

親密度の低い 2 字漢語と 4 字漢語短縮形のアクセントの生成*

The Accent Production of Unfamiliar Two-Character Sino-Japanese Words and Truncated Sino-Japanese Compounds

李墨彤

Motong LI

北京外国語大学

Beijing Foreign Studies University

ABSTRACT. This paper investigates the accent distributions of two-character Sino-Japanese words and truncated Sino-Japanese compounds, using data from unfamiliar word read-aloud tasks participated by Tokyo metropolitan dialect speakers. The results show that the accent distribution of two-character Sino-Japanese words highly depends on the mora number of the second character, which coincides with previous studies examining accent dictionaries, and truncated Sino-Japanese compounds are generally more likely to be pronounced unaccented than two-character Sino-Japanese words. An analysis based on Stochastic Optimality Theory is proposed to explain different accent distributions and variations observed in the task results.

Keywords: accent distribution, Sino-Japanese word, truncated compound, Stochastic Optimality Theory

1. はじめに

日本語のアクセント型はその音韻・形態構造から見ると非対称的な分布を成している場合が多く、アクセント辞典の項目を統計した研究ですでに明らかとなっている。特に外来語の場合、構造別にどのようなアクセント型がより好まれるかは無意味語・親密度の低い単語などを利用した読み上げ調査の結果によって実証されている。一方、日本語の語彙を多く占める漢語も、その語の音韻・形態構造によって特定のアクセント型に偏る傾向があるが、読み上げ調査など直接母語話者のアクセント感覚に対する調査がほとんどない。産出調査は母語話者が持つ音韻知識を反映する有力な証拠となるため、研究する価値が大きいと考えられる。

また、実際の言語データはバリエーションを伴うことが普通であり、近年制約ベースモデルに基づく一連の研究で取り上げられている。日本語漢語のアクセントの場合、そのバリエーションは制約ベースモデルにどう説明されるかは非常に興味深い。

2. 先行研究

漢語のアクセント型がアクセント辞典に占める割合について、最上他(1999)は『NHK 日本語発音アクセント辞典 新版』(1998 年版；『NHK98』)のデータをモーラ数別に統計している(図 1；2~4 モーラ語のみ抜粋)。図 1 を見ると、2 モーラ語と 4 モーラ語はそれぞれ①型¹と②型に偏っているのに対して(どちらも 80%以上)、3 モーラ語は②型と①型の割合の差が小さい。ただし、最上他(1999) (1)と小川(2006)のデータが示しているように、3 モーラの漢語はその形態構造

□②型 ■①型 ▨③型 □その他

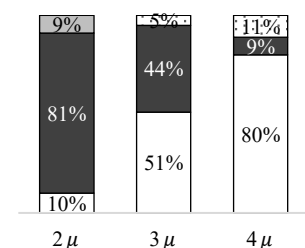


図 1 モーラ数別から見た漢語のアクセント型の割合

を見ると、 $\mu\#\mu$ は②型、 $\mu\mu\#$ は①型に偏っているという傾向が観察できる($\#$: 形態素境界)。

(1) 形態構造別に見た 3 モーラの漢語の各アクセント型の割合(最上他 1999 より)

構造	全語数	①型			②型			③型		
		語例	語数	割合	語例	語数	割合	語例	語数	割合
$\mu\mu\#$	4573	温度	3050	66.7%	会社	1347	29.5%	得意	98	2.1%
$\mu\#\mu$	4465	家族	986	22.1%	地球	3276	73.4%	自由	196	4.4%
								道具	78	1.7%
								地獄	7	0.2%

2 字漢語のアクセント型の傾向をまとめると、第 2 文字が 1 モーラ(1 軽音節)であれば①型、2 モーラ(2 軽音節か 1 重音節)であれば②型の割合が多い。ただし、辞書データに観察されているこの傾向は、母語話者が今まで見たことのない漢語を読み上げた時のアクセント型の傾向と一致するかどうかは不明である。

一方、一般の 2 字漢語と似たような構造として、4 字漢語の 2 字短縮形(例：音楽^{おん だい}大学；—：略された文字)がある。漢語の複合名詞の短縮形のアクセント規則として、秋永(1981)は(2)のようにまとめている。

(2) (漢語) 複合名詞の短縮形のアクセント規則(秋永 1981)

2 モーラ語：頭高型(例：市議^{し ぎ}会議員①)

3 モーラ語：省略意識の薄れるものが多く、2 字漢語の一般的な法則に準じる(例：健康^{けん ぼ}保険①、労働^{ろう せ}組合①、化学^{か せん}繊維②、空中^{くう ちゅう}輸送②)

4 モーラ語：原則として平板型(例：特別^{とく べつ}急行②、高等^{こう とう}学校②)

(2)のアクセント規則は 2 字漢語のアクセント型の傾向に似ているが、母語話者が今まで見聞きしたことのない、省略意識の高い短縮語に対するアクセント付与の傾向を報告した研究は管見の限りまだないため、それは 2 字漢語のアクセント型とまったく同様の傾向になるかどうかは検証が必要である。実際に、外来語複合語の短縮形は多くの場合平板型になること(秋永 1981、森 2002、Ito & Mester 2016；例：パソコン②、ジープ②)を考えると、4 字漢語短縮形が 2 字漢語よりも全般的に平板型になりやすい可能性は十分にある。

先行研究を踏まえ、本稿は親密度が低いと考えられる 2 字漢語と 4 字漢語短縮形の調査語を用いて、首都圏方言話者に読ませ、アクセント付与の傾向がどうなるかを語の音韻・形態構造ごとに調べる。その際、2 字漢語と 4 字漢語短縮形のアクセント型の傾向に差が生じるかどうかに注目する。また、制約ベースモデルを利用して、調査結果に観察された各構造の差異とバリエーションを統一的に説明する。

3. 読み上げ調査

3.1. 調査語

まず、調査語²の音韻・形態構造について、2 字漢語と 4 字漢語短縮形はどちらも 9 種のものに限定している(H：重音節；L：軽音節；-：4 字漢語短縮形の形態的境界)。2 字漢語の親密度について、『日本語の語彙特性』(天野・近藤 1999, 2000)における「文字音声単語親密度」の指標(7 段階)が 2.5 以下のものに限定している(平均 1.59；標準偏差 0.31)。短縮語のもととなる 4 字漢語について、一部インターネット上にあるものに加えて、基本的には『情報・知識 imidas' 98』(Imidas 編集部 1997)と『新和英大辞典第 5 版』(渡邊他 2003)から選出した。筆者が短縮語の各構造に合わせて、4 字漢語の第 1・第 2 要素の第 1 文字をそれぞれ取って短縮語を作成した。すべての短縮語を検索エンジン Google で調べ、

ヒット数が高くないことを事前に確認した。

また、漢語の場合、語全体のアクセント型が特定の形態素(特に第 2 要素)によって指定されることがある(最上他 1999; 例:「接写」「映写」「実写」など「写」を第 2 要素とする語は基本的に①型)。この影響を避けるために、調査語の第 2 要素について、その第 2 要素を持つ 2 字漢語の総数が『日本語の語彙特性』において 2 語以下か、単語のグループが特定のアクセント型に偏っていないものに限定了した。

3.2. 調査の概要

読み上げ調査は、2018 年 8～9 月に行われた。調査参加者は全員首都圏方言話者(12 歳まで東京都・神奈川県・埼玉県・千葉県 of いずれかの地域で生育した人)であり、合計 22 人いる(男性 9 人、女性 13 人; 平均 33 歳)。調査語を 2 グループに分け(1 グループに約 36 語の 2 字漢語と 33 語の短縮語が入る; 各構造の項目数をバランスよく配置)、参加者 1 人につき、1 グループの調査語を提示文の中に入れてランダム順に読ませた。場面設定として、2 字漢語は「あまりなじみのない漢語を友達に説明します」、短縮語は「このように長い単語を略す人がいます。あなたはこの略語をどう読みますか」にした。提示文について、2 字漢語の場合は「○○は△△△△という意味／のこと／…です」のような漢語の意味を解釈する文(例:「^{じゃひ}邪飛は^{やきゅう}野球のファウルのことです」)、短縮語の場合は△△△△を略して、○○です>の文(例:「^{かとうきょうそう}過当競争を略して、^{りやく}過競です」^{かきょう})になっている(すべての漢字に振り仮名付き)。

調査が始まる前に、アクセントとは何かについて簡単に説明した。録音の手順として、調査語および提示文をランダムに調査対象に示し、それらを一通り確認した後、提示文を読み上げて録音するように求めた。その際、赤く表記されている調査語を自然だと思ふアクセントで発音するように求めた。録音が終わった後に、筆者が録音データを聞いて、2 字漢語と短縮語のアクセント型を書き取った。

3.3. 調査結果と考察

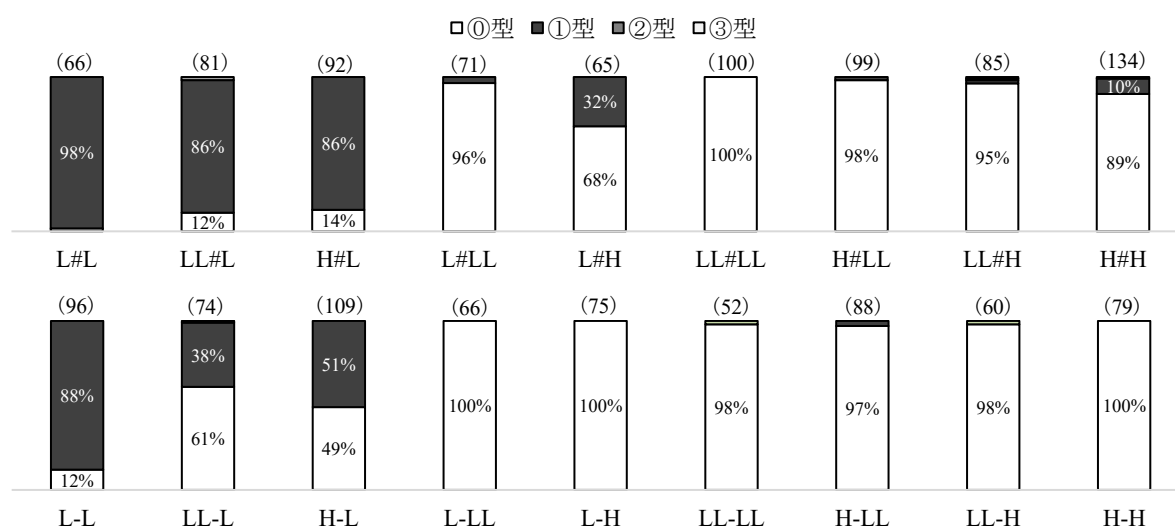


図 2 2 字漢語(上図)と短縮語(下図)の調査結果

音韻・形態構造ごとに、調査参加者が産出したアクセント型の割合(各構造のサンプル数(括弧内の数字)を分母としたもの)を図 2 に提示している。2 字漢語の結果を見ると、第 2 要素が 1 モーラであれば①型、2 モーラであれば②型になりやすいという傾向が再び観察されている。詳しく比較するため

に、『NHK98』における L#L の統計データに加え、『日本語発音アクセント辞典 改定新版』（1985 年版；『NHK85』）における 3・4 モーラの 2 字漢語の各構造の統計データを(3)に提示する。L#L の場合、読み上げ調査は辞書データよりさらに①型に偏っている。3 モーラ語の場合、LL#L と H#L の①型、L#LL の①型は辞書データよりも優位性が顕著に現れているが、L#H に関しては読み上げ調査では①型が 32%を占め、①型(68%)の優位性は辞書データ(83%)ほど顕著ではなかった。4 モーラ語の場合、H#H(①型の割合が辞書データと同様)を除き、LL#LL、H#LL、LL#H の読み上げ調査の結果はいずれも①型の割合が 95%以上であり、辞書データよりも偏りが顕著に出ている。まとめると、2 字漢語の各構造の調査結果における優勢なアクセント型は辞書データと一致した傾向を示しながらも、多くの構造において、優勢なアクセント型への偏りは辞書データよりもさらに顕著に現れている。

L#H の①型の産出が辞書データほど観察されなかった原因として、まず、ほかの 2 字漢語の構造と比べて、L#H のサンプル数が比較的少なく、アクセント型の割合が①型の数に影響されやすいことが考えられる。また、2 字漢語で前部が 1 モーラの場合、名詞的意味を持つ語には頭高型が多いと秋永(1981)が述べていることから、L#H のアクセント付与は調査語の意味にも干渉されている可能性がある。さらに、2 音節であれば、①型が優勢な L#H と H#H においても(程度の差はあるが)①型が一部産出されていることから、音節数による影響が考えられる。

(3) 辞書における 2 字漢語のアクセント型の分布(L#L のデータは最上他(1999)より抜粋；3・4 モーラ語のデータは小川(2006)より抜粋)

	L#L	LL#L	H#L	L#LL	L#H	LL#LL	H#LL	LL#H	H#H
①型	875(82%)	187(50%)	1750(78%)	138(19%)	230(15%)	0(0%)	156(10%)	6(1%)	326(8%)
②型	87(8%)	134(36%)	494(22%)	586(80%)	1298(83%)	381(96%)	1460(90%)	488(89%)	3832(89%)
その他	103(10%)	54(14%)	13(1%)	11(1%)	44(3%)	17(4%)	12(1%)	54(10%)	169(4%)

一方、短縮語の結果を見ると、L-L、LL-L、H-L、L-H、H-H の 5 構造の場合、それぞれ対応した 2 字漢語の構造よりも①型の割合が多く、特に 3 モーラ語の LL-L、H-L、L-H において①型の割合が 2 字漢語との差が大きいことがわかる。秋永(1981)は 3 モーラの短縮語は省略意識の薄れたものが多く、2 字漢語の一般的な法則に準じると述べているが((2)を参照)、省略意識の高い 3 モーラの短縮語は 3 モーラの 2 字漢語よりも①型に偏りやすいことは今回の読み上げ調査ではっきりとなっている。ただし、LL-L(①型が①型より割合が顕著に多い)と H-L(①型と②型がほぼ同様の割合を占めている)を除けば、短縮語の各構造に観察された優勢なアクセント型は 2 字漢語と一致した傾向があると言ってよからう。

以上、親密度の低い調査語を利用した読み上げ調査を通じて、首都圏方言話者が初めて見る 2 字漢語と短縮語に対して、アクセントをどのように付与するかを調べた。辞書データにおけるアクセント型の分布を利用した今までの研究と違い、これらの調査語は基底にアクセント型が指定されていないため、それらに対するアクセント付与は母語話者が持つ音韻グラマーの作用を示す直接的な証拠であると考えられる。以下、調査結果に見られた音韻・形態構造による差とバリエーションについて、制約ベースモデルに基づいて分析を行う。

4. 制約ベースモデルに基づく分析

4.1. 制約群³

第2要素が2モーラの2字漢語と多くの短縮語の構造に観察された①型の優位性を説明するために、最適性理論(Prince & Smolensky 1993/2004 ; 以下、OT)に基づいて外来語 LLLL、HLL と外来語複合語短縮形(パソコン、ジーパン)の①型の産出を分析した Ito & Mester (2016) の提案を利用する。ここで重要となる局所的なランキングは $\text{NONFIN}(\text{Ft}') \cdot \text{RIGHTMOST} \gg \text{WDACC}$ である。韻律語がフットを2つ以上持ち、かつ1つのフットが語末に位置するような韻律構造であれば、 $\text{NONFIN}(\text{Ft}')$ と RIGHTMOST により、候補 $\text{ft}' \dots \text{ft}']_0$ と候補 $\text{ft}' \dots \text{ft}]_0$ はどちらも勝つことができず、結果としてアクセント核を持たない $\text{ft} \dots \text{ft}]_0$ が最適な候補として選ばれる(ft : フット)。一方、2字漢語と短縮語の後部要素(韻律語の語末)がフットに解析されず、かわりに前部要素が最右端の(非語末)フットとして解析されれば、 $\text{NONFIN}(\text{Ft}')$ と RIGHTMOST に違反することなく、アクセント核を担うことができる。

また、第2要素が1モーラの2字漢語と短縮語の構造に多く観察された①型の産出、および第2要素が1モーラの短縮語(L-L、LL-L、H-L)に観察された①型の産出などを説明するために、OT に基づいた李(2017)の提案を利用できる。この場合、形態素ごとにフットを必要とする MORFt 、語ごとにフットを必要とする LEXFt と NONFINALITY 制約群(語末軽音節がヘッドであってはならない $\text{NONFIN}(\text{L})$ と語末音節がヘッドであってはならない $\text{NONFIN}(\sigma)$)の位置関係が特に関与している。2字漢語の場合、 $\text{NONFIN}(\text{L}) \gg \text{MORFt} \gg \text{NONFIN}(\sigma)$ の順位関係が重要となり、第2要素が1モーラ(1軽音節)であれば、(MORFt には違反するが) $\text{NONFIN}(\text{L})$ に違反しないためにフットに解析されず、 $(\text{L}')\#\text{L}$ 、 $(\text{H}')\#\text{L}$ のように①型が勝つ。一方、第2要素が2モーラであればフットに解析され(MORFt に対する違反は回避される)、 $(\text{H})\#(\text{H})$ 、 $(\text{LL})\#(\text{LL})$ のように①型が最適な候補として産出できる。短縮語の場合、 LEXFt が $\text{NONFIN}(\text{L})$ の上位に来れば、第2要素が1軽音節であってもフットに解析され、 $(\text{LL})-(\text{L})$ 、 $(\text{H})-(\text{L})$ のように①型が産出されるが、制約のランキングを変化させるモデルの仕組み(4.2 節で説明する)によって LEXFt が $\text{NONFIN}(\text{L})$ の下位に来れば、 LEXFt の効力が発揮できなくなるため、 $(\text{L}'\text{L})-\text{L}$ 、 $(\text{H}'\text{L})-\text{L}$ のように①型が産出される。

4.2. モデルとアルゴリズムの概要

実際の調査結果に観察されたバリエーション(たとえば H-L の場合、①型が 49%、②型が 51%とそれぞれ産出されたような関係)をモデリングするために、本稿は Stochastic Optimality Theory (Boersma 1997、Boersma 1999、Boersma & Hayes 2001 ; 以下、SOT)を利用する。SOT は確率的な手法を取り入れた制約ベースモデルの1種であり、言語データのバリエーションをノイズ付きのグラマーのゆれによる結果だと解釈する。SOT におけるランキングは連続的で、各制約を具体的な数値(ranking value ; 以下、RV)で表示することができる。評価するに当たって、評価器は「評価ノイズ」というパラメーターに基づいてランキング全体にノイズを与え、各 RV を変化させる。各 RV はその影響を受けて、一時的に RV' に変化した後に、入力に対して評価を行う。この場合、値の変化は正規分布(平均=RV、標準偏差=評価ノイズ)に従う。

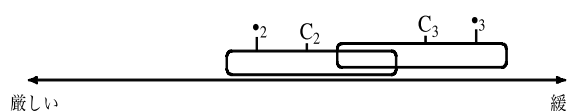


図3 通常のランキング($C_2 \gg C_3$)

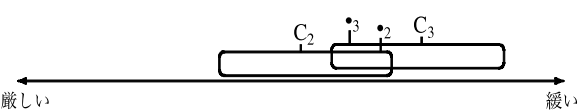


図4 稀なランキング($C_3 \gg C_2$)

SOT における制約間の相対的なランキングがどのように変化するかにについて、Boersma & Hayes (2001) の例を用いて説明する。図 3 と図 4 が示しているように、制約 C_2 と C_3 の値 (RV) はそれぞれ同じ標準偏差 (評価ノイズ) を持った正規分布を成しており (大まかな範囲は角丸長方形で示されている)、そこから評価時に制約の具体的な値 (RV') がランダムに抽出され、一時的に \cdot_2 と \cdot_3 になる。 C_2 の RV が 5.0、 C_3 の RV が 1.0 (RV が大きいほど制約が厳しい)、評価ノイズが 2.0 に設定された場合、 \cdot_2 が \cdot_3 よりも値が大きい確率は 0.921 と計算されるため⁴、多くの場合 C_2 が C_3 の上位に位置し、図 3 のような通常のランキングになる。一方、 \cdot_3 が \cdot_2 よりも値が大きい確率は $1 - 0.921 = 0.079$ であるため、 C_2 と C_3 の相対的な順位が逆転して、図 4 のような稀なランキングにもなりうる。もし C_2 が候補 a、 C_3 が候補 b をそれぞれ支持するとすれば、サンプリングを 1000 回行った場合、上の確率に基づき、候補 a が約 921 回、候補 b が約 79 回勝つことになる。このように、SOT の枠組みにおいて制約間のランキングが変化すると、最適な候補も変わってくるため、ある入力における出力のバリエーションを説明することができる。

SOT に基づいた学習アルゴリズムとして、Gradual Learning Algorithm (Boersma 1997、Boersma 1999、Boersma & Hayes 2001 ; 以下、GLA) がある。GLA は既存の言語データ (たとえば実際のアクセント型の分布) を入力としてそこから学習し、目標言語のグラマー (制約のランキング) に近づくことができる。具体的には、制約の降格と昇格両方を用いて、学習者の間違った出力と実際の言語形式に共有される違反印を消した上で、間違った出力を支持するすべての制約を降格させ、実際の言語形式を支持するすべての制約を昇格させる。その際、降格と昇格の程度 (値の変動幅) は学習率 (ϵ で表す) というパラメーターに基づいている。

最後に、フットのような曖昧性のある構造を推測・学習するためには、(4) に示した 3 レベルの音韻表示 (Apoussidou 2007、Boersma & van Leussen 2017) が必要である (以下、各レベルをそれぞれ英字の略語で指す)。UF (縦棒 | で括る) は従来の OT の入力に当たるレベルで、音節の情報のみ含むとする。今まで見たことのない語など、本来アクセント素性が基底に存在しないと考えられる場合、アクセントはグラマーによって付与されるため、その情報はこのレベルに含まれていない。SF (スラッシュ / で括る) は従来の OT の出力に当たるレベルで、音節とアクセントだけでなく、音声的な証拠が顕著でないフットも含むとする。OF (角括弧 [] で括る) は SF からフットの情報を取り除き、曖昧性が生じないと考えられる音節とアクセントの情報を持つとする。フット情報を推測するために、Expected Interpretive Parsing (Jarosz 2013 ; 以下、EIP) が提案され、ほかの解析アルゴリズムより目標グラマーへの収束率が高いというメリットがある (Jarosz 2013 ; Li 2018)。紙面の制限により、各モデルとアルゴリズムの詳細はそれぞれの先行研究を参照されたい。

(4) 3 レベルの音韻表示

レベル	表示の例	韻律情報
深層形 (Underlying form ; UF)	σσσ	音節
表層形 (Surface form ; SF)	/ (σ'σ) σ /	音節・フット・アクセント
顕在形 (Overt form ; OF)	[σ'σσ]	音節・アクセント

4.3. モデルの作成

以下、読み上げ調査の結果に対して、SOT×GLA×EIP の組み合わせを利用してモデルを作る。使用するソフトは Praat (Ver.6.0.52 ; Boersma & Weenink 2019) である。(5) は GLA の各パラメーターの設定

を示している。

(5) GLA のパラメーター設定

アップデートルール	評価ノイズ	初期の ε	ε の数	ε ごとの実行数	ε の減衰
Symmetric all	2.0	1.0	4	10000	0.1

「Symmetric all」は、 ε の値に準じて、loser(現時点のグラマーによって仮産出されるが、実際の言語データと一致しないもの)を支持するすべての制約を降格させ、winner(解析アルゴリズムによって実際の言語データの SF とされたもの)を支持するすべての制約を昇格させるアルゴリズムである。「評価ノイズ」が 2.0 に設定されていることは、毎回の評価に当たって、各制約の値がランダムに変化するとき、その変化が従う正規分布の標準偏差が 2.0 ということである。 ε は 4 つあり、それぞれ 1 万回実行される(合計 4 万回)。「 ε の減衰」が 0.1 に設定されていることは、この場合、最初の 1 万回の ε は 1.0、次の 1 万回は 0.1、3 番目の 1 万回は 0.01、最後の 1 万回は 0.001 だということを意味する。

制約の初期値はすべて 100 とする。目標データに指定された構造とアクセント型の各組み合わせの割合に準じて、学習データ(OF)をランダムに生成し、モデルに学習させる。学習ステップを 4 万回実行したら、それを 1 回のシミュレーションと見なす。学習が終わった後に、最終のランキングに基づき、各々の入力(UF)に対して、10 万個のデータ(SF)をランダムに産出させ、そこから OF を抽出して、各アクセント型が各構造に占める割合を計算する。この結果を目標データの割合と比較し、両者の相関および平均絶対誤差(mean absolute error ; 以下、MAE ; $MAE = \frac{1}{n} \sum_{t=1}^n |e_t|$)を計算してモデルを評価する。相関が高いほど、また、MAE が小さいほどモデルの説明力が高い。

4.4. モデルの結果と考察

以上の設定を用いて、シミュレーションを 50 回行った。結果として、平均相関が 0.994、平均 MAE が 0.024 となっている。平均相関が高く、また、平均 MAE が小さいため、精度よく調査結果をモデリングできていると言えよう。各シミュレーションの結果は僅差しかいないため、代表として、この中で MAE が最も小さいグラマー(各制約の値)のみ取り上げ、(6)に提示する。

(6) MAE が最も小さいグラマー

制約	値	制約	値	制約	値	制約	値
NONFIN(Ft')	135.698	RIGHTMOST	129.152	FtBIN	123.814	MT	100.000
LEXFt	132.211	INITFt	129.151	WSP	106.373	PARSE- σ	60.732
NONFIN(L)	131.927	MORFt	128.648	WDACC	105.569		
MINWDACC	131.206	NONFIN(σ)	125.554	NO LAPSE	102.000		

(6) のグラマーがどのように目標データを説明するかを検討するために、まず、平板式アクセントと起伏式アクセントをそれぞれ産出する動機となる制約群を見る。ここではヘッドフットの位置に関する制約群 NONFIN(Ft')・RIGHTMOST および特定の韻律語構造がアクセント核を持つことを求める制約群 WDACC・MINWDACC が関与しているため、それらの順位関係を表す概念図を図 5 に提示する。この場合、Ito & Mester (2016) の OT 分析における NONFIN(Ft')・RIGHTMOST \gg WDACC の絶対的な順位関係は SOT に基づいた(6)のグラマーにおいても実質的に保たれ、評価ノイズが 2.0 のとき、WDACC が逆に NONFIN(Ft')か RIGHTMOST の上位に来る確率はほぼ 0 に近い(確率の計算方法は注 4 を参照)。ただし、MINWDACC が RIGHTMOST の上位に来る確率は 0.766 であるため、2 モーラ語(最小韻律語)である L#L

と L-L は MORFt と LEXFt の効力によって 2 フットとして解析された場合でも (通常①型として産出されるフット構造)、約 0.766 の確率で前部要素がアクセント核を担った形 ((L')#(L) と (L')-(L)) で産出できる (MINWdACC を守るために、RIGHTMOST に違反した形となっている)。

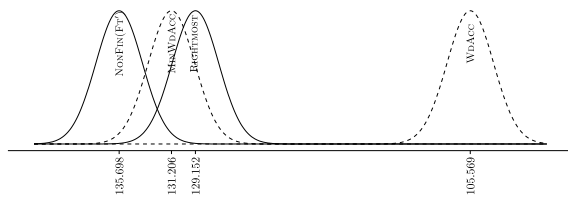


図 5 ヘッドフットの位置に関する制約群(実線)と韻律語・アクセントに関する制約群(破線)の順位関係

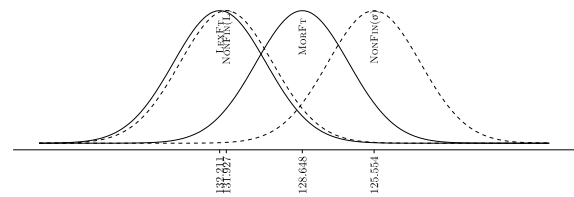


図 6 形態・音韻インターフェイスに関する制約群(実線)と NONFINALITY 制約群(破線)の順位関係

音韻・形態構造による違いについては、形態・音韻インターフェイスに関する制約群 LEXFt・MORFt と NONFINALITY 制約群 NONFIN(L)・NONFIN(σ) が関与しているため、それらの順位関係を表す概念図を図 6 に提示する。NONFIN(σ) は LEXFt と離れているため、逆に LEXFt の上位に来る確率は 0.009 しかなく、第 2 要素が 1 重音節の短縮語が LEXFt に違反しないために、ほぼ (LL)-(H)、(H)-(H) のように①型で産出されることを説明している。一方、NONFIN(σ) が MORFt の上位に来る確率は 0.137 であるため、2 字漢語の第 2 要素が 1 重音節のとき、約 0.137 の確率で (L')#H、(H')#H など NONFIN(σ) に違反しない構造が勝ち、WdACC の効力によって①型が産出されることになる。また、NONFIN(L) の位置に注目すると、LEXFt とは非常に近く、NONFIN(L) ≫ LEXFt の確率は 0.460 であるため、短縮語 LL-L、H-L においてアクセントの割合の偏りが顕著でないことを説明している。それに対して、NONFIN(L) が MORFt の上位に来る確率は 0.877 であるため、2 字漢語の第 2 要素が 1 軽音節であれば、高い確率で①型が産出されることになる。

以上、読み上げ調査の結果を目標データとして、SOT の枠組みに基づいてモデルを作成し、説明を行った。特に調査結果に観察されたバリエーションに関しては、OT に基づいた従来の研究では説明できないのに対して、SOT は評価ノイズを利用して、確率的な手法でバリエーションに近似した結果を産出することができ、より説明力が高いと考えられる。

5. おわりに

本稿は従来辞書データの研究やアクセント法則で言われていた 2 字漢語と 4 字漢語短縮形のアクセントの傾向が母語話者のアクセント感覚にも同様に反映されるかどうかを検証するために、首都圏方言話者を対象に、親密度が低いと考えられる調査語を用いて読み上げ調査を行った。結果として、2 字漢語は辞書データの傾向と一致し、第 2 要素が 1 モーラであれば①型、2 モーラであれば②型が優勢であった。短縮語の場合、2 字漢語よりも②型の割合が全般的に増える傾向が観察され、特に第 2 要素が 1 モーラである 3 モーラの短縮語 (LL-L、H-L) は①型の優位性がなくなり、2 字漢語との差が顕著であった。

また、調査結果に見られた音韻・形態構造による差とバリエーションについて、Ito & Mester (2016) と李 (2017) の提案を利用し、確率的な手法を取り入れた制約ベースモデル SOT に基づいて分析を行った。アクセント核が産出されるか否かについては、ヘッドフットの位置に関する制約群 NONFIN(Ft)・RIGHTMOST とアクセント核に関する制約群 WdACC・MINWdACC の相対的な位置関係、また、語の音

韻・形態構造による違いおよびバリエーションについては、形態・音韻インターフェイスに関する制約群 LEXFT・MORFT と NONFINALITY 制約群 NONFIN(L)・NONFIN(σ) の相対的な位置関係によって説明できた。SOT の説明力が高いことから、漢語だけでなく、和語・外来語のアクセント型の分布も網羅した分析が期待でき、今後の課題としたい。

注

* 本稿は、2019 年度春季研究発表(2019 年 6 月 21 日、首都大学東京・秋葉原サテライトキャンパス)における口頭発表の内容に加筆修正を加えたものである。貴重なコメントをくださった聴衆の皆様、および査読委員の方々に深謝する。言うまでもなく、全ての誤りは筆者の責任によるものである。

¹ 丸付き数字は単語の何番目のモーラにアクセント核が来るかを表している(例:「家族」は①型、「自由」は②型)。①型は単語がアクセント核を持たず、平らのままに発音されることを意味する(例:「安心」「会社」)。

² 各構造の調査語を下表に提示する。ただし、実際の調査では一部の短縮語の発音が促音化し、単語の音節構造が変わったため(たとえば「匿広」が「とっこう」と発音され、構造が H-H になっている)、それらのデータを除外して統計処理を行った。

構造	調査語	構造	調査語
L#L	邪飛・地祇・婆娑・举措・獻歆・夢寐	L-L	異種寄生・機密書類・予防措置・科学汚染・社長秘書・自己犠牲・古語辞典・普通預金・遺族扶助
LL#L	曲浦・直披・舳艫・国帑・発兌・木履・略綬	LL-L	結合距離・実質預金・活物寄生・国語辞典・悪性腫瘍・國平刺激・着信履歴
L#LL	舐犢・斧鑿・蘆荻・斧鉞・塗抹・蛇蝎・披瀝	L-LL	社会秩序・疑問副詞・皮膚剝離・試行錯誤・自己克服・路上喫煙
H#L	代赭・官衙・陟阻・頤志・親炙・注疏・盤踞・檻樓・囹圄	H-L	走行距離・環境汚染・電磁遮蔽・安産祈願・残留孤兒・検索履歴・信用危機・神経刺激・民間企業・交通違反
L#H	素馨・華胃・世尊・弥縫・捕繩・油單	L-H	避難勧告・家庭看護・過当競争・写真凹版・外科救急・舞台挨拶・家庭排水
LL#LL	尺牘・鬱勃・博奕・踟躕・落莫・逐鹿・絡繹・剝削・寂寞	LL-LL	質量輻射・石油頁岩・液体窒素・植物毒素・疾病特約
H#LL	駐劄・荊棘・戒飭・救恤・鷄肋・警蹕・謙抑・剛愎・繁縷	H-LL	法律秩序・飽和液体・精神錯乱・網膜剝離・人關国宝・入院特約・人權抑壓・細菌毒素
LL#H	僻陬・卜筮・束脩・疫癘・目睫・駛遁・拔錨・博搜	LL-H	角膜潰瘍・滑降競技・特惠貿易・匿名廣告・植物採集・物理探査・雜損排除
H#H	均霑・退嬰・藩屏・因循・冤枉・奸佞・繭紉・殷賑・秋霖・少輔・端倪・蹢躅	H-H	海外貿易・內陸鹹湖・生活排水・救急看護・全称肯定・飽和蒸氣・海底探査

³ 各制約の定義: LEXFT: すべての語彙素は最小限にフットに解析される。MORFT: すべての形態素は最小限にフットに解析される。MT: フット構造は(H)、(LL)、(L)のみ許す。NONFIN(σ): 語末音節はヘッドであってはならない。NONFIN(L): 語末軽音節はヘッドであってはならない。NoLAPSE: 音節は最大限にフットに解析される。フットに解析されない音節が 2 つ連続した場合に違反となる。MINWDACC: 最小韻律語はアクセント核を持つ。RIGHTMOST: ヘッドフットにほかのフットが続くと違反となる。WSP: 重音節はフットヘッドである。FTBIN: フットはモーラレベルか音節レベルの解析でバイナリーである。INITFT: 韻律語はフットで始まる。NONFIN(FT): ヘッドフットが韻律語の語末にある場合に違反となる。WDACC: 韻律語はアクセント核を持つ。PARSE-σ: すべての音節はフットに解析される。

⁴ 制約間の相対的な順位関係の確率は正規分布の再生性(正規分布に従う確率変数 X_1 と X_2 が互いに独立である場合、その差の分布も正規分布に従う： $X_1 - X_2 \sim N(\mu_1 - \mu_2, \sigma_1^2 + \sigma_2^2)$)に基づいて計算されている。たとえば C_2 の値が5.0、 C_3 の値が1.0、評価ノイズが2.0に設定された場合、 C_2 と C_3 の差は平均 $= 5.0 - 1.0 = 4.0$ 、標準偏差 $= \sqrt{2.0^2 + 2.0^2} = 2.828$ の正規分布に従うため、 C_2 が C_3 の上位に来る確率 $P(C_2 \gg C_3)$ は $1 - \text{pnorm}(0, \text{mean}=4.0, \text{sd}=2.828) = 0.921$ で計算される(`pnorm()`は指定した値(この場合は0)の正規分布の下側確率 P を返すR言語の関数である)。なお、この場合、 C_2 と C_3 それぞれの絶対値ではなく、 C_2 と C_3 の差の値(両者がどれだけ離れているか)と指定された評価ノイズの値が確率の計算に影響していることに注意されたい。

参考文献

- 秋永一枝 1981「アクセント習得法則」『明解日本語アクセント辞典 第二版』東京：三省堂。
- 天野成昭・近藤公久 1999『日本語の語彙特性』東京：三省堂。
- 天野成昭・近藤公久 2000「NTT データベースシリーズ「日本語の語彙特性」について」『音声研究』4(2).44-50.
- Apoussidou, Diana. 2007. *The learnability of metrical phonology*. Ph.D. dissertation, University of Amsterdam.
- Boersma, Paul. 1997. How we learn variation, optionality, and probability. *Proceedings of the Institute of Phonetic Sciences of the University of Amsterdam* 21.43-58.
- Boersma, Paul. 1999. Optimality-Theoretic learning in the Praat program. *IFA proceedings* 23.17-35.
- Boersma, Paul and Bruce Hayes. 2001. Empirical tests of the Gradual Learning Algorithm. *Linguistic Inquiry* 32(1).45-86.
- Boersma, Paul and Jan-Willem van Leussen. 2017. Efficient evaluation and learning in multilevel parallel constraint grammars. *Linguistic inquiry* 48(3).349-388.
- Boersma, Paul. and David Weenink. 2019. Praat: Doing phonetics by computer [Computer program] Version 6.0.52. Online: <http://www.praat.org/>.
- Imidas 編集部(編) 1997『情報・知識 imidas' 98』東京：集英社。
- Ito, Junko and Armin Mester. 2016. Unaccentedness in Japanese. *Linguistic Inquiry* 47(3).471-526.
- Jarosz, Gaja. 2013. Learning with hidden structure in Optimality Theory and Harmonic Grammar: beyond Robust Interpretive Parsing. *Phonology* 30(1).27-71.
- Li, Motong. 2018. The learnability of the accentuation of Sino-Japanese words and loanwords: The hidden structure problem. *Phonological Studies* 21.31-40.
- 最上勝也・坂本充・塩田雄大・大西勝也 1999『『日本語発音アクセント辞典』～改訂の系譜と音韻構造の考察～』『NHK 放送文化調査研究年報』44.97-157.
- 森庸子 2002「3 モーラ複合語略語の生成要因：若者のキャンパスことばから」『音声研究』6(1).121-137.
- NHK 放送研究所(編) 1998『NHK 日本語発音アクセント辞典新版』東京：NHK 出版。
- 小川晋史 2006「日本語諸方言の 2 字漢語アクセント」神戸大学修士論文。
- Prince, Alan and Paul Smolensky. 1993/2004. *Optimality Theory: Constraint interaction in generative grammar*. Malden, MA & Oxford, UK: Blackwell.
- 李墨彤 2017「日本語漢語の優勢なアクセント型の分布 ―外来語と比較して―」『音韻研究』20.11-20.
- 渡邊敏郎・Edmund R. Skrzypczak・Paul Snowden(編) 2003『新和英大辞典 第5版』東京：研究社。