

機械学習を用いた日本語アクセント型の分類 — 母語話者と学習者による単語発話と朗読発話の比較 —

波多野 博顕 (国際交流基金日本語国際センター) アルビン エレン・王睿来 (神戸大学)
石井カルロス寿憲 (株式会社国際電気通信基礎技術研究所)
Hiroaki_Hatano@jpf.go.jp

1. 背景と先行研究

日本語教育では音声評価に自信のない教師が多いことを背景に、学習者が音声教育を受ける機会は少ない。海外のノンネイティブ (NNS) 日本語教師の多くは、「自分が知らない」「自信が無い」という理由から、授業で音声を扱っていない (磯村 2001)。しかし、自然な発音で話したいという学習者のニーズは高い (佐藤 1998)。また、ビジネスや国内長期生活では発音が問題となり (小河原 2001)、評価には韻律が与える影響が大きい (佐藤 1995)。ところが、国内のネイティブ (NS) 日本語教師であっても、単音等と比較して韻律 (アクセント) は音声教育項目としての重要度の認識が相対的に低い (轟木・山下 2009)。このような背景から、自学自習が可能な音声自動評価法の開発が求められている (松崎 2016)。

日本語アクセント (共通語) は単語や文節にかかる韻律特徴であるため、発話音声全体の自然性に大きく影響する。アクセントに注目した研究は多いものの、音響特徴から型の推定を行う研究の多くは 80% 以下の精度に留まっている (石井・他 2001、広瀬 2005、波多野・他 2014、Hatano, et al. 2018)。先行研究では 1 つの特徴量のみを推定に用いているが、実際の発話では「おそ下がり」(杉藤 1969) が生じることもあり、特に連続発話では基本周波数 (fo) の動態が不安定な場合が多い。

本研究では、観測した fo に基づいて全体の軌跡を再構築し、そこから抽出した複数の特徴量によって機械学習を行うことで、従来よりも高精度なアクセント型の推定を試みた。

2. データの概要と分析方法

2.1. データの概要

分析には、NS および NNS による単語発話と朗読発話を用いた。NS だけでなく NNS の発話も対象としたのは、音響的に様々な程度で NS と異なる音声に対しても、本手法が頑健かどうか検討するためである。また、対象を単独で発話した単語発話では明瞭な fo が期待できる一方、朗読発話は連続発話中の要素を対象とするためイントネーション成分の影響が相対的に強く表れる。後者も分析することで、本手法の有効性を検証することができると考えた。なお、本研究では分析対象を 3 モーラに限定した。

2.1.1. 単語発話データ

NS による単語発話は、Albin (2017) のデータを用いた。3 名の NS (女性) が 160 語の 3 モーラ無意味語を単独で読み上げている (3 名 * 160 語 = 480 語)。無意味語は 87.5% が CVCVCV の音節構造 (長音、促音、撥音はなし) で、分節音の無声/有声阻害音や共鳴音

の比率および読み上げる際のアクセント型がほぼ均等になるように構成されている。NSは設定されたアクセント型通りに読み上げるよう指示され、異なった場合は再度発話した。

NNSによる単語発話は、王・他（2018）のデータを用いた。52名の初級中国人日本語学習者（男性9・女性43、学習歴約6ヶ月）が、2モーラ有意味語+助詞「が」を単独で9つ読み上げている（52名 * 9語 = 468語、本研究ではこれを単語発話に含める）。有意味語は母音の無声化や特殊拍を避けたものが選定された。学習者が読み上げる際にはアクセント核の位置に記号が付けられた状態で呈示されたが、それと異なる型で発話されても修正されなかった。各読み上げ音声のアクセント型は、2名の評価者（NSとNNS）が全て聴取して判定し、一致しなかった場合はもう1名の評価者（NS）が聴取して最終的なアクセント型を決定した。全員の評価者が日本語音声学の知識を有した日本語教育の経験者である。

2.1.2. 朗読発話データ

『日本語学習者による日本語／母語発話の対照言語データベース モニター版（2005）』のDVDに収録されている朗読発話を用いた。日本語・中国語・韓国語・タイ語の母語話者各10名（うち男性は日1、中3、韓2、泰1）が朗読する課題のうち、「コマ」（17文633モーラ）と「タバコ」（11文592モーラ）を用いた。各NNSの平均日本語学習月数（sd）は、中67.1（30.9）、韓43.5（32.9）、泰34.8（9.2）である。テキストから、「1つの自立語に0以上の付属語が後続する」という条件で文節境界を定め、3モーラの文節を計24抽出した（「コマ」18、「タバコ」6）。ただし、言い直しやポーズの挿入による分割などの影響で、全員が同じ文節数ではない。単語発話と揃えるため、特殊拍が含まれる文節は除外した。最終的に、全体で1,152文節が得られた（日250、中273、韓295、泰334）。全文節を1名のNS（日本語音声学の知識がある日本語教育経験者）が聴取し、アクセント型を判定した。

全2,100データ（単語発話968、朗読発話1,152）のアクセント型分布を表1にまとめる。なお、朗読発話のNNSには3つの母語が含まれるが、全てまとめて分析を行った。

表1 単語発話と朗読発話におけるNSとNNSのアクセント型分布

アクセント 型	単語発話		朗読発話	
	NS	NNS	NS	NNS
0型	168	177	89	319
1型	156	143	111	335
2型	156	148	50	248
計	480	468	250	902

2.2. 分析方法

音素単位でfoの代表値を求めるため、まず、各音素のアライメントとfoの抽出を行った。その後、fo抽出の欠損や不安定さに対処するため、fo形状のモデリング（再構築）を行った。それに基づいてアクセント型を判定するための特徴量を抽出し、機械学習を行った。

2.2.1. 音素のアライメントと fo の抽出

NS の単語発話データは既に音素境界のアライメントが行われていたので、本分析ではそれを利用した。それ以外のデータには、音声認識エンジン Julius (ver. 4.4.2) の音素セグメンテーションキット (ver. 4.3.1) を使用して、音素境界のアライメントを行った。言語モデルと音響モデルは、キットに含まれているデフォルトのものを使用した。朗読音声では、アライメント情報に基づいて、3 モーラ文節部分の切り出しも行った。

全 2,100 の音声ファイルに対し、Praat (ver. 6.0.36) の To Pitch コマンドを用いて fo を抽出した。抽出パラメータは、Time step を 0.001 sec、Pitch floor を 75 Hz、Pitch ceiling を 600 Hz に設定し、Kill octave jumps の機能を使用した。

2.2.2. fo 形状のモデリング

まず、各音素区間の fo と時間情報から、音素ごとに回帰直線を計算した。ただし、fo 抽出の安定性を考慮して、対象は母音/a, i, u, e, o/、鼻音/m, n/、わたり音/w, y/、撥音/N/とした。これらの音素であっても、区間全体で fo 抽出されたフレームが 20 %以下であった場合は、信頼性の観点から回帰直線を計算しなかった。これらの条件は Albin (2017) を参考にした。さらに、区間内の fo から四分位数を求め、「第 1 四分位数 * 1.5 以上」「第 3 四分位数 * 1.5 以下」に該当した fo は外れ値として、回帰直線の計算に含めなかった。

次に、算出した回帰直線の時間方向で 25 % と 75 %の地点に制御点を設定し（先頭音素は 0 %、末尾音素は 100 %の地点にも）、それらを線形補間することで全体の fo 形状を再構築した。なお、末尾の母音音素が無声化などによって回帰直線が計算できなかった場合、直近の回帰直線の 100 %地点の fo を、末尾母音音素の 100 %地点まで延伸させた (図 1 a)。また、朗読発話で頻出した「末尾上げ」に対処するため、最終音素で回帰直線の傾きが正であった場合は、その 25 %地点における fo を 100 %地点まで延伸させた (図 1 b)。このように fo をモデリングすることで、fo 抽出の欠損や不安定さに対処した。

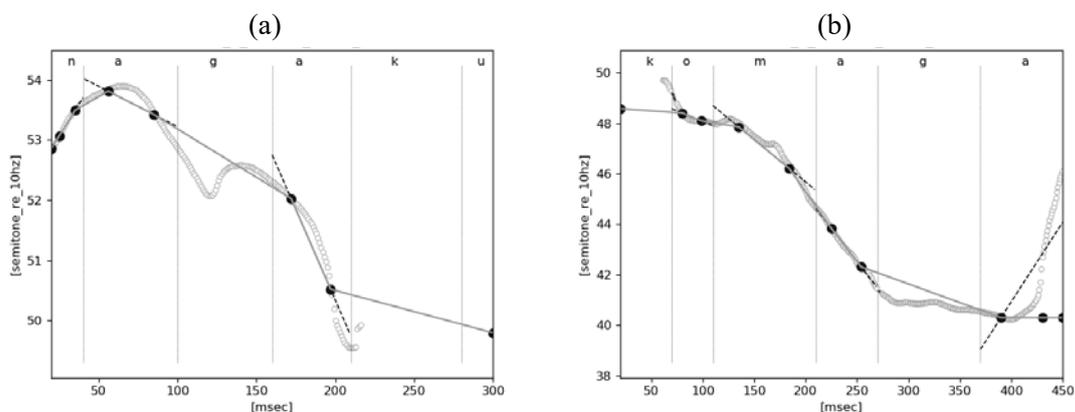


図 1 fo 形状のモデリング概要 (いずれも朗読発話より、(a)「長く」と(b)「コマが」)

白抜き円は fo の値を、点線は回帰直線を、黒丸は制御点を、実線は線形補間の結果を示す。

2.2.3. 特徴量の抽出

5 種類・7 項目の特徴量を抽出した。線形補間した fo 軌跡からは、各母音音素の 50 %地点の fo (=中央値) から求めた「V₁V₂」「V₂V₃」の差、および、100 %地点の fo (=終端値) から求めた「V₁V₂」「V₂V₃」差を計算した (V = 母音, semitone)。その他、対象単位冒頭の fo 値 (Hz)、対象単位全体の fo レンジ (semitone)、発話速度 (msec) も求めた。

母音音素の fo 中央値の差分は、波多野・他 (2014) のアクセント型判定において高い推定精度を示したことから、本研究でも特徴量として採用した。fo 終端値の差分は、おそ下がりに対処するための特徴量として採用した。また、冒頭 fo 値は性差を、fo レンジは学習者母語の特性を、発話速度は学習者の習熟度を考慮するため、特徴量に含めた。

2.2.4. 機械学習の実行

機械学習には、Python (ver. 3.6.5) の Scikit-learn (ver.0.19.1) パッケージを使用した。機械学習に用いる推定器は、正解ラベル (今回はアクセント型) に基づいた「分類」であるという点や、データ数を考慮し、SVC (Support Vector Classification) とした。

データ全体を、分類モデルを学習するための「トレーニングセット」と、そのモデルの精度を検証するための「テストセット」に分割し (割合は 8 対 2)、層化 k 分割交差検証 (stratified k-fold cross validation) を行った。k 分割交差検証では、トレーニングセット全体を非復元抽出で k 個のサブセットに分割した上で、k - 1 個のサブセットをモデルのトレーニングに使用し、残りのサブセットでそのモデルの検証を行うことを k 回繰り返すことで、最終的な分類モデルを学習する (Raschka 2015)。分析データはアクセント型の分布が不均等なため (表 1)、アクセント型で層化抽出を行うことで、各サブセットにおける型の比率が維持されるようにした。今回は k = 10 とし、10 回の交差検証で推定平均性能を計算した後、トレーニングセットからは独立したテストセットで最終的な分類性能の評価を行った。

なお、SVC を行う際、各特徴量の値は標準化 (平均値 0、標準偏差 1) し、ハイパーパラメータはグリッドサーチによって最適な組み合わせを選択した。

3. 結果と考察

3.1.1. 単語発話データの分類結果

表 2 に、NS および NNS による単語発話データの分析結果を示す。いずれもテストセットの分類結果が 90 %を超えており、NS では 100 %となった。これらはアクセントを意識した発話のため、各母音間の fo 差が比較的明瞭であった可能性がある。

表 2 単語発話データにおけるアクセント型の分類結果

データ	n	トレーニングセット		テストセット	
		N	平均性能 (sd)	N	分類結果
NS	480	384	99.7 % (0.008)	96	100.0 %
NNS	468	374	94.1 % (0.023)	94	92.6 %

3.1.2. 朗読発話データの分類結果

表 3 に、NS および NNS による朗読発話データの分析結果を示す。NNS ではテストセットの分類結果が 80 %を超えたものの、NS では 76 %に留まった。

単語発話と比較して低い精度に留まったのは、1 名の評価者によるアクセント型判定の信頼性が影響している可能性がある。今後は評価者を増やす等をして正解ラベルの信頼性を高める必要がある。また、NS の分類結果が NNS よりも悪かったのは、過小なサンプル数で十分な分類モデルの学習ができなかったことによると考えられる。

表 3 単語発話データにおけるアクセント型の分類結果

データ	n	トレーニングセット		テストセット	
		N	平均性能 (sd)	n	分類結果
NS	250	200	83.0 % (0.085)	50	76.0 %
NNS	902	721	82.1 % (0.029)	181	84.0 %

3.1.3. アクセント型ごとの分類性能

単語・朗読発話のテストセットにおけるアクセント型ごとの分類結果について、正解ラベルとの混同行列を図 2 に示す。図 2 から、朗読発話で 2 型の分類精度が良くないことがわかる (NS 50.0 %、NNS 71.4 %)。

朗読発話では、イントネーション成分の影響で 2 型の特徴である V_2V_3 の fo 差が明瞭に現れなかった可能性がある。また、朗読発話では 2 型の出現数がそもそも少ない (表 1)。トレーニングセットに分割した際には更に少なくなるため、2 型で特に精度が悪かった原因として以上が考えられる。

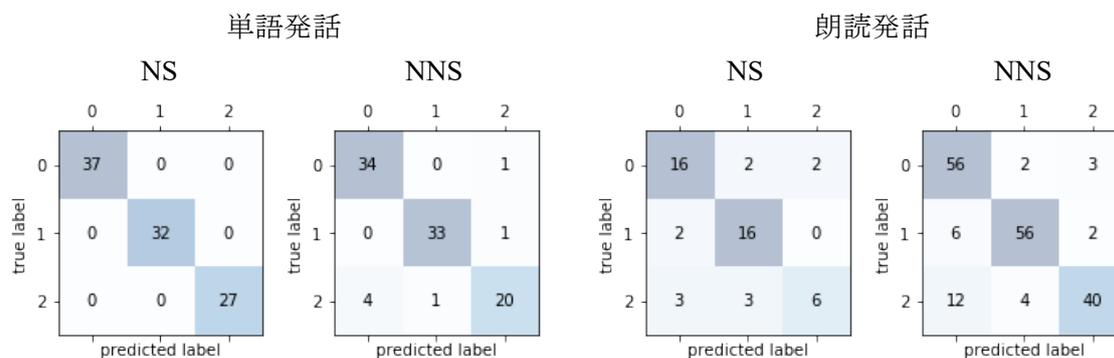


図 2 単語発話と朗読発話のテストセットにおけるアクセント型の混同行列

true label は正解のアクセント型を、predicted label は機械学習によるアクセント型の分類を示す

4. まとめ

機械学習を用いた本研究の手法は、単語発話では極めて高い精度でアクセント型分類が可能であり、朗読発話ではデータ拡充により更なる精度向上が期待できることが示された。今後は更にデータを増やすとともに、本研究で明らかになった問題を改善していきたい。

謝辞

本研究の単語発話データで使用した王・他（2018）は JSPS 科研費 17H02352 の助成を受けたものである。また、朗読発話データで使用した DVD は JSPS 科研費 14380121 の研究成果報告書の付録である。

参考文献

- Albin, Aaron (2017) “F0 Contour Parameterization Using Optimal Regression Chains” 『第 31 回 日本音声学会全国大会予稿集』 79-84.
- Hatano, Hiroaki., Ishi, Carlos Toshinori., Song, Cheng Chao., Matsuda, Makiko (2018) “Automatic evaluation of accentuation of Japanese read speech,” In M. Ueyama and I. Srdanović (eds.) *Digital resources for learning Japanese*, 55-71, Bologna: Bononia University Press.
- Raschka, Sebastian (2015) *Python Machine Learning*. Packt Publishing.(株式会社クイープ (訳) (2016) 『Python 機械学習プログラミング』 株式会社インプレス.)
- 石井カルロス寿憲・西山隆二・峯松信明・広瀬啓吉 (2001) 「日本語のアクセント・イントネーションを対象とした発音教育システム構築に関する検討」 『信学技報』 100(594), 33-40.
- 磯村一弘 (2001) 「海外における日本語アクセント教育の現状」 『日本語教育学会秋季大会予稿集』 211-212.
- 王睿来・林良子・磯村一弘・新井潤 (2018) 「中国語母語話者による日本語アクセントの習得—知覚と生成の関係に着目して—」 『ことばの科学研究』 19, 81-96.
- 小河原義朗 (2001) 「日本語非母語話者の話す日本語の発音に対する日本人の評価意識：社会人の場合」 『日本語教育方法研究会誌』 8(2), 10-11.
- 佐藤友則 (1995) 「単音と韻律が日本語音声の評価に与える影響力の比較」 『世界の日本語教育』 5, 139-154.
- 佐藤友則 (1998) 「韓国および台湾の日本語学習者のニーズ調査」 『言語科学論集』 2, 49-60.
- 杉藤美代子 (1969) 「動態測定による日本語アクセントの解明」 『言語研究』 55, 14-39.
- 轟木靖子・山下直子 (2009) 「日本語学習者に対する音声教育についての考え方：教師への質問紙調査より」 『香川大学教育実践総合研究』 18, 45-51.
- 波多野博頭・石井カルロス寿憲・松田真希子 (2014) 「日本語朗読音声を対象にしたアクセント型自動判定方法の検討」 『日本音響学会秋季研究発表会講演論文集』 363-364.
- 広瀬啓吉 (2005) 「音声の韻律と CALL」 『音声研究』 9(2), 38-46.
- 松崎寛 (2016) 「日本語音声教育における韻律指導」 『日本音響学会誌』 72(4), 213-220.
- 『日本語学習者による日本語発話と母語発話との対照データベース—開発・応用のための研究—』 (平成 14-16 年度科学研究費補助金 基板研究 (B) (2) 研究成果報告書、研究代表者：宇佐美洋)

母語話者シャドーイングとそれに基づく「聞き取り易さ」の客観的計測

峯松信明・井上雄介・椛島優・齋藤大輔（東京大学大学院）

金村久美（名古屋経済大学）山内豊（創価大学）

{mine, inoue0124, kabashima, dsk_saito}@gavo.t.u-tokyo.ac.jp,
kanamura@nagoya-ku.ac.jp, yutaka@soka.ac.jp

1 はじめに

外国語音声学習の目標は、多くの場合（母語話者のような発音ではなく）、実用的に十分伝わる発音の習得であると言える。音声技術を用いて学習者音声を評価する場合、母語話者の発音モデルとの比較を通してスコア付けをすることが一般的であり（河原・峯松, 2013）、原理的に、母語話者の発音に近いほど高得点となる。本研究では、可解性（comprehensibility）の高低を技術的に自動計測するための手段として、母語話者による学習者音声シャドーイングを提案し、母語話者によるシャドーイング音声を用いた可解性の自動予測に関して、その妥当性を実験的に検証した。

2 了解性・可解性と外国語訛り

応用言語学では、学習者音声の中の個々の単語が明確に聞き取れるか否かに着目した尺度を了解性（intelligibility）と呼び、例えば、母語話者の書き取りテストで定量化される。一方、その音声の意味理解が容易か否か（理解するための認知タスクが低いかな）に着目した尺度を可解性と呼んでおり、主観的な定量化や、理解度テストにより定量化される（Derwing & Munro, 2015）。可解性は、各単語の知覚のみならず、統語・意味・談話解析が容易か否かも考慮され、学習者音声に対するより総合的な評価基準と言える。外国語訛りが了解性や可解性に与える影響について多くの先行研究があり、母語話者発音からのずれがあっても十分了解性・可解性は高いが（Munro & Derwing, 1995）、聴取者がその訛りに慣れていない場合は、低い了解性となる（Mineamtsu et al., 2011）。

了解性・可解性に基づく発音評価を技術的に実装する場合、母語話者発音モデルとの近接性ではなく、聴取者の許容量（に相当するもの）を計測・モデル化する必要がある。これは、学習者発音に対する調音的・音響的分析では計測できず、聴取者の認知プロセスが計測の対象となり、客観的計測そのものが困難となる。筆者らの知る限り、特に可解性に対して、聴取者の許容量を、客観的かつ容易な計測方法を通して検討した研究例はない。筆者らは予備的検討として、ストレス感の自動計測を謳っている、「感性アナライザ」（種々の機能が付与された簡易脳波計）（電通, 2016）を試用した。強い外国語訛りの音声や、雑音下音声を用いた意味理解テストを構成し、聴取者の様子を計測したが、音声以外の刺激（視覚刺激など）によってもセンサー値が変動し、安定した計測は困難であった。

3 了解性の客観的な計測と可解性計測へのアプローチ

了解性（可解性ではない）を客観的に計測する場合、学習者音声を母語話者に提示し、書き取らせる／復唱させるなどして単語単位の正解率を見ることが多い。例えば、Mineamtsu et al. (2011) では、日本人の読上げ英語音声を対象とした了解性の客観的測定が行なわれている。この場合、学習者音声を提示し、書き取らせる／復唱させる訳だが、その際に許される認知的作業量を制限するなどの条件は課していないため、どの程度努力したのか（頑張ったのか）は聴取者依存である。例えば、学習者音声の提示後に復唱させれば、呈示内容の推測に十分な時間が与えられることになる。

復唱に対して時間的制限を設け、聴取と同時に復唱を行なわせれば、それは追唱（シャドーイング）となる。時間的制約をかけることで、当該単語の同定や、更には単語間の統語的、意味的、談話的關係を捉える認知的作業に割く時間が限られてくるため、より容易に（迅速に）処理が行なえなければ、シャドーの精度に影響を与えると期待される。筆者らはこれまで、母語話者音声に対する学習者シャドーイング音声（外国語教育の現場で広く行われる通常のシャドーイング）の「崩れ」を自動的に計測する手段として、音素事後確率に基づく定量化を検討してきた（Yue et al., 2017）。本研究で



図 1: 学習者相互シャドーイング

はこの技術を母語話者による学習者音声シャドーイングに適用し、母語話者にとってのシャドーイングの容易さ (shadowability) を、可解性と解釈することの妥当性を実験データを通して論じる。

4 学習者相互シャドーイング

議論を進める前に、学習者音声の (可解性) 評価のために、母語話者シャドワーを事前に用意する必要性について言及する。Lang-8 という外国語教育支援サイトでは、学習者が学ぶ言語で書いた文章を、その言語の母語話者が添削し、その母語話者がある言語を学んでいる場合は、その言語の作文を、更にもその言語の母語話者が添削する、という、学習者相互依存型の支援を実現している。どの学習者も第一言語を持っており、母語話者として他者を支援し、学習者として他者から支援される訳だが、この枠組みをシャドーイングに導入する (図 1 参照)。即ち、学習者相互のシャドーイングインフラ (互いが互いをシャドーし、シャドーされる¹) を構築すれば、学習者音声の可解性自動評価において、母語話者シャドワーが必要となる前提は、決して大きな問題ではないと考えている。

更に、学習者音声と母語話者シャドーイング音声の対が大量に入手できれば、任意の新たな学習者音声に対して、どのようなシャドーイング音声を得られるのか、それを予測する技術的枠組みも検討可能となる。この段階まで来れば、母語話者シャドワーは不要となる。

以下、母語話者シャドーイングにおけるシャドーの精度 (shadowability) と、その母語話者がシャドー時に主観的に感じた可解性 (comprehensibility) との関係を実験的に検討する。

5 母語話者による学習者音声シャドーイング実験

5.1 ベトナム人学習者によるカラオケ式読上げ日本語音声の収録

L2 を日本語、L1 をベトナム語として学習者音声の収録を行なった。母語話者にシャドーイングを課す場合、提示音声の話速が遅すぎると (ベトナム語訛りが強くても) 可解性は高くなるため、話速を統制しつつ音声収録を行なった。具体的には中級レベルの音読用教科書 (松浦・福池・河野・吉田, 2014) に添付されている音声 CD を用い、モデル話者の話速に従って読上げテキストの文字色がかわる、カラオケ式音声収録ソフトを作成し、それを用いて学習者からの音声を集めた。

教科書から、固有名詞等のない 10 文章を抜き出し、6 名のベトナム人学習者 (3 名は上級, 3 名は中級) と 6 名の母語話者から、カラオケ音読ソフトを使って音声収録した。最終的に、学習者一人当たり約 100 音声、母語話者一人当たり 164 音声を得られた。この中から、学習者の習熟度を考慮し、ベトナム人日本語 96 音声 (VJ) と、母語話者 68 音声 (NJ) をシャドーイング用提示音声として用いた。これらは、フレーズ (句) を単位とした音声であり、全て、異なるフレーズである。

5.2 シャドーイング精度に関する三つの指標

VJ96 音声、NJ68 音声に対する母語話者シャドーイング音声に対して音声分析・音声認識技術を適用し、shadowability に関係すると思われる、下記の音声特徴量を抽出した。

¹学習者が自身の母語を学ぶ外国人に音声指導することは、専門知識がなければ困難である (故に Speech-8 は存在しない)。しかし、一般の母語話者が学習者音声をシャドーする場合に、特別な専門知識が必要となることはない。

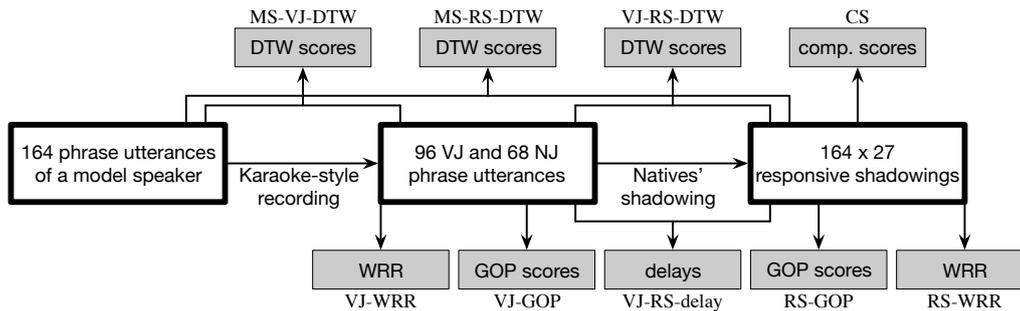


図 2: 母語話者による学習者音声シャドーイング実験と算出した各種音声特徴量

5.2.1 音素事後確率に基づく GOP スコア (DNN-GOP)

提示音声に対して単語の同定ができない場合（理解性が低い部位があると）、シャドーイングが崩れる（不適切な調音制御となる）ことが多い。調音制御の正確さを示す尺度として音素事後確率を導入する。ある時刻 t のスペクトル特徴量 o_t に対して、「 o_t が母語話者の発声だとするとどの音素 c_i が意図されたのか」を確率分布として表現したのが音素事後確率 ($P(c_i|o_t)$) である。深層学習に基づく音声認識技術では、その front-end モジュールにて、スペクトル特徴量を音素事後確率に変換しており、このモジュールを利用する。事後確率化することで、話者や年齢など非言語的特徴を凡そ抑制できる。シャドワーに提示した音声の音素表記を用いてシャドーイング音声から音素境界を推定し、個々の音素区間に対して当該音素の事後確率を求め、それを音素数だけ累積して平均化する。即ち、学習者音声評価タスクにおいて標準的技術として使われる GOP (Goodness Of Pronunciation) スコアを、母語話者シャドーイング音声に適用する。もちろん、GOP スコアは学習者音声（即ちシャドーイングの提示音声）からも算出である（本来この目的のために提案された技術である）。

5.2.2 事後確率ベクトルに基づく発話比較 (DNN-DTW)

GOP 計算では、提示音声の音素表象は用いるが、提示音声とシャドーイング音声を直接比較することはない。シャドーイング音声を音素事後確率分布の時系列とすることができるよう、提示音声も同様に音素事後確率分布の系列に変換できる。この二つの系列を、時系列の自動対応付け技術である、Dynamic Time Warping (DTW) で対応づけ、両者の差異を定量化する。今回の実験の場合、1) 教科書 CD のモデル音声、2) カラオケ式収録で得られたシャドーイング対象の音声、3) 母語話者によるシャドーイング音声と 3 種類の音声があるため、任意の 2 対間で DTW 距離を算出できる。

5.2.3 シャドーイング遅れ

DNN-GOP, DNN-DTW とともにシャドー時の調音制御の乱れを定量化することを意図しているが、シャドーイングの遅れについても自動計測した。提示音声、シャドーイング音声両者から音素境界を推定し、その時間差を音素境界の数だけ累積し、平均化する。この平均遅れを、そのシャドーイング音声のシャドー遅れとして定義する。

5.3 母語話者シャドーイング実験

27 名の正常な聴力をもつ成人母語話者がシャドーイング実験に参加した。シャドーイングは web 上に構築した専用のプログラムを通して実施した。クリック後、1 秒して音声提示が始まり、被験者はシャドーする。シャドーイング音声は、イヤーフックマイクを通して収録した。シャドーに関しては、「意味を捉え、ベトナム語特有の訛りを真似せず、母語話者として日本語でシャドーしよう」指示した。また、各フレーズ音声のシャドー後に、提示された音声の理解しやすさについて 7 段階で評価させた (comprehensibility, 可解性の主観的評価)。実験全体の流れと、自動計測した音声特徴量を図 2 に示す。MS は Model Speech, RS は Responsive Shadowing, CS は Comprehensibility Score の略である。また、WRR とは Word Recognition Rate を意味し、深層学習型音声認識の標準ツールキッ

表 1: 各種音声特徴と可解性 (CS) との相関

VJ-GOP	VJ-WRR	MS-VJ-DTW		
0.58	0.47	-0.52		
RS-GOP	RS-WRR	VJ-RS-DTW	MS-RS-DTW	VJ-RS-delay
0.74	0.53	-0.55	-0.58	-0.59

トである KALDI を、CSJ コーパスに適用して構築された CSJ-KALDI を用いた精度を意味する。

従来学習者音声の評価に使われてきた音声特徴量は、VJ-GOP, VJ-WRR, MS-VJ-DTW である。本研究では、母語話者発音との近接性ではなく、学習者音声の可解性 (CS) を主眼に置いた評価を検討している。この場合、VJ-GOP, VJ-WRR のように学習者音声から得られる特徴量と、RS-GOP や RS-WRR, 更には VJ-RS-delay や MS-RS-DTW など母語話者シャドーイング (RS) があって初めて計測可能となる特徴量のどちらがより可解性と相関が高いのか、が検討の中心的課題となる。

5.4 結果と考察

VJ のフレーズ単位の 96 発声の各々について、GOP スコア (VJ-GOP), 音声認識精度 (VJ-WRR), モデル音声との DTW 距離 (MS-VJ-DTW) が算出される。また、27 名の被験者から CS スコアが得られる。この CS スコアの被験者間平均を、当該 VJ 音声の CS スコアと定義する。一方、母語話者シャドーイング音声の各々について、GOP スコア (RS-GOP) や、音声認識精度 (RS-WRR), 更には VJ 発声からの遅れ (VJ-RS-delay), VJ 発声やモデル発声からの DTW 距離 (VJ-RS-DTW, MS-RS-DTW) が計測される。即ち、母語話者シャドーイング音声の各々に対して、これら五種類の音声特徴の被験者間平均が計算できる。各 VJ 発声の CS スコアに対する相関値を表 1 に示す。母語話者シャドーに提示した VJ 音声よりも、彼らのシャドーイング音声からの音声特徴の方が、VJ 音声に対する主観的可解性と、高い相関を示していることが分かる。従来学習者音声を自動評定する場合、学習者音声を分析対象としていたが、上記の結果は、学習者音声よりもそれを聴取及びシャドーした、母語話者音声の方が分析対象としてより適切であることを意味している。評価基準を母語話者音声との近接性とすれば、学習者音声を分析対象とすべきだが、評価基準を可解性とした場合は、母語話者シャドーイング音声の方へ着目すべきである。更に、単なる音声認識結果よりも、音素事後確率化した方が、主観的な可解性と相関がより高くなっている。

これらの音声特徴を説明変数として用い、Lasso 回帰²を用いて、CS スコアを予測する実験を行なった。交差検定の結果、予測値と実測値 (CS 平均値) との相関は 0.81 となった。これは、26 名の被験者の CS 平均値と残り一人の被験者の CS スコアとの相関値の平均、0.66 を大きく上回った。

5.5 Shadowability は online intelligibility なのか、それ以上なのか？

学習者音声よりも、その音声を母語話者にシャドーさせ、そのシャドーイング音声の崩れ (調音的崩れ、シャドーの遅れ) を測定した方が、聴取者の主観的可解性と高い相関を示すことが実験的に示された。これら shadowability は、その測定方法を考えれば、comprehensibility よりも、時間制約付き intelligibility, 即ち online intelligibility と解釈すべき指標とも言える。intelligibility と comprehensibility は個々の単語の同定のみに着目するのか、単語間の関係の把握 (即ち意味の把握) までに着目するのか異なる。shadowability が intelligibility と comprehensibility のどちらに近いのかを考える場合、shadowability が (提示音声の) 意味の把握の容易さによって左右されるのか否か、を検討することになる。以下、筆者らの検討 (Trisitchoke et al., 2018) を紹介する。

本節では外国語訛りが混入された提示音声を用い、それによって shadowability が変わる様子を示した。一方 Trisitchoke et al. (2018) では、プロのナレータによる読上げ音声を用いており、読上げ文の内容を制御することで、意味理解の難易度が定性的に異なる文章音声に対する母語話者シャドー

²過学習とならないよう、正則化による制限を導入した線形回帰モデル

表 2: 6 種類の読上げ音声

A	桃太郎
B	NHK NEWS WEB EASY (NWE) 中の文章
C	NHK NWE 中の内容語のランダム列
D	NHK NWE 記事に対する元原稿
E	日経サイエンスからの記事
F	無意味モーラ (清音のみ) 列

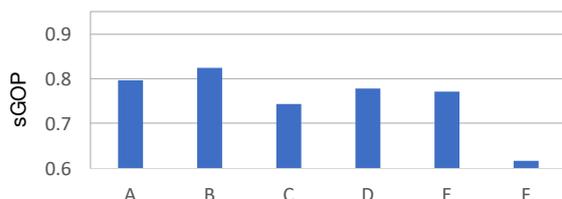


図 4: 母語話者シャドーイング音声の GOP

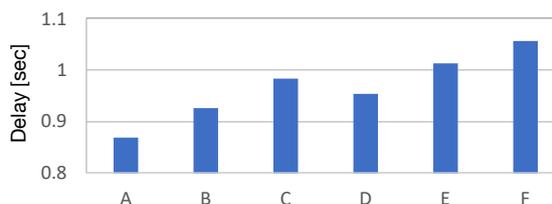


図 3: 母語話者シャドーイングの遅れ

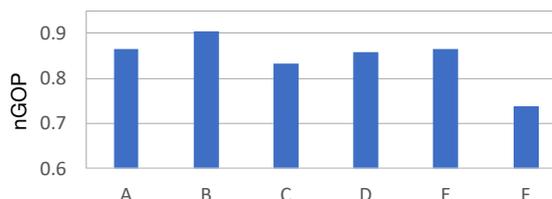


図 5: プロのナレータ読上げ音声の GOP

イングを検討している。準備した文章音声を表 2 に示す。平易な日本語文として、日本語学習者向けに用意されたニュース文 (B) を用い、一般的な日本語文として、内容は異なるが NWE 記事に対するオリジナル原稿 (D) を用いた。また、語彙は平易だが統語構造も意味構造もない刺激として、NWE に出現する内容語のランダム列 (C) を用意した。A は、次の単語やフレーズが容易に予測できるが、日常語以外の語彙も含まれる文章を意図しており、桃太郎の読上げ音声を用いた。

詳細な実験条件や有意差検定結果は Trisitichoke et al. (2018) を参照して載せたいが、7 名の母語話者シャドワーが示した shadowability として、遅れと GOP の (被験者間及びセット内) 平均を文セット毎に図 3, 図 4 に示す。更にはプロのナレータによるモデル音声に対する GOP スコアも図 5 に示す。表 2 に示した文セットの意味的難易度にほぼ従う形で、シャドーイング音声の遅れや GOP が分布している。興味深いことに、被験者の GOP スコアと非常に類似した変化パターンが、プロのナレータ音声にも観測された。ナレータ音声の収録は事前のリハーサル後に行なっており、図 5 の GOP スコアの変動は、各文章セットの意味理解の難易度に起因するものと考えられる。当然、通常母語話者が読み上げる場合でも、意味理解の困難さが、その音声に影響を与えていることが示唆されるが、タスクを読上げから復唱 (repetition) や追唱 (shadowing) に変えても、同様の傾向は観測されるであろう。図 4 はシャドー時の結果である。Mineamtsu et al. (2011) では、母語話者の復唱結果を書き起こし、その書き起こし結果に基づいて了解性を算出している。書き起こしではなく、復唱音声を分析すれば、可解性と類似した計量が可能であったと推測される。タスクを復唱から追唱にすれば、より可解性へ迫ることが可能になると考えられる。

6 学習者相互シャドーイングが可能にするもの

本研究では、学習者音声を母語話者にシャドーさせ、そのシャドーイング音声の崩れを計測することで、聴取者が学習者音声に対して感じる可解性を効果的に予測できることを実験的に示した。また、シャドーイングというタスクそのものが、単に提示音声の中の語の同定及びその復唱のみを意味せず、提示音声の意味理解に依存した発話活動となっている様子も実験的に示した。

本節では、この学習者相互シャドーイングを通じた音声評定方式が有する、幾つかの興味深い側面について論じる。一般の外国語教育現場では、学習者の動機付けを大切にすあまり、学習意欲が阻害されないよう、各学習者への教示内容 (言葉遣い) に配慮することがある。所謂、学習者 (教える側にとっては、学習者は顧客となる場合もある) への付度である。初級者に対しては有効な教示戦略であると思われるが、学習過程の進捗と共に、より厳しい、本音の教示を返す必要もあろう。学習者音声をシャドーイングする場合、付度とは、上手なシャドーを返してあげることを意味するが、

事前に学習者の発声内容を知らない限り、このような付度シャドーイングは不可能である。即ち、学習者間の相互シャドーイングは、母語話者が聴取時に感じる本音の「聴き取り易さ」(可解性)を直接的に返す場を学習対象言語に依らず、全外国語学習者に対して導入する、という意味を持つ。

了解性にせよ、可解性にせよ、これらは、十分聴き取り易い発音であればどれだけ母語話者発音からずれていても是とする教育戦略と関係する。さて、ある学習者の外国語発音を、最も了解性・可解性が高いと評価する聴取者は誰だろうか？筆者らが考えるに、それは学習者本人である。「十分聞き取れるから、発音を改善する必要性が理解できない」という学習者がいても不思議ではない。了解性・可解性を強調すればするほど、そのような学習者は増えるのかもしれない。学習者相互シャドーイングは、他者となる母語話者にとっての可解性を直接的に返す枠組みであるが、それ以上の機能を持つ。日本人が自身の英語音声をもとに母語話者にシャドーしてもらおうと共に、日本語学習者の日本語音声をシャドーする。二種類のシャドーイング音声を使えば、以下が可能となる。その学習者の英語音声(母語話者)シャドーイングをどの程度崩すのか、その「崩れ」と等価な「崩れ」をもたらす外国人の日本語音声を、その日本人英語学習者に提示できる。即ち、自身の英語音声と(可解性という尺度において)等価な外国人の日本語音声を提示できる(外国語訛りの母語音声を通して、自身の英語音声を把握させる)。可解性がレベルCとか、30点とか論理的に示すよりも、学習者の感覚に訴えられる、説得力のある教示となるかもしれない。なお、機械による合成音声を品質評価する場合、評価者が対象言語の母語話者なのか否かは明確に区別して行われる。その理由は、母語話者の方が不自然さに対する感度が極めて高いからである。この「不自然さに対する高い感度」をもって、自身の英語発音の現状を把握させることが可能となるだろう。理論的には学習者を、極めて甘い評価者から、極めて厳しい評価者へと変遷させることになる訳だが、まだ何ら実験的検証結果はない。現在、実験的検証に向けて、データ収集の準備を行なっている段階である。

7 まとめ

本研究では、学習者音声を母語話者にシャドーさせ、その崩れを計測することが、可解性に基づく評価を、客観的に、定量的に、かつ、容易に実現できることを示した。様々な言語の学習者を繋いで、学習者間の相互シャドーイングを音声学習インフラとして実現した場合に期待される効果についても記述した。後者については何ら実験的検証を行っていないが、ベトナム・ハノイにて日本語学習者音声の大規模収録を計画しており、実験的検証を行なう環境を整えつつある。

本研究は、JSPS/MEXT 科研費 JP26118002, JP26240022, JP18H04107 の支援を受けた。

参考文献

- Derwing, T. M., & Munro, M. J. (2015) *Pronunciation fundamentals: Evidence-based perspectives for l2 teaching and research*. Amsterdam: John Benjamins Publishing.
- Mineamtsu, N., et al. (2011) "Measurement of objective intelligibility of japanese accented english using erj database." In *Proc. INTERSPEECH* (pp. 1481-1484).
- Munro, M. J., & Derwing, T. M. (1995) "Foreign accent, comprehensibility, and intelligibility in the speech of second language learners." *Language Learning*, 45, 73-97.
- Trisitichoke, T., et al. (2018) "Influence of content variations on native speakers' performance of shadowing." In *Proc. ASJ autumn meeting*.
- Yue, Y., et al. (2017) "Automatic scoring of shadowing speech based on dnn posteriors and their dtw." In *Proc. INTERSPEECH* (pp. 1422-1426).
- 電通サイエンスジャム (2016) *Kansei analyzer*. <https://kansei-analyzer.com>.
- 松浦 真理子・福池 秋水・河野 麻衣子・吉田 佳世 (2014) 『日本語音読トレーニング』東京: アスク出版.
- 河原 達也・峯松 信明 (2013) 「音声情報処理技術を用いた外国語学習支援」『電子情報通信学会論文誌』96: 7, 1549-1565.

短期集中コースにおける意識化・気づきを活用した日本語の発音指導

赤木浩文（専修大学）

1 研究目的

研究の目的は大学の短期集中日本語コースで行っている意識化・気づきを活用した日本語の発音指導の結果を分析し、短期間で行える効果的な発音指導について考察することである。

2 研究の背景

第二言語の発音の学習や習得は母語などの影響による個人差が大きく、クラスでの指導には工夫が必要である。日本語学習者の中には、長年日本語を学習していても、コミュニケーションを阻害する発音上の課題が残る学習者や発音の指導を受けたことがない学習者が見受けられる。このような学習者の発音によるコミュニケーションの失敗を減らすために、当大学の日本語コースでは、体系的な発音指導をルーティンワークとして行っている。当コースは短期集中、多国籍、少人数という特徴があり、クラス単位の指導は時間的制約、発音の習熟度、母語の影響による個人差を考慮した指導が要求される。そのため、自律学習に結びつく指導として、第二言語学習において重要性が注目されている意識化と気づきを利用した発音指導を採用した。その結果、学習項目に効果が現れやすいものや共通した学習ストラテジーなどが観察された。本研究ではそれを整理し、短期集中コースにおける、より効果的、効率的な発音指導を検証し、今後の発音指導について考察した。

3 意識化・気づきと発音指導

3.1 第二言語習得論と意識及び気づき

第二言語習得理論では、インプット、インプットの気づき、理解、内在化（インテイク）、統合といった流れの連続によってアウトプットが可能になると考えられている（Gass & Selinker 1994）。インプットの中で学習者に認識され、意味内容と言語形式の関係が理解されたものが「インテイク」として学習者に内在化され、学習者自身の言語システムとの統合が起きる。この統合によって発達する言語システムからアウトプットが引き出されると言われている。この理論で重要なものが意識である。

Schmidt（1990）は、第二言語習得において学習者に意識的な注意を喚起させること、または「気づき」を起こさせることが重要であると述べ、「気づき仮説」（noticing hypothesis）を提示した。「気づき仮説」とは、インプットのうち学習者が意識的に注意を向けたものだけが、インテイクとして取り入れられ、そのインプットをインテイクに変えるのに「気づき」が必要かつ十分条件であるという考え方である。

また、村野井（2006）は、自分の第二言語能力の「穴」に気づくこと、目標言語と中間言語のギャップに気づくことなど、アウトプットの気づきの重要性について言及している。さらに、Long（1996）は、言語習得は対象言語を相互交流の中で使用することによって促進されるとし、学習者がインタラクションで負のフィードバックを受けた場合、学習者の理解しようとする意志や意図を理解してもらおうとして行う試行錯誤などの努力によって習得が進むと主張した（インタラクション仮説）。意識的な側面から、負のフィードバックに対する気づきが習得を促進する重要な要素だと考えられる。

クラスで教師が学習者に意識化を促し、気づきを起こさせる工夫として「インプット強化」がある（Sharwood Smith, 1991）。これは、学習項目の言語形式を目につきやすいように、ハイライトにする、色付けするなど視覚的に強調し、インプットがインテイクとして取り込まれる率を高める工夫である。Doughty（1991）は、学習項目への高頻度の接触と学習者の注意を引く工夫が習得に有効に働くと述べている。また、Schmidt（1990）は、言語形式に気づかせる有効な方法として focus on form を取り上げている。そして、学習項目にインプット強化を施し、言語形式に気づかせるのは、教師の役割であるとしている。

3.2 発音指導及び発音学習と意識化

発音学習と意識化の有益性を述べた研究に、磯村（1996）がある。磯村（1996）は、日本語学習者に対して日本語の韻律の理論的知識と練習法を提示する指導を行い、アクセント型の違いを実現していなかった学習者が、アクセント型の違いを意識し、発話しようとしたこと、イントネーションとの練習によって型の違いを韻律によって区別しようとし、以前より正しい韻律に近づいたことを確認した。そして、学習者が日本語の韻律を習得するには、韻律に関する理論的知識と目標言語の音声の明確な意識化が有益であるとした。

発音指導における学習者の気づきの有効性を分析した研究に赤木（2011, 2013）がある。赤木（2011）では、学習項目を焦点化した発音指導において、学習者が学習理由、学習項目、課題を意識するよう工夫し、インプット強化として音声モデル、視覚情報、リズムビート、対話におけるフィードバックを活用した。その結果、学習過程で気づきが促進され、発音の改善につながったと述べている。特にインタラクションで学習者がインパクトを受けたときに学習者効果が高まった例を報告している。さらに、赤木（2013）では、上級学習者を対象に気づきを促す発音指導を行い、気づきと発音の向上の関連性を分析した。そして、気づきを①学習目的に関する気づき、②言語形式に関する気づき、③できないことに関する気づき、④課題に関する気づき、⑤ストラテジーに関する気づきに分類し、発音習得にはストラテジーに関する気づきが重要であるとした。

4 発音指導の方法

発音クラスの指導目標は、コミュニケーションに必要な発音の習得、つまり話者が伝えたい意味や意図が正しく伝わる発音の習得である。指導期間は7週間（全9回～12回）ま

の3クラスの合計34名の学習者である。参加した日本語学習者は表2の通りである。学習期間にばらつきがあるが、中級前期で1年半～2年程度、中級は2年～3年程度である。各クラスによって共通する発音上の課題、個人別の課題があり、授業で強調する箇所、個人へのフィードバックには異なる部分もあるが、基本的には同様の指導法を行った。

	レベル	期間	人数	出身
1	中級前	7週	5 (女3男2)	チリ2、豪、韓、米
2	中級前	12週	4 (女3男1)	豪、台、米、マレーシア
3	中級前	7週	7 (女6男1)	米4、加、中、韓
4	中級	7週	9 (女)	韓5、印、中、伊 越
5	中級	7週	9 (男3女6)	韓 新、イラン、英、加、中、米 越2

表2 対象クラスと学習者

6 分析項目と方法

クラスの指導結果を発音の改善、意識化や気づきによる学習効果という観点から検証し、その関係を考察した。発音については、コース開始時と終了時の発音チェック（朗読文の読み上げ、自己紹介、経験に関するモノログ）の録音を用いて、学習者の発音上の課題や特徴を具体的に記述し、聞きやすさを5段階で評価した。評価者は40代と50代日本語教師男女2名である。また、実際に使用場面での応用を見るために、スピーチ、会話テスト、プレゼンテーションの発音に対する評価やコメントを整理して、学習者の課題や改善点を整理し、改善した項目と学習者の人数をまとめた。改善度は、①変わらない、②やや改善した、③かなり改善した、④非常に改善したの4段階で評価した。

一方、意識的側面に関しては、アンケート、インタビュー、クラス観察記録のデータを用いた。まず、コース開始時と終了時に、選択肢式と自由回答式のアンケートを行い、日本語の発音学習に対する意識や姿勢、その変化を調べた。また、発音の授業の様子をできるだけ具体的に記録した。さらに、インタビューを行い、意識化、気づきと発音習得、発音学習との関係を分析した。インタビューは、4ステップコーディング法（SCAT）（大谷2007）による質的分析を行い、学習者ごとに学習の過程で気づきがどのようなタイミングでどのような学習項目にどのように起こったかを整理した。

7 結果と分析

指導後に改善したという評価が多かった学習項目は、文の区切り（28名中25名）と文末イントネーション（19名中17名）、促音（25名中21名）であった。やや改善したが、安定しないものとして長音が挙げられた。長音の失敗例には、「女性」や「旅行」などの、短・長の組み合わせの語は前の拍が長くなる傾向がレベルや母語に関係なく見られた。撥音や単音は、個人による差が大きかった。アクセントはコース終了時まで課題として残る項目であった。意識的側面では、どのクラスでも学習者のモチベーションが高まったことがわかった。コース前には発音指導や発音の規則の学習に対して具体的なイメージがなか

った学習者が、コース終了後は学習目的や学習項目などが具体的に明確になった例が見られた。学習過程で意識化・気づきによる学習効果として、①目標言語への試行錯誤、②自己修正、③自分の穴に対する気づき、④発音ストラテジーの利用が観察された。気づきを促進したインプットとして、音声情報の視覚化、記号化が有効であり、これらはアウトプットにも活用されていることがわかった。自律的学習として発音練習用テキストへの視覚情報（記号）の付加が見られた。インタラクションでは、負のフィードバックによる自分自身の穴に対する気づきなどが観察され、学習者自身の記憶にも強く残っていた。発音の改善と意識的な要因の関係を見ると、区切り、文末イントネーションには、記号を用いた自発的な練習が見られ、意識化と学習の関係が示唆された。また、音声的側面の改善が顕著ではなかったが、長音は、学習課程の観察を見ると、最も試行錯誤、自己修正、視覚情報の利用などが見られ、意識化による学習が最も進んだ項目であった。どの項目も視覚情報が手がかかりとして用いられ、意識化を促進するインプット強化として記号等の視覚情報が有効なことがわかった。クラス及びレベルによる大きな差は出なかったが、学習者の学習意欲、学習者の性格や学習背景によって、意識化、気づきに影響が観察され、発音の習熟度が高い学習者の方がアウトプットの気づきが観察される例が多かった。また個人別では、音の聞き分け能力が低い学習者はモチベーションが低い傾向が見られた。

8 考察

結果と分析から、意識化、気づきを活用した日本語の発音指導で、短期間に指導効果が現れたと判定された学習項目は、①区切り、②文末イントネーション、③促音で、長音は、短期間では安定しないものの意識化が進むことがわかった。学習者の意識化の促進には、音声情報の視覚化（記号化）、気づきの喚起にはインタラクションにおける負のフィードバックが有効であることがわかった。視覚情報は、学習者の多くが目標言語の発音の手がかかりとして有用だと考えており、目標言語に到達するために行う試行錯誤や自己訂正は自己修正能力の養成に大きな役割を果たすと考えられる。インプット強化としての視覚情報は、意識化されると、音声とともにインテイクとして取り入れられ、音声情報と統合されれば、音声の習得の促進が期待される。統合が完了するまでには、失敗、試行錯誤、自己修正の試みの繰り返しが必要となり、その過程で起こる気づきが学習を促進すると考えられる。しかし、内在化と正確なアウトプットは、学習項目、学習者によってかかる時間が異なる。

今回、区切りやイントネーションカーブの記号の付加は、発音の習熟度や聞き分けの能力に関係なく、自分自身でコントロールできる作業であることから、学習者の多くに活用され、効果的な学習につながったと考えられる。また評価対象の音声がモノログ中心であったことも要因かもしれない。長音に関しては、情報統合に時間はかかるが、学習者の意識化が進み、積極的な学習が見られることから、将来的な効果が予測できる。これらのことから、視覚情報の使用に対する工夫、学習項目の提出順序、学習者に課す課題などにいくつかの示唆が考えられる。短期コースで行えることは、学習者のインテイクを促進す

るためのインプット強化と気づきを誘発するインタラクションの場の提供、自律学習に結びつく発音ストラテジーを取得させることだと言える。そういった意味で意識化・気づきを利用した発音指導は短期の指導に有効だと言えるだろう。今回の調査から、これまでの指導に加えて、例えば、記号があるものとなないもので交互に練習させるなどのタスクや、学習項目や学習方法が習得しやすいものから先に提出するなどが考えられる。

参考文献

- Doughty,C.(1991) Second language instruction does make a difference:Evidence from an empirical study of relativation. *Studied in Secound Language Acquisition.13*, 431-470.
- Ellis,R. (1995) Interpretation tasks for grammar teaching. *TESOL Quarterly*, 29/1, 87-105.
- Gass, S. (1997) *Input, Interaction, and the Second Language Learner*. Lawrence Erlbaum Associates.
- Krashen S.D.(1985) The input hypothesis:Issues and implications. Harlow,Essex:Longman House.
- Long,M.H.(1996) The role of the linguistic environment in second language acquisition, In W.Ritchie & Batia (Eds), *Handbook of research on second language acquisition* (pp.413-468), New York:Academic.
- Sharwood Smith, (1991) Sharwood Smith, M. 1991. Input enhancement in instructed SLA: theoretical bases. *Studies in Second Language Acquisition*, 15, 165-179.
- Schmidt,R.W.(1990) The role of consciousness in second langage learning. *Applied Linguistics*,11/2,129-158.
- _____. (1995) Consciousness and foreign language learning: A tutorial on the role of attention and awareness in learning. In R. Schmidt (ed.) *Attention and awareness in foreign language learning*. University of Hawaii Press, 1-63.
- Swain,M(1993) The output hypothesis:Just speaking and writing aren't enough. *The Canadian Modern Language Review*,51,158-164.
- 赤木浩文 (2011) 「日本語の発音指導～学習項目の意識化と焦点化」, 『外国語教育研究』 第 11 号 53 - 66.外国語研究学会.
- _____(2013) 「気づきを促す発音指導」 『外国語教育研究』 第 16 号, 37-54.
- 赤木浩文 古市由美子 内田紀子 (2010) 『毎日練習！リズムで身につく日本語の発音』 スリーエーネットワーク.
- 磯村一弘 (1996) 「アクセント型の意識化が外国人日本語学習者の韻律に与える影響」 『日本語国際センター紀要』 6、1-18.
- 大谷尚(2007) 「4 ステップコーディングによる質的データ分析手法 SCAT の提案—着手しやすく小規模データにも適用可能な理論化の手続き—」, 『名古屋大学大学院発達科学研究科紀要』 第 54 巻第 2 号, 27 - 44, 名古屋大学.
- 酒井英樹(2002) 「誤りの種類と気づき—フィードバックの役割と第 2 言語学習者の発話修正」, 『上越教育大学研究紀要』 21 (2) ,741-755 上越教育大学.
- 村野井仁(2006) 『第二言語習得から見た効果的な英語学習法・指導法』, 大修館書店.

自発発話データから見たアクセントの遅下がり現象

佐藤 大和 (東京外国語大学)
sato.hirokazu@tufs.ac.jp

1. はじめに

日本語 (共通語) のアクセントの型は, 単語を構成する拍の「高」「低」の配置形式, もしくはピッチが「高」から「低」へ降下する際の「高」の拍 (アクセント核) の位置によって記述される. これらの型は, 拍を意識した比較的丁寧な発音もしくはその内省によって把握される. 一方, 実際の発話における音調の動的様相を見てみると, こうした規範的なアクセントの型がそのまま音響特性として実現されているわけではなく, アクセントがあるとされる拍と音響上アクセントがあると推定されるピッチの下降時点との間に乖離が見られることがある. 特にアクセントのある「高」に続く「低」の拍でピッチが一部または全体に高く留まって, ピッチ下降時点が遅れる現象はアクセントの「遅下がり」(杉藤 1980, Hasegawa and Hata 1995) として知られている.

「遅下がり」現象には, 二つの問題点がある. 一つは, 生成的側面から, ピッチ降下が遅れる要因や仕組みは何かということである. 拍を意識した丁寧な発音の音声と発話制約のない連続音声との間で, アクセント感を生み出す音調の動的特性に違いがあるのではないかと考えられる. 二つ目は, 知覚的側面から, アクセントがあるとされる拍の後続拍が先行拍より必ずしも低くはないのに, 本来のアクセント型として知覚されるのは何故かということである. 音調の動態パターンの中に, 「遅下がり」が生じても安定なアクセント認知を保証する特性が存在すると考えられる. 本報告は, 自発発話音声データに基づき, 「遅下がり」現象の多様な実現形態を明らかにして, 主に生成的側面から考察することを目的としている.

2. 音声データと分析方法

2.1. 分析用音声データ

分析のための音声資料として, 「日本語話し言葉コーパス (CSJ)」における東京方言話者 (女性) 1 名の独話資料 (模擬講演) を用いた. 従来の研究に多かった語の単独発話より多様な音調動態に基づいて検討するためである. 発話時間は 11 分ほどであり, この中でアクセントのある約 520 個の音声単位を基本データとした.

2.2. 分析方法

この話者の発話は, 発話末の終結ピッチ周波数 (F0) が 130 Hz 程度であることから, この値を基準値とする Semitone (ST) を求め, 主に ST から音調特性を見ることとした. ST 上では, 基準値より 1 オクターブ高い 12 ST が 260 Hz, 2 オクターブ高い 24 ST が 520 Hz に相当する. 発話データのピッチ範囲は, 2 オクターブ, すなわち 520 Hz 以内に収まっている. 話し言葉コーパスのデータから, 10 msec のフレーム (Frame (FR)) 毎に, 時間・ピッチ周波数・ピッチ ST・音声セグメント情報等の時系列を取り出し, 各音声セグメント (主に母音) におけるピッチ変化率 (F0 変化率 $\Delta f : \text{Hz}/\text{FR}$, ST 変化率 $\Delta \text{ST} : \text{ST}/\text{FR}$) を区分内直線近似で求めた.

アクセントに関しては、音声データの聞き取りによってアクセント型の判断を行うとともに、ピッチ周波数特性に基づいてアクセント位置を定めた。前者の判断によるアクセントのある拍を「アクセント拍」、音響特性から設定したアクセントの時間軸上の位置を「アクセント位置」と呼ぶ。「アクセント位置」は、CSJのドキュメント（五十嵐・菊池・前川(2006)）の記載に準じており、アクセント拍およびこれに後続する拍のピッチ周波数パターンに基づき、上昇ののち下降する特性においてはそのピークを、緩やかな変化から急峻な下降がある場合は下降の開始点を、下降特性のみの場合はその開始時点「アクセント位置」とした。

一つのアクセントを有する音声単位はアクセント句呼ばれるが、ここで分析される音声単位はこれより狭い単位であり、接続助詞、音調上昇を伴い易い副助詞などを除いた文構成上の基本的単位（文節のコア部分）であって、原則1個のアクセントを有する音声区分である。ここではこれを「アクセント単位」と呼ぶことにする。

次に、アクセント位置はアクセント拍の音調形式と関連することから、各アクセント拍の拍内のピッチ周波数を直線近似し、以下の音調形式を求めた(佐藤 2018)。

- ・下降音調(Falling Pitch: F) 下降ピッチの音調形式
 - ・平坦音調(Level Pitch: L) 拍内ピッチの傾きの絶対値が 0.1 ST/FR 未満の場合を平坦のピッチとした。
 - ・上昇音調(Rising Pitch: R) 上昇ピッチの音調形式
 - ・上昇・下降音調(Rising+Falling Pitch:RF) 拍または音節内で上昇・下降する音調
- この他、平坦・下降音調なども設定したが、これらは若干数なので特に記述しない。

さらに、「遅下がり」の判定に関しては、「アクセント拍」に後続する拍内に「アクセント位置」あるものは「遅下がり」現象とした。引く音、撥音、下り二重母音を含む長音節にアクセントがある場合で、これら後部音素にアクセント位置がある場合も「遅下がり」に含めた。

3. 分析結果

3.1. 各種条件下での「遅下がり」の生起数とその割合

アクセント位置の分析の結果、合計 81 例の「遅下がり」が見出された。これは、全アクセント単位数の 15.6%に当たる。表 1 にアクセント型毎の生起数、および全遅下がり数に対する割合と当該アクセント型生起数に対する割合を示す。アクセント型は、先頭から数えたアクセント拍の位置で示した。

表 1 から、「遅下がり」は 1 型アクセントで著しく多く(72.8%)、また 1 型アクセントの抽出例のうち、1/4 が「遅下がり」であった。2 型アクセントがこれに続き、3 型以上

表 1 アクセントの「遅下がり」生起数と生起割合

アクセント型	1 型	2 型	3 型以上	計
遅下がり生起数	59	11	11	81
全遅下がり数に対する割合	72.8 %	13.6 %	13.6 %	
アクセント型生起数	235	96	188	519
アクセント型生起数に対する割合	25.1 %	11.5 %	5.9 %	

表2 アクセント単位の拍数別「遅下がり」生起数

アクセント単位拍数	2拍	3拍	4拍	5拍	6拍	計
1型アクセント	6	17	26	9	1	59
2型アクセント		3	2	5	1	11
3型アクセント			3	1	5	9

表3 アクセント拍における音調形と「遅下がり」生起数

アクセント拍音調形	1型アクセント			2型アクセント		
	F	L	R & RF	F	L	R & RF
遅下がり生起数	3	7	47	0	2	9
割合	5.1%	11.9%	79.7%	0%	18.2%	81.8%

では「遅下がり」は起りにくくなる。

表2は、各アクセント型におけるアクセント単位の「遅下がり」生起数を拍数別に見たものである。1型では4拍の単位に多く、2型では5拍の単位に多い。これらは「遅下がり」拍のあと2拍程度の後続部の続くものである。

表3は、アクセント拍における音調形と「遅下がり」数との関係をまとめたものである。ここで(R&RF)の欄は上昇に関わる二つの音調（上昇と上昇・下降）の生起数である。1型および2型アクセントとも、上昇に関わる音調(R&RF)での生起数が全体の約8割を占める。

3.2. 「遅下がり」の音調パターンとその生起要因

以下、具体的に「遅下がり」を示す音調パターンからその生起要因について考える。

(1) アクセント拍の上昇調との関係

「遅下がり」は1型アクセントで多く、またアクセント拍が上昇調の場合に多いことはすでに述べた。1型アクセントでかつ上昇拍の場合の例を図1に示す。図中2例とも「我が家の」という発話であり、(1)は卓立型発話で「遅下がり」が観測され、(2)は非卓立型発話で「遅下がり」が見られない。アクセント拍が上昇調の場合、次の拍への下降のためアクセント位置は当該拍末近傍にくるか、場合によっては後続拍の子音・母音境界付近にくる

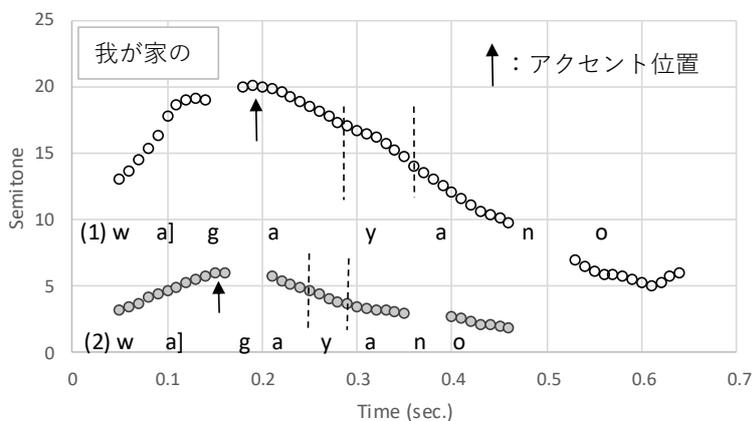


図1 語頭上昇とアクセント「遅下がり」の例(1)

(佐藤 2018)。後者が「遅下がり」である。つまり、「遅下がり」はピッチ上昇制御の overshoot 現象と捉えることもできる。一方、図2はアクセント位置が「遅下がり」拍とさらに後続の拍との境界付近にくる場合であり、「遅下がり」拍内のピッチ降下は殆どない例である。

1型アクセントが上昇調となるのは、句頭の最初の拍で

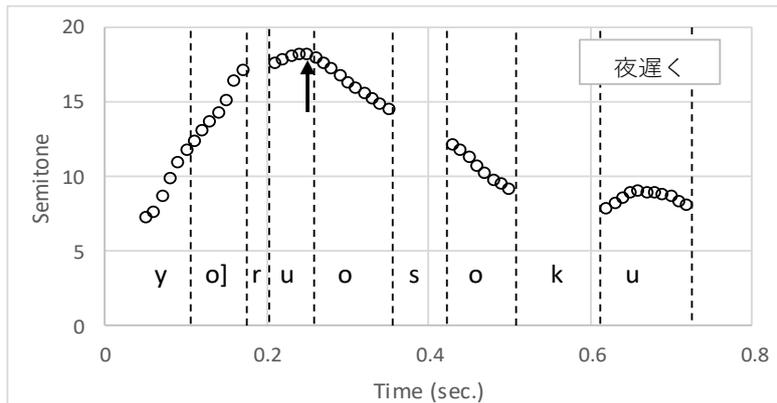


図2 語頭上昇とアクセント「遅下がり」の例(2)

「高」の音調が要請されるため、これを上昇調で実現するからであろう。2型アクセントの場合は、2拍目で「高」が要請されるので、1型と同様に上昇調が観測され、「遅下がり」との関連が見られることもあるが、1拍目から緩やかな上昇になるなど、1型ほど急峻な特性にはならない。

1拍目で上昇調となる1型アクセント語のアクセント知

覚に関する実験では、1拍目で十分急峻なピッチ上昇特性場合、許容される後続拍のピッチ下降には大きな幅があり、わずかな下降でも1型の知覚が得られること、また比較的緩やかな上昇特性の場合、後続拍に一定以上の下降特性が求められることが分かっている(佐藤 2016, 2017)。このことは、1拍目の急峻なピッチ上昇があれば、それがアクセント知覚の Cue となっており、後続拍が「遅下がり」でも構わないということを示している。従来から、アクセント感覚に関してはピッチの「下降」に着目して論じられてきたが、「上昇」特性の役割についても目を向ける必要があると思われる。

(2) 強調的発話との関係

特定の句を意図的にあるいは対比的に強調して発話する場合、アクセント拍ではピッチ

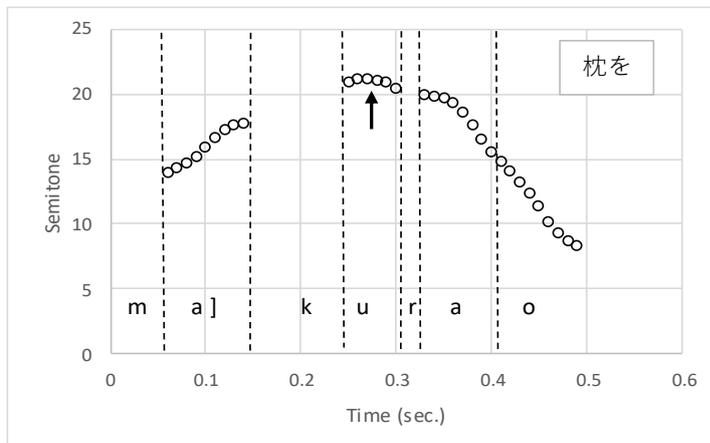


図3 強調的発話における「遅下がり」

上昇が生じやすく、かつアクセント位置でのピッチの高さは大きくなり、ピッチ上昇幅や下降幅は1オクターブまたはそれ以上(12 Semitone 以上)になる。上昇が大きくなるとその影響は後続拍に及び、ほとんどの場合「遅下がり」現象が観測された。1例として、「私の枕を使って」における「枕を」の部分を経験的に発話した場合を図3に示す。

(3) フット単位の発話との関係

長音、撥音等を含む長音節にアクセントがある場合は、前部拍にアクセントが置かれる。しかし、語としての発話の際には、この2拍の音節は一体となって発音されるため、アクセント位置がしばしば後部拍にくる。図4は2型アクセントの「デパ]ートの」の場合の音調パターンである。アクセント位置は長音節/pa] a/の音節末近傍に置かれ、ピッチはアクセントがあるとされる拍より2拍遅れて下降する「遅下がり」となる。

日本語の長音節は2拍を単位とするフットの単位ともなっている。通常の短音節(軽音

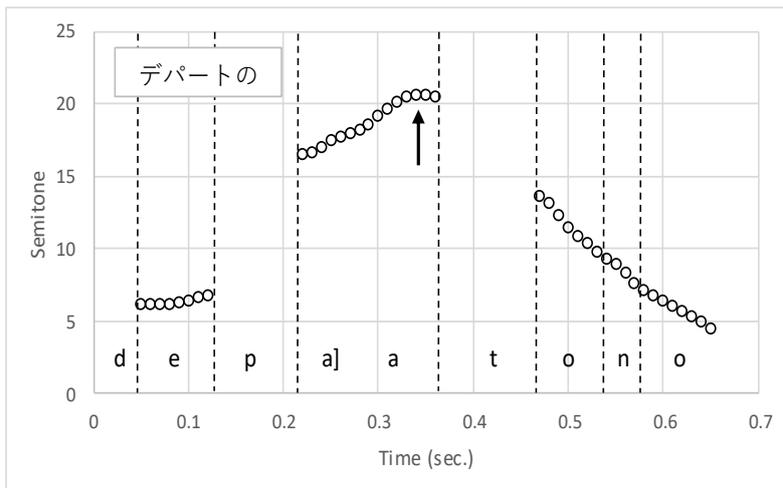


図4 長音節にアクセントがある場合の「遅下がり」

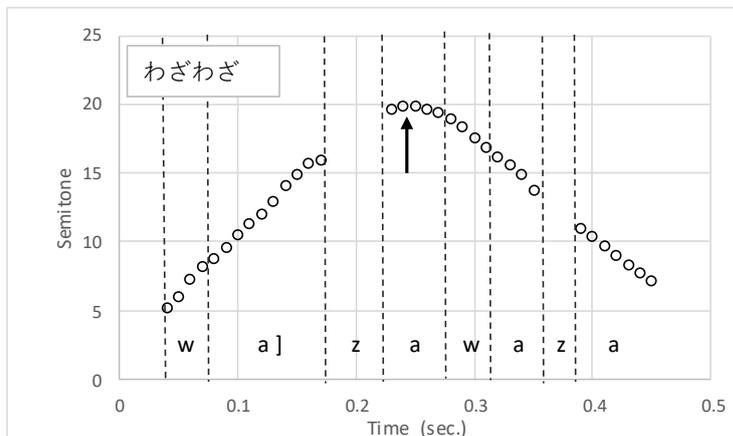


図5 フットリズムの発話における「遅下がり」

(4) アクセント拍でのピッチ上昇がない場合の「遅下がり」現象

これまで示してきた「遅下がり」の実例では、アクセント拍が上昇音調の場合のみであった。表3で示されているように、80%は上昇調であるが、下降調(F)、平坦調(L)の場合も

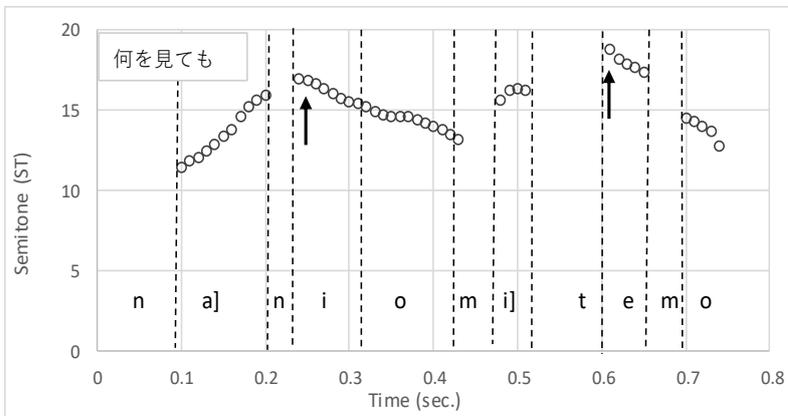


図6 ピッチ上昇がない場合の「遅下がり」例(1)

節)の連続も、2拍ごとにまとまりやすく、フット単位で発話される傾向がある。図5は「わ]ざわざ」と発話された例であるが、(わ]ざ)(わざ)のようなフット単位の発音であることが伺える。ピッチは最初のフット(わ]ざ)で上昇し、次の(わざ)で下降する「遅下がり」となる。こうした2拍単位の音調制御が「遅下がり」をもたらしているのではないかと考えられる。単に音調制御の overshoot が原因というだけではなく、フットという発話様式がアクセント拍に続く拍でピッチを降下させない原因になっている。

これまで挙げた「遅下がり」の例においても、図2:「夜遅く」では、(よ]る)(おそ)(く)、図4:「デパートの」では、(デ)(パ]ー)(トの)のようなフット発話と見なすことができる。

ここでは上昇調以外の場合の実例を見ていく。図6は、「何を見ても」の例であり、音声区分「何を」と「見ても」のいずれも「遅下がり」である。後者の「見ても」ではアクセントのある /mi/ の母音でピッチ上昇がないため、後続拍では高いピッチからの降下が

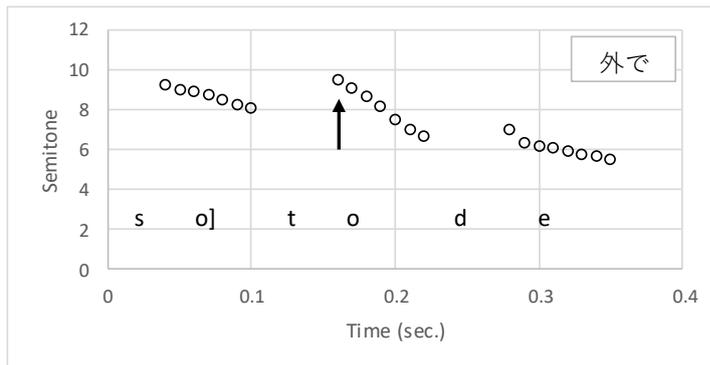


図7 ピッチ上昇がない場合の「遅下がり」例(2)

必要となる。これがアクセント拍の次の拍にアクセント位置がくる原因となっている。アクセント拍が上昇調のときは、後続拍でピッチ降下が認められない場合もあるが、明瞭な上昇調でない場合は後続拍の降下特性が必要になるのであろう。図7は、非卓立型発話の「外で」の例である。この場合は、2拍目母音 onset の位置が高くなって「遅下がり」的現象の様相を

呈する。これは、無声閉鎖音では開放後の母音のピッチが高まるという **micro prosody** 効果によるものと思われるが、アクセント拍に続く拍での急峻な降下は、アクセント感の付与にも寄与しているであろう。

4. 結びと今後の課題

アクセントの「遅下がり」現象に関して、自発発話音声データを分析した結果を報告した。「遅下がり」は、主にアクセント拍がピッチの上昇調で実現されることと関係が深く(それ故1型アクセントと)、また2拍がまとまったフット単位でのアクセント上昇音調が「遅下がり」の原因となっている。アクセント拍での十分なピッチ上昇があれば、後続拍でのピッチ降下は少なくてもよく、ピッチ上昇がなければ後続拍でのピッチ降下は必要になると考えられる。今後は、アクセント単位全体に渡るピッチの上昇、下降の動態の分析が必要であると考えている。

本研究は、科研費(基盤C)「声調言語と非声調言語のリズムに関する研究(代表者:益子幸江)」の支援を受けた。また、国立国語研究所コーパス開発センターの共同研究プロジェクト「コーパスアノテーションの拡張・統合・自動化に関する基礎研究」の共同研究者としても実施した。

参考文献

- Hasegawa, Y. and Hata, K. (1995). "The function of f0-peak delay in Japanese", Proceedings of the 21st Annual Meeting of the Berkeley Linguistics Society, pp.141-151
- 五十嵐陽介・菊池英明・前川喜久雄(2006).「報告書 日本語話し言葉コーパスの構築法」『第7章 韻律情報』, URL: pj.ninjal.ac.jp/corpus_center/csj/document.html
- 佐藤大和(2016).「共通語における動的音調形式とアクセント知覚」, 日本音声学会 第334回研究例会
- 佐藤大和(2017).「アクセント核のあとピッチの急峻な降下はあるか?—ピッチの動態特性とアクセント知覚—」, 2017日本音響学会春季研究発表会
- 佐藤大和(2018).「アクセント音調の諸相とその動態形式」, 「言語資源活用ワークショップ2018」予稿集, 2018年9月, 於国立国語研究所(発表予定)
- 杉藤美代子(1980).「“おそ下がり”考—動態測定による日本語アクセントの研究」, pp.201-229, 徳川宗賢編「論集日本語研究2 アクセント」, 有精堂

後部要素が3・4拍の複合名詞の核位置保存について —連濁との関係を中心に—

劉 汝源 (神戸市外国語大学大学院)
liu15645578@yahoo.co.jp

1. はじめに

東京方言における複合名詞のアクセント(以下A)は、Aの面から、「1単位の複合名詞」(以下A型)、「不完全複合名詞」(以下B型)、「2語連続」の3タイプに分類できる。A型は前部要素(以下X)と後部要素(以下Y)ともにA核が消え、全体のA核をYの語頭拍に置く。それに対し、「2語連続」はXとYの各語のA核位置が保存される。中間的なB型は、XのA核が消え、Yの核位置が保存される。

上野(1999)は窪菌(1997)の観点を補足し、Yの核位置保存はYの拍数や形態素数によって分類できると主張している。(1)Y \geq 5拍或いはY \geq 3形態素であれば、B型になる。例：南カリフォルニア \rightarrow 市立図書館。(2)ただし、Yが3・4拍かつ中高型であれば、B型になる場合がある。例：大和ナデ \rightarrow シユ。

このうち、(2)について、松森他(2012)によればYが4拍語の場合には、3拍語よりもB型はるかに多い。しかし、Yが4拍でも、外来語以外の中高型ではYの語末が特殊拍の場合は、A型となることが多い。特に、Yが「4拍」の「-2型」で、かつ語末が特殊拍の場合には、Yの中高型が保存されないことが多い。例：ホーゲ \rightarrow ン(方言) \rightarrow チバ \rightarrow ホーゲン(千葉方言)。(以下「語末特殊拍要因」)

また(2)について、『新明解日本語アクセント辞典』(2015)によると、Yが漢語二字の結合名詞であれば、「後部が中高型の語の場合に限り、もとの高さの切れめまで高い。但し、この場合も拍数の多いものや、無声化で中高型になった語は、高さの切れめが前にずれる」と記載している。例：地方 \rightarrow ほーの語頭に無声化が起って \rightarrow ほーになり、複合して奥羽地方 \rightarrow おおわ \rightarrow ほーになる。(以下「語頭拍母音無声化要因」)

上野(1999)は複合名詞のA保存について、一度「複合語化」(複合名詞A規則の適用)によりAが与えられた単語は、上位の2次複合語にYとして組み込まれても、そのA型が保存されると主張している。例：針 \rightarrow 師、庭 \rightarrow 師 \rightarrow 女針 \rightarrow 師、見習い庭 \rightarrow 師。(以下「複合回数要因」)

先行諸研究以外に、本発表では(2)について、「連濁」という要因を新たに指摘する。例：心 \rightarrow こ \rightarrow ろ、こ \rightarrow ろ \rightarrow 歌心 \rightarrow うた \rightarrow ご \rightarrow ろ、砂糖 \rightarrow さと \rightarrow う \rightarrow 角砂糖 \rightarrow かく \rightarrow ざ \rightarrow とう。(以下「連濁要因」)

2. 検証結果と分析

「連濁要因」を検証するため、『日本語発音アクセント新辞典』(2016)と『新明解日本語アクセント辞典』(2015)からYが3・4拍かつ中高型の複合名詞(Yが漢語または和語の合計786語、ただしYが3形態素以上の場合は除く)を抽出・分析し、「連濁要因」が正しいことを実証した。

抽出したすべての語を連濁・不連濁・連濁不関与の3種類に分ける：

このうち、不連濁語の「語頭拍母音無声化要因」によって、語頭拍母音の無声化がYの

核位置保存に影響を与えるため、「語頭拍母音が無声化する語」と「語頭拍母音が無声化しない語」に分類できる。不連濁語には「ライマンの法則」によって連濁しない語が約3分1を占めているが、連濁不関与ではなく、不連濁に入れた。

ア辞典の記載では、Yが中高型のみの場合と、中高型と平板型、頭高型、尾高型と共存する場合がある。元のYのアが複合名詞のアに影響を与える可能性があるため、この2つのタイプに分けて統計した。

複合名詞のアは以下の3つに分類できる。したがって、連濁とYの核位置と複合名詞のアの関係は表1のように示される。(A型: ~+○]○○, ~+○]○○○ B型: ~+○○]○, ~+○○]○○, ~+○○○]○ AB型: 1つの複合名詞にA型とB型が共存する)

表 1: 連濁と核位置の関係

		タイプ I : Y が中高型のみ			(e)タイプ II : Y が中高型とその他の型が共存する		
		A 型	B 型	AB 型	A 型	B 型	AB 型
総体数		206 (26.2%)	126 (16.0%)	125 (15.9%)	200 (25.5%)	3 (0.3%)	126 (16.0%)
連濁	(a)	65 (8.3%)	0	5 (0.6%)	121 (15.5%)	0	5 (0.6%)
不連濁	(b)Y の語頭拍母音が無声化する	1 (0.1%)	7 (0.9%)	28 (3.6%)	16 (2.0%)	0	67 (8.5%)
	(c)Y の語頭拍母音が無声化しない	62 (7.9%)	51 (6.5%)	51 (6.5%)	34 (4.3%)	0	28 (3.6%)
連濁不関与	(d1)Y の語頭が濁音	44 (5.6%)	5 (0.6%)	20 (2.5%)	18 (2.3%)	0	4 (0.5%)
	(d2)Y の語頭が母音	7 (0.9%)	28 (3.6%)	13 (1.7%)	2 (0.2%)	1 (0.1%)	14 (1.8%)
	(d3)Y の語頭子音が /n//m//j//r//w/	27 (3.4%)	35 (4.4%)	8 (1.0%)	9 (1.2%)	2 (0.2%)	8 (1.0%)

表1から以下のことがわかる。

- (a) **連濁** 連濁が起った語において、A型が圧倒的に多く (A型 94.9%)、タイプ I (A型 34.9%)・II (A型 65.1%) とともに関係なく B型が存在しない。AB型も極小 (AB型 5.1%)。
- (b) **Y の語頭拍が無声化** 不連濁語の中、Y の語頭拍母音が無声化する場合、AB型がもっとも多く、タイプ II に集中する (AB型 70.5%)。
- (c) **Y の語頭拍が無声化しない** Y の語頭拍母音が無声化しない不連濁語 (c) では、連濁語に比べると B型がかなり存在する (B型 51個 22.6% (連濁語 B型 0個))。
- (d) **連濁不関与** 連濁不関与語全体 (d) では B型の割合が Y の語頭拍が無声化しない語 (c) の割合より少し大きい (B型 71個 28.9%)。
- (e) **タイプ II** 連濁と語頭拍無声化を除くと、タイプ II では B型がほとんど存在しない。(120例中 B型 3例 2.5%)

以上の (a) ~ (e) について考察を加える。

(a) **連濁** B型が存在しないことから連濁が Y の核位置保存に影響することがわかる。

表 2 は『日本語発音アクセント新辞典』(2016) から抽出した連濁語 (109 個) の一部である。両辞典の連濁語の AB 型語が重複しているため、『新明解日本語アクセント辞典』(2015) から抽出した連濁語 (87 個) はここで省略する。Y の語境界を判断する基準は『明鏡国語辞典 第二版』(2011) による。

表 2:『日本語発音アクセント新辞典』(2016) から抽出した(a)連濁語

Y 語彙	Y 読み	Z 語彙	Z 読み	Y の保存	Y 語種	Y 語境界	Y タイプ
境	さか]い	国/県/潮/地境	~+ざ]かい	A 型	和語	3	I
月夜	つき]よ	星/朧月夜	~+づき]よ ~+づ]きよ	AB 型	和語	2+1	I
一つ	①と]つ	一つ一つ	①と]つび]とつ	A 型	和語	2+1	I
砂糖	さと]ー	赤/角/黒/氷/白砂糖	~+ざ]とー	A 型	漢語	1+2	I
胡椒	こしょ]ー	柚子胡椒	ゆずこ]しょー ゆずご]しょー	A 型	漢語	1+2	I
化粧	けしょ]ー	厚/薄/早/冬/夕/雪化粧	~+げ]しょー	A 型	漢語	1+2	I
焼酎	しょーちゅ]ー	芋焼酎	いもじょ]ーちゅー	A 型	漢語	2+2	I
提灯	ちよーち]ん	絵/小田原/岐阜/白張り/高/高張り/箱/酸漿/盆/弓張り提灯	~+ちよ]ーちん	A 型	漢語	2+2	I
正面	しょーめ]ん	裏正面	うらじょ]ーめん	A 型	漢語	2+2	I
沢山	たの]さん	盛り沢山	もりだ]の]さん もりだく]の]ん	AB 型	漢語	2+2	I
		子沢山	こだ]の]さん	A 型	漢語	2+2	I
心	ここ]ろ こころ]	遊び/魚/歌/絵/幼/男/乙女/親/気/匂/恋/子供/里/静/下/好き/漫ろ/旅/出来/手/情け/盗み/人/二/隔て/仏/真/水/娘/物/大和/世心	~+ご]ころ 気心 きごころ= もある	A 型	和語	3	II
鉢	はさ]み はさ]み	糸切/金/紙/刈り込み/木/裁ち/花鉢	~+ば]さみ	A 型	和語	3	II
袴	はか]ま はか]ま	行燈/伊賀/革/素襖/長/羽織/藤/雪袴	~+ば]かま 羽織袴 はおりばかま] もある	A 型	和語	3	II
刀	かた]な かた]な	押っ取り/小/血/手/鈍/腹切り/懐/守り/山刀	~+ちが]たな 血刀 ちがたな]	A 型	和語	2+1	II
		小刀	こが]たな こがたな] こがた]な	AB 型	和語	2+1	II
二重	②た]え ②たえ]	羽二重	はぶ]たえ は]ぶたえ	AB 型	和語	2+1	II
年寄り	としよ]り としよ]り]	奥/若年寄	~+ど]しより	A 型	和語	2+2	II
金持ち	かねも]ち かねも]ち]	大金持ち	おおが]ね]もち おおが]ねもち]	AB 型	和語	2+2	II
繕い	つくろ]い= つくろ]い]	毛/羽繕い	~+づ]くろい	A 型	和語	2+2	II

所帯	しょた]い しょ]たい	大/男/女/新/貧乏/寄合所帯	~+じょ]たい	A型	漢語	1+2	II
普請	しん= し]ん	仮/川/数寄屋/本/道/安普請	~+ぶ]しん	A型	漢語	1+2	II
巾着	きんちやく きんちやく]	磯/腰巾着	~+ぎ]んちやく	A型	漢語	2+2	II

表2の特殊な例は以下の3語である:羽二重 はぶ]たえ, は]ぶたえ Xの直後にア核がくる. 小刀 こがたな] は, 中高型以外の尾高型のア核位置保存が見られる. 大金持ち おおがね]もち は, A型にならないがア核位置保存もしない.

「語末特殊拍要因」 例外はあるが, 全体としてB型が少ない. 原因としては, 語末が重音節の場合, 語末音節の核を避けてA型になると考えられる. (1) Yが「4拍」の「-2型」で, かつ語末が特殊拍の場合には:A型176個74.6%, B型13個5.5%, AB型47個19.9%. B型例:若先生 わかせんせ]ー. (2) Yが「3拍」の中高型の場合:A型90個44.6%, B型19個9.4%, AB型93個46.0%. B型例:北日本 きたにほ]ん, 絶対多数 ぜったいたす]ー. 連濁語はA型と極小のAB型の組み合わせである. (A型94.9%, AB型5.1%)

「複合回数要因」 大金持ち おおがね]もち おおが]ねもち, 奥年寄 おくど]しよりのような例では, Yがア上の複合名詞だが, 連濁によってYのアが保存されなくなる.

削除例 1つだけの削除例がある. 連濁語としての 夕月夜 ゆーづき]よ は連濁してもYのア位置が保存される. 『明鏡国語辞典 第二版』(2011)によると, 夕月夜の語構成は ゆー・づきよ である. 40人の日本人大学生を対象に「夕月夜」のアと語境界についてアンケート調査を行った. 結果, 語境界について: ゆー・づきよ 18人45%, ゆーづき・よ 16人40%, どちらでも6人15%;アについて: ゆーづ]きよ 15人37.5%, ゆーづき]よ 21人52.5%, どちらでも4人10%. 4割りの人は 夕月夜 を ゆーづき・よ の語構成だと認識した. こうして, Yが3拍の「月夜 つきよ」ではなく1拍の「夜 よ」になる可能性が十分高い. 「ゆーづき・よ」の語構成では, 「ゆーづき]よ」になると考えられる. 本発表の「Yが3・4拍」の限定範囲から外れるため, ここでは総体数に入れず削除した.

(b) Yの語頭拍が無声化

表3では, Yの語頭拍母音が無声化する不連濁語のすべてのA型とB型の語, そして代表的AB型の語の一部を示している.

表3: 両ア辞典から抽出した(b)Yが語頭拍無声化する語の一部

Y語彙	Y読み	Z語彙	Z読み	Yの保存状況	Y語種	Y語境	Yタイプ	出典辞典
一つ	①と]つ	身一つ	み①と]つ み①]とつ	AB型	和語	2+1	I	NHK
宿舍	②ゆく]しゃ	国民宿舍	こくみん②ゆく]しゃ	B型	漢語	2+1	I	NHK
地方	③ほ]ー	カシミール地方	カシミール③ほ]ー	A型	漢語	1+2	I	NHK
試験	④け]ん	予備試験	よび④]けん よび④]けん	AB型	漢語	1+2	I	NHK
躑躅	⑤つ]じ ⑥つじ=	山躑躅	やま⑤]つじ やま⑥]つじ	AB型	和語	3	II	NHK
司教	⑦きよ]ー ⑧]きよー	大司教	だい⑦]きよー	A型	漢語	1+2	II	NHK
資金	⑨]きん ⑩]きん	回転資金	かいてん⑨]きん かいてん⑩]きん	AB型	漢語	1+2	II	NHK

機関	㊦かん ㊦かん	下部機関	かぶ㊦かん	AB型	漢語	1+2	II	NHK
期間	㊦かん ㊦かん	在学期間	ざいがく㊦かん ざいがく㊦かん	AB型	漢語	1+2	II	NHK
不孝	㊦こー	親不孝	おや㊦こー	B型	漢語	1+2	I	新明解
資金	㊦きん	越年資金	えつねん㊦きん えつねん㊦きん	AB型	漢語	1+2	I	新明解
畜生	㊦くしょー	犬畜生	いぬ㊦くしょー	B型	漢語	2+2	I	新明解
躑躅	㊦つじ ㊦つじ=	山躑躅	やま㊦つじ やま㊦つじ	AB型	和語	3	II	新明解
地方	㊦ほー ㊦ほー	奥羽地方	おおわ㊦ほー おおわ㊦ほー	AB型	漢語	1+2	II	新明解
試験	㊦けん ㊦けん	期末試験	きまつ㊦けん きまつ㊦けん	AB型	漢語	1+2	II	新明解
機関	㊦かん ㊦かん= ㊦かん	金融機関	きんゆう㊦かん きんゆう㊦かん	AB型	漢語	1+2	II	新明解
期間	㊦かん ㊦かん= ㊦かん	有効期間	ゆうこう㊦かん ゆうこう㊦かん	AB型	漢語	1+2	II	新明解

表3からわかるように、不連濁のYが語頭拍母音無声化の場合において、Yが1+2漢語なら、「語頭拍母音無声化」説に当てはまりやすく、AB型になる。例外もある：不孝 ㊦こーなどの語は、ア核位置が保存される場合がある。試験 ㊦けん (NHK) と試験 ㊦けん, ㊦けん (新明解) のような違いがあるが、NHKは生産性に忠実であることに対し、新明解が規則による対応関係に忠実であることが予測される。例：試験 ㊦けん→期末試験 きまつ㊦けん, ㊦けん→きまつ㊦けん (新明解)。躑躅 ㊦つじ ㊦つじ= は和語であり頭高型ではなく平板型を持つため、「語頭拍母音無声化要因」と無関係であると考えられる。

(c) Yの語頭拍が無声化しない 「連濁不関与語」と「Yが語頭拍母音無声化の不連濁語」を除き、連濁語と一般的な不連濁語(Yが語頭拍母音無声化しない不連濁語)を比較する。連濁語ではB型数が0個である。一方、一般的な不連濁語ではB型がかなり存在する(B型51個22.6%)。つまり、連濁が生じることによってYのア核位置が保存されなくなる傾向が見られる。

(d) 連濁不関与 連濁語の0個0%に比べるとかなり割合が大きいの(B型71個28.9%)。これも「連濁要因」の正しさを証明する。

(e) タイプII Yの中高型とその他のA型の併用によって、A型又はAB型になる(Yが頭高型、尾高型や平板型であれば、複合名詞のア規則によりA型になるのが普通である)。

3. 連濁と意味のまとまりとア

前節に連濁とアの関係を紹介したが、本節では他の関連する現象から、連濁がYのアに影響する理由について考察する。

3.1 格関係と連濁とア

表4は鈴木(2008)が奥村(1984)、金田一(1976)、佐藤(1989)、中川(1966)の研究に基づいて、まとめた連濁細則をさらに整理した表である。

表4: 格関係と連濁とアの関係(Yが2拍動詞連用形転成名詞の複合名詞)

	連濁を生ずる		連濁を生じない
	(2) 目的格・主格		
	(2a) その動作の結果生ず	(2b) 「…ヲ…スル」	(2c) 「…ヲ…スルコト/モノ」

	(1) 連用修飾格	る具体物・対象を示すとき	コト」の意識が薄く、 一語としての意識が 強いもの	ヒト」の目的格・主格
書き		人相書き にんそーがき＝		物書き ものかき] ものか]き
張り			目張り めばり＝	傘張り かさ]はり
きり	辻斬り つじぎり＝			爪切り つめき]り つめきり]

表4の(2c)の目的格・主格では連濁を受けない起伏型になる。一方、意味のまとまりが強い、(2a) (2b)の目的格・主格と(1)の連用修飾格では、連濁を生じて、アも平板型になる。

3.2 Yが2拍の同音異義語

窪菌(1995)は「日本語の複合語を特徴づける音韻現象である連濁も意味論による意味制約を受けている」と述べている。尾鰭 お]ひれ は「尾と鰭」の意味であり、おびれ＝が「尾の鰭」の意味である。山川 やま]かわ は「山と川」の意味であり、やまがわ＝が「山の中での川」の意味である。この2つの例から、連濁と不連濁の対立が意味の区別に繋がっていることが分かる。おひれ、やまかわ は並列複合名詞であり連濁を生じず、アが起伏型になる。一方、まとまりが強い おびれ、やまがわ は連濁してアも平板型になる。

4. まとめ

後部要素が3・4拍かつ中高型語の場合において、連濁を生じると、後部要素のアクセント核位置が保存されず、アクセント核は後部要素の語頭拍に置くという傾向がみられる。すなわち、複合名詞が「不完全複合名詞」から「1単位の複合名詞」に変わる。これは、連濁が起こると、複合名詞の意味のまとまりがより強くなるのがその原因である。

参考文献

- 秋永一枝(編)(2015)『新明解日本語アクセント辞典』三省堂
- 上野善道(1999)「複合名詞後部要素のアクセント型保存」『言語と文化の諸相』岩手大学人文社会科学部 pp.159-212
- 奥村三雄(1984)「連濁」『日本語学』3-5 明治書院 pp.89-98
- 北原保雄(編)(2011)『明鏡国語辞典 第二版』大修館書店
- 金田一春彦(1976)「連濁の解」*Sophia Linguistica II* 上智大学大学院言語学研究室 pp.1-22
- 窪菌晴夫(1995)『語形成と音韻構造』くろしお出版
- 窪菌晴夫・伊藤順子・A. Mester(1997)「音韻構造から見た語と句の境界—複合名詞アクセントの分析—」『文法と音声』くろしお出版 pp.147-166
- 佐藤大和(1989)「複合におけるアクセント規則と連濁規則」『日本語の音声・音韻(上)』明治書院 pp.233-265
- 鈴木豊(2008)「動詞連用形転成名詞を後部要素とする複合語の連濁」『文京学院大学外国語学部文京学院短期大学紀要』8 pp.212-234
- 中川芳雄(1966)「連濁連清(仮称)の系譜」『国語国文』35-6 京都大学文学部 国語学国文学研究室 pp.302-314
- 松森晶子・新田哲夫・木部暢子・中井幸比古(2012)『日本語アクセント入門』三省堂
- NHK放送文化研究所(2016)『日本語発音アクセント新辞典』NHK出版

アクセント句頭の Fo 上昇は条件異音ではない

前川 喜久雄 (国立国語研究所・音声言語研究領域)
kikuo@ninjal.ac.jp

1. 問題のありか

標準日本語（東京語）のアクセント句（以下 AP と略することがある）はいわゆる句頭の上昇によって特徴づけられる。そして、その上昇のあり方に句頭の音節の特性が影響を及ぼすこともまた広く知られている。この問題を最初に指摘したのはおそらく服部(1955)である。服部は「コマギレ」「コマゴマ」と「コーバン」「コンダン」を比較すると、後者ではすべてのモーラを高く発音する話者が少なくないと報告している。この指摘はその後の日本語韻律研究で、理論的な立場を越えて広く受容された。生成音韻論による日本語アクセントの記述でも、この指摘に従った定式化が施されている(Haraguchi 1977)。なお服部は、句頭音節が長母音ないし撥音の場合を例としていたが、現在では、句頭音節が sonorant な重音節（すなわち長母音・二重母音・撥音）である場合に生じる現象という理解が定着していると思われる。本稿でもその立場をとる。

ところで、Haraguchi による定式化は、句頭（正確には語頭）の H tone を L tone に書き換える音韻規則の形をとっていた。Pierrehumbert & Beckman (1988 以下 P&B)はこれを問題視して、句頭の境界 L tone (L%) の条件異音として扱うべきであると主張した。すなわち H が L に、あるいは L が H に記号レベルで書き換えられるのではなく、句頭が sonorant な重音節であると、本来のはっきりと低い L (strong L) が高めの L (weak L) として音声的に実現されるとの主張である。

P&B の主張は実験的な証拠に基づいたものであり、服部の主観的な観察よりも信頼性が高いことはたしかである。しかし P&B の実験では句頭の音節を、sonorant な重音節かそれ以外かの二分法でしか処理しておらず、それ以外の分類の可能性を最初から排除している点に問題がある。今回この点をより詳細に検討した結果、Haraguchi とも P&B とも異なる解釈に到達したので以下に報告する。

2. データ

『日本語話し言葉コーパス』のコア部分を RDB 化した CSJ-RDB (Koiso et al. 2014) を利用して分析を進める。CSJ-RDB には東京地域出身話者による音声約 44 時間分（約 50 万語に相当）が記録されており、X-JToBI 韻律ラベル(Mackawa, et al. 2002)を含む豊富なアノテーション情報を SQL 言語で検索できる。今回はモノローグ（学会講演および模擬講演の 2 種類）を対象を絞って分析することにした。

Fig. 1 は X-JToBI による AP アノテーションの概要を示している。AP の始端は %L、終端は L% で区切られており、いわゆる句末イントネーション (boundary pitch movements) は L% に後続する H%, HL% などで表現される。H-は句頭の上昇の頂点を示す tone であり、有核

AP においては核によるピッチ下降に含まれるピッチの頂点ないし変曲点に A が付与される。以上は X-JToBI で利用されている記号であるが、本稿ではこれらに替えて Fig. 1 で X-JToBI 記号の上に表示されているアルファベット 3 文字の記号を用いる。ILT は initial low tone、IHT は initial high tone、Acc はアクセント、FLT は final low tone の略称である。

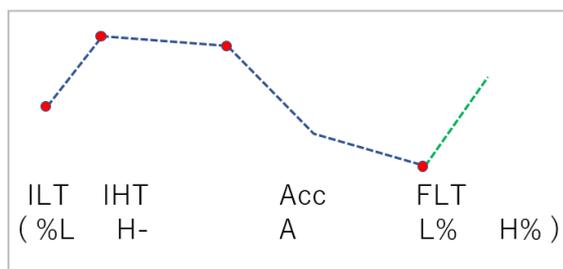


Fig. 1: X-JToBI による AP アノテーションと本研究で用いる記法

本研究の手段となる句頭の上昇は ILT と IHT との関係として表現される。

以下で分析するデータからは、AP 長が 1 モーラの AP (主にフィラー) を除外し、Acc が第 1 ないし第 2 モーラに付与されているサンプルも除外した。さらに IHT ないし Acc に対応する Fo 値が母音の無声化などによって欠損値となっているサンプルも除外した。最終的に分析対象として残ったのは 29,257 個の AP である。これらの AP に記録された IHT および Acc の Fo 値は 10 を底とする対数に変換した後、話者ごとに Z スコアに変換した。この処理によって性別による Fo 差はほぼ完全に正規化されるので、以下では男女話者のデータをプールして分析する。

3. 分析

3.1. P&B 分析の追試

P&B の分析はかなり特殊なものなので、最初に CSJ のデータを用いた追試を行う。P&B は、データ中の連続する 2 アクセント句を対象として、第 1 句における IHT (phrasal H) の高さを横軸に、2 句の境界における ILT (boundary L%) の高さを縦軸にとった散布図を作成して、句頭に sonorant な重音節をもつサンプルの ILT (weak L) とそうでないサンプルの ILT (strong L) とでは、前者の方が縦軸の下方に位置することを確認し、そのうえでデータ全体の回帰直線からの残差を話者ごと・音声特徴ごとに計算し、4 名中 3 名の話者で平均残差が strong L では負の値、weak L では正の値をとることを報告している。

これと同じ分析を CSJ データに適用した結果を Fig. 2, 3 に示す。追試のためにこれらの図では Fo の単位を P&B 同様 Hz に設定した。Fig. 2 は第 1 句が無核の場合、Fig. 3 第 1 句が有核の場合である。両図とも上部パネルに散布図と回帰直線 (および 95% 信頼区間)、下部パネルに Strong L と weak L の回帰直線からの平均残差 (および 95% 信頼区間) を示している。散布図では第 1 句第 2 句の境界に生じる L における strong/weak の別を ○ と △ で表示している。これらの図を P&B 第 2 章の図 (p.31f) と比較すると、CSJ のデータを用いた場合も P&B の分析と同一といってよい結果が得られたことが分かる。

なお CSJ コアモノログ部分の話者数は 177 名であるが、個人レベルで Fig. 2-5 と同じ分析を行えるだけのサンプルを持つ話者は少ない。N>15 のサンプルが得られた 4 名を分析したところ、3 名で P&B の主張に一致する結果が得られた。残り 1 名のデータには大きな外れ値が含まれており、それを除外すればやはり P&B の主張にかなう結果が得られた。

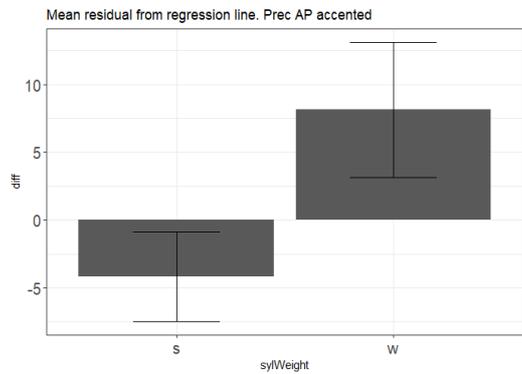
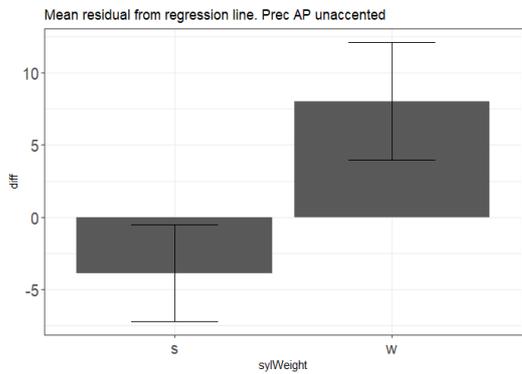
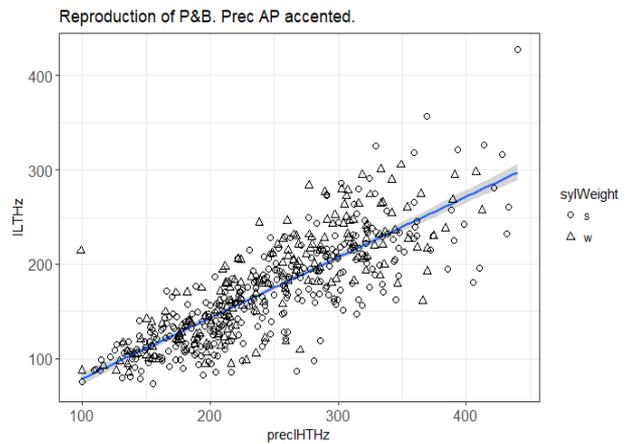
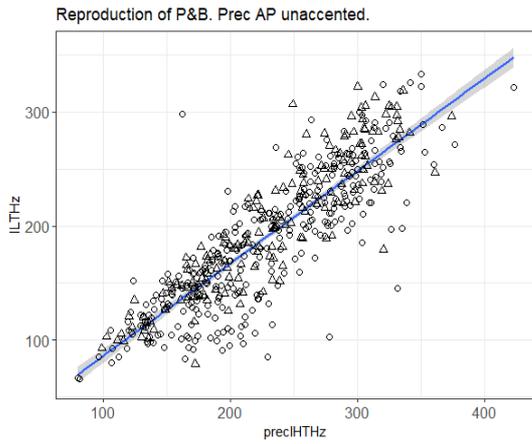


Fig. 2: 無核第1APのIHT(横軸)とAP境界のILT(縦軸)の散布図および回帰直線(上パネル). StrongLおよびweak Lの回帰直線からの残差(下パネル). 単位は Hz. 散布図の○と△は strong/weak に対応.

Fig. 3: 有核第1APのIHT(横軸)とAP境界のILTの散布図および回帰直線(上パネル). StrongLおよびweak Lの回帰直線からの残差(下パネル). 単位は Hz. 散布図の○と△は strong/weak に対応.

3.2. 句頭音節の細分類

前節の分析で CSJ コアが P&B のデータと質的に共通したものであることが確認できた。本節では音節の分類が句頭の上昇量にどの程度影響するかという本来の問題を分析する。P&B による分析における重要な問題点は句頭音節の特性を sonorant な重音節かそれ以外かという二分法でしか分類していない点にある。そこでまず比較のベースラインを設定するために、AP の句頭音節に sonorant な重音節(H~Heavy)かそれ以外(L~Light)かという二項分類を適用した場合の句頭上昇の平均値を比較した。結果を Fig. 4 に示す。縦軸は IHT - ILT の値(対数 Z スコア)であり、エラーバーは 95%信頼区間である。予想どおり Light > Heavy という関係が生じている。

次に「sonorant な重音節」を「長母音・二重母音・撥音」に下位分類し、またそれ以外の軽音節から促音を別扱いすることで句頭音節を 5 種に分類し、それぞれの平均上昇量を比較した。二重母音の定義には諸説あるが、ここでは/C₀ai/のみを二重母音に認定した。

Fig.5 に分析結果を示す。横軸は句頭音節のクラスであり、句頭上昇の平均値の下降順にソートされている。H は長母音 (Heavy でないことに注意)、D は二重母音、N は撥音、Q

は促音、そしてLはすべての軽音節である。

Fig.5からはふたつの重要な事実が読みとれる。第一に sonorant な重音節である N, D, H の句頭の上昇量に対する効果は均一とは言い難い。分散分析と下位検定(Tukey HSD)を行うと、すべての水準対のうち有意差(5%)が認められないのは H と D の対だけであった。

第二に句頭音節が促音 Q を含む重音節である場合には、軽音節全般(L)よりもさらに大きな上昇が生じている。これらの事実は P&B 流の二分法の妥当性を疑わせるものである。

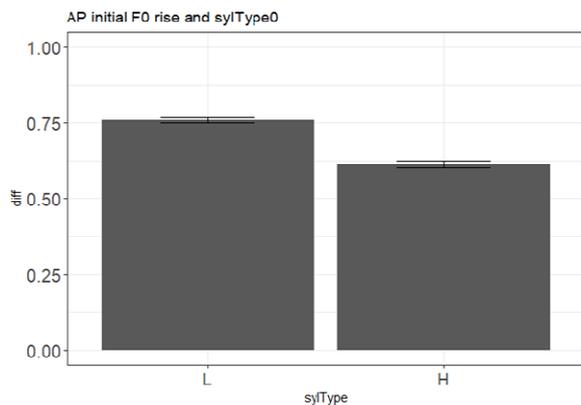


Fig.4: AP 頭の音節を sonorant な重音節 (H~Heavy)とそれ以外(L~Light)に分類した場合の句頭の Fo 上昇量の比較. 縦軸は Z スコア化された対数 Fo 値.

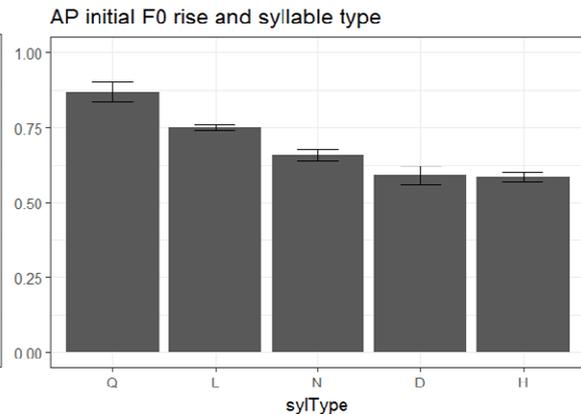


Fig.5: Sonorant な重音節を長母音(H), 二重母音(D), 撥音(N)に細分化した場合. Qは促音. 縦軸は Z スコア化された対数 Fo 値.

3.3. 句頭音節の一層の細分類

そこで軽音節全般 (Fig. 4, 5 の L) にも下位分類を施したうえで再度分析を実施した。Table 1 の第 2 列に下位分類の定義を示す。これは句頭から 2 番目のモーラの属性 (いわゆるモーラ音素にあってはその種類、それ以外のモーラにおいてはモーラ頭子音の特性) に関わる分類である。Table 1 の第 3 列には該当する AP の例が 2 例ずつ示されており、第 4 列にはサンプル数が示されている。Fig. 6 に音節種別ごとの句頭の上昇平均量を示す。表示方法は Fig. 5 と同一である。

Table 1: 句頭音節の一層の細分類

記号	説明	該当する AP の例	N
Q	促音	発表します／実験を通して	1,787
N	撥音	音楽を／弁別が	3,965
H	長母音	乳児が／報告が	4,944
D	二重母音	海外の／対象とした	1,171
Vwl	母音連鎖	変えた／二オクターブ	857
App	接近音	では／というより	877
Nasal	鼻音	どのように／つまり	3,560
Fric_u	無声摩擦音・無声破擦音	一段階の／マスキング用の	3,194
Fric_v	有声摩擦音・有声破擦音	気づいて／馴染みがあると	939
Stop_u	無声閉鎖音	聞き分けに／求めました	5,657
Stop_v	有声閉鎖音	普段から／全ての	1,794
Flap	弾き音	あるいは／彼らは	3,483

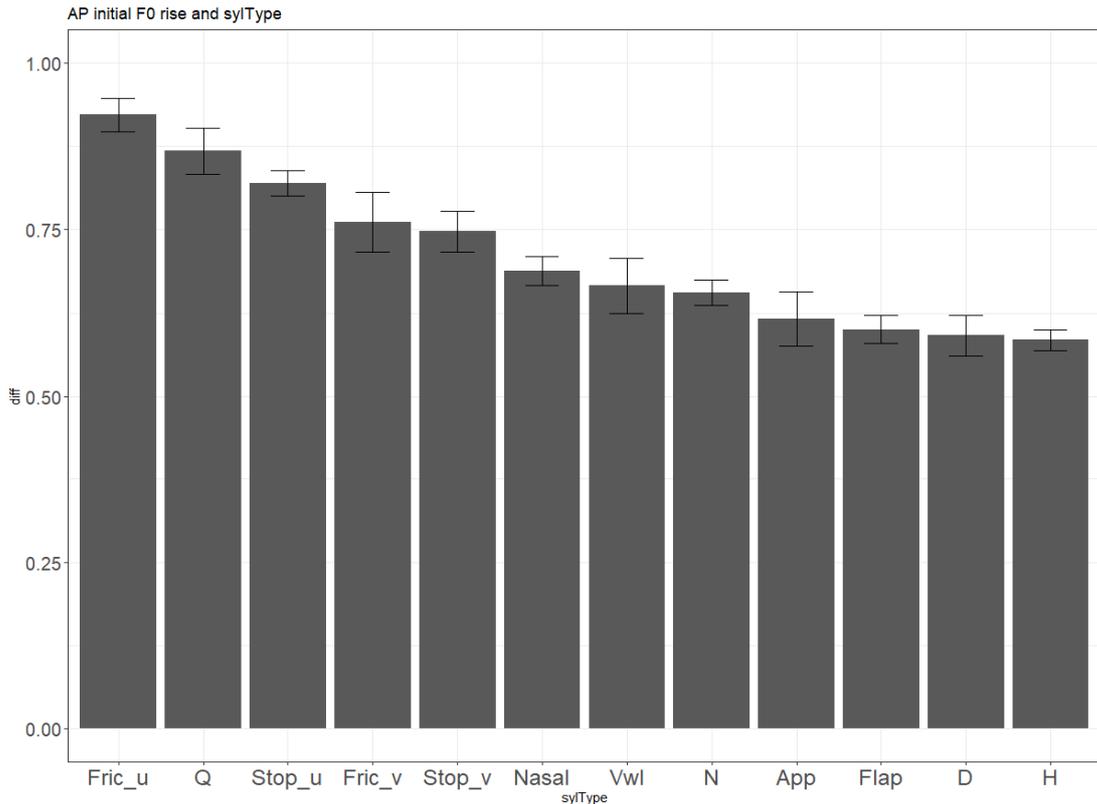


Fig. 6: AP 頭の 2 モーラ境界の特性を一層詳細に分類した場合の句頭の Fo 上昇量の比較. 縦軸は Z スコア化された対数 Fo 値.

ここでは注意すべき事実を三つ指摘できる。まず句頭の上昇量は音節の種類によって変化しているが、その変化は連続的であって、二項分類の存在を正当化できるような不連続性は見あたらない。分散分析と下位検定の結果、Fig. 6 の横軸において隣接する音節対 (H と D, D と Flap 等々) には有意差(5%)を示す対がひとつもないことが確認できている。

次に sonorant な重音節における句頭の上昇量に注目すると、長母音と二重母音における上昇量は最小となっているが、撥音における上昇量は、弾き音(Flap)や接近音(App)よりも大きくなっている。これは sonorant な重音節を特別視する二分法に対する明瞭な反証である。

最後に図の横軸における分節音の配置に注目すると、図の右半分 (H から Nasal まで) に sonorant な分節音が位置し、左半分 (Stop_v から Fric_u まで) には obstruent な分節音が位置している。句頭上昇量の平均値の下降順にソートされている横軸がこのような特性を示すことは、句頭上昇の量が当該 AP の第 1 モーラと第 2 モーラの境界に位置する分節音の阻害性に影響されている可能性を示唆している。

4. 議論：モーラ境界の明瞭性

Fig.6 に観察された連続的な変化は音韻規則でも条件異音でもなく、むしろ調音結合現象である可能性が高い。つまり、句頭の上昇の弱化ないし消失は AP の第 1 モーラにリンクされた L tone が第 2 モーラにリンクされた H tone による逆行同化を被り、L 本来の低さを失

って H に漸近していく現象とみなすことができる。今回の分析結果は、この同化が音節の軽重にではなく、AP の冒頭の第 1 第 2 モーラ間の境界の音声学的な明瞭性に深く関わっていることを示唆しているように思われる。以下、この可能性について論じる。

Fig.6 の右端には長母音と二重母音が位置している。これらの重音節内部のモーラ境界が明瞭でないことは明かである。長母音・二重母音と撥音の間には弾き音(Flap)と接近音(App)が位置している。このうち接近音は音声学的には母音に近い性格の音であるから、やはりモーラ境界が明瞭とはいいいにくい。弾き音は日本語子音中最も持続時間の短い音であり、その調音は前後の母音連続の調音のうえに舌尖のすばやい上下運動が重畳されたものと考えられている。従ってやはりモーラ境界の音声学的特徴は顕著とはいいいにくい。

次に Fig. 6 横軸の左端に注目する。最左翼に位置するのは無声摩擦音(Fric_u)であり、これに撥音(Q)と無声閉鎖音(Stop_u)が続いている。無声摩擦音と無声閉鎖音とを比較すると、sonority は無声閉鎖音の方が低いとみなすのが一般的である。しかしモーラ境界の明瞭性という観点からすれば、無声ノイズとして実現される無声摩擦音の方が、大部分が無音区間として実現される無声閉鎖音よりも、より明瞭に境界の存在を示している可能性がある。促音には無声閉鎖音と無声摩擦音の両タイプが存在するから、促音が両者の中間に位置するのは自然である。また摩擦音と閉鎖音の双方で、無声音の方が有声音よりも左端寄りに位置しているのも子音の音声学的ないし音響学的性質から考えて妥当な結果である。このように Fig. 6 の横軸は AP 冒頭のモーラ境界の明瞭性の尺度を構成していると考えられる。

最後に残る問題は、AP 冒頭のモーラ境界の明瞭性が、なぜ ILT の IHT への同化に影響するのかという根本問題である。この問題について筆者は現時点ではっきりとした解答を持ちあわせていない。推測を述べれば、第一に検討すべきは AP 冒頭 2 モーラの持続時間長が関係している可能性であろう。今後の課題としたい。

謝辞: 本稿の内容は東京外国語大学大学院で連携大学院教授として行っている日本語韻律全般に関する講義の一部である。未整理な内容の多い講義を毎回聴講してくれる学生諸君に感謝する。

参考文献

斎藤純男『日本語音声学入門【改訂版】』三省堂, 2006.

服部四郎「音韻論から見た国語のアクセント」国語研究, 2, 1955.

Haraguchi, Shosuke. *The Tone Pattern of Japanese: An Autosegmental Theory of Tonology*. Tokyo: Kaitakusha, 1977.

Koiso, Hanae, Yasuharu Den, Ke'nya Nishikawa and Kikuo Maekawa. "Design and development of an RDB version of the Corpus of Spontaneous Japanese." *Proc. LREC 2014*, pp. 311-315, 2014.

Maekawa, Kikuo, Hideaki Kikuchi, Yosuke Igarashi and Jennifer Venditti. "X-JToBI: An extended J_ToBI for spontaneous speech." *Proc. ICSLP2002*, Denver, pp.1545-1548, 2002

Pierrehumbert Janet B. and M. E. Beckman. *Japanese Tone Structure*. The MIT Press, 1988.