Society of America*, 126 (3), 1500-1510.

Silverman, K., M. Beckman, J. Pitrelli, M. Ostendorf, J. Pierrehumbert, J. Hirschberg, and P. Price (1992). TOBI: A Standard Scheme for Labeling Prosody, *Proc. ISCLP92*, Banff, 867-870.

Stubbs, M. (1996). *Text and corpus analysis*. Oxford: Blackwell.

Svartvik, J. and Randolph Quirk (1980). *A Corpus of English Conversation*. Liber Läochromedel, Lund.

't Hart, Johan, René Collier and A. Cohen (1990). *A perceptual study of intonation: an experimental-phonetic approach*. Cambridge University Press.

Trask, R. L. *A dictionary of phonetics and phonology* (Routledge, London 1996).

Tsuchida, Ayako (1997). *Phonetics and phonology of Japanese vowel devoicing*. PhD Dissertation, Cornell University.

Umeda, Noriko (1982). "F0 declination" is situation dependent." *Journal of Phonetics*. 10, 279-290.

Vance, Timothy, J. (2008). *The sounds of Japanese*. Cambridge University Press.

Venditti, Jennifer (1995=1997). "Japanese ToBI Labelling Guidelines." In K. Ainsworth–Darnell M. D'Imperio (eds.) *Papers from the Linguistics Laboratory*, Ohio State University Working Papers in Linguistics 50, 127–162. [First distributed in 1995 at a web document].

Venditti, Jennifer (2000). *Discourse structure and attentional salience effects on Japanese intonation*. PhD thesis, The Ohio State University.

Venditti, J. J., Maeda, K., and van Santen, J. P. H. (1998). "Modeling Japanese boundary pitch movements for speech synthesis." In *Proceedings of the 3rd ESCA Workshop on Speech Synthesis*, Jenolan Caves.

Venditti, J.J., K. Maekawa and M. E. Beckman (2008). "Prominence marking in the Japanese intonation system." In S. Miyagawa and M. Saito (eds.) *The Oxford Handbook of Japanese Linguistics*, Oxford University Press, 456-512.

Watanabe, Michiko (2009). *Features and Roles of Filled Pauses in Speech Communication—A corpus based study of spontaneous speech.* Tokyo: Hituzi.

Weintraub, M., K. Taussig, K. Hunicke-Smith, and A. Snodgrass. (1996). Effect of speaking style on LVCSR performance. *Proc. ICSLP 1996*, Philadelphia, 16-19.

Yaeger-Dror, M. (1996). "Intonation and register variation: The case of the English negative." In Arnold, J., Blake, R., Davidson, B., Schwenter, S., Solomon, R. (eds.) *Sociolinguistic variation: Data, theory, and analysis*. CSLI, Standord, 243-260.

Yaeger-Dror, M. (2002). "Register and prosodic variation, a cross language comparison." *Journal of Pragmatics*, 34, 1495-1536.

秋永一枝(2001).『新明解日本語アクセント辞典』三省堂.

青木繁伸(2009).『R による統計解析』オーム社.

秋田祐哉, 河原達也(2005).「話し言葉音声認識のための汎用的な統計的発音変動モデル」電子情報通信学会論文誌, Vol. J88-D-II, No. 9, 1780-1789.

秋田祐哉, 三村正人, 河原達也 (2010).「会議録作成支援のための国会審議の音声認識システム」電子情報通信学会論文誌, Vol.J93-D, No.9, 1736-1744.

秋葉友良, 相川清明, 伊藤慶明, 河原達也, 南條浩輝, 西崎博光, 安田宜仁, 山下洋一, 松井知子, 胡新輝, 中川聖一, 伊藤克亘(2008).「SLP 音声ドキュメント処理ワーキンググループ活動報告」情報処理学会研究報告. SLP, 音声言語情報処理, 115-120.

天沼寧・大坪一夫・水谷修(1978).『日本語音声学』くろしお出版.

有坂秀世(1940).『音韻論』三省堂.

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(2006)所載, 347-453 (http://www.ninjal.ac.jp/products-k/katsudo/seika/corpus/csj_report/).

伊藤 慶明, 堀 貴明(2009).「音声認識の応用システム : 音声ドキュメント検索・音声翻訳・音声対話の新たな展開」日本音響学会誌, 66(1), 36-40.

上野善道(1977).「日本語のアクセント」『岩波講座日本語５音韻』 282-321, 岩波書店.

大石初太郎(1959).「プロミネンスについて―東京語の観察にもとづく覚え書―」『ことばの研究』(国立国語研究所論集 1), 87-102.

奥田浩三, 河原達也, 中村哲(2001).講演音声認識における発話速度の変動を考慮した音声認識手法, 電子情報通信学会技術報告, NLC201-60/SP2001-104, 13-18.

小椋秀樹(2006).「形態論情報」国立国語研究所(2006)所載, 133-186 (http://www.ninjal.ac.jp/products-k/katsudo/seika/corpus/csj_report/).

籠宮隆之, 山住賢司, 槙洋一, 前川喜久雄(2007).「聴取実験に基づく講演音声の印象評定データの構築とその分析」社会言語科学, 9 (2), 65-76.

片桐恭弘(1997).「終助詞とイントネーション」音声文法研究会編『文法と音声』くろしお出版, 235-256.

川上蓁(1953).「「花高し」と「鼻高し」－東京アクセント段階観の限界―」音声学会会報, 82, 6-9.
川上蓁(1957a).「東京語の卓立強調の音調」国語研究, 6, 21-31.
川上蓁(1957b).「準アクセントについて」国語研究, 7, 4-60.
川上蓁(1963).「文末などの上昇調について」『国語研究』16, 25–46.
川上蓁(1977) .『日本語音声概説』桜楓社.
河津宏美,前川喜久雄(2009).「子音の調音様式が母音の無声化におよぼす影響 ―『日本語話し言葉コーパス』の分析―」日本音響学会 2009 年春季研究発表会講演論文集, 443-444.
菊池英明, 前川喜久雄(2007).「韻律研究のための日本語話し言葉コーパス XML 文書作成」人工知能学会研究会資料 SIG-SLUD-A603, 3-8.
菊池英明, 前川喜久雄, 五十嵐陽介, 米山聖子, 藤本雅子(2003).「日本語話し言葉コーパスの音声ラベリング」音声研究, 7 (3), 16-26.
金田一春彦(1965).「丁寧な発音の弁」国語国文, 34 (2), 15-35.
金田一春彦(1966).『新日本語論―私の現代語教室』筑摩書房.
小磯花絵, 西川賢哉, 間淵洋子(2006).「転記テキスト」国立国語研究所(2006)所載, 23-132 (http://www.ninjal.ac.jp/products-k/katsudo/seika/corpus/csj_report/).
郡史郎(1997).「日本語のイントネーション －型と機能－」『日本語音声 2 アクセント・イントネーション・リズムとポーズ』三省堂.
郡史郎(2006).「日本語の「口調」にはどんな種類があるか」音声研究, 19 (3), 52-68.
国立国語研究所(1955).『談話語の実態』秀英出版.
国立国語研究所(1960).『話しことばの文型(1)』秀英出版.
国立国語研究所(1963).『話しことばの文型(2)』秀英出版.
国立国語研究所(2006).『日本語話し言葉コーパスの構築』国立国語研究所.
小山哲春(1997).「文末詞と文末イントネーション」音声文法研究会編『文法と音声』くろしお出版, 97-119
佐久間鼎(1929).『日本音聲學』京文社.
佐藤大和(1987).『規則による音声合成の研究』北海道大学学位論文.
武田一哉, 桑原尚夫(1987).「母音無声化の要因分析と予測手法の検討」日本音響学会 1987 年度秋季研究発表会講演論文集, 105-106.
轟木靖子(2008).「東京語の終助詞の音調と機能の対応について―内省による考察―」音声言語, 5, 5-28.
中川聖一(2009).「特集「音声ドキュメント処理」の編集にあたって」情報処理学会誌論文誌, 50 (2).
南條浩輝, 加藤一臣, 李晃伸, 河原達也(2003).「大規模な日本語話し言葉データベースを用いた講演音声認識」電子情報通信学会論文誌, Vol.J86-DII, No.4, 450-459.
服部四郎(1951).『音聲學』岩波書店.

服部四郎(1973).「アクセント素とは何か？そしてその弁別特徴とは？」『言語の科学』4, 1-61.

文化庁(2001).『平成 12 年度国語に関する世論調査〔平成 13 年 1 月調査〕』文化庁文化部国語課.

藤崎博也, 須藤寛(1971).「日本語単語アクセントの基本周波数パタンとその生成機構のモデル」日本音響学会誌, 27 (9), 445-453.

藤本雅子, 菊池英明, 前川喜久雄(2006).「分節音情報」国立国語研究所(2006)所載, 323-346 (http://www.ninjal.ac.jp/products-k/katsudo/seika/corpus/csj_report/).

藤本雅子, 桐谷滋(2003).「東京方言と近畿方言における母音の無声化の比較」音声研究, 7 (1), 58-69.

古井貞熙, 前川喜久雄, 井佐原均(2000).「科学技術振興調整費開放的融合研究制度：大規模コーパスに基づく『話し言葉工学』の構築」日本音響学会誌, 56(11), 752-755.

前川喜久雄(1989).「母音の無声化」講座日本語と日本語教育２ 日本語の音声・音韻, 明治書院, 135-153.

前川喜久雄(1999).「韻律とコミュニケーション」日本音響学会誌, 55, 119-125.

前川喜久雄(2001).「言語研究における自発音声」日本音響学会 2001 年春季研究発表会講演論文集, 19-22.

前川喜久雄(2002).「『日本語話し言葉コーパス』を用いた言語変異研究」 音声研究, 6(3), 48-59.

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, 111-133.

前川喜久雄(2005).「自発音声とデータベース」日本音響学会誌,61(9), 544-549.

前川喜久雄(2006).「日本語発音辞書の改良」土岐哲先生還暦記念論文集編集委員会編『日本語の教育から研究へ』くろしお出版, 119-132.

前川喜久雄(2007).「内省からコーパスへ」文部科学教育通信,169, 22-23.

前川喜久雄(2008).「話し言葉と書き言葉」日本語学, 27(5), 23-33, 2008.

前川喜久雄(2010).「日本語有声破裂音における閉鎖調音の弱化」音声研究, 14-2, 1-15.

前川喜久雄(投稿中).「PNLP の音声的形状と言語的機能」

前川喜久雄, 五十嵐陽介(2006).「２モーラ有核助詞の韻律上の独立性―『日本語話し言葉コーパス』の分析―」音声研究, 10 (2), 33-42.

前川喜久雄, 籠宮隆之, 小磯花絵, 小椋秀樹, 菊池英明(2000).「日本語話し言葉コーパスの設計」音声研究, 4 (2), 51-61.

前川喜久雄, 菊池英明(2007).「アクセント句を単位としてみた自発音声の韻律特徴―韻律境界強度の予備的分析―」第 21 回日本音声学会全国大会予稿集, 117-122.

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), 25-30.

前川喜久雄, 北川智利(2002).「音声はパラ言語情報をいかに伝えるか」 認知科学, 9(6), 46-66.

丸山岳彦, 谷口未希(2001).「文の焦点構造と局所的卓立」*KLS*, 22, 18-28.

丸山岳彦, 柏岡秀紀, 熊野正, 田中英樹(2003).「節境界自動検出ツールの作成と評価」言語処理学会第 9 回年次大会発表論文集, 517-520.

丸山岳彦, 高梨克也, 内元清貴(2006).「節単位情報」国立国語研究所(2006)所載, 255-322 (http://www.ninjal.ac.jp/products-k/katsudo/seika/corpus/csj_report/).

南不二男(1974).『現代日本語の構造』 大修館書店.

山住賢司, 籠宮隆之, 槙洋一, 前川喜久雄(2005).「講演音声の印象評価尺度」 日本音響学会誌, 61(6), 303-311.

山住賢司, 籠宮隆之, 槙洋一, 前川喜久雄(2007).「講演音声の音声的特徴とその印象に対する評価構造モデル」日本官能評価学会誌, 11(1), 30-36.

吉田夏也 (2002). 「音声環境が母音の無声化に与える影響について」国語学, 53 (3), 34-47.

吉田夏也・匂坂芳典(1990).「母音無声化の要因分析」ATR テクニカルレポート(TR-I-0159), 1-9.

発表文献

[1] Maekawa, Kikuo (1990). “Production and perception of the accent in the consecutively devoiced syllables in Tokyo Japanese.” *Proceedings of International Conference on Spoken Language Processing (ICSLP)*, Kobe, 517-520.

[2] Maekawa, Kikuo (2003). “Corpus of Spontaneous Japanese: Its Design and Evaluation.” *Proceedings of ISCA and IEEE Workshop on Spontaneous Speech Processing and Recognition (SSPR2003)*, Tokyo, 7-12. （招待講演）

[3] Maekawa, Kikuo (2004a). “Production and Perception of 'Paralinguistic' Information”, *Proceedings of International Conference: Speech Prosody 2004*, Nara, 367-374. （招待講演）

[4] Maekawa, Kikuo (2004b). “Design, Compilation, and Some Preliminary Analyses of the Corpus of Spontaneous Japanese.” In K. Yoneyama and K. Maekawa (eds.) *Spontaneous Speech: Data and Analysis,* Tokyo: The National Institute for Japanese Language, 87-108.

[5] Maekawa, Kikuo (2005). “Toward a Pronunciation Dictionary of Japanese: Analysis of CSJ.” *Proceedings of Symposium on Large-Scale Knowledge Resources (LKR2005)*, Tokyo Institute of Technology 21st Century COE Program, 43-48.

[6] Maekawa, Kikuo (2009). “Analysis of Language Variation Using a Large-Scale Corpus of Spontaneous Speech.” In Shu-Chuan Tseng (ed.) *Linguistic Patterns in Spontaneous Speech* (Language and Linguistic Monograph Series A25), Institute of Linguistics, Academia Sinica, Taipei, 27-50. （査読有）

[7] Maekawa, Kikuo (2010a). “Coarticulatory reinterpretation of allophonic variation: Corpus-based analysis of /z/ in spontaneous Japanese.” *Journal of Phonetics*, 38 (3), 160-174. （査読有）

[8] Maekawa, Kikuo (2010b). “Final lowering and boundary pitch movements in spontaneous Japanese." *Proceedings of DiSS-LPSS Joint Workshop 2010*, Tokyo, 47-50. （査読有）

[9] Maekawa, Kikuo (Submitted). “Prosodic registers in spontaneous speech.”

[10] Maekawa, Kikuo, Hanae Koiso, Sadaoki Furui and Hitoshi Isahara (2000). “Spontaneous speech corpus of Japanese.” *Proceedings of the Second International Conference of Language Resources and Evaluation (LREC2000)*, Athens, 947-952.

[11] Maekawa, Kikuo, Hanae Koiso, Hideaki Kikuchi and Kiyoko Yoneyama (2003). “Use of a large-scale spontaneous speech corpus in the study of linguistic variation.” *Proceedings of the 15th International Congress of Phonetic Sciences (ICPhS 2003)*, Barcelona, 643-646.（査読有）

[12] Maekawa, Kikuo, Hideaki Kikuchi, Yosuke Igarashi, and Jennifer Venditti (2002). “X-JToBI: An extended J_ToBI for spontaneous speech.” *Proceedings of 7th International Conference on Spoken Language Processing (ICSLP2002)*, Denver, 1545-1548. （査読有）

[13] Maekawa, Kikuo and Hideaki Kikuchi (2005). “Corpus-based analysis of vowel devoicing in spontaneous Japanese: An interim report.” In J. van de Weijer, K. Nanjo, and T. Nishihara,(eds.), *Voicing in Japanese*, Mouton de Gruyter, 205-228. （査読有）
[14] Venditti, J.J., Kikuo Maekawa and Mary Beckman (2008). “Prominence marking in the Japanese intonation system.” In S. Miyagawa and M. Saito (eds.) *The Oxford Handbook of Japanese Linguistics*, Oxford University Press, 456-512. （査読有）
[15] 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『日本語話し言葉コーパスの構築法』, 347-453.
[16] 籠宮隆之, 山住賢司, 槙洋一, 前川喜久雄(2007).「聴取実験に基づく講演音声の印象評定データの構築とその分析」社会言語科学, 9 (2), 65-76. （査読有）
[17] 河津宏美,前川喜久雄(2009).「子音の調音様式が母音の無声化におよぼす影響 ―『日本語話し言葉コーパス』の分析―」日本音響学会 2009 年春季研究発表会講演論文集, 443-444.
[18] 菊池英明, 前川喜久雄(2007).「韻律研究のための日本語話し言葉コーパス XML 文書作成」人工知能学会研究会資料 SIG-SLUD-A603, 3-8.
[19] 菊池英明, 前川喜久雄, 五十嵐陽介, 米山聖子, 藤本雅子(2003).「日本語話し言葉コーパスの音声ラベリング」音声研究, 7 (3), 16-26. （査読有）
[20] 小磯花絵, 土屋菜穂子, 間淵洋子, 斉藤美紀, 籠宮隆之, 菊地英明, 前川喜久雄(2001).「『日本語話し言葉コーパス』における書き起こしの方法とその基準について」日本語科学, 9, 43-58. （査読有）
[21] 藤本雅子, 菊池英明, 前川喜久雄(2006).「分節音情報」国立国語研究所『日本語話し言葉コーパスの構築法』, 323-346.
[22] 古井貞熙, 前川喜久雄, 井佐原均(2000).「科学技術振興調整費開放的融合研究制度：大規模コーパスに基づく『話し言葉工学』の構築」日本音響学会誌, 56(11), 752-755.
[23] 前川喜久雄(1989).「母音の無声化」講座日本語と日本語教育２ 日本語の音声・音韻, 明治書院, 135-153.
[24] 前川喜久雄(1999).「韻律とコミュニケーション」日本音響学会誌, 55, 119-125. （招待原稿）
[25] 前川喜久雄(2001).「言語研究における自発音声」日本音響学会 2001 年春季研究発表会講演論文集, 19-22.
[26] 前川喜久雄(2002).「『日本語話し言葉コーパス』を用いた言語変異研究」 音声研究, 6(3), 48-59. （査読有）
[27] 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, 111-133. （査読有）
[28] 前川喜久雄(2005).「自発音声とデータベース」日本音響学会誌,61(9), 544-549. （招待原稿）
[29] 前川喜久雄(2006).「日本語発音辞書の改良」土岐哲先生還暦記念論文集編集委員会編『日本語の教育から研究へ』くろしお出版, 119-132. （査読有）

[30] 前川喜久雄(2007).「内省からコーパスへ」文部科学教育通信,169, 22-23.
[31] 前川喜久雄(2008).「話し言葉と書き言葉」日本語学, 27(5), 23-33, 2008.
[32] 前川喜久雄(2010).「日本語有声破裂音における閉鎖調音の弱化」音声研究, 14-2, 1-15.（査読有）
[33] 前川喜久雄(投稿中).「PNLP の音声的形状と言語的機能」
[34] 前川喜久雄, 五十嵐陽介(2006).「２モーラ有核助詞の韻律上の独立性―『日本語話し言葉コーパス』の分析―」音声研究, 10 (2), 33-42.（査読有）
[35] 前川喜久雄, 籠宮隆之, 小磯花絵, 小椋秀樹, 菊池英明(2000).「日本語話し言葉コーパスの設計」音声研究, 4 (2), 51-61.（招待原稿）
[36] 前川喜久雄, 菊池英明(2007).「アクセント句を単位としてみた自発音声の韻律特徴―韻律境界強度の予備的分析―」第 21 回日本音声学会全国大会予稿集, 117-122.
[37] 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), 25-30.
[38] 前川喜久雄, 北川智利(2002).「音声はパラ言語情報をいかに伝えるか」 認知科学, 9(6), 46-66.（査読有）
[39] 山住賢司, 籠宮隆之, 槙洋一, 前川喜久雄(2005).「講演音声の印象評価尺度」 日本音響学会誌, 61(6), 303-311.（査読有）
[40] 山住賢司, 籠宮隆之, 槙洋一, 前川喜久雄(2007).「講演音声の音声的特徴とその印象に対する評価構造モデル」日本官能評価学会誌, 11(1), 30-36.（査読有）