

東京方言のデフォルトアクセントに関する予備的調査

り ぼくとう
李 墨彤

ribokuto@gmail.com

大阪大学大学院

PAIK

2018 年 5 月 26 日

神戸大学 文学部 C 棟 5 階会議室

研究の目的

- 東京方言のデフォルトアクセントのバリエーションに注目する
- できるだけ多くの語種・音節構造について調べる
- 生成と知覚の両方からデフォルトアクセントを考察する
- データを踏まえて、多様な制約ベースモデル¹を用いて分析し、より精度の高いモデルを作る

¹ Stochastic OT (Boersma 1997, 1999; Boersma & Hayes 2001) ・ Harmonic Grammar (Legendre et al. 1990) ・ Maximum Entropy Model (Jäger 2007)などを指す。

前提

- レキシコンと親密度の関係を考慮し、語彙の使用頻度などの情報を利用して、アクセント型のレキシコンにおける度合い²を表す
 - 度合いの高いもの：レキシコンにあるアクセント型の情報を直接参照して、アクセント型を産出する
 - 度合いの低いもの・無意味語：アクセント型がグラマー³によって産出される
- レキシコンの度合いは多値を取りうる
- グラマーは語彙の使用とともに形成される
- デフォルトアクセントのゆれは、グラマーのゆれによって解釈される

²Zuraw (2000)の Listedness など。

³制約ベースモデルの場合、制約のランキング、または個々の制約の値を指す。

音節構造

- Ito & Mester (2016)は以下 23 個の音節構造（外来語）の優勢なアクセント型を分析している

L ド・レ・ミ	LL ガム	LLL バナナ	LLLL アルバム	LHLL コマーシャル	H L H チャンピオン
H H シャンプー	HL ブーケ	LH プリン	H パン	HLL ゴンドラ	LLH キャプテン
LHL ケチャップ	LLLLL クリスマス	LLLLLL アスバラガス	LLLLLLL アナクロニズム	HLLLL ジャーナリスト	HHL ハンバーグ
LLLH エネルギー	LLHH カレンダー	LL+H ⁴ パソコン	H+H ジープ	LL+L ファミマ	

- Ito & Mester (2016)の調べたアクセント型を含め、李 (2017)は以下個の音節構造（漢語）の優勢なアクセント型（1～4 モーラ語）も分析している

L 可	LL 悪	L#L 以後	LL#L 握手	L#LL 異質	H 運
H#L 安易	L#H 亜鉛	LL#LL 確實	H#LL 閥節	LL#H 液体	H#H 強引

⁴Ito & Mester (2016)では+は形態素境界を表しているが、今回の研究では+は単語境界（word boundary）、#は形態素境界をそれぞれ示している。

- 今回はさらに以下 27 個の音節構造を追加している（合計 57 個⁵⁾）
（外来語：19 個）

HHH
コンピューター

HHHL
コンポーネント

HHLL
バイナッブル

HLHH
インフォメーション

HLHL
パンフレット

HLLLH
ノンフィクション

HLLHL
コンサルタント

HLLL
ヨーグルト

LLLLH
ヘリコプター

LLLLHL
アシスタント

LHHL
クリーニング

LHLH
プライバシー

LHLHL
カウンセリング

LLLLHL
コレステロール

LHLLL
リバーシブル

LLHH
アナウンサー

LLHLL
スチュワード

LLLLHH
コミュニケーション

LLHL
プレゼント

（和語：8 個）

HH
幸い

HL
女

HLL
たんぼ

LH
住まい

LL
酒

LLH
あじさい

LLL
子供

LLLL
かまきり

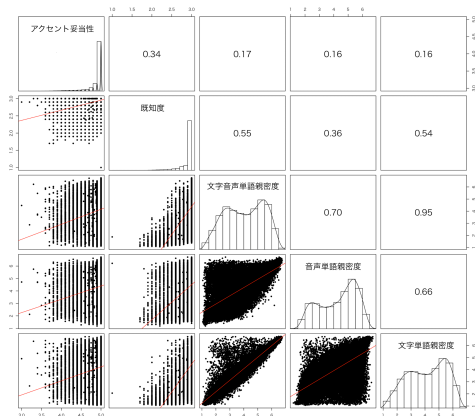
- 音節構造を選ぶ基準
 - 辞書における所属語彙数が 10 語以上
 - 和語は 4 モーラ語まで
 - 外来語は 7 モーラ語まで

⁵⁾ モーラ語 (L) は今回の調査に含まれていない。

辞書（データベース）

- 『日本語の語彙特性』（天野・近藤 1999）
 - 理由
 - 親密度の情報を利用できる
 - 整形可能な電子ファイルを入手できる
 - 天野・近藤（1999）は『新明解国語辞典第四版』（金田一他 1989）の見出し語を利用し、以下の5つの項目について評定実験を行った（項目数 88569、単語数 69084）：
 - **既知度**：評定対象単語を知っている度合い（3段階）
 - **アクセント妥当性**：アクセントが妥当だと思う度合い（5段階；評定者：東京方言話者10名、平均年齢（当時）22.5歳）
 - **文字音声親密度**：文字と音声を同時に見たり聞いたりした場合における単語親密度（7段階）
 - **文字親密度**：文字を見た場合における単語親密度（7段階）
 - **音声親密度**：音声を聞いた場合における単語親密度（7段階）

- 天野・近藤 (1999) のデータから、以下の条件に基づいてデータを選出した (41039 語)
 - 名詞
 - 複数のアクセント型を持つ単語について、アクセント妥当性のもっとも高い単語
 - UniDic 2.1.2 (伝他 2007) に語種情報のある単語
- 各項目間の相関

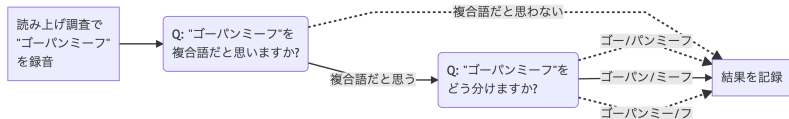


予備調査：読み上げ調査

- 予備調査の対象：東京方言話者 1 名（50 代男性）
- 調査語（合計 206 語）
 - 外来語の各音節構造に対して、無意味語を 2～6 語用意した（合計 129 語）
 - 漢語の各音節構造に対して、漢字を組み合わせた新造語を 4 語用意した（合計 44 語）
 - 和語の各音節構造に対して、文字音声親密度の低い単語を 2～7 語用意した（合計 33 語）
- 手順
 - 調査対象にできるだけ自分にとって自然なアクセント型で調査語を発音するように指示した
 - モニターで調査語を示し（無意味語：カタカナ表記；漢語と和語：漢字表記+振り仮名）、調査対象の発音を録音した
- 調査時間：約 25 分

予備調査：単語分割調査

- 調査語：読み上げ調査で使われた、4 モーラ以上の無意味語
- 手順
 - 読み上げ調査と同時に実施した
 - 該当の単語を録音した後に、「この単語を複合語だと思いますか？」という質問を用意した
 - 複合語だと思われる語に対して、「この単語をどう分けますか？」という質問と可能な選択肢を用意し、調査対象に単語を分割させた
 - 例（調査語が「ゴーパーンミーフ」の場合。実線は調査対象が実際に選んだルートを示している）：



- 調査時間：約 15 分

予備調査：自然性評価調査

- 調査語：読み上げ調査で使われた各音節構造から 1～2 語を選び、その単語の可能なアクセント型⁶を研究者が録音した（合計 224 項目）
- 手順
 - 録音した音声を調査対象に聞かせ、以下の 7 段階で音声のアクセント型の自然性を評価させた

とても不自然

不自然

やや不自然

どちらとも言えない

まあまあ自然

自然

とても自然

- 結果を 1（とても不自然）～7（とても自然）の数値に変換した
- 辞書や読み上げ調査の結果（割合）などと比較できるようにするため、softmax 関数で自然性を割合に変換した

今回使っている softmax 関数⁷：
$$p(\text{Acc}_j) = \frac{c^{\text{nt}r(\text{Acc}_j)}}{\sum_{k=1}^K c^{\text{nt}r(\text{Acc}_k)}}$$

- 調査時間：約 30 分

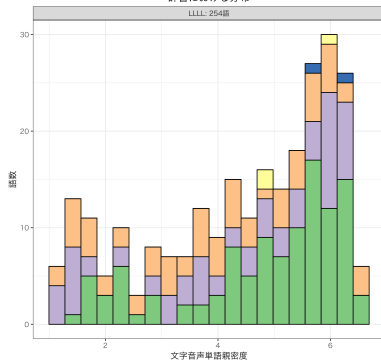
⁶尾高型を除く。

⁷ $\text{nt}r()$ はアクセント型の評価結果を数値に変換する関数で、 c は 3 で指定している。

調査結果：外来語 LLLL の場合

辞書における分布

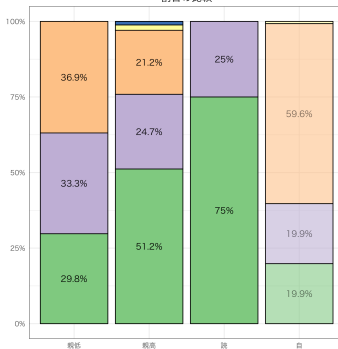
LLLL: 254語



親密度が上位に来る単語

表記	ア型	親密度
アルバム	0	6.594
アクセル	1	6.375
ストレス	2	6.562
ネグリジェ	3	5.844

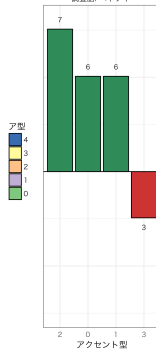
割合の比較



調査語（無意味語・読み上げ）

表記	ア型（読）	複合語？
クリームア	0	NO
シェギネカ	0	NO
ベネゾキ	0	NO
ダリシブ	1	NO

自然性評価の結果
調査語: ベネゾキ



相関



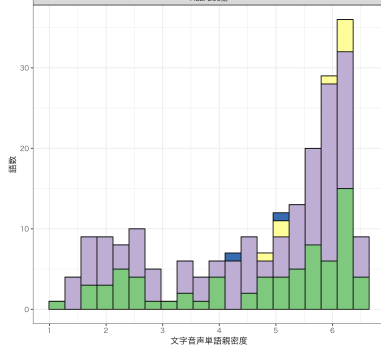
平均絶対誤差 (MAE)



調査結果：外来語 HLL の場合

辞書における分布

HLL: 205語



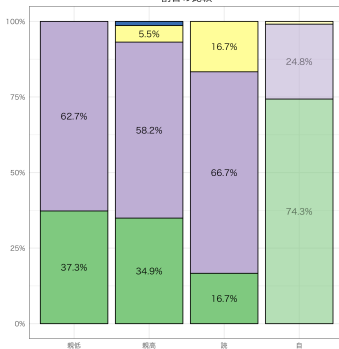
親密度が上位に来る単語

表記	ア型	親密度
てんぶら	0	6.594
ボーナス	1	6.625
ハンカチ	3	6.312

調査語（無意味語・読み上げ）

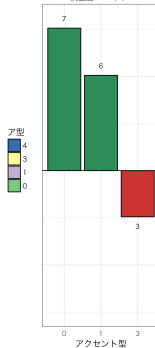
表記	ア型（読）	複合語？
ニンギョ	3	NO
パンギナ	0	NO
ジャーナヘ	1	NO
サッカム	1	NO
ノンタタ	1	NO
ヒュッピラ	1	NO

割合の比較



自然性評価の結果

調査語: ノンタタ



相関



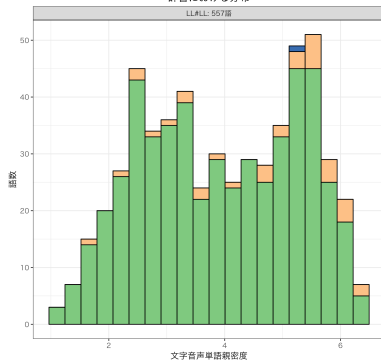
平均絶対誤差 (MAE)



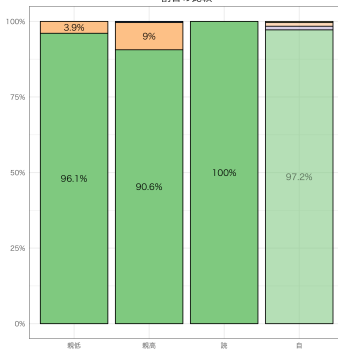
調査結果：漢語 LL#LL の場合

辞書における分布

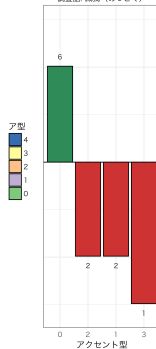
LL#LL: 557語



割合の比較



自然性評価の結果
調査語: 減濁 (めつどく)



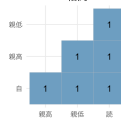
親密度が上位に来る単語

表記	読み	ア型	親密度
目的	モクテキ	0	6.344
植物	ショクブツ	2	6.375
一躍	イチヤク	4	5.156

調査語 (読み上げ)

表記	読み	ア型 (読)
責様	セキレキ	0
減濁	メツドク	0
遮冠	テキアク	0
湿落	シツラク	0

相関



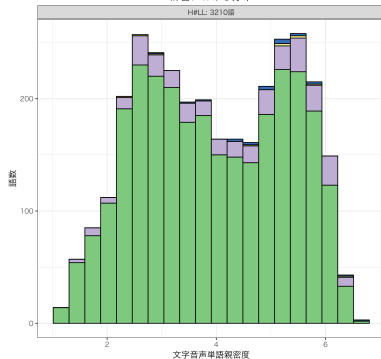
平均絶対誤差 (MAE)



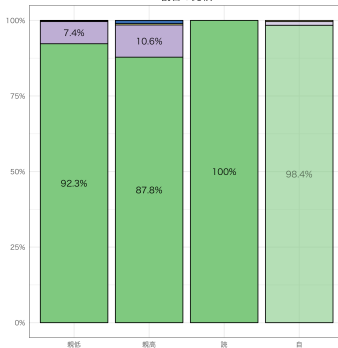
調査結果：漢語 H#LL の場合

辞書における分布

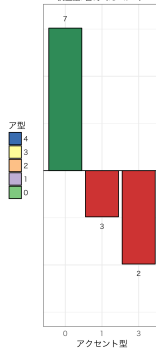
H#LL: 3210語



割合の比較



自然性評価の結果
調査語: 呂葛 (えいかつ)



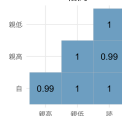
親密度が上位に来る単語

表記	読み	ア型	親密度
大学	ダイガク	0	6.625
音楽	オンガク	1	6.469
唯識	ユイシキ	2	2.188
正直	ショウジキ	3	6.344
正月	ショウガツ	4	6.562

調査語 (読み上げ)

表記	読み	ア型 (読)
瞬副	シュンフク	0
傑筑	ケツチク	0
宮葛	エイカツ	0
航背	コウセキ	0

相関



平均絶対誤差 (MAE)



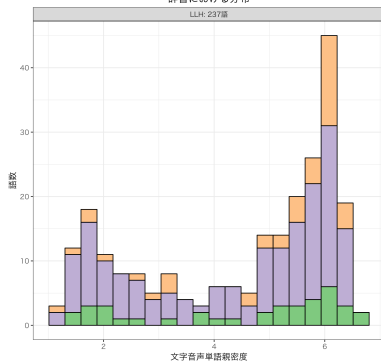
LLで終わる4モーラの外来語と漢語について

- 先行研究
 - Kubozono (1996)はLLで終わる4モーラの外来語は0型になりやすい
 - Ito & Mester (2016)、李 (2017)は最適性理論を用いて外来語LLLLとHLLの0型の産出を分析している
 - 漢語について、塩田 (2016b)が行った調査結果では1型と0型が同程度
- 今回の調査結果
 - 外来語LLLL：2型・0型・1型とも自然だと評価される
 - 外来語HLL：自然性評価においては0型が優勢であるが、辞書と読み上げ調査においては1型が優勢
 - 漢語：ゆれ無し

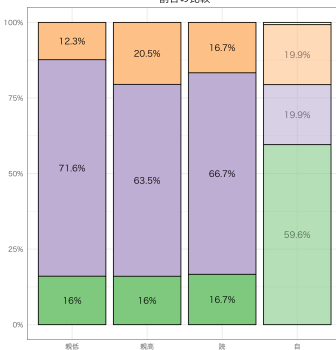
調査結果：外来語 LLH の場合

辞書における分布

LLH: 237語

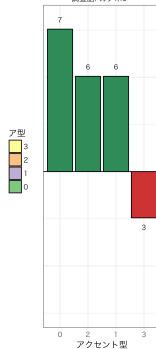


割合の比較



自然性評価の結果

調査語: カグボン



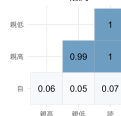
親密度が上位に来る単語

表記	A型	親密度
パソコン	0	6.594
キャプテン	1	6.469
ビタミン	2	6.438

調査語 (無意味語・読み上げ)

表記	A型 (読)	複合語?
キャカルー	1	NO
ロブラー	0	NO
ミシャノー	2	NO
ニュシアン	1	NO
カグボン	1	NO
ジェリバー	1	NO

相関



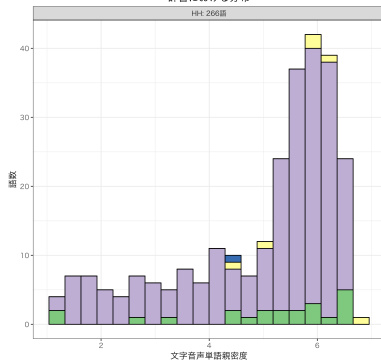
平均絶対誤差 (MAE)



調査結果：外来語 HH の場合

辞書における分布

HH: 266語



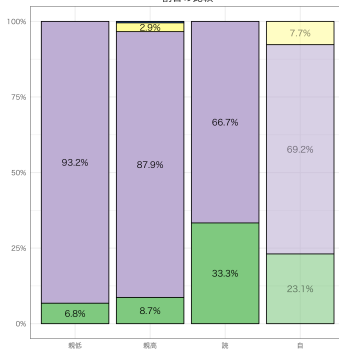
親密度が上位に来る単語

表記	ア型	親密度
アイロン	0	6.500
シャンプー	1	6.625
コーヒー	3	6.719

調査語（無意味語・読み上げ）

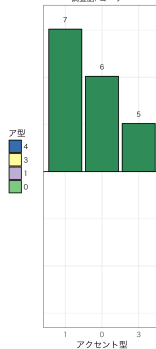
表記	ア型（読）	複合語？
ヨーサー	1	NO
ジャーチン	1	NO
ハッピーン	0	NO

割合の比較



自然性評価の結果

調査語: ヨーサー



相関



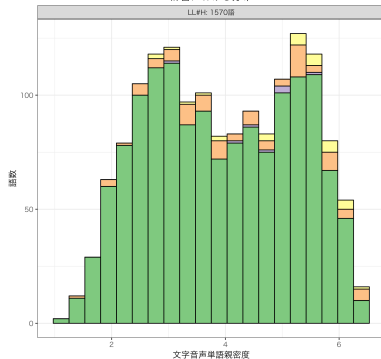
平均絶対誤差 (MAE)



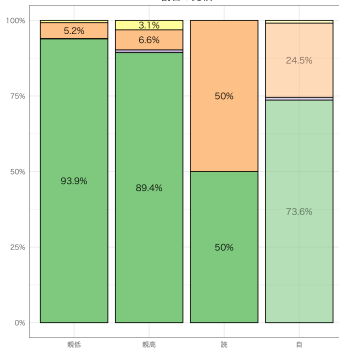
調査結果：漢語 LL#H の場合

辞書における分布

LL#H: 1570語

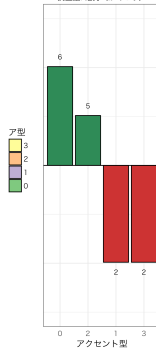


割合の比較



自然性評価の結果

調査語: 確拠 (かくこう)



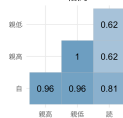
親密度が上位に来る単語

表記	読み	ア型	親密度
必要	ヒツヨウ	0	6.531
白状	ハクジョウ	1	5.688
一番	イチバン	2	6.406
日曜	ニチヨウ	3	6.469

調査語 (読み上げ)

表記	読み	ア型 (読)
折腕	タクワン	2
確拠	カクゴウ	2
夜眠	エキメイ	0
設問	セツドウ	0

相関



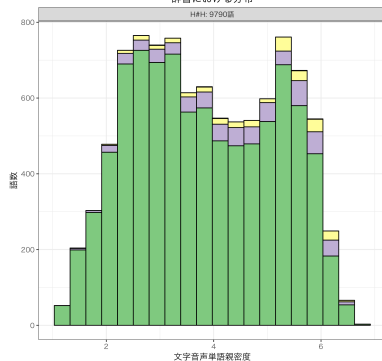
平均絶対誤差 (MAE)



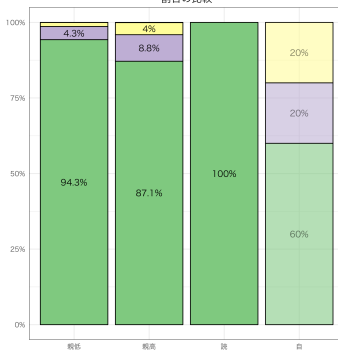
調査結果：漢語 H#H の場合

辞書における分布

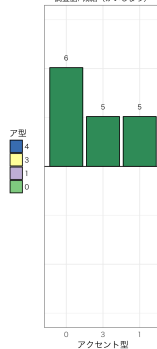
H#H: 9790語



割合の比較



自然性評価の結果
調査語: 械紹 (かいしょう)



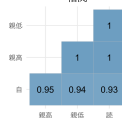
親密度が上位に来る単語

表記	読み	A型	親密度
人間	ニンゲン	0	6.688
両親	リョウシン	1	6.594
台風	タイフウ	3	6.500
大体	ダイタイ	4	5.906

調査語 (読み上げ)

表記	読み	A型 (読)
族會	シツキョウ	0
紫胤	キンガイ	0
症殉	ショウジュン	0
械紹	カイショウ	0

相関



平均絶対誤差 (MAE)



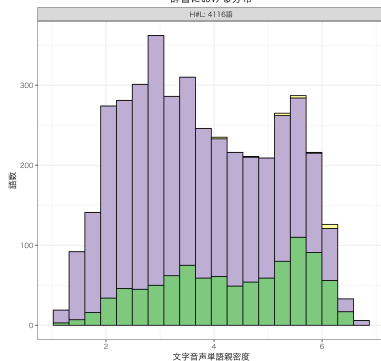
Hで終わる4モーラの外来語と漢語について

- 先行研究
 - 外来語 LLH：塩田 (2016a)は CVCVCaR が平板化する傾向があると報告している
 - 漢語 LL#H：塩田 (2016b)によると、第4モーラが撥音か「い (の長音か二重母音)」の場合、1型が優勢
 - 漢語 H#H：塩田 (2016b)によると、第2モーラが促音か撥音、また第4モーラが撥音か「い (の長音か二重母音)」の場合、1型が優勢
- 今回の調査結果
 - 外来語 LLH：0型・2型・1型とも自然だと評価される；辞書と読み上げ調査においては1型が優勢
 - 外来語 HH：1型・0型・3型とも自然だと評価される；全体的に1型が優勢
 - 漢語 LL#H：読み上げでは2型と0型が各50%；辞書と自然性評価においては0型が優勢
 - 漢語 H#H：0型・3型・1型とも自然だと評価される；辞書と読み上げ調査においては0型が優勢

調査結果：漢語 H#L の場合

辞書における分布

H#L: 4116語



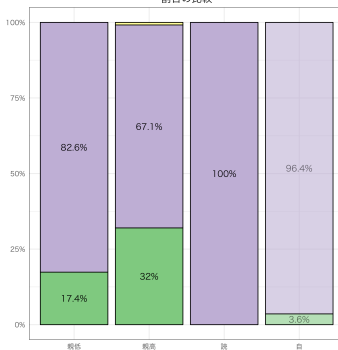
親密度が上位に来る単語

表記	読み	ア型	親密度
会社	カイシャ	0	6.531
今夜	コンヤ	1	6.714
返事	ヘンジ	3	6.188

調査語（読み上げ）

表記	読み	ア型（読）
漢供	カンク	1
疾進	シツシュ	1
準備	ジュンギ	1
製菓	セイカ	1

割合の比較

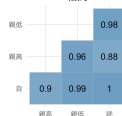


自然性評価の結果

調査語: 準備 (じゅんぎ)



相関



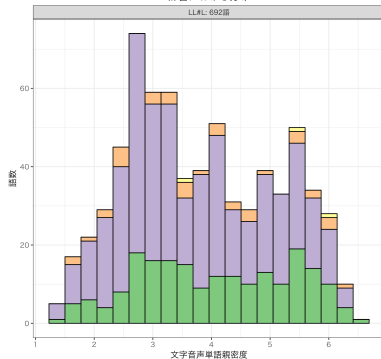
平均絶対誤差 (MAE)



調査結果：漢語 LL#L の場合

辞書における分布

LL#L: 692語



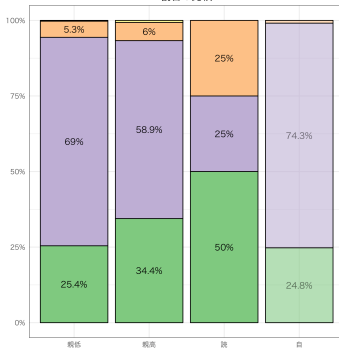
親密度が上位に来る単語

表記	読み	ア型	親密度
食事	ショクジ	0	6.531
読書	ドクショ	1	6.312
昨夜	サクヤ	2	6.156
役所	ヤクショ	3	5.971

調査語（読み上げ）

表記	読み	ア型（読）
複歯	フクロ	2
哲論	テツユ	1
祝儀	シュキガ	0
適娯	テキゴ	0

割合の比較



自然性評価の結果

調査語: 適娯（てきご）



相関



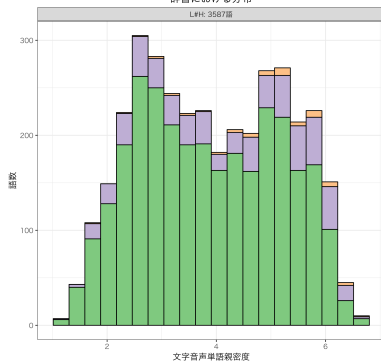
平均絶対誤差 (MAE)



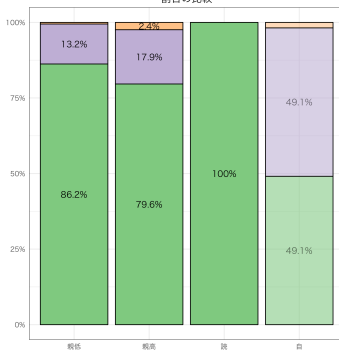
調査結果：漢語 L#H の場合

辞書における分布

L#H: 3587語

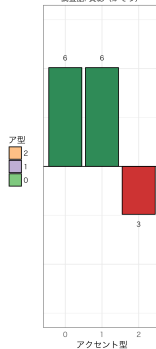


割合の比較



自然性評価の結果

調査語: 貸総 (かそう)



親密度が上位に来る単語

表記	読み	A型	親密度
自分	ジブン	0	6.656
氏名	シメイ	1	6.625
自由	ジユウ	2	6.688

調査語 (読み上げ)

表記	読み	A型 (読)
貸総	カソウ	0
戯永	ギエイ	0
魅昆	ミコン	0
須湘	スショウ	0

相関



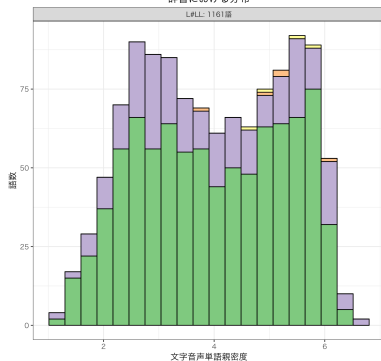
平均絶対誤差 (MAE)



調査結果：漢語 L#LL の場合

辞書における分布

L#LL: 1161語



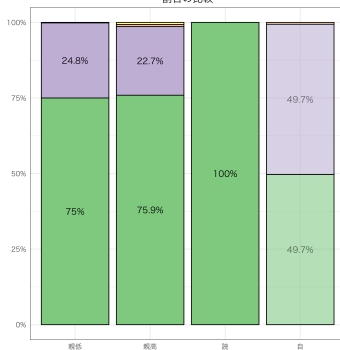
親密度が上位に来る単語

表記	読み	ア型	親密度
自宅	ジタク	0	6.344
家族	カゾク	1	6.594
個数	コスウ	2	5.938
地獄	ジゴク	3	5.875

調査語（読み上げ）

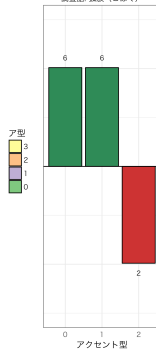
表記	読み	ア型（読）
孤寝	コバク	0
笊敷	ドカク	0
庶務	ショカツ	0
汽薩	キサツ	0

割合の比較

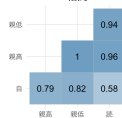


自然性評価の結果

調査語: 孤寝（こばく）



相関



平均絶対誤差 (MAE)



3 モーラの漢語について

- 先行研究

- 塩田 (2016b)によると、すべての構造において 1 型と 0 型のゆれが多い
- H#L：0 型が優勢：第 2 モーラ＝長音(塩田 2016b)
- L#H：1 型が優勢：第 3 モーラ＝「い (の長音か二重母音)」；0 型が優勢：第 3 モーラ＝撥音(塩田 2016b)
- L#LL：0 型が優勢：第 3 モーラ＝自立拍(塩田 2016b)

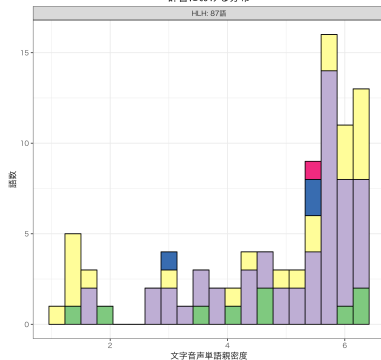
- 今回の調査結果

- H#L：辞書はゆれがあるが、読み上げと自然性評価はゆれがない
- LL#L：1 型・0 型とも自然だと評価される；辞書の場合 1 型が優勢だが、読み上げの場合 0 型が優勢
- L#H と L#LL：0 型・1 型とも自然だと評価される；辞書と読み上げの場合 0 型が優勢

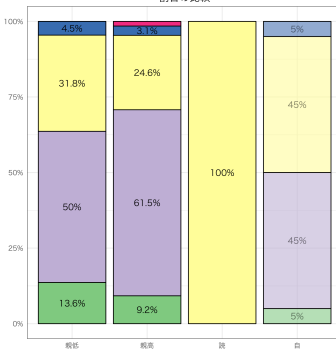
調査結果：外来語 HLH の場合

辞書における分布

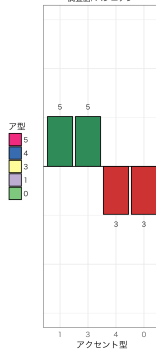
HLH: 87語



割合の比較



自然性評価の結果
調査語: ハンニナン



親密度が上位に来る単語

表記	ア型	親密度
レントゲン	0	6.344
チャンピオン	1	6.375
オートバイ	3	6.344
サイドカー	4	5.469

調査語 (無意味語・読み上げ)

表記	ア型 (読)	複合語?
アーシャゾン	3	アー シャゾン
カンリューズ	3	NO
ハンニナン	3	NO
レッタセン	3	NO
パーチャョー	3	パー チャョー

相関



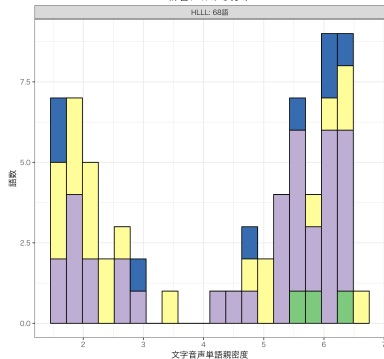
平均絶対誤差 (MAE)



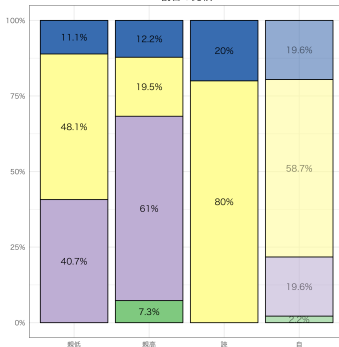
調査結果：外来語 HLLL の場合

辞書における分布

HLLL: 68語

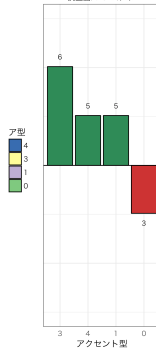


割合の比較



自然性評価の結果

調査語: ノンムクボ



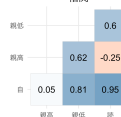
親密度が上位に来る単語

表記	A型	親密度
ハーモニカ	0	6.281
コンタクト	1	6.344
ヨーグルト	3	6.500
ランドセル	4	6.250

調査語 (無意味語・読み上げ)

表記	A型 (読)	複合語?
ムーロカブ	3	ムーロ カブ
リョーケブヒ	3	リョー ケブヒ
ヒンツボス	4	ヒンツ ボス
ノンムクボ	3	ノン ムクボ
ヒップシャリ	3	ヒップ シャリ

相関



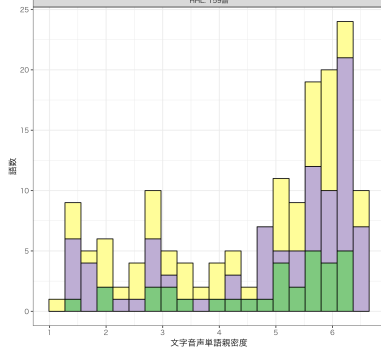
平均絶対誤差 (MAE)



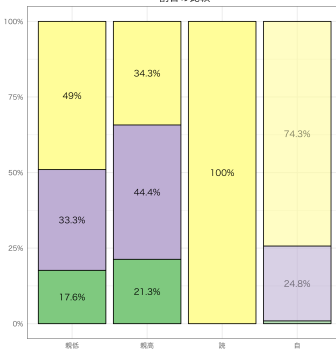
調査結果：外来語 HHL の場合

辞書における分布

HHL: 159語



割合の比較



自然性評価の結果
調査語: ギンターフ



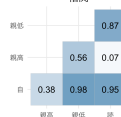
親密度が上位に来る単語

表記	A型	親密度
タイミング	0	6.250
チューリップ	1	6.562
ハンバーグ	3	6.625

調査語 (無意味語・読み上げ)

表記	A型 (読)	複合語?
ギンターフ	3	ギン ターフ
シャーレッチャ	3	シャー レッチャ
パンヘージャ	3	パン ヘージャ
ズーミョーバ	3	ズー ミョーバ
リューフンク	3	NO
ツッパンザ	3	NO

相関



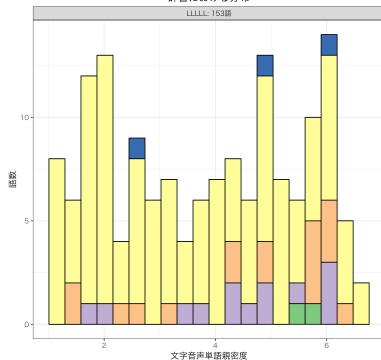
平均絶対誤差 (MAE)



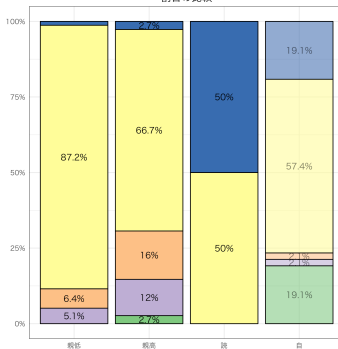
調査結果：外来語 LLLLL の場合

辞書における分布

LLLLL: 153語



割合の比較



自然性評価の結果

調査語: シザクブレ



親密度が上位に来る単語

表記	A型	親密度
プレミアム	0	5.719
ファクシミリ	1	6.062
オリジナル	2	6.219
クリスマス	3	6.625
ソフトウェア	4	5.938

調査語 (無意味語・読み上げ)

表記	A型 (読)	複合語?
ロメビレチョ	4	ロメ ビレチョ
シザクブレ	3	シザク ブレ

相関



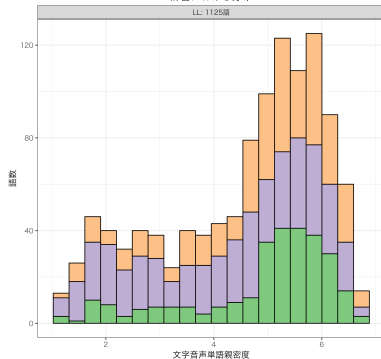
平均絶対誤差 (MAE)



調査結果：和語 LL の場合

辞書における分布

LL: 1125語



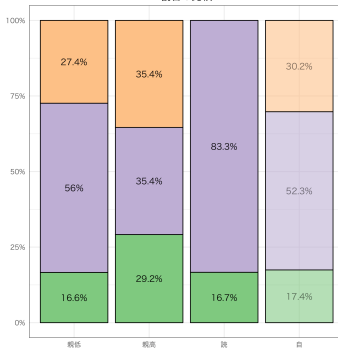
親密度が上位に来る単語

表記	読み	ア型	親密度
酒	サケ	0	6.594
父	チチ	1	6.719
耳	ミミ	2	6.656

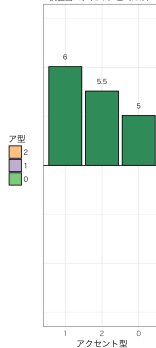
調査語（読み上げ）

表記	読み	ア型（読）
矩	かね	0
魁	たぼ	1
勃	ぶと	1
塙	とや	1
鰯	いな	1
蝶	にし	1

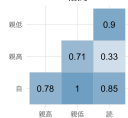
割合の比較



自然性評価の結果
調査語: 塙（とや）魁（たぼ）



相関



平均絶対誤差 (MAE)



まとめ

- 読み上げ調査は一方のアクセント型に偏っているのに対して、自然性評価は複数のアクセント型が容認可能な場合が多い
- 無意味語はモーラ数が多くなるにつれて、より複合語だと判断される傾向がある（モーラ数と「複合語だと思う」傾向の相関は 0.806）
- 複合語の分け方は単独の音節構造においてもバリエーションが生じる

問題点

- 調査語の数が限られているため、アクセント型は特殊な語の音配列に影響されやすい（特に自然性評価の場合）
- 調査時間が長い（1 時間前後）
- 東京方言話者を集めることが簡単ではない

参考文献

- Boersma, Paul (1997) "How we learn variation, optionality, and probability," in *Proceedings of the Institute of Phonetic Sciences of the University of Amsterdam*, Vol. 21, pp. 43–58, Amsterdam.
- Boersma, Paul (1999) "Optimality-theoretic learning in the Praat program," in *IFA proceedings*, Vol. 23, pp. 17–35.
- Boersma, Paul & Bruce Hayes (2001) "Empirical Tests of the Gradual Learning Algorithm," *Linguistic inquiry*, Vol. 32, No. 1, pp. 45–86.
- Ito, Junko & Armin Mester (2016) "Unaccentedness in Japanese," *Linguistic Inquiry*, Vol. 47, No. 3, pp. 471–526.
- Jäger, Gerhard (2007) "Maximum entropy models and stochastic Optimality Theory," *Architectures, rules, and preferences: variations on themes by Joan W. Bresnan. Stanford: CSLI*, pp. 467–479.
- Kubozono, Haruo (1996) "Syllable and Accent in Japanese: Evidence from Loanword Accentuation," *Onsei-Gakkai-Kaiho*, Vol. 211, pp. 71–82.
- Legendre, Géraldine, Yoshiro Miyata, & Paul Smolensky (1990) *Harmonic grammar: A formal multi-level connectionist theory of linguistic well-formedness: Theoretical foundations*: Citeaser.
- Zuraw, Kie Ross (2000) "Patterned exceptions in phonology," Ph.D. dissertation, UNIVERSITY OF CALIFORNIA Los Angeles.
- 天野成昭・近藤公久 (1999) 『日本語の語彙特性』, 三省堂, 東京.
- 金田一京助・柴田武・山田明雄・山田忠雄 (編) (1989) 『新明解国語辞典第四版』, 三省堂, 東京.
- 塩田雄大 (2016a) 「NHK アクセント辞典 “新辞典” への大改訂 ④外来語のアクセントの現況 ～在来語化する外来語～」, 『放送研究と調査』, 第 66 巻, 第 10 号, 84–102 頁.
- 塩田雄大 (2016b) 「NHK アクセント辞典 “新辞典” への大改訂 ⑥漢語のアクセントの現況 ～変化の「背景」を探る～」, 『放送研究と調査』, 第 66 巻, 第 12 号, 64–85 頁.
- 伝康晴・小木曾智信・小椋秀樹・山田篤・峯松信明・内元清貴・小磯花絵 (2007) 「コーパス日本語学のための言語資源: 形態素解析用電子化辞書の開発とその応用」, 『日本語科学』, 第 22 巻, 101–122 頁.
- 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』, 第 20 号, 11–20 頁.