

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

黒田 航¹, 阿部 慶賀², 横野 光³, 土屋 智行⁴, 小林 雄一郎⁵, 金丸 敏幸⁶, 浅尾 仁彦⁷, 田川 拓海⁸

Kow Kuroda, Keiga Abe, Hikaru Yokono, Tomoyuki Tsuchiya, Yuichiro Kobayashi, Toshiyuki Kanamaru, Yoshihiko Asao, Takumi Tagawa

杏林大学¹, 岐阜聖徳学園大学², 富士通研究所³, 九州大学⁴, 日本大学⁵, 京都大学⁶, NICT⁷, 筑波大学⁸
Kyorin University, Gifu Shotoku University, Fujitsu Laboratory, Kyushu University, Nippon University, Kyoto University, NICT, Tsukuba University
contact: kow.k@ks.kyorin-u.ac.jp

Abstract

Acceptability Rating Data for Japanese (ARDJ) project ran Survey 2 in 2018. 300 tokens (under 282 types) of sentence were used for stimulus. Raters were asked to give one of the following 4 categories to the stimuli: **0.** natural and easy to understand; **1.** more or less deviant but comprehensible; **2.** deviant and difficult to understand; **3.** quite unnatural and incomprehensible. Responses by 1,880 raters (sum of phrases 1 and 2) were collected. Rater were also asked to answer the following 11 questions for attributes: **Q1:** Age, **Q2:** Gender, **Q3:** Nativity, **Q4:** Lived places, **Q5:** Experience of life abroad over a year, **Q6:** Total number of languages learned, **Q7:** Total years of learning foreign languages, **Q8:** Daily contact with foreigners, **Q9:** Number of books read in a month, **Q10:** Years of education after elementary school, **Q11:** Choice among i) strongly scientifically oriented, ii) more or less scientifically oriented, iii) neutral, iv) more or less literary oriented, v) strongly literary oriented. Layered analysis of them was conducted to get quantitative assessment of their influence on acceptability rating. For each attribute, stimuli were marked significantly different from others when their Fisher exacts returns p -values less than 0.05. This criterion gave the following magnitude hierarchy of influence: 52.2 (23.5%) cases for Q1, 23 (10.7%) cases for Q4, 22 (9.9%) cases Q7, 20 (9.0%) cases for Q2, 19.9 (9.0%) cases for Q9, 17 (7.7%) cases for Q3, 16.3 (7.4%) cases for Q5, 15 (6.8%) cases for Q10, 13 (5.9%) cases for Q8, 12.2 (5.9%) cases for G6, and 11.2 (5.1%) cases for Q11.

キーワード: **acceptability rating, acceptability judgement, layered analysis, magnitude hierarchy**

1. はじめに: ARDJ とは何か?

日本語容認度評定データ (Acceptability Rating Data of Japanese (ARDJ) [5, 6, 14]) は, 理論的バイアス (特に確証バイアス [1, 2, 4]) を避けて構築した, なるべく数多くの刺激文に対して, なるべく多様な評定者から,

高い解像度の評定値を収集する企画である¹⁾.

当データを構築した主な目的は2つある: 第一に理論言語学の研究の中核をなす容認度判断の概念の実質化のため, 第二に理論言語学の確証バイアスに強く影響された論証法を是正する方法の基礎データを提供するためである.

1.1 構築済みのデータ

ARDJ は 2017 年に 200 種類の刺激文を使って予備実験 (survey 1) を, 2018 年に 282 種類の刺激文²⁾ を使って本実験 (survey 2) を, 相 1 と相 2 に分けて実施した. 相 1 は大学生や大学院生を対象にした小規模のデータ収集 ($n = 201$) で, 相 2 は大規模な web 基盤のデータ収集 ($n = 1,679$) である³⁾.

2つの相は 282 種類の刺激文を共有しているが, 後述のように実施条件が同一ではない (相 1 は 10 種類, 相 2 は 11 種類の属性を取得). 整合性を保持する形で相 1, 2 のデータを統合し, GitHub サイトで survey 2 unified (s2u) として公開した⁴⁾.

相 2 の概要は [6] で報告した. 本研究はそれとは別に, s2u に評定値と合わせて収集した 11 属性を使った層別分析を行ない, その結果を報告する.

本論文は層別解析で有意差が認められた実例の提示を, 結果の分析より優先した. 具体事例を共有する事の方が, 興味深い事例を割愛しつつ中途半端な説明を試みるより重要だと考えたからである. 分析から得られている結果の考察は, 稿を改めて行いたい.

¹⁾本研究は JSPS の挑戦的萌芽研究 (課題番号: 16K13223) の支援を受けて実施された.

²⁾surveys 1, 2 の反応の等化を見込んで一部は survey 1 から流用.

³⁾事後的に判った事だが, 2 相で評定者の年齢は重なっていない. phase 2 の web 調査には 20 才代の参加者がほとんどいなかった.

⁴⁾<https://kow-k.github.io/database-of-japanese-acceptability> データは無償で公開しているが, 利用には管理者が発行するパスワードが必要. 入手したい方は第一著者に問い合わせを.

1.2 ARDJ の方法論に関する注意

1.2.1 なぜ容認度評価データベースなのか？

言語学(特に理論言語学と呼ばれる領域)の研究課題の一つは、普通に使われうる文と使われない文の境界を確定する事である。前者を容認度可能(acceptable)な文と呼び、容認不可能(unacceptable)(か困難)な文と呼ぶ。厄介なのは2つの種類の表現の区別が文法理論から中立に行なえない事である。

言語学者の判断の非中立性は次の2つの形で具現化する。第一に、大きな(メタ)理論的利害がかかった事例ばかりが研究で取り上げられる。第二に、個々の研究者が、自説に都合の良い判断をする。このような背景があつて、言語研究の専門家である言語学者の判断は、しばしば分野の外の非専門家の判断から乖離する。こうなると言語学者が研究しているのは、研究者集団内でのみ共有されている虚構でしかないという疑いを晴らせない。

これを回避するのは、理論的立場に影響されない文を十分な数用意し、それらに非専門家が与える評価値を基本反応として収集しておき、理論的利害が関わる事例の容認度評価の際に参照する事である。ARDJはそのような参照データを構築する目的で実施された。

1.2.2 容認度判定の実態調査: 理論的前提の見直し

何人かの人に同じ文の容認度を判定/評価させても、皆が同じ結果を返す訳ではない。これは小規模な実験でもすぐに確かめられる。だが、理論言語学は単純化のために(1)を想定している:

- (1) a. 外的条件が同じなら、得られる反応が常に一様で、
- b. その反応を有能な言語学者が模倣/エミュレートできる

これら2つの想定の見直しは十分に確かめられていない。(1b)に関しては、Experimental Syntaxの名プログラムで議論される機運が高まっている[11, 12]⁵⁾。だが、(1a)の内実を確かめた研究は世界的に存在しない。

(1a)が成立しないなら、どんな反応が実際に得られるのかを確かめる必要がある。それがARDJの第一の目的である⁶⁾。

⁵⁾Experimental Syntaxの先駆的な研究は[3, 9, 10]であり、範囲を統語論に限定しなければ、[8]に含まれた研究の一部も同じような研究基盤の見直しのカテゴリーに入るだろう。

⁶⁾[7]は、ARDJの調査結果に基づいて(1a)を評価した研究であるが、本論文では紙面の都合で内容に触れない。

1.3 評価課題と提示文の作成法

容認度評価は、それぞれの参加者に30文が提示され、それぞれの文に次の4つ評価のいずれかと与える課題だった。調査は調査会社の協力を得てwebで実行された。詳細は[6]を参照されたい。

- (2) 0. 違和感がなく自然に理解できる文
 1. 違和感を感じるが理解可能な文
 2. 違和感を感じて理解困難な文
 3. 不自然で理解不能な文

この課題に使われた282種類の刺激文がどのように作成されたかを簡単に説明する。これは§2.-§3.で述べる結果の解釈に関係する。

§1.2.1で述べた理由で、本調査は確認型ではなく探索型である。確認型でないという事は、理論言語学が特定の文に与える容認度の予想を意図的に排除しているという事である。ARDJの出発点は、理論言語学が流派を問わず、言語表現の変異を十分に広い範囲の見ておらず、確認バイアス[1, 2, 4]に陥っているという反省である。確認バイアスは仮説検証型の研究をする場合には避けがたい落とし穴であり、それを避けるには探索型の研究をするしか方法がない。だが、どうやってそれを実現するか？

ARDJの相1, 2で使用された刺激文は、遺伝子の突然変異を模して作成された⁷⁾。具体的に言うと、健全な数の健全な文を人が用意し、それぞれの一部にランダムに変異を発生させて逸脱文の候補を生成し、それからサンプリングとフィルタリングによって刺激文集合を得ている(技術的詳細は[5]に委ねる)。こうして得られた刺激文は結果的に、文法理論の観点から見て興味の対象にならない逸脱文がそれなりに多く含まれる。この手法に難があるとすれば、それがrecall重視precision軽視である事だろう。実際、言語学におもしろい現象を見つけようと思えば、この手法は非常に効率が悪い。だが、これは意図された効果であり、容認度の可能性空間の全体探索が目的ならば、短所ではなく長所である。

2. s2u データの層別分析

2.1 前処理

外れ値を除去としてs2uの1,880の反応のうち、Step 1) $0.6 < (\text{評価値のsd}) < 1.5$ であり、かつStep 2) Mahalanobis距離が0.95の範囲内にある反応を選別した。Step 1の結果は1,736、Step 2の結果は1,649である。これを以下の解析で使った。

⁷⁾2016年の認知科学会33回大会で日高昇平氏(JAIST)から、容認度の可能性空間の探索するのにランダムウォークを使うと良いと示唆を受けた事が本実装の動機である。

2.2 本実験で収集した 10 (+1) 種類の属性

(3) Survey 2 で収集した 10 (+1) 属性のための質問:

- Q1. 年齢 (age);
- Q2. 性別 (gender);
- Q3. 日本語が母語か (nativity);
- Q4. 県単位での出身地 (place);
- Q5. 非日本語圏での 1 年以上の生活の有無 (abroad);
- Q6. これまでに学んだ異国語の概数 (nLangLearned);
- Q7. 異国語を学んだ年の概数 (yLearnLang);
- Q8. 日常的に異国人と接するかどうか (foreignContact);
- Q9. 一月当りの読書量 (冊数) (nBooks);
- Q10. 小学校より後に受けた教育年数 (yEdu);
- Q11. 自分は (強く) 文系か (強く) 理系か? (orientation)

Q11 は相 2 のみで収集し, Q4 の回答は相 1 で択一選択, 相 2 で複数選択だった. これらが 300 種類 (タイプ数 282) の文のそれぞれの容認度評定に影響する程度を調べた結果を以下に示す.

2.3 層別分析の結果の概要

Q4 と他の属性は有意性の評価法が異なる. Q1-Q3, Q5-Q11 では $s001, \dots$ の評定値を属性の値で再分類し, その分布の Fisher exact test で得た p 値が特定の値 (e.g., 0.01, 0.05, 0.10) を下回った場合に有意な影響があったと考える. Q4 は $s001, \dots$ の評定値の分布を地名の非排他的共起表で線型回帰し, 得られた p 値が特定の値を下回った場合に有意な影響があったと考える. どの p 値を基準として選ぶべきかは明らかでない (データスパースネスの問題があり, 単純に p が小さければ良いという訳ではない). そのため, 三つの条件で得られた結果を比較してどの値が妥当か判断する.

2.3.1 3 種類の p 値に対応する事例数とそれらの比較

図 1 のグラフは, Fisher exact test の $p = 0.01$ 未満の有意差条件で得られた Q1-Q11 の該当事例数のデータ (表 1) をグラフにしたものである. 表 1 は異なるデータ ($d1, d2, \dots$) と実行回 ($r1, r2, \dots$) から得られた該当事例数を示している. Q1. Age, Q7. yLearnLang, Q11. Orientation で無視できない該当数の変動が認められる. これは Fisher exact test の p 値をシミュレートする場合, $p = 0.01$ の判定条件が厳し過ぎる可能性を示唆する.

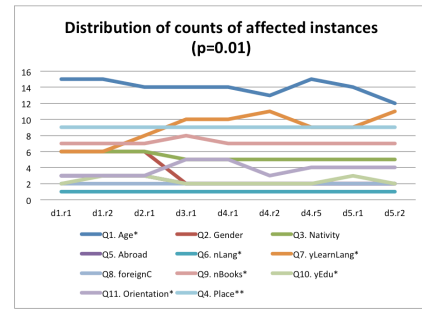


図 1 Q1-Q11 の有意性該当事例数の変動 ($p = 0.01$)

図 2 のグラフは, Fisher exact test の $p = 0.05$ 未満の有意差条件で得られた Q1-Q11 の該当事例数のデータ (表 2) をグラフにしたものである. 表 2 は異なるデータ ($d1, d2, \dots$) と実行回 ($r1, r2, \dots$) から得られた該当事例数を示している. Q1. Age についてのみ大きな該当数の変動が認められる.

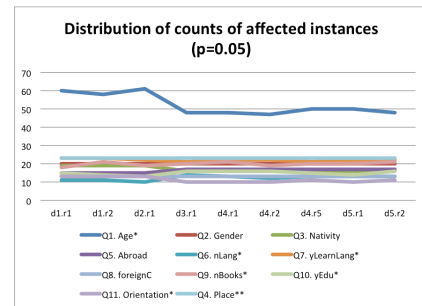


図 2 Q1-Q11 の有意性該当事例数の変動 ($p = 0.05$)

図 3 のグラフは, Fisher exact test の $p = 0.10$ 未満の有意差条件で得られた Q1-Q11 の該当事例数のデータ (表 3) をグラフにしたものである. 表 3 は異なるデータ ($d1, d2, \dots$) と実行回 ($r1, r2, \dots$) から得られた該当事例数を示している.

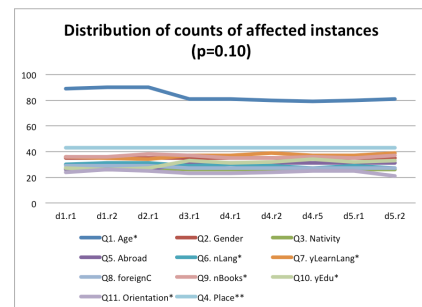


図 3 Q1-Q11 の有意性該当事例数の変動 ($p = 0.10$)

図 4 のグラフは, 三つの p 値ごとの平均事例数 (表 4) の推移をグラフにしたものである. この図からわかるように, 該当数は p 値の増加に対しほぼ線型で増加している.

表1 該当事例数の変動 ($p = 0.01$ の場合)

| Attribute | d1.r1 | d1.r2 | d2.r1 | d3.r1 | d4.r1 | d4.r2 | d4.r5 | d5.r1 | d5.r2 | av. | stdev | max | min |
|-------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-----|-------|-----|-----|
| Q1. Age* | 15 | 15 | 14 | 14 | 14 | 13 | 15 | 14 | 12 | 14 | 1 | 15 | 12 |
| Q2. Gender | 6 | 6 | 6 | 2 | 2 | 2 | 2 | 2 | 2 | 3.3 | 2 | 6 | 2 |
| Q3. Nativity | 6 | 6 | 6 | 5 | 5 | 5 | 5 | 5 | 5 | 5.3 | 0.5 | 6 | 5 |
| Q5. Abroad | 3 | 3 | 3 | 2 | 2 | 2 | 2 | 2 | 2 | 2.3 | 0.5 | 3 | 2 |
| Q6. nLang* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 |
| Q7. yLearnLang* | 6 | 6 | 8 | 10 | 10 | 11 | 9 | 9 | 11 | 8.9 | 1.9 | 11 | 6 |
| Q8. foreignC | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 0 | 2 | 2 |
| Q9. nBooks* | 7 | 7 | 7 | 8 | 7 | 7 | 7 | 7 | 7 | 7.1 | 0.3 | 8 | 7 |
| Q10. yEdu* | 2 | 3 | 3 | 2 | 2 | 2 | 2 | 3 | 2 | 2.3 | 0.5 | 3 | 2 |
| Q11. Orientation* | 3 | 3 | 3 | 5 | 5 | 3 | 4 | 4 | 4 | 3.8 | 0.8 | 5 | 3 |
| Q4. Place | 9 | 9 | 9 | 9 | 9 | 9 | 9 | 9 | 9 | 9 | 0 | 9 | 9 |

表2 該当事例数の変動 ($p = 0.05$ の場合)

| Attribute | d1.r1 | d1.r2 | d2.r1 | d3.r1 | d4.r1 | d4.r2 | d4.r5 | d5.r1 | d5.r2 | av. | stdev | max | min |
|-------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|------|-------|-----|-----|
| Q1. Age* | 60 | 58 | 61 | 48 | 48 | 47 | 50 | 50 | 48 | 52.2 | 5.7 | 61 | 47 |
| Q2. Gender | 20 | 20 | 20 | 20 | 20 | 20 | 20 | 20 | 20 | 20 | 0 | 20 | 20 |
| Q3. Nativity | 19 | 19 | 19 | 16 | 16 | 16 | 16 | 16 | 16 | 17 | 1.5 | 19 | 16 |
| Q5. Abroad | 15 | 15 | 15 | 17 | 17 | 17 | 17 | 17 | 17 | 16.3 | 1 | 17 | 15 |
| Q6. nLang* | 11 | 11 | 10 | 14 | 13 | 12 | 13 | 13 | 13 | 12.2 | 1.3 | 14 | 10 |
| Q7. yLearnLang* | 23 | 23 | 22 | 21 | 22 | 22 | 21 | 22 | 22 | 22 | 0.7 | 23 | 21 |
| Q8. foreignC | 13 | 13 | 13 | 13 | 13 | 13 | 13 | 13 | 13 | 13 | 0 | 13 | 13 |
| Q9. nBooks* | 18 | 21 | 19 | 20 | 21 | 19 | 20 | 20 | 21 | 19.9 | 1.05 | 21 | 18 |
| Q10. yEdu* | 15 | 14 | 13 | 16 | 16 | 16 | 15 | 14 | 16 | 15 | 1.2 | 16 | 13 |
| Q11. Orientation* | 13 | 13 | 13 | 10 | 10 | 10 | 11 | 10 | 11 | 11.2 | 1.4 | 13 | 10 |
| Q4. Place** | 23 | 23 | 23 | 23 | 23 | 23 | 23 | 23 | 23 | 23 | 0 | 23 | 23 |

表3 該当事例数の変動 ($p = 0.10$ の場合)

| Attribute | d1.r1 | d1.r2 | d2.r1 | d3.r1 | d4.r1 | d4.r2 | d4.r5 | d5.r1 | d5.r2 | av. | stdev | max | min |
|-------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|------|-------|-----|-----|
| Q1. Age* | 89 | 90 | 90 | 81 | 81 | 80 | 79 | 80 | 81 | 83.4 | 4.72 | 90 | 79 |
| Q2. Gender | 35 | 35 | 35 | 35 | 35 | 35 | 35 | 35 | 35 | 35 | 0 | 35 | 35 |
| Q3. Nativity | 27 | 27 | 27 | 26 | 26 | 26 | 26 | 26 | 26 | 26.3 | 0.5 | 27 | 26 |
| Q5. Abroad | 26 | 26 | 26 | 31 | 31 | 31 | 31 | 31 | 31 | 29.3 | 2.5 | 31 | 26 |
| Q6. nLang* | 30 | 31 | 31 | 29 | 30 | 29 | 27 | 29 | 27 | 29.2 | 1.5 | 31 | 27 |
| Q7. yLearnLang* | 36 | 35 | 34 | 37 | 37 | 39 | 37 | 37 | 39 | 36.8 | 1.6 | 39 | 34 |
| Q8. foreignC | 29 | 29 | 29 | 27 | 27 | 27 | 27 | 27 | 27 | 27.7 | 1 | 29 | 27 |
| Q9. nBooks* | 36 | 36 | 38 | 37 | 35 | 35 | 36 | 35 | 37 | 36.1 | 1.1 | 38 | 35 |
| Q10. yEdu* | 27 | 26 | 27 | 33 | 31 | 32 | 34 | 32 | 33 | 30.6 | 3.0 | 34 | 26 |
| Q11. Orientation* | 24 | 26 | 25 | 23 | 23 | 24 | 25 | 25 | 21 | 24 | 1.5 | 26 | 21 |
| Q4. Place** | 43 | 43 | 43 | 43 | 43 | 43 | 43 | 43 | 43 | 43 | 0 | 43 | 43 |

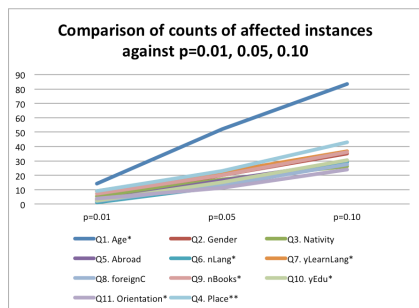


図4 Q1-Q11の該当事例数の p 値による推移 ($p = 0.01, 0.05, 0.10$)

表4 3種類の p 値の該当事例数(平均値)の比較

| Attribute | $p = 0.01$ | $p = 0.05$ | $p = 0.10$ |
|-------------------|------------|------------|------------|
| Q1. Age* | 14 | 52.2 | 83.4 |
| Q2. Gender | 3.3 | 20 | 35 |
| Q3. Nativity | 5.3 | 17 | 26.3 |
| Q5. Abroad | 2.3 | 16.3 | 29.3 |
| Q6. nLang* | 1 | 12.2 | 29.2 |
| Q7. yLearnLang* | 11 | 22 | 36.8 |
| Q8. foreignC | 2 | 13 | 27.7 |
| Q9. nBooks* | 7 | 19.9 | 36.1 |
| Q10. yEdu* | 2 | 15 | 30.6 |
| Q11. Orientation* | 4 | 11.2 | 24 |
| Q4. Place** | 9 | 23 | 43 |

2.4 影響の強さの階層

Fisher exact test の $p = 0.05$ 未満の条件で評定値の分布に有意差を認めた場合、事例数(複数回の計算の平均値)の概要は次の通り:

- (4) Q1: 年齢の影響で 52.2 事例 (23.5%)⁸⁾.
- Q2: 性別の影響で 20 事例 (9.0%).
- Q3: 母語の影響で 17 事例 (7.7%).
- Q4: 地域の影響で 23 事例 (10.7%).
- Q5: 非日本語圏での生活歴の影響で 16.3 事例 (7.4%).
- Q6: 異国語を学んだ数の影響で 12.2 事例 (5.5%).
- Q7: 異国語を学んだ期間の影響で 22 事例 (9.9%).
- Q8: 異国語との日常接触の影響で 13 事例 (5.9%).
- Q9: 読書量の影響で 19.9 事例 (9.0%).
- Q10: 教育歴の影響で 15 事例 (6.8%).
- Q11: 理系・文系の違いで 11.2 事例 (5.1%).

一つの p 値 (e.g., $p = 0.05$) を代表値として選ぶ代わりに、三つの p 値の評価で得られた該当数の平均値を代表値とする事もできる。この基準だと Q1-Q11 は次の順に並ぶ:⁹⁾

- (5) (群 1) age >> (群 2) place, yLearnLang, nBooks > (群 3) gender, nativity > (群 4) abroad, orientation, yEdu > (群 5) foreignContact, nLangLearned

⁸⁾ この評価では区間を 15 年幅で (10,25] (25,40] (40,55] (55,70] としたが、20 年幅にしても結果は大きく変わらない。

⁹⁾ データの性質により、母語性の影響が少な目に評価されている可能性が大きい(日本語を母語としない評定者はほとんどいない)、エンコードの方式により、地域の影響にも同じ事が言える。

年齢の影響が抜群に大きい、群 2 の要因がそれに迫る影響力を持っている。群 2, 3 の差は明確ではないが、群 4, 5 は他の要因より明らかに影響が小さい。

地域の影響は意外に小さいように見える(詳細は §2.6.1 にある)が、評価の方法が異なっているため直接比較が困難である事は指摘しておくべきだろう。実際、決定木を使って評定値を全属性を合わせて回帰した場合、特定の地域が年齢のグループと同様かそれ以上に階層の上位に表れる傾向が確認されている。

2.5 影響の相互関係

先に述べたように、同じ p 値でも該当事例が変動する。PCA の結果は少なからず変動の影響を受ける。これは要因の相互作用に決定的な解釈を与えるのが難しい事を意味する。この理由があり、 $p = 0.01, 0.05, 0.10$ の三つの結果を比較する。

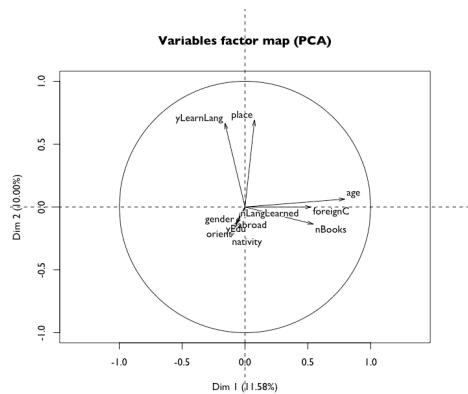


図5 有意差分布を元にした Q1-Q11 の PCA ($p = 0.01$)

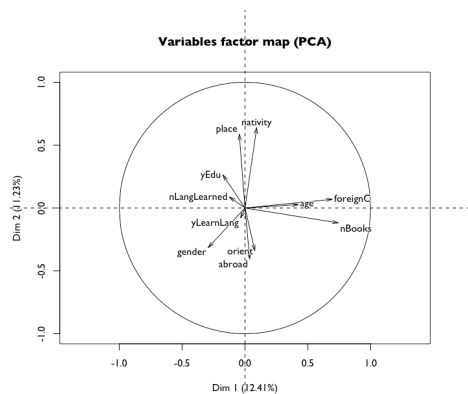


図6 有意差分布を元にした Q1-Q11 の PCA ($p = 0.05$)

p 値が小さいほど、要因の説明力の大きさの差が明確になる。図5 と他の場合を比較すると、yLearnLang,

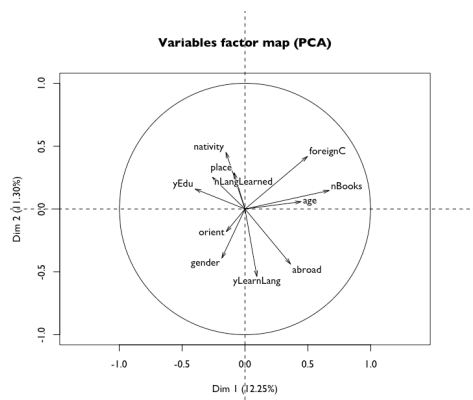


図7 有意差分布を元にした Q1-Q11 の PCA ($p = 0.10$)

place, age, nBooks, foreignC の説明力が相対的に高いのがわかる。ただ、 p ごとに要因の向きが意外と違っている。これには変動の影響もあり、同一の p 値でも異なる得る該当数は変動し、それが PCA の配置に違いとなる。このため、PCA の解釈に関して確定的な事は述べられない。

$p = 0.01$ の場合はスパースネスの影響を受けていると思われる。そのため、変動の小さい $p = 0.05$ の結果 (図6) に基づいて PCA の解釈を行うのが穏当だろう。この条件では、PC1 は後天的 (+方向)/ 先天的 (-方向) の対比を、PC2 は地域の影響 (+方向) に対応しているように思える¹⁰⁾。

興味深いのは、i) 幾つかの要因が相殺している事 (ただし age, nBook, foreignC を効果的に相殺する属性は存在していない)、ii) yLearnLang, nLearnedLang がそれほど強い影響を持っていない事の二点である。前者について言えば、例えば

(6) {place, nativity} \leftrightarrow {abroad, orientation, (gender)} の相殺関係が認められる。

次の点は指摘する価値があるだろう:

(7) 相殺の要因対の存在は、反応の層別化に意味がある事を、つまり反応の平均化による単純化は過度の一般化に伴う危険性を示唆する。

これが ideal speaker-hearer¹¹⁾ を想定して話を始める理論言語学の方法論に意味する事は軽微ではないだろう。

¹⁰⁾ PC が素直に解釈できるのは、Mahalanobis 距離を使った外れ値除去の効果だと思われる。sd のみの外れ値除去では別の結果が得られ、この場合には PC1, PC2 の解釈に苦しんだ。

¹¹⁾ 生成言語学の ideal speaker-hearer の想定の実現性は経済学の homo economicus の非現実性に良く似ている。

2.6 属性の個別の評価

紙面の都合で $p = 0.05$ 未満の有意差条件で該当した事例のみを検討する。Q4 の影響は評価法が異なるので、先に別枠で扱う。

2.6.1 Q4: 住んだ地域の影響

Q4 への回答は、北海道, 青森, ..., 沖縄, その他の 49 種類の地名に該当するかどうかを非排他的にエンコードし、その表データを評定値を線型回帰した結果、次の 23 事例が該当した († の付いた 9 事例は $p = 0.01$ 未満の水準でも有意差あり)。

- (8) a. s001: 幼児が肩車で茎に手を届いた (宮城, 山形*, 茨木, 栃木*, 群馬*, 愛知, 奈良, 宮崎**);
- b. s021†: 弟が家で妹と料理を習った (-切片-, 栃木*, 東京, 奈良, 岡山***, 高知***, 福岡*);
- c. s032: 携帯で担当者が出張先から電話を入れた (福井*, 兵庫*, 和歌山*, 鳥取*, 福岡**, 佐賀, 宮崎, 沖縄);
- d. s061: 子供が宿題で足し算と分数を間違えた (-切片-, 千葉*, 石川**, 福井*, 三重*, 奈良, 香川, 大分);
- e. s087†: 男性がスプーンで角砂糖をコーヒーに入れた (福島, 三重, 大阪*, 和歌山, 愛媛*, 佐賀***, 沖縄*);
- f. s091: 男性がスプーンを角砂糖を紅茶に入れた (-切片-, 茨木*, 栃木**, 群馬*, 岐阜*, 静岡, 大阪**, 和歌山*, 愛媛***, 長崎);
- g. s100: 暴漢が鋭利な刃物で背後から人を襲った (埼玉**, 東京**, 大阪, 兵庫, 岡山*, 広島*, 佐賀*, 長崎**);
- h. s116: 新婚さんが役所で窓口に転居届けと届けた (青森*, 山形, 茨木**, 神奈川**, 石川*, 山梨, 岐阜*, 静岡*, 大阪, 鳥取, 愛媛, 高知*, 大分);
- i. s129: 波紋が投石で落下点から水面を表れた (山形**, 神奈川, 静岡, 兵庫, 奈良*, 和歌山*, 島根*, 佐賀*, 大分*);
- j. s146†: 無線で司令官が本部から命令を伝えた (青森**, 秋田*, 神奈川*, 静岡*, 滋賀***, 奈良*, 大分*, 宮崎**);
- k. s150: 失業者が盗った自動車で仲間とコンビニを襲った (岩手*, 愛知*, 大阪*, 兵庫*, 奈良*, 岡山**, 愛媛*, 長崎);
- l. s161†: 部下が給湯室で上司から噂を聞いた (北海道, 茨木, 山梨***, 岐阜*, 京都, 鳥取*, 香川*, 佐賀**, 大分**);
- m. s196: 通り魔が休日の路上で通行者を不意に襲った (岩手, 秋田***, 東京, 神奈川, 新潟**, 石川*, 滋賀, 兵庫*, 長崎, その他*);
- n. s213: 医師に解剖実習で医学生が看護師と習った (岩手*, 宮城, 茨木, 千葉*, 岐阜, 三重*, 島根**, 岡山*, 宮崎*);
- o. s219: 熊が四重奏で自転車をトレーナーに習った (岩手, 神奈川, 長野**, 奈良*, 山口, 福岡*, 大分**);
- p. s235†: 子供が宿題で足し算とかけ算を分かった (北海道*, 宮城, 栃木***, 埼玉, 愛知*, 三重*, 滋賀*, 大阪*, 鳥取*, 香川*, 高知*, 鹿児島*);
- q. s242: 先生が赤ペンで一から文章を直した (北海道**, 京都*, 奈良**, 香川*);

- r. s246†: 先生が赤ペンで一から文章を直した (-切片*, 青森*, 秋田***, 茨木**, 栃木*, 新潟***, 愛知**, 大分, 宮崎***);
- s. s248: 広場で市民が銃声と叫び声を聞こえた (岩手*, 栃木, 山梨, 長野, 鳥取, 島根, 広島, 香川, 愛媛*, 大分*, 宮崎*, その他)
- t. s250: 医学生が解剖実習で看護師と医師に習った (岩手, 宮城**, 東京, 神奈川*, 静岡*, 愛知*, 滋賀, 山口*, 福岡***)
- u. s265†: 伝書鳩が戦地で戦況を司令分掌に届けた (岩手*, 秋田**, 東京**, 神奈川*, 福井, 岡山*, 愛媛**,)
- v. s281.6†: 夫が職場で真夜中に妻へ知り合った (宮城**, 千葉*, 山梨**, 京都**, 兵庫*, 鳥取*, 岡山**, 山口*, 愛媛, 長崎*, 鹿児島**)
- w. s282.5†: 学生が合格発表の場で足下から幸福を感じた (茨木, 栃木**, 千葉***, 静岡**, 三重, 滋賀, 京都*, 大阪*, 兵庫*, 奈良, 和歌山**, 広島, 香川, 鹿児島*);

***: $p = 0.001$; **: $p = 0.01$, *: $p = 0.05$; no mark: $p = 0.10$ は事例内での有意性水準を示す。この表記は以下同様である。

2.6.2 Q1: 年齢の容認度への影響

年齢の容認度への影響が $p = 0.05$ 水準で有意だった 49 事例を示す († の付いた 15 事例は, $p = 0.01$ 未満の水準でも有意差あり):¹²⁾

- (9) s001†: 幼児が肩車で茎に手を届いた。
s005: 担当者が携帯で出張さきから電話を入れた。
s006: 青年が震災が救出時に飼い犬と助かった。
s008†: 先生が一から赤ペンで文章を直した。
s023: 暴漢が鋭利な刃物で人を背後から襲った。
s029†: 部下が北海道で温泉と同僚と遊べた。
s033: 水を動物園でシロクマが足元に遊んだ。
s038: 人が河原で洪水から子猫が助けた。
s046: 戦況を戦地で伝書鳩が司令官に届けた。
s049: 市民が銃声と叫び声を広場で聞こえた。
s053†: 上司が給湯室で上司から噂を聞いた。
s055†: 司令官が無線で本部から命令をくみ取った。
s059: ランナーが悲鳴を路上で夕暮れ時に聞こえた。
s065: 彼女が手紙で真実を不意に話しかけた。
s067: 娘が病院で医者に重症と訊いた。
s082†: 職員が美術館で絵画でまっすぐに直した。
s089†: 子供が宿題で足し算とかけ算に間違えた。
s092†: 学的が論文賞で顔写真を広報誌に載った。
s110: 仲間とコンビニを失業者が盗んだ自動車で襲った。
s111: シロクマが動物園で水と足元に遊んだ。
s113: 司令官が無線で本部から命令を読みとった。
s120: ランナーが路上で悲鳴を夕暮れ時にかき消した。

- s123: 司令官が無線で本部が命令を伝えた。
s127†: 高得点者が掲示を一位から十位を載った。
s143†: 職員が美術館で絵画をまっすぐに直せた。
s159: 幼児が肩車が枝に手を届いた。
s162: 雇用者が盗んだ自動車で仲間とコンビニを襲った。
s169: 伝書鳩が戦地で戦況を司令官に届けた。
s172†: 青年が震災で救出時に飼い犬と救った。
s180: 失業者を盗んだ自動車で仲間とコンビニを襲った。
s186: 医者に病院で娘が重症と聞いた。
s190: ランナーが路上で悲鳴を夕暮れ時に聞こえた。
s196: 通り魔が休日の路上で通行者を不意に襲った。
s210: 宿敵が続編で苦境に主人公と助けた。
s215: 新婚さんが役所で窓口に転居届と受け取った。
s233†: 伝書鳩が戦地で戦況を司令官に持ち帰った。
s235†: 子供が宿題で足し算とかけ算を分かった。
s236: 高得点者が掲示で一位から十位を書けた。
s241†: チーターが全速力で見事に井倉を襲った。
s242: 失業者が盗んだ自動車で仲間とコンビニを悩ました。
s243: 男性がスプーンで角砂糖を紅茶に取り出した。
s246: 先生が赤ペンで一から文章を直した。
s254: 娘が病院で重症と医者に聞いた。
s262: 幼児が肩車で枝に手を送った。
s272: 捕虜が忠誠心から秘密を尋問で黙った。
s281.0: 夫が職場で真夜中に妻へ知り合った。
s281.1: 夫が職場で真夜中に妻へ知り合った。
s281.7†: 夫が職場で真夜中に妻へ知り合った。
s282.4: 学生が合格発表の場で足下から幸福を感じた。

2.6.3 Q2: 性別の容認度への影響

性別の容認度への影響が $p = 0.05$ 水準で有意だった 20 事例を示す († の付いた 2 事例は $p = 0.01$ 水準でも有意差あり):

- (10) s001: 幼児が肩車で茎に手を届いた。
s019: 先生が赤ペンで文章を一から直した。
s028: 医学生が解剖実習で医師に看護居士と習った。
s040†: 職員がまっすぐに絵画を美術館で直した。
s041: 祖母がキッチンで早朝にラジオを聞いた。
s052: 高得点者が掲示で一位から十位を載った。
s071: 抗議をデモで政府に大衆が繰り返した。
s096: 脚本家が話し合いで前日に台詞を直した。
s110: 仲間とコンビニを失業者が盗んだ自動車で襲った。
s117†: チーターから全速力で見事にガゼルを襲った。
s134: 医師が手術で血管と神経を組んだ。

¹²⁾他の場合でも同じ事だが, Fisher exact test の p 値をシミュレーションで産出している場合, 該当事例数は変動する。

- s149: 命令を無線で本部から司令官が伝えた。
- s154: 学者が論文賞で顔写真を広報誌に書いた。
- s177: 母がキッチンが早朝にラジオを聞いた。
- s212: 子供が宿題で足し算とかけ算を忘れた。
- s232: 新婚さんが役所で窓口で転居届で届けた。
- s260: 見解をその芸能人が質問者に会見で答えた
- s264: ランナーが路上で悲鳴から夕暮れ時に聞こえた。
- s277: 青年が震災で救出時に飼い主と助かった。
- s278: 社長が口頭で社員に解雇を伝えた。

- s109: 男性がスプーンで角砂糖を紅茶に入れた。
- s110: 仲間とコンビニを失業者が盗んだ自動車で襲った。
- s111: シロクマが動物園で氷と足元に遊んだ。
- s167: 弟が家で妹と料理まで習った。
- s188: 宿敵へ続編でピンチに主人公と助けた。
- s197: 大衆が抗議を政府にデモで繰り返した。
- s213†: 医師に解剖実習で医学生が看護師と習った。
- s282.0: 学生が合格発表の場で足下から幸福を感じた。
- s282.1: 学生が合格発表の場で足下から幸福を感じた。

2.6.4 Q3: 母語性の容認度への影響

母語が何かは容認度に影響があつて当然であるが、 $p = 0.05$ 未満の条件で該当したのは次に示す 16 事例のみである († の付いた 5 事例は $p = 0.01$ 水準でも有意差あり):

- (11) s030†: 伝書鳩が戦地で進攻を司令官に届けた。
- s032†: 携帯で担当者が出張先から電話を入れた。
- s049: 市民が銃声と叫び声を広場で聞こえた。
- s100: 暴漢が鋭利な刃物で背後から人を襲った。
- s116: 新婚さんが役所で窓口で転居届けと届けた。
- s162: 雇用者が盗んだ自動車仲間とコンビニを襲った。
- s168: 司令官が無線で本部から命令を伝えた。
- s178: 幼児が肩車で枝に手を届いた。
- s196: 通り魔が休日の路上で通行者を不意に襲った。
- s201: 暴漢が鋭利な刃物で背後から人を悩ました。
- s221†: 職員がギメで絵画をまっすぐに直した。
- s222†: 先輩から現場で後輩が基礎を習った。
- s249: 学者が論文賞で顔写真を広報誌に載った。
- s251: 市民が広場で銃声と叫び声で聞こえた。
- s252: 休日の路上で通り魔が通行者を不意に襲った。
- s276†: 脚本家が話し合いで前日にフレーズを直した。

2.6.5 Q5: 非日本語圏での 1 年以上の生活の容認度への影響

非日本語圏での 1 年以上の生活の容認度への影響が $p = 0.05$ 水準で有意だった 17 事例を示す († の付いた 2 事例は $p = 0.01$ 水準でも有意差あり):

- (12) s003: 担当者が携帯で出張先を電話を入れた。
- s005: 担当者が携帯で出張先から電話を入れた。
- s010: 担当者が携帯で出張先から電話を入れた。
- s011: 同僚が北海道で温泉に同僚と遊んだ。
- s068: 宿敵が続編でピンチに主人公と助け出した。
- s070: 続編で宿敵がピンチに主人公と助けた。
- s071: 抗議をデモで政府に大衆が繰り返した。
- s080†: 娘が病院で医者に軽症と聞いた。

2.6.6 Q6: 異国語を学んだ数の容認度への影響

異国語を学んだ数の容認度への影響が $p = 0.05$ 水準で有意だった 13 事例を示す († の付いた 1 事例は $p = 0.01$ 水準でも有意差あり):

- (13) s018: オコジョが動物園で氷を足元に遊んだ。
- s040: 職員がまっすぐに絵画を美術館で直した。
- s051: 娘が病院で医者に重症と聞いた。
- s061: 子供が宿題で足し算と分数を間違えた。
- s065: 彼女が手紙で真実を不意に話しかけた。
- s088: 人が河原で洪水から子猫を助け出した。
- s133: 熊がトレーナーに自転車をサーカスで習った。
- s139†: 男性がスプーンで角砂糖を紅茶に放り込んだ。
- s140: 弟が家で妹と料理を習わせた。
- s142: 通り魔が休日の路上で通行者を不意に襲った。
- s156: 大衆がデモに政府に抗議を繰り返した。
- s239: チーターが全速力で見事にガゼルを見舞った。
- s258: 新婚さんが役所で窓口で転居届と届けた。

2.6.7 Q7: 異国語を学んだ期間の容認度への影響

異国語を学んだ期間の容認度への影響が $p = 0.05$ 水準で有意だった 23 事例を示す († の付いた 9 事例は $p = 0.01$ 水準でも有意差あり):

- (14) s001: 幼児が肩車で茎に手を届いた。
- s002: 通り魔が休日の路上で通行者を不意に襲った。
- s005: 担当者が携帯で出張先から電話を入れた。
- s027: 社長が口頭で社員に解雇を伝えた。
- s037†: コミュニティが広場で銃声と叫び声を聞こえた。
- s047: 学者が論文賞で顔写真を広報誌に載った。
- s077: 青年が震災で救出時正午に飼い犬と助かった。
- s087†: 男性がスプーンで角砂糖をコーヒーに入れた。
- s094†: 伝書鳩が戦地で戦況を司令官で届けた。
- s097†: 医学生が解剖実習で医師に看護師で習った。
- s101: 青年が震災で救出時に飼い犬と助かった。
- s112†: 初戦を優勝候補がトーナメントで突然に負けた。
- s122: 学者が論文賞で顔写真から広報誌に載った。

s151: 通り魔が休日の路上で不意に通行者を襲った。
 s153: 医学生が解剖実習で医師に看護師と習った。
 s162: 雇用者が盗んだ自動車仲間とコンビニを襲った。
 s172: 青年が震災で救出時に飼い犬と救った。
 s200†: 脚本家が話し合いで翌日に台詞を直した。
 s202†: 子供が宿題で足し算とかけ算から間違えた。
 s253: 高得点者が掲示で一位から十ワーストを載った。
 s264†: ランナーが路上で悲鳴から夕暮れ時に聞こえた。
 s281.5: 夫が職場で真夜中に妻へ知り合った。
 s282.1†: 学生が合格発表の場で足下から幸福を感じた。

2.6.8 Q8: 異国語を話す人との日常的な接触の容認度への影響

異国語を話す人との日常的な接触の容認度への影響が $p = 0.05$ 水準で有意だった 13 事例を示す († の付いた 2 事例は $p = 0.01$ 水準でも有意差あり):

- (15) s005: 担当者が携帯で出張さきから電話を入れた。
 s063: 後輩が現場で先輩から基礎を通った。
 s092†: 学的が論文賞で顔写真を広報誌に載った。
 s142: 通り魔が休日の路上を通行者を不意に襲った。
 s144: 熊がサーカスで自転車トレーナーに習った。
 s234: 十位を掲示で一位から高得点者が載った。
 s236: 高得点者が掲示で一位から十位を書けた。
 s246: 先生が赤ペンで一から文章を直した。
 s249: 学者が論文賞で顔写真を広報誌に載った。
 s263: 医学生が解剖実習で医師に看護師と教わった。
 s281.2: 夫が職場で真夜中に妻へ知り合った。
 s281.8: 夫が職場で真夜中に妻へ知り合った。
 s281.9†: 夫が職場で真夜中に妻へ知り合った。

2.6.9 Q9: 読書量の容認度への影響

読書量の容認度への影響が $p = 0.05$ 水準で有意だった 22 事例を示す († の付いた 7 事例は $p = 0.01$ 水準でも有意差あり):

- (16) s003: 担当者が携帯で出張先を電話を入れた。
 s005: 担当者が携帯で出張さきから電話を入れた。
 s029†: 部下が北海道で温泉と同僚と遊べた。
 s038: 人が河原で洪水から子猫が助けた。
 s059†: ランナーが悲鳴を路上で夕暮れ時に聞こえた。
 s075: シロクマが足元に氷を動物園で遊んだ。
 s089†: 子供が宿題で足し算とかけ算に間違えた。
 s118: 飼い犬と震災で救出時に青年が助かった。
 s166: 脚本家まで話し合いで前日に台詞を直した。
 s171†: 絵画を美術館で職員がまっすぐに直した。
 s189†: 娘が病院で医者に重症と聞いた。

s213: 医師に解剖実習で医学生が看護師と習った。
 s218: 講座が受験対策で女学生に良い固有値を教えた。
 s226: 医学生が解剖実習で医師に看護師と習わせた。
 s228: 路上でランナーが悲鳴を夕暮れ時に聞こえた。
 s236: 高得点者が掲示で一位から十位を書けた。
 s249†: 学者が論文賞に顔写真を広報誌に載った。
 s259†: ランナーが路上で悲鳴を夕暮れ時に聞こえた。
 s281.8: 夫が職場で真夜中に妻へ知り合った。
 s281.9: 夫が職場で真夜中に妻へ知り合った。
 s282.0: 学生が合格発表の場で足下から幸福を感じた。
 s282.6: 学生が合格発表の場で足下から幸福を感じた。

2.6.10 Q10: 教育期間の容認度への影響

教育期間の容認度への影響が $p = 0.05$ 水準で有意だった 18 事例を示す († の付いた 2 事例は $p = 0.01$ 水準でも有意差あり):

- (17) s043: 暴漢が鋭利な刃物で背後から人を襲いかかった。
 s051: 娘が病院で医者に重症と聞いた。
 s081: 医学生が解剖実習で医師に看護院と習った。
 s091: 男性がスプーンを角砂糖を紅茶に入れた。
 s101: 青年が震災で救出時に飼い犬と助かった。
 s107: 前日に話し合いで脚本家が台詞を直した。
 s130: 先生が赤ペンで一から詩を直した。
 s131†: 肩車で幼児が枝に手を届いた。
 s153: 医学生が解剖実習で医師に看護師と習った。
 s161: 部下が給湯室で上司から噂を聞いた。
 s163: 顔写真を論文賞で学者が広報誌に載った。
 s192: シロクマが動物園で氷を足元に楽しんだ。
 s201†: 暴漢が鋭利な刃物で背後から人を悩ました。
 s218: 講座が受験対策で女学生に良い固有値を教えた。
 s234: 十位を掲示で一位から高得点者が載った。
 s267: 医師が手術で血管と神経をつないだ。
 s279: 母がキッチンで早朝にラジオを聞いた。
 s281.4: 夫が職場で真夜中に妻へ知り合った。

2.6.11 Q11: 理系/文系の違いの容認度への影響

理系/文系の違いの容認度への影響が $p = 0.05$ 水準で有意だった 10 事例を示す († の付いた 3 事例は $p = 0.01$ 水準でも有意差あり):

- (18) s011: 同僚が北海道で温泉と同僚と遊んだ。
 s112: 初戦を優勝候補がトーナメントで突然に負けた。
 s120: ランナーが路上で悲鳴を夕暮れ時にかき消した。
 s214: 担当者が電話を出張先から携帯で入れた。
 s218†: 講座が受験対策で女学生に良い固有値を教えた。
 s247†: シロクマがほ乳類園で氷を足元に遊んだ。
 s256: 宿敵が続編からピンチに主人公と助けた。

- s270: 家で弟が妹と料理を習った。
 s280: 新婚さんが役所が窓口に移居届と届けた。
 s281.1†: 夫が職場で真夜中に妻へ知り合った。

3. 属性間の相互作用の予備的解析

評定値の決定木分析 (= 全要因を組み合わせて回帰分析) を実行すれば, 属性間の相互作用を確認できる。解析は終わっているが, 紙面の都合で詳細を割愛する。分析結果の本格的な検討は稿を改めて行いたい。

3.1 結果の概要

概要のみを簡単に述べると (19) にある通り:

- (19) a. Q1-Q11 の要因の効き方は (特に優先順位の点で) 事例ごとに大きく異なるが,
 b. 地域や年齢の影響は決定木の階層の上に現れる = 優先度の高い要因である (地域の方が年齢よりも優先度が高い場合も多い)。

4. 結論と今後の展望

本研究は理論的バイアスの少ない大規模データを使って, 容認性判断に与える社会的要因の強さを定量的に評価した初めての研究である。それにより, 容認度評定が質的に異なるグループから構成されている事が示され, 理論言語学の単純化が素朴過ぎる可能性が示唆された。

解析に使った属性は十分に網羅的とは言えず, 一部の属性のサンプル数が不足しており, 決して現時点で完成度が高い研究とは言えない。それでも, 容認度に顕著な影響を与える要因を幾つか (例えば年齢, 異国語の学習期間, 読書量, 性別), その強さと共に確認できた。年齢の影響が強い事の確認は定説の確証の意味しかないが, 同じ程度の強さで性別と教育年数の影響が出る事は (計測誤差の範囲でないならば) 予測されていた事実はと言えない。加えて, 地域の影響は通説で言われているより弱い可能性も示唆されている。このような結果から見て, 本研究の結果は一部で確証的, 一部で探索的であるが, 将来なされるべき研究の先駆として一定の成果を挙げていると結論したい。説明が後知恵で「常識」と思える時に, それが実際には単なる勘違いである事は, 社会科学では一般的である [13]。

今後の展開は次の通りである。まず, 有意性が生じる原因の解明を行う— これは本発表では意図的に避けている。更に異なる種類の反応データとの連結を目指す。具体的には, 評定値を取得済みの刺激文に対し, 反応時間データや眼球運動データを取得し, それらに対応させる。観測値の充実により, 心理プロセスとしての容認度評定の実態に迫りたい。これらに加えて, 反応の個人差の分析も充実させる必要があり, 刺

激文の変異を拡張し, 探索範囲を広げる必要もある。やるべき事は, まだまだある。ないのは資金と時間だけである。

謝辞

本研究は JSPS 挑戦的萌芽研究 (課題番号 16K13223) の支援を受けた。

参考文献

- [1] Jonathan Baron. *Thinking and Deciding*. Cambridge University Press, 2000 [1988, 1994].
- [2] H. W. Bierhoff and R. Klein. Expectations, confirmation bias, and suggestibility. In V. A. Gheorghiu, P. Netter, H. J. Eysenck, and R. Rosenthal, editors, *Suggestion and Suggestibility*, pp. 337–346. New York: Springer, 1989.
- [3] Wayne Cowart. *Experimental Syntax: Applying Objective Methods to Sentence Judgments*. Sage Publishing, Inc., 1997.
- [4] Joshua Klayman. Varieties of confirmation bias. *Psychology of Learning and Motivation*, 32:384–418, 1995.
- [5] 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 Natural Language Processing*, pp. 65–68, 2018.
- [6] 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 Natural Language Processing Society*, pp. 253–256, 2019.
- [7] 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*, to appear.
- [8] Ira A. Noveck and Dan Sperber, editors. *Experimental Pragmatics*. Palgrave Macmillan, 2005.
- [9] Gary Dean Prideaux, Bruce L. Derwing, and William J. Baker. *Experimental Linguistics: Integration of Theories and Applications*. Studies in the Sciences of Language. John Benjamins, 1979.
- [10] Carson T. Schütze. *The Empirical Base of Linguistics*. Language Science Press, 2016 [1996].
- [11] Jon Sprouse. Three open questions in experimental syntax. *Linguistics Vanguard*, 1(1):89–100, 2015.
- [12] Jon Sprouse and Norbert Horstein. *Experimental Syntax and Island Effects*. Cambridge University Press, 2013.
- [13] Duncan J. Watts. *Everything is Obvious* [*Once You Know the Answer]: How Common Sense Fails*. Atlantic, 2012. [翻訳: ダンカン・ワッツ: 「偶然の科学」. 早川書房.]
- [14] 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.