ARDJ で使用した文に対する反応時間の取得とその多変量解析
ARDJ is coupled with RT data: Preliminary MVA

黒田 航[1], 阿部 慶賀[2], 粟津 俊二[3], 寺井 あすか[4], 土屋 智行[5]
Kow Kuroda, Keiga Abe, Shunji Awazu, Asuka Terai, Tomoyuki Tsuchiya

[1]杏林大学, [2]岐阜聖徳学園大学, [3]実践女子大学, [4]公立はこだて未来大学, [5]九州大学
Kyorin Univ., Gifu Shotoku Univ., Jissen Woman's Univ., Future Univ. of Hakodate, Kyushu Univ.
kow.k@ks.kyorin-u.ac.jp, abekeiga@gifu.shotoku.ac.jp, awazu-shunji@jissen.ac.jp, aterai@fun.ac.jp, tsuchiya@flc.kyushu-u.ac.jp

概要

日本語容認度評定データ (ARDJ) 構築の第一期と第二期の調査で刺激文に使われた 466 文の読み時間データを追加収集し，評定値データと対応づけた．そのデータの多変量解析と回帰分析の結果から，容認性判断とそれに要する時間は，刺激文を分割された部分への反応時間からは予測できない事が示唆された．ただし元になった反応時間データに代表性が保証されていないため，結果の一般性には自ずから限界がある．

1 はじめに

日本語容認度評定データ (Acceptability Rating Data for Japanese: ARDJ) [11] は，大規模かつ確証バイアスの少ない日本語の文の容認度評定値のデータベースである．これは**証拠に基づく言語学** (Evidence-based Linguistics: EBL) [10] の実現を可能にする基礎データとして企画され，これまでに第一期 [3]，第二期 [4, 5] の調査を実施し，成果を公開した[1)]．構築は現在も継続している．

ARDJ の大規模性には二つの意味がある．第一に，評定された刺激文の数が多い (第 1 期と第 2 期を合わせて異なり数 466)，第二に，評定者の数が多い事 (第二期では一文について 70 名以上の評定者から評定を得ている)．ARDJ の脱確証バイアス性には二つの意味がある．第一に，刺激文の構築が特定の言語理論を確証，ないしは反証するために行なわれたものではない．理論的には刺激文の集合は，可能性空間の部分空間のランダムサンプリングの結果として得られている．詳細は §2.2 を参照．第二に，評定者を状況が許す限り無作為化している．評定者は，性別，年齢，居住地の属性値の選択でなるべく片寄りが生じないように選ばれている．

更に ARDJ は部分的に社会調査として実施されている．それは調査の際に，単に日本語の文の容認度を評定して貰っているだけでなく，評定者の社会的属性 (年齢，性別，異国語学習歴，異国での生活歴，暮らした地域，教育歴，読書量，文系理系の程度など) を一緒に収集している．これにより，容認度に影響を与える要因の層別解析が可能になる．結果は [12] で簡単に報告した．

ARDJ は無償で利用可能なデータである．申し出があれば誰でも利用できる．これは得られた結果の再現可能性と反証可能性を高める．

1.1 本研究の意義

読み時間データは，特定の統語理論の検証 (通常は確証) を目的として実施される心理言語学実験で良く利用される．ARDJ は 466 種類の文に対する容認度評定値データを公開している．これらの刺激文に評定値の他に読み時間データを対応させる事は，データの利用可能性の向上させると期待できる．

本報告には得られた結果の多変量解析を含めるが，検討範囲と精度が十分とは言えない．特に統制されていない実験条件の異なりが結果にバイアスを生じさせている可能性は大きい．だが，この難点は研究の目的から見て本質的ではない．本研究は検証型の研究ではなく探索型の研究であり，これまでに認知心理学や認知科学で探索されなかった可能性空間を探索するのが目的だからである．本研究には結論と言えるものが伴っていないが，それは特定の仮説を確証していないからである．

2 RT データ収集

2.1 課題

調査の目的は RT の取得であるが，それを self-paced reading のパラダイムで実装した．実験参加者は，刺激文が事前に指定されている区切り (表 1 の S (segmented) で "/" で示されている単位) ごとに段階的に提示される．適

[1)] ARDJ データ公開用サイト https://kow-k.github.io/Acceptability-Rating-Data-of-Japanese/

表 1: RT 実験で使った刺激文 (gr0) の見本 [S (segmented) の部分を使用]

S.ID	RT.S.ID	V.ID	Pattern	Type	S (segmented)	#seg
s1-016	474	44	p3	o	高校生が/デートの場で/しらじらしさを/恋人に/感じた。	5
s1-062	270	338	p1	s	ころっと/相手に/大事な試合で/有望選手が/負けた。	5
s1-114	240	131	p3	o	刑事が/捜査で/手がかりを/手当たり次第に/探した。	5
s1-117	66	326	p1	n	客が/そのスーパーで/店員に/文句を/言うと/黙った。	6
s1-122	354	1197	p1	p	私が/遊園地で/インフルエンザに/家族に/感染した。	5
s1-144	114	326	p4	s	不安から/妊娠を/次女が/実家で/黙った。	5
s1-186	342	22	p3	v	船が/遠回りで/海路を/安全に/来た。	5

当なキー (例えばスペースバー) を押すと，次の区切りが提示される．このように実験参加者は自分のペースで読み進め，全体の提示終了後に，文を (1) のいずれかに判定するように求められた:[2)]

(1) 1: 違和感がなく自然に理解できる文
2: 不自然で理解不能な文

この課題の結果，(rt1, rt2, …, rt5, RT, response) という数値が得られる．rt$i$ は i 番目の区画と $i+1$ 番目の区画の反応時間の差であり，次の区画に移動するまでの所要時間を意味する．RT は最後の区画を見た後に (1) の容認度判断を下すまでの時間で，response の値 (1 か 2) が判断の結果である[3)]．

2.2 刺激

ARDJ はこれまでに調査 1 (survey 1) と調査 2 (survey 2) を実施し，2 回の調査で延べ 466 種類の刺激文の評定値を得ている．その内訳は次の通り: 調査 1 では 200 種類の文を刺激に使った．調査 2 では 300 種類の文を刺激に使ったが，調査 1 から 12 種類を再利用した．そのため二つの調査で使った刺激文は 466 (=188 + 280 − 2) 種類[4)]．

表 1 に RT 取得に使われた刺激の見本を示す．刺激文の分節数は 5 か 6 である．大半を 5 の場合が占めるが少数ながら 6 個の刺激がある[5)]．表 1 の変数の簡単な説明は次の通り:[6)]

(2) a. S.ID は刺激文の ID (s1- は第 1 期のみで使われた刺激，s2- は第 2 期で使われた刺激を意味する).
b. RT.S.ID は S.ID とは別に今回の実験でデータの無作為化のために利用した ID.
c. V.ID は刺激文を作成する際に種文に使われた動詞の ID (NINJAL-LWP for BCCWJ[7)] の動詞の頻度順位に弱く対応).
d. Pattern は動詞の値に拠らずに事前に決めた 5 種類の文の雛型
e. Type は §2.3 で説明する刺激文の編集の型.
f. S (segmented) は刺激文 (分割箇所は “/” で補助的に示している).
g. #seg は分割数

2.3 刺激文の作成手順

刺激文は 65 種類の原文 (originals) に (「変異」と呼ぶ) 無作為な編集を適用して生成された．変異は i) 動詞の置換 (mutated verb)，ii) 名詞の置換 (mutated nominal)[8)]，iii) 格助詞の置換 (mutated positional)，iv) 句の入れ替え (phrase swapping) である (詳細は [3, 4] を参照)．変異の分布は表 2 の通り．評定者の属性が状況が許す限り無作為化されている事に加えて，変異が本質的に無作為である事が，確証バイアスの最小化に貢献する．

表 2: 変異の割合

code	type of mutation	count	ratio
o	original [no mutation]	65	0.139
v	mutated verb	90	0.193
n	mutated nominal	108	0.232
p	mutated postpositionl	95	0.204
s	swaped phrases	108	0.232
	sum	466	1.00

2.4 反応

合わせた 466 種類の刺激文をランダムに 6 つのグループ (gr0, gr1, …, gr5) に分割した (1 グループに含まれるのは約 80 文)．それらのグループをなるべく重複しないように被験者に割り当てた．

[2)] 調査 1, 調査 2 の容認度評定は [0: 違和感がなく自然に理解できる文; 1: 違和感を感じるが自然に理解できる文; 2: 違和感を感じ理解が困難な文; 3: 不自然で理解不能な文] の 4 件法だったが，カテゴリー判断を模するように，両端の二値を使った.

[3)] それぞれの実験で使われた数値は異なるが，ここでは {1,2} に統一した．また，回帰解析では 1 → 0，2 → 1 の数値変換を施し，逸脱の程度を [0,1] の値に正規化してある.

[4)] s2u データで得られる異なり数が 468 でなく 466 なのは，s1-010=s2-010=s2-281 と s1-127=s2-127=s2-282 が別の文として扱われているため.

[5)] これは意図した事ではなくて，どちらかと言えば刺激作成の際の不注意による.

[6)] 詳細は [3, 4] を参照されたい.

[7)] http://nlb.ninjal.ac.jp/search

[8)] nominal は形容動詞を含む.

反応の取得は，函館[9]，東京[10]，岐阜[11]の 3 ヶ所で大学生を対象に行った．函館では 10 名から (1 名当たり 90 文への反応で) 合計 900 反応を，東京では 15 名から (1 名当たり 80–83 文への反応で) 合計 1,223 反応を，岐阜では 10 名から (1 名当たり 76–79 文への反応で) 合計 778 名から反応を得た．こうして合計 2,901 反応を収集した．

実施条件の制約から被験者の属性の無作為化はできていない (そもそも，十分な数の反応を集められていない)．得られたデータの rt$i$, RT の分布を見ると，データ収集がすべての場所で同じ条件で実施されたと言いがたい[12]．これらの意味で，本調査で得た反応は代表性を持つとは言いがたく，予備実験的な位置づけを免れない．解析結果は示唆的であるとは言え，一般性に限界がある．

3 解析

3.1 前処理: はずれ値除去

はずれ値の除外では，SD 濾過と Mahalanobis 距離濾過の 2 つを検討した．結果として SD < 3 ではずれ値を除外する事にした．詳細は次の通り．

3.1.1 SD 濾過

表 3: SD を使った濾過の有効行数と含有率

#rows	inclusion rate	upper bound
1691	0.582902	sd < 1
2170	0.748018	sd < 1.5
2436	0.839710	sd < 2
2601	0.896587	sd < 2.5
2697	0.929679	sd < 3
2771	0.955188	sd < 3.5

SD ごとの事例数と含有率を参考値として表 3 に示す．

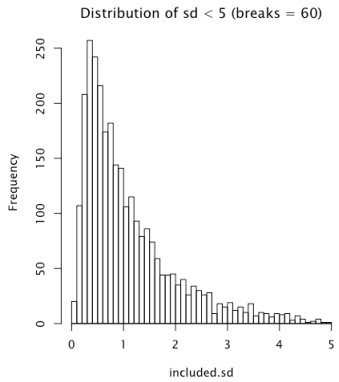


図 1: Histogram of SD's under 5

また，RT, rt1–rt5 の SD 値 (5 未満に限定) のヒストグラムは図 1 にある通り．

3.1.2 Mahalanobis 距離濾過

表 4: Mahalanobis 距離での濾過の有効行数と含有率

#rows	inclusion rate	upper bound
1266	0.436401	M.dist < 1
2403	0.828335	M.dist < 4
2599	0.895898	M.dist < 7
2689	0.926922	M.dist < 10
2749	0.947604	M.dist < 13

rt5 が NA 値を含むため，Mahalanobis 距離の算出では RT, rt1–rt4 のみを用いている．参考値を表 4 に示す．

3.1.3 効果の比較

数値の上では，SD < 3 と M-dist < 10 の外れ値除去の効果が相応するように見えるが，効果はかなり異なる．

図 2 に SD <3 の条件で濾過した反応を，図 3 に Mahalanobis 距離 < 10 の条件で濾過した反応を示す．図では，左端に函館で取得された反応が，右端に岐阜で取得された反応が，中央に東京で取得された反応が位置している．

明らかに 2 つの濾過法ははずれ値の認定法が違う．最大の違いは (Mahalanobis 距離の計算で除外された rt5 ではなく) RT への選別のかかり方である．Mahalanobis 距離では RT の極端な値が除外されていない．この事から判断して，結果的に SD < 3 の条件でのはずれ値除去を採用した[13]．これにより，2697 個 (全体の約 93%) の反応が有効と判断され，以下の解析の対象となった．

3.2 データ取得の反応バイアス

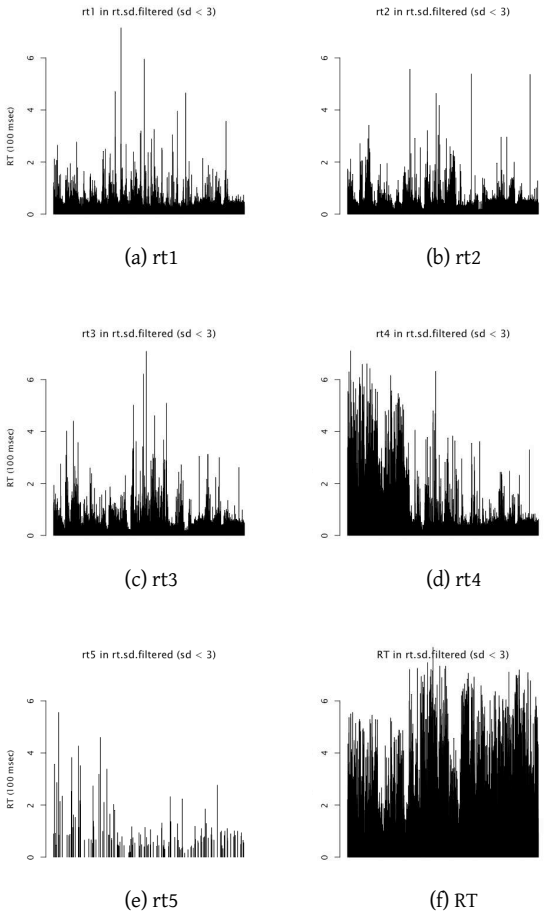
函館，東京，岐阜の 3 ヶ所で同じ条件でデータを取得したが，一部で傾向が違っている．rt4 についてのみ，函館の被験者の反応がおしなべて長い．これは実験の実施条件に違いに拠るものかも知れない．違いを生んでいる要因の検討の必要がある．

3.2.1 生反応の PCA

はずれ値を除去した反応の PCA を図 4 に示す．図 4 に表わされているのは，反応の配置と分類である．これは同一の刺激の集合に対する反応の類型化を与える．X-means 法[14]により，4 つのクラスターが認定されている．密度を考慮すると，分布の中心にあるのはクラスター 2 であり，それから PC1 の負方向にクラスター 1 が，PC2 の正方向にクラスター 3, 4 が発展している．クラスター 3 と 4 の区別は PC1, PC2 平面では表現されていない．

このように 4 つのクラスターが認定されるのは自明で

9) データは MatLab Psychotoolbox で取得.
10) データは SuperLab 4 .5 で Self-Paced Reading の設定で取得.
11) データは PsychoPy3 (Windows 10) で取得.
12) 例えば，函館で実施された実験で明らかに rt4 が長く，RT が短か目である.
13) その一方で，SD による濾過では，rt1, rt2, rt3 の極端な値の濾過が十分でない可能性がある.
14) X-means 法 [7] は k-means 法の一種だが，k を自動推定する.

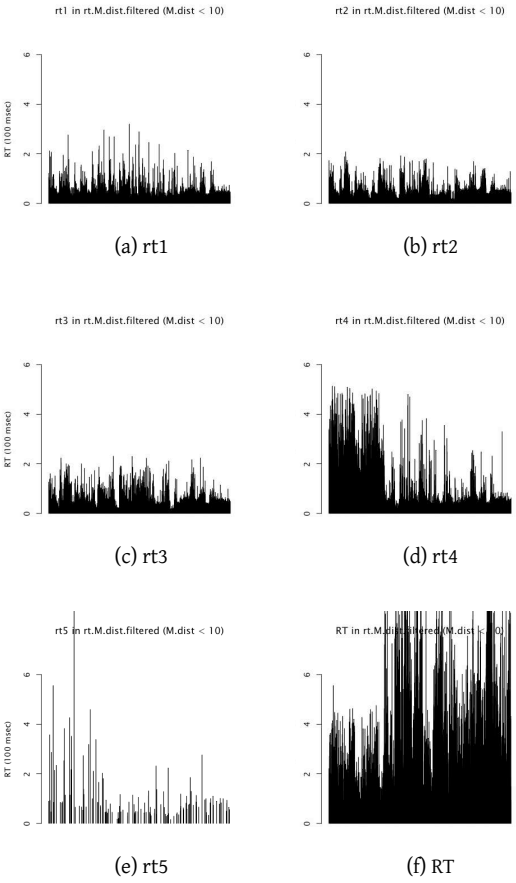


(a) rt1 (b) rt2

(c) rt3 (d) rt4

(e) rt5 (f) RT

図 2: sd-filtered rt values



(a) rt1 (b) rt2

(c) rt3 (d) rt4

(e) rt5 (f) RT

図 3: M-dist-filtered rt values

はなく，有用な解析結果である[15]が，今検討したいのは個々の反応そのものではなく，それらに分散的に表現された刺激文の潜在表現である．それを推定する手法が必要である．その方法として代表値による集約を利用する．

rt5 を持つ刺激 (31 種類) と持たない刺激 (435 種類の) は同様に扱えるか怪しいので，分離した．以下の解析では rt5 を持たない 435 種類の刺激のみを対象としている．

3.3 刺激の表現の見本

刺激文ごとに反応の集合をヴァイオリンプロット[16]をすると，図 5 のようになる．

3.4 RT の集約

刺激文への反応は (rt1, rt2, rt3, rt4, rt5, RT) というベクトルで表される．それぞれの刺激文はこれらのベクトルの集合として表現される．ベクトルの集合を一つの値で代表させるため，複数の値を集約した．集約には median を使った[17]．

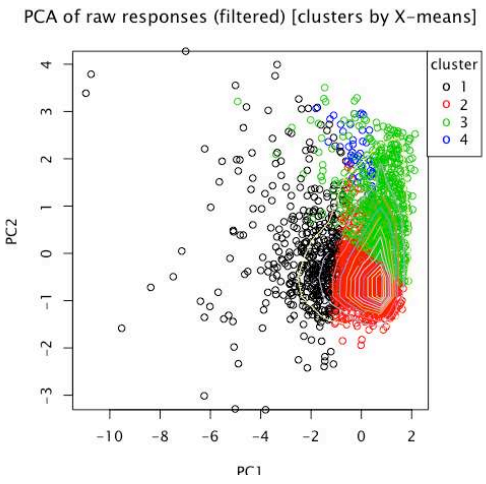


図 4: PCA of raw responses (sd filtered)

3.5 刺激文の PCA

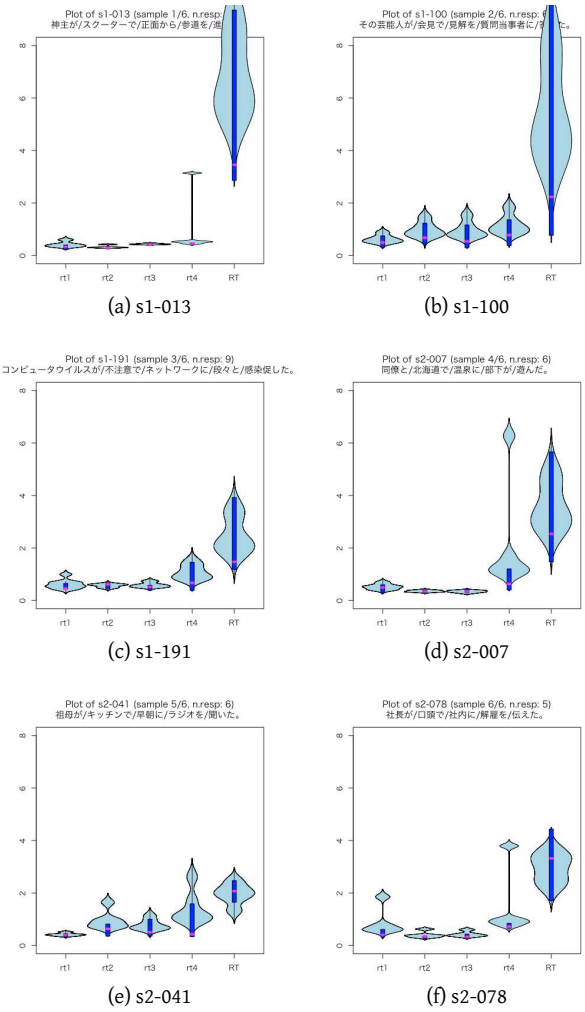
集約された RT のベクトルを PCA で処理し，その結果を X-means でクラスタリングしたところ，図 6 に示すように 4 つのクラスターが認識された[18]．この図にはデー

15) これは被験者の分類に役立つ情報である．

16) 可視化でヴァイオリンプロットを使った理由は，複数の反応に分散的に表現された反応ポテンシャルを，単純なプロットよりうまく集約するからである．

17) 集約の方法で幾何平均も試したが大きな違いがなかった．

18) ここではデータ集約に PCA を使っているが，これは唯一の選択肢ではないし，最良の選択肢でもない．Isometric Mapping [8] や t-SNE [9] を使った集約の方が洞察をもたらすと期待できる [2] が，解釈が難しいため，PCA に基づく結果を紹介している．



(a) s1-013 (b) s1-100

(c) s1-191 (d) s2-007

(e) s2-041 (f) s2-078

図 5: Plots of sample stimuli

タ点の密度を表す等高線を加えてある．密度情報を手がかりにして，次のように言える:

(3) a. クラスター 2 が分布の中心にあり，
b. PC1 と相関して伸張するのがクラスター 4 (と 3)
c. PC2 と相関して伸張するのがクラスター 1 (と 3)
d. クラスター 4 はクラスター 1 と 4 の中間
にそれぞれ対応する (が，PC1–PC2 の平面ではクラスター 4 とクラスター 3 が未分離).

3.5.1 クラスターの内実

認識された 4 つのクラスターの事例をバイオリンプロットで集約して表現すると，次の図 7 のようになる．

3.6 クラスターの解釈

クラスター 2 は標準的な反応を集約的に表現している．rt1–rt3 にはほとんど時間を要さず，rt4 で少し余計に時間がかかり，RT でもう少し時間がかかっている．

クラスター 4 はクラスター 2 に似ているが，rt1–rt3 で少し余計に時間がかかっている．とは言え，クラスター 2

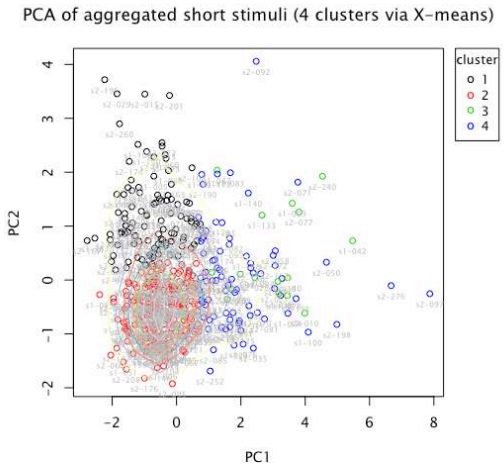
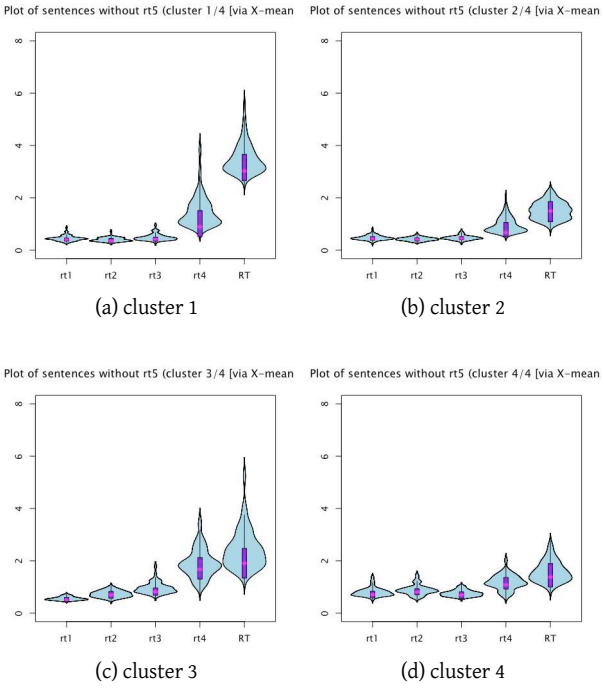


図 6: PCA of stimuli as aggregated rts



(a) cluster 1 (b) cluster 2

(c) cluster 3 (d) cluster 4

図 7: Plot of rt-based clusters

はクラスター 4 の差は小さい．単にグラフで見比べる限りでは区別できない程である．これは逸脱が早い段階で予見でき，RT まで持ち越されない場合だと考えられる．

クラスター 1 はクラスター 2 やクラスター 4 に比べて RT の反応が明らかに長い．RT では容認度判断を求められる．それにかかる負荷が RT に現われていると考えられる．ただ，rt1–rt4 に比べて RT が極端に長引いているとも言えない．逸脱が生じていない限り，容認度評定は負荷の高い課題ではない事が窺える．

クラスター 3 は rt4, RT が長目である点でクラスター 1 に似ているが，クラスター 4 とも共通性がある．ク

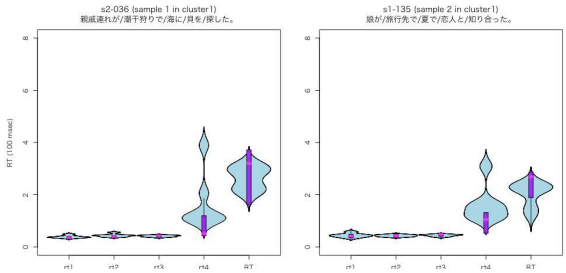
ラスター 3 はクラスター 1 に比べて RT が短い．これは rt1–rt3 の早い段階で逸脱性が予測できて，RT まで持ち越されない点で，そうである．

RT の段階で (予想外の逸脱などにより) 負荷が生じていると言えるのはクラスター 1 のみで，他の 3 つではそれが生じていないと言える (クラスター 3 はクラスター 1 に準じるが，負荷は軽減されている)．

3.7 クラスターごとの事例

3.7.1 クラスター 1 の事例見本

クラスター 1 の事例見本を図 8 に示す．

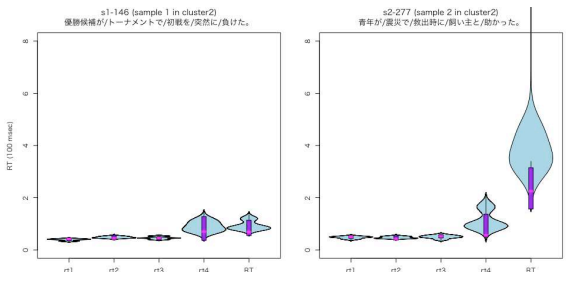


(a) sample 1 (b) sample 2

図 8: Samples of PCA cluster 1

3.7.2 クラスター 2 の事例見本

クラスター 2 の事例見本を図 9 に示す．



(a) sample 1 (b) sample 2

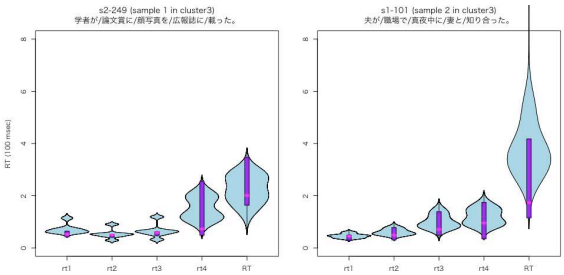
図 9: Samples of PCA cluster 2

3.7.3 クラスター 3 の事例見本

クラスター 3 の事例見本を図 10 に示す．

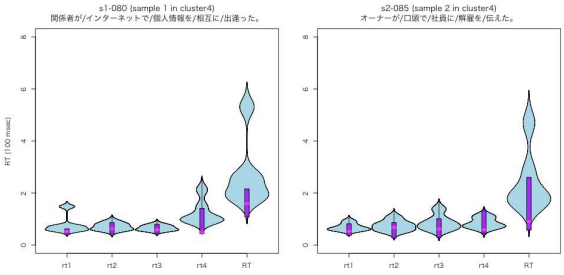
3.7.4 クラスター 4 の事例見本

クラスター 4 の事例見本を図 11 に示す．



(a) sample 1 (b) sample 2

図 10: Samples of PCA cluster 3



(a) sample 1 (b) sample 2

図 11: Samples of PCA cluster 4

4 反応時間と容認度評定値の関係を探索する

4.1 変数の回帰分析

今回の実験では，容認度の評定値である resp が調査 1，調査 2 で取得した評定値の粗い近似になっている．この関係を考え，変数 resp, rt1, …, rt4, RT と区分ごとの文字数 seg1.size, …, seg6.size の間の回帰分析を行った結果を以下に示す (seg*n*.size は区間の表記文字数で，音韻的単位 (例えばモーラ) 数の粗い近似として使った)．回帰分析に用いたモデルは Generalized Linear Model [6] である (link 関数は Gaussian)．表示で用いる有意差のコードは通常通り: $p < 0.001$: “***”; $p < 0.01$: “**”; $p < 0.05$: “*”; $p < 0.1$: “.” である．

4.1.1 resp の回帰

表 5: resp の他の変数による回帰

	Estim.	Std. Err	t-value	Pr(>|t|)	Sign.
(切片)	0.417	0.083	5.02	7.6e-07	***
rt1	-0.039	0.096	-0.41	0.68	
rt2	0.013	0.078	0.16	0.87	
rt3	0.039	0.070	0.56	0.58	
rt4	0.042	0.028	1.52	0.13	
RT	0.004	0.010	0.38	0.70	
seg1.size	0.016	0.007	2.18	0.03	*
seg2.size	0.004	0.008	0.54	0.59	
seg3.size	0.004	0.008	0.50	0.62	
seg4.size	-0.005	0.009	-0.58	0.56	
seg5.size	0.001	0.013	0.12	0.91	

表 5 に resp の値を他の変数で回帰した結果を示す．

以下の解析結果に共通する事だが，回帰で大きな負の値は促進度が大きい=反応時間が早まる事を，大きな正の値は阻害度が大きい=反応時間が遅くなる事を意味する．

resp は解像度の粗い容認度評定である．resp 値を強く

予測する変数は切片で，弱く予測する変数が seg1.size である．これは，容認度判断の結果は実質的に反応時間からは予測できないという事を意味しており，意外と言えば意外な結果である．切片の予測力が高い事は，回帰で使われている変数で表現されていない因子の影響が有意だという事である．それがどんな因子なのかは本研究では明らかにされない．

容認度の評定に時間がかかる場合とかからない場合があるが，所要時間と判定結果の間に対応はない．

4.1.2 rt1 の回帰

表 6: rt1 の他の変数による回帰

	Estim.	Std. Err.	t-value	Pr(>|t|)	Sign.
(切片)	0.285	0.039	7.22	2.2e-12	***
resp	-0.009	0.023	-0.41	0.6838	
rt2	0.307	0.035	8.69	<2e-16	***
rt3	0.108	0.034	3.19	0.0015	**
rt4	-0.011	0.014	-0.81	0.4203	
RT	0.001	0.005	0.18	0.8592	
seg1.size	0.007	0.004	2.02	0.0439	*
seg2.size	-0.000	0.004	-0.04	0.9690	
seg3.size	-0.002	0.004	-0.38	0.7048	
seg4.size	-0.006	0.004	-1.34	0.1816	
seg5.size	-0.005	0.006	-0.80	0.4259	

表 6 に rt1 の値を他の変数で回帰した結果を示す．

rt1 の値を予測するのは，切片, rt2, rt3, seg1.size である．切片の予測力が高いのは resp と同じである．rt2, rt3 の順に影響が大きいのは，rt1 増大の影響が逓減的に事後に及ぶ (が rt4 には及ばない) 事を意味している．seg1.size の影響があるのは，自然な結果であるが，rt1 に特化した特徴で一般性はない．

4.1.3 rt2 の回帰

表 7: rt2 の他の変数による回帰

	Estim.	Std. Err.	t-value	Pr(>|t|)	Sign.
(切片)	4.09e-02	5.11e-02	0.80	0.4242	
resp	4.61e-03	2.80e-02	0.16	0.8696	
rt1	4.62e-01	5.32e-02	8.69	< 2e-16	***
rt3	3.31e-01	3.90e-02	8.47	3.4e-16	***
rt4	6.40e-02	1.63e-02	3.92	0.0001	***
RT	-7.43e-03	5.93e-03	-1.25	0.2107	
seg1.size	6.29e-03	4.31e-03	1.46	0.1446	
seg2.size	7.22e-03	4.82e-03	1.50	0.1348	
seg3.size	-1.24e-02	4.95e-03	-2.51	0.0124	*
seg4.size	8.95e-05	5.29e-03	0.02	0.9865	
seg5.size	7.94e-03	7.48e-03	1.06	0.2894	

表 7 に rt2 の値を他の変数で回帰した結果を示す．

rt2 に影響するのは，rt1, rt3, rt4, seg3.size である．このうち，rt1, rt3, rt4 の影響は大きい．これは rt2 の値が前後反応と強い相互作用をしている事を示唆している．seg3.size が rt2 を説明しているのは自然な結果であるが特筆すべき結果ではない．逆に，seg2.size が rt2 を説明していない事の方が特筆すべき結果である．

特筆すべき特徴として，他の変数を予測した場合と違って，rt2 への影響は値が非常に小さい．また resp, rt1 の場合と異なり，切片の値は rt2 の値を予測しない．

4.1.4 rt3 の回帰

表 8: rt3 の他の変数による回帰

	Estim.	Std. Err.	t-value	Pr(>|t|)	Sign.
(切片)	0.147	0.057	2.59	0.010	**
resp	0.017	0.031	0.56	0.577	
rt1	0.202	0.063	3.19	0.002	**
rt2	0.410	0.048	8.47	3.4e-16	***
rt4	0.072	0.018	3.98	8.0e-05	***
RT	-0.006	0.007	-0.91	0.365	
seg1.size	-0.006	0.005	-1.27	0.206	
seg2.size	0.002	0.005	0.45	0.651	
seg3.size	0.026	0.005	4.84	1.7e-06	***
seg4.size	-0.005	0.006	-0.78	0.434	
seg5.size	-0.005	0.008	-0.63	0.526	

表 8 に rt3 の値を他の変数で回帰した結果を示す．

rt3 に影響する変数は，rt2, rt4, seg3.size, 切片, rt1 である．

rt2 の場合と異なり，resp, rt1 の場合と同じで，切片の値は rt3 の値をそれなりに予測する．影響の強さの順序は rt3 の値が前後と強い相互作用をしている事を示唆している．seg3.size が rt3 を説明しているのは自然な結果であるが，特筆すべき結果ではない．

4.1.5 rt4 の回帰

表 9: rt4 の他の変数による回帰

	Estim.	Std. Err.	t-value	Pr(>|t|)	Sign.
(切片)	0.199	0.144	1.38	0.168	
resp	0.120	0.079	1.52	0.128	
rt1	-0.130	0.162	-0.81	0.420	
rt2	0.509	0.130	3.92	0.000	***
rt3	0.463	0.117	3.98	8.0e-05	***
RT	-0.016	0.017	-0.94	0.347	
seg1.size	-0.021	0.012	-1.69	0.091	.
seg2.size	-0.005	0.014	-0.37	0.709	
seg3.size	-0.012	0.014	-0.86	0.389	
seg4.size	0.068	0.015	4.63	4.8e-06	***
seg5.size	-0.002	0.021	-0.08	0.936	

表 9 に rt4 の値を他の変数で回帰した結果を示す．

rt4 への影響が有意な変数は，rt2, rt3, seg4.size, seg1.size である．resp, rt1, rt3 の場合と異なり，rt2 の場合と同じで，切片の値は rt4 の値を予測しない．rt2, rt3 は rt4 に強い影響を持っているのがわかるが，rt1 からの影響は有意ではない．これは rt4 の値が短い範囲で前と相互作用をしている事を示唆している．

seg4.size が rt4 を説明しているのは自然な結果であるが，seg3.size の rt2, rt3 への影響より大きい．この違いが何に起因するのかは，検討の余地がある．

4.1.6 RT の回帰

RT は最後の区間の提示が終わってから容認度評定が行なわれるまでの時間である．表 10 に RT の値を他の変数で回帰した結果を示す．

RT への影響が有意な変数は実質的に切片のみである．これは，容認度評定にかかる時間は局所的な処理時間

表 10: RT の他の変数による回帰

	Estim.	Std. Err.	t-value	Pr(>|t|)	Sign.
(切片)	2.041	0.391	5.22	2.7e-07	***
resp	0.084	0.221	0.38	0.704	
rt1	0.080	0.452	0.18	0.859	
rt2	-0.461	0.368	-1.25	0.211	
rt3	-0.230	0.330	-0.91	0.365	
rt4	-0.123	0.130	-0.94	0.347	
seg1.size	-0.039	0.034	-1.15	0.249	
seg2.size	-0.004	0.038	-0.16	0.876	
seg3.size	0.072	0.039	1.84	0.066	.
seg4.size	0.049	0.042	1.18	0.240	
seg5.size	0.058	0.059	0.98	0.327	

rt1–rt4 からは予測できない事を意味している．

seg3.size から弱い影響があるのは興味深い．これは §4.3.3 で触れる rt2 の特殊性を関係していると思われる．

4.1.7 単純回帰の結果のまとめ

rt1 の値を (切片の他には) rt2, rt3 が予測する．rt2 の値を rt1, rt3, rt4 が予測するが，切片の予測はない．rt3 の値を (切片の他には) rt2, rt4, rt1 が予測する．rt4 の値を rt2, rt3 が予測するが，切片の予測はない．rt1, ..., rt4 の受ける影響は基本的に隣接性の効果だと解釈できる．

これに対し，resp 値と RT 値を予測する変数は実質的に切片の他にない．これは，rt1–rt4, seg1–seg6.sizes の resp と RT への影響は独立性が高いという事である．別の言い方をすると，rt1–rt4 や seg1.size–seg6.size から resp や RT の値は単純に決まらない．

4.2 容認度評定値と反応時間データの相互回帰

先の回帰解析は今回の実験の中で閉じたものだった．ARDJ は調査 1 と調査 2 で容認度評定値を高い分解能で刺激文に対応づけている．先行調査 1, 2 で独立に取得した評定値と反応時間がどう関係しているかを調べられる．

以下の変数を拡大した回帰解析には，edit.type, r01, r12, r23, r3x を独立変数に追加した．

edit.type は刺激文の生成に使われた変異のタイプで，次の値を持つ: o(riginal): 変異なし; v(erbal): 動詞を一定の範囲の文脈類似度を持つ別の動詞に置換; p(ositional): 格助詞を粗っぽく出現頻度を反映するようにランダムに置換; n(ominal): 名詞 (形容動詞語幹を含む) を一定の範囲の文脈類似度を持つ別の名詞に置換; s(wapped): 分節の対のランダムな入れ替え．

r*ij* は逸脱度の評定値 0, 1, 2, 3 を区間 [0,1), [1,2), [2,3), [3,∞) の代表値と解釈した上で，それぞれの区間が評定者に選択される確率である (そのため，すべての刺激文で r01+r12+r23+r3x ≃ 1.0 となる)．

4.3 変数選択に関する注意

以下の解析には先の解析で独立変数に使った segN.size を含めていないないが，それは主に紙面の都合に拠る．

以下のすべての結果で r3x の影響は singularity 故に除外された．これは，r3x が他の変数 (おそらく r01, r12, r23) から予測できるためで，意外ではない．

以下のすべての結果で edit:n の影響が表われていない．これが R の glm 関数のバグなのか，他の要因 (例えば singularity) に拠るかは現時点で明らかではない．

4.3.1 resp の拡大回帰

表 11: resp の拡大変数組みによる回帰

	Estim.	Std. Err.	t-value	Pr(>|t|)	Sign.
(切片)	0.724	0.082	8.86	< 2e-16	***
rt1	-0.061	0.088	-0.70	0.484	
rt2	0.119	0.071	1.68	0.094	.
rt3	-0.096	0.062	-1.54	0.125	
rt4	0.026	0.024	1.06	0.290	
RT	0.002	0.010	0.16	0.871	
edit:o	0.006	0.031	0.19	0.848	
edit:p	0.077	0.028	2.76	0.006	**
edit:s	0.009	0.026	0.35	0.724	
edit:v	0.056	0.028	2.03	0.043	*
r01	-0.455	0.070	-6.53	1.7e-10	***
r12	-0.247	0.076	-3.27	0.001	**
r23	-0.063	0.146	-0.43	0.667	

表 11 に resp の値を他の変数で回帰した結果を示す．容認度評定値 resp を，切片, r01, r12, edit:p, edit:v, rt2 がこれらの順に強く予想する．

resp の場合に限らず，切片が影響する変数が多い．これらの変数には回帰で明示化されていない変数の影響が大きいという事である．

他の変数の影響について言うと，r01, r12 が負の方向に大きい (= 逸脱度が小さい) 事が容認度を高め，格助詞の変異と先行している名詞句と動詞の意味的整合性の不足が容認度を低めると解釈でき，自然な結果である．

rt2, rt4 が阻害的に働いていて，rt2 のみが有意な効果をもたらしている．この結果は日本語の文の統語構造を考えると，示唆的である．rt2 は 3 つ目の分節を見る段階 =3 つ目の格要素を見る段階であり，格要素動詞の不整合が最初に露呈する位置である．これは動詞を見る前に容認度の概要が決まる場合があるという事を意味する．

4.3.2 rt1 の拡大回帰

表 12 に rt1 の値を他の変数で回帰した結果を示す．rt1 値の予測する変数は，切片, rt2, rt3, edit:p であり，切片, rt2 の予測力が強く，rt3 と edit:p の予測力はあると言えばある程度である．rt2 の影響は隣接性を考えれば自然

表 12: rt1 の拡大変数組みによる回帰

	Estim.	Std. Err.	t-value	Pr(>|t|)	Sign.
(切片)	0.312	0.046	6.84	2.5e-11	***
resp	-0.018	0.025	-0.70	0.484	
rt2	0.315	0.035	8.89	< 2e-16	***
rt3	0.056	0.034	1.65	0.099	.
rt4	0.002	0.013	0.12	0.908	
RT	-0.000	0.006	-0.08	0.940	
edit:o	0.018	0.016	1.11	0.267	
edit:p	0.025	0.015	1.65	0.099	.
edit:s	0.016	0.014	1.14	0.257	
edit:v	0.018	0.015	1.19	0.237	
r01	-0.027	0.039	-0.68	0.499	
r12	-0.020	0.041	-0.50	0.620	
r23	-0.098	0.078	-1.24	0.215	

な事である．edit:p の影響は多かれ少なかれ他のすべての変数の回帰で認められるので，rt1 の回帰に特化したものではない．

4.3.3 rt2 の拡大回帰

表 13: rt2 の拡大変数組みによる回帰

	Estim.	Std. Err.	t-value	Pr(>|t|)	Sign.
(切片)	0.0143	0.0585	0.24	0.8074	
resp	0.0518	0.0309	1.68	0.0937	.
rt1	0.4706	0.0529	8.89	< 2e-16	***
rt3	0.3274	0.0382	8.57	< 2e-16	***
rt4	0.0548	0.0157	3.49	0.0005	***
RT	-0.0075	0.0068	-1.10	0.2711	
edit:o	-0.0379	0.0201	-1.89	0.0596	.
edit:p	-0.0463	0.0184	-2.52	0.0120	*
edit:s	-0.0312	0.0172	-1.81	0.0704	.
edit:v	-0.0295	0.0184	-1.60	0.1093	
r01	0.1117	0.0478	2.34	0.0199	*
r12	-0.0211	0.0504	-0.42	0.6756	
r23	0.1737	0.0957	1.81	0.0702	.

表 13 に rt2 の値を他の変数で回帰した結果を示す．rt2 値に影響するのは，rt1, rt3, rt4, edit:p, r01, rt1, edit:o, edit:s, r23 である．強い影響をもつのは，rt1, rt3, rt4 である．それより少し弱い影響をもつのは，edit:p, r01 である．更に弱い影響をもつのは，resp, edit:o, edit:s, r23 である．切片の影響は認められない．

rt2 に影響する変数は多い．今回検討した組み合わせの中では，もっとも多い．これは一見すると意外であるが，良く考えると妥当な結果である．先にも触れたが，rt2 は日本語の文の統語構造を考えると特殊な環境であり，それが結果に表れていると思われる．rt2 は 3 つ目の分節を見る段階 = 3 つ目の格要素を見る段階であり，格要素動詞の不整合が最初に露呈する位置であると考えられる．これが後に現れる不整合と連動しているため，このような結果が得られていると考えるのが適当だろう．

4.3.4 rt3 の拡大回帰

表 14 に rt3 の値を他の変数で回帰した結果を示す．rt3 の値に強く影響するのは，切片, rt2, rt4, r01 であり，切片, rt2 の影響が顕著である．rt2 からの影響は隣接性からすれば当然で，特筆すべきものではない．rt3 と r01 の間に連動があるという事は，刺激文の容認度は rt3 の段階，つまり rt4 か rt5 で動詞を見る前におおよそ決まっているという事である．この効果は rt2 でも弱いながら，すでに検出されている．

表 14: rt3 の拡大変数組みによる回帰

	Estim.	Std. Err.	t-value	Pr(>|t|)	Sign.
(切片)	0.3661	0.0644	5.69	2.3e-08	***
resp	-0.0540	0.0352	-1.54	0.1254	
rt1	0.1077	0.0651	1.65	0.0988	.
rt2	0.4249	0.0496	8.57	< 2e-16	***
rt4	0.0554	0.0180	3.08	0.0022	**
RT	-0.0042	0.0077	-0.54	0.5913	
edit:o	-0.0067	0.0230	-0.29	0.7714	
edit:p	0.0382	0.0210	1.82	0.0696	.
edit:s	-0.0014	0.0197	-0.07	0.9421	
edit:v	-0.0136	0.0210	-0.65	0.5160	
r01	-0.1594	0.0543	-2.94	0.0035	**
r12	-0.0404	0.0574	-0.70	0.4817	
r23	-0.1767	0.1091	-1.62	0.1061	

4.3.5 rt4 の拡大回帰

表 15: rt4 の拡大変数組みによる回帰

	Estim.	Std. Err.	t-value	Pr(>|t|)	Sign.
(切片)	0.4136	0.1708	2.42	0.0159	*
resp	0.0963	0.0909	1.06	0.2901	
rt1	0.0194	0.1686	0.12	0.9082	
rt2	0.4743	0.1361	3.49	0.0005	***
rt3	0.3694	0.1198	3.08	0.0022	**
RT	-0.0100	0.0199	-0.50	0.6177	
edit:o	-0.0637	0.0592	-1.08	0.2825	
edit:p	-0.1484	0.0539	-2.75	0.0061	**
edit:s	0.0559	0.0508	1.10	0.2709	
edit:v	-0.1175	0.0539	-2.18	0.0298	*
r01	-0.1645	0.1412	-1.16	0.2447	
r12	-0.1106	0.1482	-0.75	0.4560	
r23	-0.0668	0.2825	-0.24	0.8132	

表 15 に rt4 の値を他の変数で回帰した結果を示す．rt4 値を予想するのは，rt2, rt3, edit:p, 切片, edit:v で，強く予想するのは rt2, rt3, edit:p である．

rt4 では動詞が見えている．rt4 に RT, resp からの影響が認められないのは，意外と言えばである．

4.3.6 RT の拡大回帰

表 16: RT の拡大変数組みによる回帰

	Estim.	Std. Err.	t-value	Pr(>|t|)	Sign.
(切片)	2.1896	0.3908	5.60	3.7e-08	***
resp	0.0348	0.2139	0.16	0.87	
rt1	-0.0297	0.3962	-0.08	0.94	
rt2	-0.3566	0.3236	-1.10	0.27	
rt3	-0.1528	0.2844	-0.54	0.59	
rt4	-0.0550	0.1102	-0.50	0.62	
edit:o	-0.1257	0.1392	-0.90	0.37	
edit:p	-0.3279	0.1268	-2.59	0.01	*
edit:s	0.0466	0.1194	0.39	0.70	
edit:v	0.1726	0.1271	1.36	0.18	
r01	-0.0907	0.3323	-0.27	0.79	
r12	0.2042	0.3484	0.59	0.56	
r23	0.7647	0.6631	1.15	0.25	

表 16 に RT の値を他の変数で回帰した結果を示す．RT

値を強く予測する変数は切片のみであり，edit:p が弱く，rt2 が更に弱く予想する事が示唆されている．他の変数がほとんど影響しないのは，意外と言えば意外である．

4.3.7 拡大回帰の結果のまとめ

以上の結果からは反応時間が容認度評定値から容易に予測可能な量だという事は示されていない．とは言え，二量間に相当に複雑な相互作用がある事は予想できる．その明示化に決定木分析 (e.g., Classification and Regression Trees) [1] が有効で，すでに一定の結果を得ているが，本稿には紙面の都合で含めない事にする．

5 終わりに

日本語の容認度評定データ (ARDJ) は大規模かつ確証バイアスの少なく，無償利用可能な日本語の文の容認度評定値のデータベースである．それは将来的に**証拠に基づく言語学** (Evidence-based Linguistics) [10] を実現するために必須の参照データだと筆者らは考える．本研究は，容認度評定の対象になった 466 種類の文を刺激に使って反応時間を取得し，容認度評定値と対応させた．このデータ拡張は ARDJ の利用価値を高めると期待される．

本研究は探索型の研究であり，得られた結果から結論と言えるものを引き出すのは難しい．ただ，強いて結論らしきものを挙げるとするならば，読み時間と容認度評定の対応関係はおそらく，これまで人が想像して来たよりずっと複雑かも知れないと言う (言語学者にとっては嬉しくない) 可能性の提示だろう．得られた結果の多変量解析からは少なくとも，容認度評定と反応時間には単純な相関がない事が示唆されている．

とは言え，今回取得した反応データは提供者の属性が片寄っており，代表性が不足しているので，得られた結論が暫定的である事は付記しておく必要がある．特に実験規模の拡大と無作為化は，結果に信頼性を求めるならば不可欠である．

容認度評定値を読み時間データと結びつけたのは，何通りもある行動データとの結びつけの一つに過ぎない．同じような趣旨の拡張として，眼球運動データとの結びつけが考えられる．機会を見つけて実現したい．

Acknowledgments

すべての解析は RStudio (1.1.x) 上の R (version 3.5.3) で実行した．

参考文献

[1] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone. *Classification and Regression Trees*. Wadsworth and Brooks, Monterey, CA, 1984.

[2] Kow Kuroda. Reassessing PCA of Acceptability Rating Data for Japanese (ARDJ) using kernel MVA. In *Proceedings of the 26th Annual Meeting of Association of NLP*, pp. 1523–1526, 2020.

[3] Kow Kuroda, Hikaru Yokono, Keiga Abe, Tomoyuki Tsuchiya, Yoshihiko Asao, Yuichiro Kobayashi, Toshiyuki Kanamaru, and Takumi Tagawa. Development of Acceptability Rating Data of Japanese (ARDJ): An initial report. In *Proceedings of the 24th Annual Meeting of the Association for NLP*, pp. 65–68, 2018.

[4] Kow Kuroda, Hikaru Yokono, Keiga Abe, Tomoyuki Tsuchiya, Yoshihiko Asao, Yuichiro Kobayashi, Toshiyuki Kanamaru, and Takumi Tagawa. Insights from a large scale web survey for Acceptability Rating Data for Japanese (ARDJ) project. In *Proceedings of the 25th Annual Meeting for the Association of NLP*, pp. 253–256, 2019.

[5] Kow Kuroda, Hikaru Yokono, Keiga Abe, Tomoyuki Tsuchiya, Yoshihiko Asao, Yuichiro Kobayashi, Toshiyuki Kanamaru, and Takumi Tagawa. Rudimentary modeling of acceptability judgement from a large scale, unbiased data. In *Proceedings of the 41st Annual Meeting for Cognitive Science Society*, 2019.

[6] P. McCullagh and J. A. Nelder. *Generalized Linear Models, Second Edition*. Chapman and Hall/CRC Monographs on Statistics and Applied Probability Series. Chapman & Hall, 1989.

[7] Dau Pelleg and Andrew Moore. *X*-means: Extending *k*-means with efficient estimation of the number of clusters. In *Proceedings of the 17th International Conference on Machine Learning*, pp. 727–734. Morgan Kauffmann, 2000.

[8] Joshua B. Tenenbaum, V. de Silva, and J. C. Langford. A global framework for nonlinear dimensionality reduction. *Science*, 290:2319–2323, 2000.

[9] Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-SNE. *Journal of Machine Learning Research*, 9:2579–2605, 2008.

[10] 黒田 航. 証拠に基づく医療 (EBM) との比較を通じて理論言語学の方法論を見直す. In **第 *16* 回日本認知言語学会発表論文集**, pp. 580–585, 2016.

[11] 黒田 航, 阿部 慶賀, 横野 光, 田川 拓海, 小林 雄一郎, 金丸 敏幸, 土屋 智行, and 浅尾 仁彦. (言語学者による) 容認度評定の認証システムを試作する構想: 入念に設計された日本語文の容認度評定データベースに基づいて. In **日本認知科学会第 *33* 回大会発表論文集**, pp. 557–562. 日本認知科学会, 2016.

[12] 黒田 航, 阿部 慶賀, 横野 光, 土屋 智行, 小林 雄一郎, 金丸 敏幸, 浅尾 仁彦, and 田川 拓海. 容認度評定に影響する要因の定量的評価: 日本語容認度評定データ (ARDJ) から得られた知見. In **日本認知科学会第 *36* 回大会発表論文集**, pp. 727–736, 2019.