A-4 日本語における語の情報量と持続時間の関係

橋本大樹（東京大学）/ daiki@phiz.c.u-tokyo.ac.jp

[要旨] コンテクストから予測可能性の高い語は、弱化した音声として音声実現することが知られている。この現象は確率性弱化（probabilistic reduction）と呼ばれる。こうした語の予測可能性と持続時間の関係は英語や英語に関わる強勢時間言語で議論されているが、日本語に関してはほとんど議論されていない（Jaeger & Buz, 2017）。日本語は所謂モーラ時間言語で英語やその関連言語とはそのリズムが大きく異なる。本研究では、諸外国語で観察されてきた予測可能性に応じた弱化が日本語に見られることを明らかにする。また予測可能性の高いコンテクストに現れやすい語は、ローカルに予測可能性が低いコンテクストであっても弱化することが知られている（Seyfarth, 2014; Sóskuthy & Hay, 2017）。これは語の発音が以前使用されたコンテクストに影響される現象で、累積使用効果（cumulative usage effect）と呼ばれるものである。本研究では、日本語において累積使用効果については明確な証拠を得ることができないことを示す。これらの現象は日本語話し言葉コーパス（前川, 2004）を用いて明らかにする。

Keywords: probabilistic reduction, cumulative usage effect, corpus, message predictability, exemplar

1. はじめに

コンテクストから予測可能性（contextual predictability）の高い言語単位（例：音素、音節、語）は、弱化した音声として実現することが知られている。この現象は確率性弱化（probabilistic reduction）と呼ばれる。弱化した音声とは音響的・調音的に明確でない音声のことで、中央化した母音や短い持続時間を持つ音声のことを指す（Aylett & Turk, 2006）。例えば Jurafsky et al. (2001) では、先行する語が与えられたときの予測可能性 $p(\text{word}_i|\text{word}_{i-1})$ が高い時、及び後続の語が与えられた時の予測可能性 $p(\text{word}_i|\text{word}_{i+1})$ が高い時、語の持続時間は短くなることが示されている。Jaeger & Buz (2017) で “the majority of existing research on phonetic reduction coming from English” と述べられている様に、予測可能性の効果は英語やそれに関連する強勢時間言語で多く議論されているが、著者の知る限り日本語（モーラ時間言語）で確率性弱化を議論している研究はほとんどない。(但し、Sano (2018) と Turnbull (2018)、 Shaw & Kawahara (2018) は文節音レベルで確率性弱化を議論している。) こうした研究背景から “モーラ時間言語の語レベルで確率性弱化は見られるのか” という一般的な疑問が生まれる。本研究ではモーラ時間言語のひとつである日本語で、この研究課題に取り組む。2 節で議論するが、この研究課題に取り組むことで “人間が情報伝達をどのように行っているのか” という理論的問題に知見を与えることができる：

[RQ1] モーラ時間言語である日本語において、語の音声実現はその予測可能性にどの程度影響を受けるか？

この研究課題に加えて、本研究では予測可能性に応じた弱化がどの程度記憶に残っているかも明らかにする。Seyfarth (2014) と Sóskuthy & Hay (2017) では、予測可能なコンテクストに現れやすい語は、予測しにくいコンテクストに現れやすい語に比べて、平均して持続時間が短いことを示している。(つまり予測しにくい環境に現れる場合であっても、予測可能なコンテクストに現れやすい語は短く発音されるという意味である。) こうした音声実現が過去の使用の影響を受ける効果のことを、累積使用効果（cumulative usage effect）と呼ぶ。この結果は話者が今まで語をどのように使ってきたかを記憶していることを意味する。これは後述の Exemplar Theory の考えに一致するもので、我々の言語知識が音声的詳細を伴うエピソード（即ち

exemplar）から形成されていることを示唆するものである。本研究では日本語という英語とは大きく異なる言語で、この効果を再現できるかを明らかにする：

[RQ2] 日本語において、語の持続時間に関する累積使用効果は再現できるか？

本稿は以下の構成になっている。まず 2 節で上述の二つの研究課題に関連する理論を紹介し、それらに関わる予測を演繹する。3 節でこの予測をどのように検証するかという方法について述べ、4 節では得られたデータに統計分析を行う。5 節でその結果を理論に照らし合わせながら議論し、結論をまとめる。

2. 理論と予測

二つの研究課題に対して、それぞれ Message-Oriented Phonology (MOP) と Exemplar Theory を用いて予測を演繹する。両理論は排除しあうものではなく、両立するものである（Hashimoto, 2019a: 1.7）。

2.1 Message-Oriented Phonology (MOP)

MOP は RQ1 に関わる理論である（Hall et al., 2016; Hume 2016; Hashimoto, 2019a）。MOP の鍵となる仮説は“音体系は情報伝達の最適化によって形成される”というものである。我々の情報伝達は図 1 のようになっている。まず話者が伝えたいメッセージを音声信号に変え、それを伝達する。その信号を聞き手が受け取り、そこから話者のメッセージを推測する。話し手の意図したメッセージと聞き手の推測したメッセージが一致するとき、情報伝達が正確に行われ成功したと呼ばれる（Shannon, 1948）。しかし情報伝達は常に完璧ではなく、この情報伝達には常に不確定さ（uncertainty）がつきまとう。例えば [wi:ʔ] のような弱化した音声信号は、複数のメッセージ（“wheat”や“weed”、“week”）。このような場合、情報伝達に曖昧さが生まれてしまう。このことは英語だけでなく、自然言語一般に言えることである。

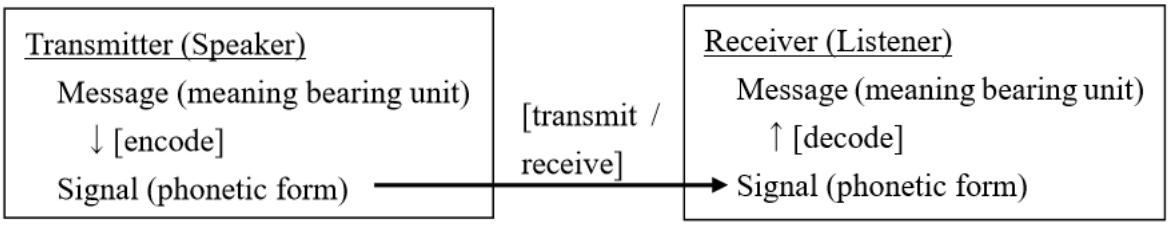

図 1. 情報伝達の経路（adapted from Hashimoto 2019a）

情報伝達の正確さを高めるために、話者は音声信号に余剰性（redundancy）を加えることができる。多くの場合余剰性の高い音声信号というのは、情報伝達を正確にする。例えば語末の閉鎖音が明確に破裂しているような [wi:tʰ] という発音は、より正確に“wheat”というメッセージにマッピングされるだろう。しかしながら音声信号に余剰性を加えると、その分情報伝達の円滑さ（efficiency）が失われてしまう。つまり情報伝達において余剰性と円滑さはトレードオフの関係になっているのである。そのため MOP は“話者は不用意な余剰性を避け、十分に正確で円滑な情報伝達を目指す”と仮説づけている。これこそが MOP の情報伝達の最適化に関わる仮説で、以下の一般的予測を演繹する：

[一般的予測] 話者は重要なメッセージには余剰性の高い音声信号を産出するが、重要でないメッセージには余剰性の低い音声信号を産出する。

ここで重要なメッセージとはどのようなメッセージであるか定義する必要がある。メッセージの重要性を

決める一つの要因は予測可能性であろう。コンテクストから予測できないメッセージは、情報量（information content）が高く、情報伝達において重要である。一方コンテクストから予測できるメッセージは、情報量が低く、情報伝達において重要でない。(information content は事象確立-$\log_2$で対数変換したもの。つまり確率とは負の相関を示す：-$\log_2$ (0.25) = 2 bits > -$\log_2$ (0.5) = 1 bit）この仮説と一般的予測を組み合わせて、語の情報量（予測可能性）に関して以下の予測を演繹できる：

[RQ1 への予測] 日本語においても、情報量が高い（予測可能性の低い）メッセージに対しては、余剰性を加えるため、そのメッセージに対応する語の持続時間は長くなる。

先行研究（Jurafsky et al., 2001; Seyfarth, 2014）に倣い、語のメッセージの情報量を以下のように隣接する語が与えられたときの条件確立として定義する：

$$\mathrm{IC}(w_i|w_{i-1}) = -log_2(p(w_i|w_{i-1})) \qquad \mathrm{IC}(w_i|w_{i+1}) = -log_2(p(w_i|w_{i+1}))$$

2.2 Exemplar Theory

RQ2 に関わるのが Exemplar Theory (Pierrehumbert, 2001; Docherty & Foulkes, 2014; Hashimoto, 2019b, c) である。Exemplar Theory の重要な仮説は“会話の中で出会った全ての言語使用は詳細な情報とともにエピソード（exemplar）として記憶される”というものである。例えば［pengiN］という音声に会話の中で出会ったとしよう。この時我々の認知スペースに［pengiN］という詳細な音声（例：フォルマントの値や、VOT の値など）が記憶され、それに伴い統語的情報や社会的情報（このトークンを産出した話者の情報）も記憶される。その結果我々の言語知識は膨大な数の exemplar で表示されることになる。本研究での Exemplar Theory は Pierrehumbert が唱えるハイブリッド型であり、“類似した exemplar は認知システムの中で塊を成し、カテゴリを形成する”という仮説も重要である。例えば日常的に［pengiN］というトークンに複数出会うことがあれば、これらの exemplar は塊を成し、語彙カテゴリとして PENGUIN が形成されるだろう。語彙カテゴリと同様に、類似した exemplar が集合することで、音韻カテゴリ（例：音素や音節など）や社会カテゴリ（例：女性や原宿系など）も形成されるだろう。実際の言語産出は、認知スペースに記憶されているカテゴリと exemplar の選択によって行われる。例えば「ペンギン」というメッセージを送りたい場合、PENGUIN というカテゴリを活性化し、そこに所属している exemplar を選択する。選択された exemplar を基に算出ターゲット（production target）を形成し、これを音声信号として産出する。

話者には言語使用に関して詳細なエピソード記憶があるのであるから、今までの会話でどのように語が使用されてきたか記憶していても驚くことではない。このことから以下の一般的予測を演繹することができる：

[一般的予測] 今までの会話において、ある語彙がどのような音声実現をしてきたか、話者は記憶している。

2.1 節である語が予測可能性の高いコンテクストに現れる時、その後は弱化して音声実現することを予測した。Exemplar Theory の一般的予測によれば、こうした音声実現は認知システムに詳細に記憶されている。つまり予測可能性の高いコンテクストに何度も現れるメッセージに対応する語彙カテゴリは、弱化した音声実現を伴う exemplar を多く含む。逆もまた真で、予測可能性の低いコンテクストに何度も現れるメッセージに対応する語彙カテゴリは、弱化しない音声実現を伴う exemplar を多く含む。この“どの程度の頻度で予測可能なコンテクストに現れるか”という性質は、informativity という単位で測ることができる。これは各語の情報量の期待値である。以下の予測を演繹できる：

[RQ2 への予測] informativity の高い（情報量の期待値の高い）語のカテゴリは、余剰性の高い exemplar がたくさん所属しているため、平均して語の持続時間は長くなる。

2.1 で定義した情報量に基づき、informativity を以下のように定義する：

$$\text{Informativity}(w_i|w_{i-1}) = \sum_{w_{i-1}} p(w_i|w_{i-1}) * IC(w_i|w_{i-1}) \qquad \text{Informativity}(w_i|w_{i+1}) = \sum_{w_{i+1}} p(w_i|w_{i+1}) * IC(w_i|w_{i+1})$$

3. 検証方法

本研究は、日本語話し言葉コーパス（前川, 2004）の CSJ-Core と呼ばれるアノテーション済みの wav ファイル・TextGrid ファイルからデータを集めた。なお本研究では学会講演と模擬講演のみからデータを集めた。具体的には、Praat (Boersma & Weenink, 2019) でスクリプトを組み、以下の情報を csv ファイルに変換した：話者（ファイル名)、語のラベル（音声表記)、語の持続時間、先行する語のラベル、後続する語のラベル、語の前後の韻律句情報（アクセント句またはイントネーション句の境目であるかどうか)、発話番号。結果 392,368 トークンの語が得られた。

この csv ファイルを基に R (R Core Team, 2019) を用いて、各語に関して以下の情報を取得した。これらの情報を統計モデルに組み込むことで、先述の予測の妥当性を確かめる。

○応答変数　*logDur*: log 変換した語の持続時間（min=-3.48, max=0.15, mean=-1.69, med= -1.65, sd=0.58）

○説明変数（下線のものは興味のある変数で、他はコントロール変数。n は数値変数 / b は二値変数）

precIC (n) 先行する語が与えられた時の語の情報量　　*folIC* (n) 後続する語が与えられた時の語の情報量

precInf (n) 先行する語に基づく語の informativity　　*folInf* (n) 後続する語に基づく語の informativity

logFreq (n) log 変換した語の頻度　　*NofMora* (n) 語中のモーラ数

speech (b) 学会講演 vs. 模擬講演　　*speechRate* (n) 発話内のモーラ数/秒

precAP (b) 語の直前に AP の境界がある vs. ない　　*folAP* (b) 語の直後に AP の境界がある vs. ない

precIP (b) 語の直前に IP の境界がある vs. ない　　*folIP* (b) 語の直後に AP の境界がある vs. ない

※ 数値変数の標準偏差 3 を超えるものは外れ値として削除したため、最終的に 370,119 トークンになった。

4. 統計分析

370,119 トークンに ① mixed-effects linear modelling (Bates et al., 2015) と ② generalized additive mixed modelling (Wood, 2011) を行った。それぞれの分析について以下で述べる。

まず ① に関して述べる。著者は backward elimination でモデルを選択しようと試みた。この手法は一番大きなモデルから、有意でない変数を一つずつ切っていく方法である。しかしながら random slope を複数フィットしたモデルは、どれも収束 (converge) しないことがわかった。この理由は、本研究で扱うデータが膨大

	Estimate	Std. Error	t value
(Intercept)	-1.2556042	0.0113411	-110.713
logFreq	-0.0319032	0.0013274	-24.034
precIC	0.0019679	0.0002042	9.637
folIC	0.0155763	0.0002058	75.677
precInform	0.0023217	0.0006925	3.352
folInform	-0.0073717	0.0006963	-10.587
NofMora	0.2991066	0.0021913	136.500
speech simulated	-0.0193475	0.0027149	-7.126
speechRate	-0.0990244	0.0003623	-273.330
precAP yes	0.0329390	0.0016074	20.491
precIP yes	0.0219554	0.0018377	11.947
folAP yes	0.1079614	0.0015195	71.052
folIP yes	0.1685684	0.0015488	108.836

表 1. Mixed-effects linear model ①

であることと、12 もの説明変数を 1 つのモデルにフィットしていることなどが考えられる。そこでこの統計分析に関しては、random slope を諦めて random intercept のみフィットすることにした。説明変数はそのまま使い、話者ごとの random intercept と語ごとの random intercept をフィットしている。なお optimizer に optimx (Nash, 2014) を含んでいる。結果は表 1 の通りで、12 の説明変数は全て有意である（p<0.05)。

次に②について述べる。この統計分析では直線関係にない変数間の関係も捉えることができる（Winter & Wieling, 2016; Sóskuthy, 2017)。①と同様に 12 の説明変数を全てモデルに組み込んだ。この分析では数値変数に関しては centred value に変換している。また全ての数値変数に、話者ごとの random smooth をかけている。図 2 が示すようにこれらの 4 つの変数は *logDuration* との関係は完全な linear でないため、random smooth によって話者ごとのバリエーションは捉えられている。①同様、12 の説明変数は全て有意（p<0.05）である。

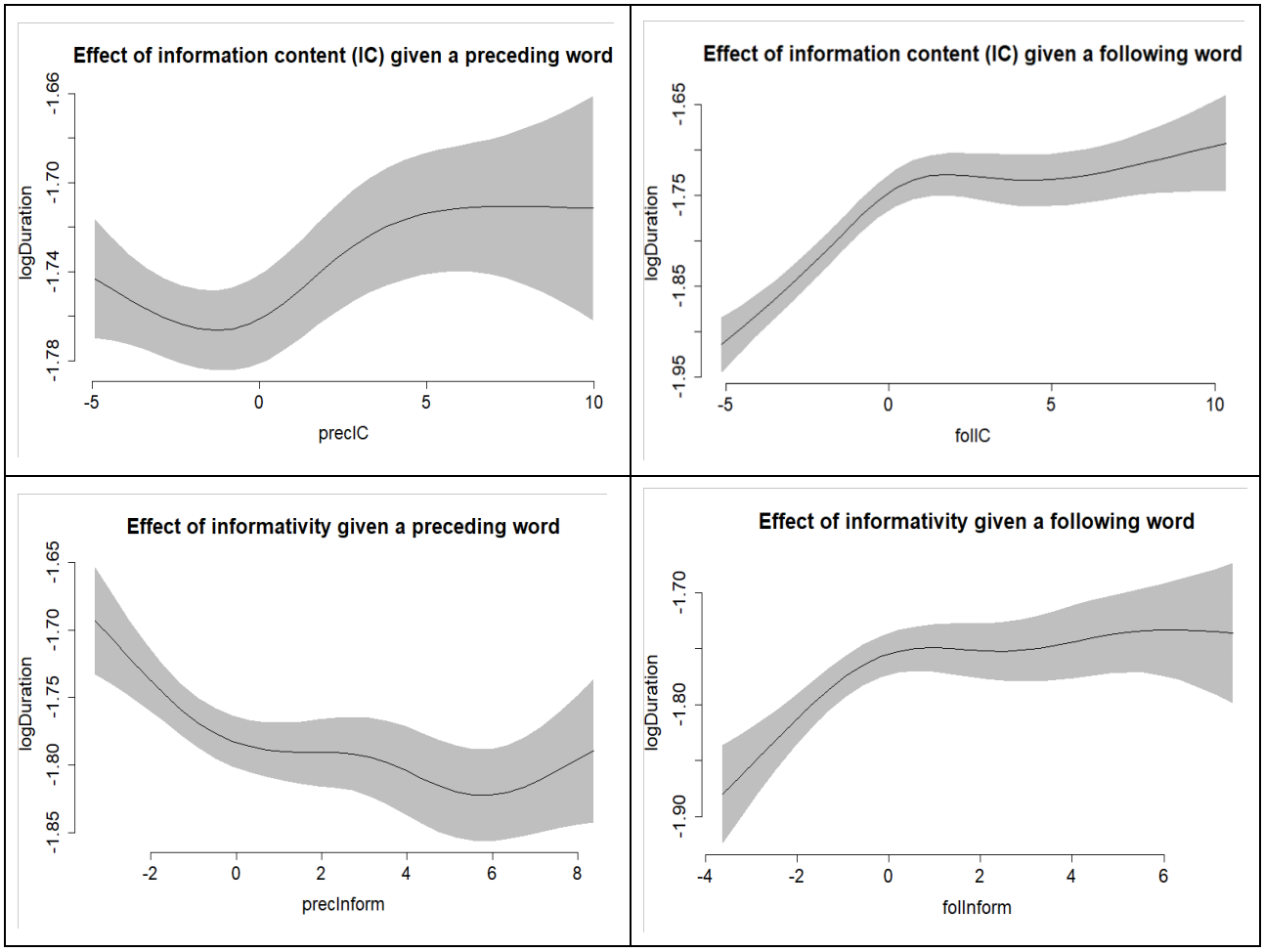

図 2. Generalized additive mixed model ②

5. 議論と結論

本節では前節で示した統計分析結果が 2 節で演繹した 2 つの予測をどの様に支持・反駁したのかについてまとめ、本研究をまとめる。

まず MOP に基づき、日本語においても“話者は重要なメッセージには余剰性の高い音声信号を産出するが、重要でないメッセージには余剰性の低い音声信号を産出する”という予測を演繹した。この予測は本研究において大いに支持されたといえる。本研究では 2 種類の統計分析結果を示したが、どちらの結果もこの

予測を裏付けている。つまり日本語話者も情報伝達の最適性という目的を達成するために、メッセージの重要性（情報量）に応じて、それを伝える音声信号の余剰性を変更している。重要なメッセージにはより多くの余剰性を音声信号に加え、重要でないメッセージには音声的余剰性を加えない。その結果重要なメッセージ（情報量の多いメッセージ）は長い持続時間で音声実現し、重要でないメッセージ（情報量の少ないメッセージ）は短い持続時間で実現する。

この予測をベースに、Exemplar Theory を用いて演繹した予測が以下である：“informativity の高い（情報量の期待値の高い）語のカテゴリは、余剰性の高い exemplar がたくさん所属しているため、平均して語の持続時間は長くなる”。(Informativity は情報量の期待値を数値化した単位であることを思い出してほしい。）普段から予測可能性の高いメッセージは重要でないため、少ない余剰性を持つ音声信号として実現しやすい。その結果これらの短い持続時間を持つ発音がエピソード（exemplar）として我々の認知体系に蓄積する。それゆえ普段から情報量の低い語彙カテゴリを産出する際には、少ない余剰性を持つ exemplar が選択されやすいはずであると予測した。この予測は本研究の結果から支持されたとは言い難い。Mixed-effects linear model の結果は、*precInform* に関しては予測通りの振る舞いをしていると解釈できるが、*folInform* に関しては逆の効果を有意に示している。一方で Generalized additive mixed model の結果に関しては、*folInform* に関しては予測通りの振る舞いをし、*precInform* に関しては予測通りでない。この結果に関しては、今後の研究で再検討していきたいと考えている。

本研究では情報量に応じて、日本語話者がメッセージの音声実現を変更していることを示した。この研究結果は、あくまで真実の一端に過ぎない。情報理論に基づく研究は今後更に増えると予想され、“日本語の音韻体系がなぜ今の形になっているか”に関する新たなアプローチを提供するだろう。本研究で見たような音声信号の持続時間といった音声的現象だけでなく、アクセント規則や連濁などの音韻現象が“なぜそうなっているのか”が明らかになっていくことが期待されている。あらゆる音声・音韻現象がコミュニケーションの最適化による産物であり、今まで提案されてきた規則や制約を全て情報伝達の最適化によって統合的に説明できるかもしれない。

参考文献

Aylett, M., & Turk, A. (2006). Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllable nuclei. *Journal of Acoustic Society of America*, 119, 3048-3058.

Bates, D., Maechler, M., Bolker, B., Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67, 1-48.

Boersma, P., & Weenink, D. (2019). Praat version 6.0.49. praat.org

Hall, K., Hume, E., Jaeger, T. F., Wedel, A. (2016). The message shapes phonology. Unpublished manuscripts. Retrieved from https://www.researchgate.net/publication/309033386_The_ Message_ Shapes_Phonology/download

Hashimoto, D. (2019a). *Loanword phonology in New Zealand English: Exemplar activation and message predictability* (Doctoral thesis, University of Canterbury). Retrieved from https://ir.canterbury.ac.nz /handle/10092/16634

Hashimoto, D. (2019b). Sociolinguistic effects on loanword phonology: Topic in speech and cultural image. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, 10(1) 11, 1-34.

Hashimoto, D. (2019c). Cumulative usage effects on lexeme-final /s/: Probability of being affixed. *Lingua*, https://doi.org/10.1016/j.lingua.2019.102739

Hume, E. (2016). Phonological markedness and its relation to the uncertainty of words. 『音韻研究』 19. 107-116.

Jaeger, T. F., & Buz, E. (2017). Signal reduction and linguistic encoding. In E. M. Fernández, & H. S. Cairns (Eds.) *The handbook of psycholinguistics* (pp. 38-81). New Jersey: WileyBlackwell.

Jurafsky, D., Bell, A., Gregory, M., Raymond, W. (2001). Probabilistic relations between words: Evidence from reduction in lexical production. In J. Bybee, & P. Hopper (Eds.) *Frequency and the emergence of linguistic structure* (pp. 229-254). Amsterdam: Benjamin.

前川喜久雄. (2004).『日本語話し言葉コーパス』の概要. 『日本語科学』15, 111-133.

Nash, John C. (2014). On Best Practice Optimization Methods in R. *Journal of Statistical Software*, 60, 1-14.

Pierrehumbert, J. B. (2001). Exemplar dynamics, word frequency, lenition, and contrast, In J. Bybee, & P. Hopper (Eds.) *Frequency and the emergence of linguistic structure* (pp. 135-157). Amsterdam: John Benjamins.

R Core Team. (2019). R version 3.5.3. https://www.r-project.org/

Sano, S. (2018). Durational contrast in gemination and informativity. *Linguistics Vanguard*, 2017-0011.

Seyfarth, S. (2014). Word informativity influences acoustic duration: Effects of contextual predictability on lexical representation. *Cognition*, 133, 140-155.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27, 329-423.

Shaw, J., & Kawahara, S. (2019). Effects of surprisal and entropy on vowel duration in Japanese. *Language and Speech*, 62, 80-114.

Sóskuthy, M. (2017). Generalized additive mixed models for dynamic analysis in linguistics: A practical introduction. arXiv: 1703.05339v1

Sóskuthy, M., & Hay, J. (2017). Changing word usage predicts changing word durations in New Zealand English. *Cognition*, 166, 298-313.

Turnbull, R. (2018). Effects of lexical predictability on patterns of phoneme deletion/reduction in conversational speech in English and Japanese. *Linguistics Vanguard*, 2017-0033.

Winter, B., & Wieling, M. (2016). How to analyze linguistic change using mixed models, Growth Curve Analysis and Generalized Additive Modeling. *Journal of Language Evolution*, 1, 7-16.

Wood, S.N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society*, (B) 73(1), 3-36.