

How nearly random mutations to sentences affect their acceptabilities:

Preliminary quantitative analyses based on ARDJ data

Kow Kuroda¹, Hikaru Yokono², Keiga Abe³, Tomoyuki Tsuchiya⁴, Yoshihiko Asao⁵,
Yuichiro Kobayashi⁶, Toshiyuki Kanamaru⁷, and Takumi Tagawa⁸

¹Kyorin University, ²Meisei University, ³Wako University, ⁴Kyushu University,

⁵National Institute of Communications and Information Technology, ⁶Nihon University, ⁷Kyoto University, ⁸Tsukuba University

Abstract

ARDJ s2u data consists of 300 sentences paired with acceptability ratings. Out of them, 36 are “originals” that were human-made and other 264 are their “mutations” that were generated either by nearly random replacement of a nominal, a verbal or a positional (e.g., case-marker) or by random swapping a pair of phrases (i.e., bunsetu) inside. This allows us to analyze effects of mutation systematically. With DBSCAN of the originals, layered analysis of mutation effects was carried out. Findings are: i) No mutations improve low acceptabilities; ii) The order of systematicity is: $p < v < n < s$; iii) The order of effect size is: $p < v, n, s$, so that n -, s -mutations often cause as large effects as v -mutation, or even larger; iv) No simple dichotomy is possible between “acceptable” and “unacceptable” sentences. Rather, their distinction is continuous and graded, so that (logistic) regression is a better modeling of the distinction.

1 Introduction

Linguistics is concerned with acceptabilities of sentences, if not exclusively. But acceptability is an elusive notion [1, 2, 3, 4, 5, 6]. Any group of linguists rarely agree on its theoretical definition, meaning that it is a highly theory-laden notion [7]. If its theoretical definition is somehow shared, its operative definition is virtually missing. The situation is still worse. Even if presume its definition is fine, we are still far from knowing what acceptabilities really are. No unbiased, large-scale collection of the required data has ever been done. In particular, we are far from fully understanding what modifications to a sentence bring about what effect, and to what degree, because data-driven study into this problem has never been done. What linguists know about acceptability suffers from “confirmation-bias” [8, 9, 10, 11, 12, 13]. This is the issue that dictates this research based on *Acceptability Rating Data for Japanese* (ARDJ for short) [14, 15, 16, 17, 18, 19].

ARDJ is a project of empirically-oriented language science, rather than mere linguistics, that attempts to implement the framework of randomized control trials in the field of linguistics, with sympathy for *Evidence-based Medicine*

(*EBM*) [20, 21]. A crucial conviction that paved the way for ARDJ is that intuitions by experts (e.g., acceptability judgments by linguists) are valuable with no doubt, but they are too weak “evidence” to be treated as “facts”; rather, they are more or less reliable “guides.”

The stimuli used in the ARDJ surveys comprise two kinds of sentences: 36 “originals” hand-crafted by humans and their “variants” generated by applying nearly random mutations on the originals. Mutations are only nearly random rather than fully random, because they have target sites, and values for replacement are chosen based on word2vec distributional similarities. See [16] for relevant details. Appendix A.1 gives a brief summary. Effects of mutation can be systematically analyzed by comparing responses to the originals and their mutations.

2 Analysis and results

To make our analysis effective, we use PCA, tSNE with DBSCAN clustering, overplot, and KL-divergence-based heatmap/clustering. In §2.3.1–§2.3.4, analysis of effects of mutation is implemented as a layered analysis in which Clusters 1, 2, 3 and 4, are taken as four separate groups of data. Discarding the originals in Clusters 2, 3 and 4, KL-divergence-based analysis is performed on mutations to the originals in Cluster 1 in §2.4. Lastly, effects are analyzed by exploring properties of mutation-wise KL-divergence heatmaps in §2.5.

2.1 36 originals and their attributes

ARDJ s2u data [17] has 36 “originals” presented in Figure 1, o01, o02, ..., o36, with relevant attributes. Details are presented in §2.1. How this data was prepared is described in Appendix A. The originals are sorted decreasingly by $r01$, $r12$ values, and increasingly by $r23$, $r3x$ values. Only 32 originals have more than 2 mutations. So, the number of effective originals is 32 rather than 36, as Figure 1 shows. The effective ones have (8.31 - 1) mutations/variants on average, with a maximum of 16 and minimum of 4.

o.id	S.TEXT	s.id	r01	r12	r23	r3x	count
o05	男性がステージンで角砂糖を紙袋に入れた。	s109	0.73	0.06	0.02	0.01	8
o08	長身は口で指しを刺さるに任せた。	s227	0.76	0.06	0.02	0.02	9
o27	先生が紙のついでに文章を書いた。	s246	0.76	0.06	0.02	0.02	6
o20	愚漢が鋭利な刀で背後から人を襲った。	s100	0.83	0.10	0.02	0.01	6
o06	担当者が携帯電話で出張先から電話を入れた。	s199	0.86	0.11	0.01	0.02	8
o34	飛脚が飛脚で先陣から長巻を穿った。	s179	0.82	0.11	0.03	0.03	8
o22	通り魔が財布の裏面に盗難者名簿を書いた。	s196	0.79	0.14	0.05	0.03	7
o03	目がキーンと早くにラジオを聞いた。	s279	0.78	0.16	0.05	0.01	7
o31	弟が家で練習料理を習った。	s021	0.78	0.16	0.04	0.02	7
o17	医師が手術で血管と神経をつないだ。	s267	0.78	0.15	0.02	0.05	8
o06	職員が事務用で書類を全部でなくした。	s245	0.79	0.16	0.05	0.02	7
o25	脚本家が話し合いで毎日に行方不明をした。	s096	0.74	0.20	0.04	0.02	10
o23	先業者が盗んだ自動車で特撮にエキストラを襲った。	s231	0.74	0.16	0.07	0.03	7
o29	伝書鳩が戦場で戦況を司令官に届けた。	s169	0.71	0.18	0.07	0.04	9
o09	司会者が観客と実演家と両方を伝えた。	s168	0.69	0.23	0.05	0.03	8
o02	部下が伝言棒で上司から噂を聞いた。	s161	0.66	0.27	0.05	0.02	7
o21	チーターが全速力で見事にガゼルを襲った。	s158	0.59	0.32	0.05	0.03	9
o16	人が所帯で湯水から子猫を助けた。	s206	0.56	0.34	0.07	0.03	9
o04	子供が幼稚園で先生が肉を噛んだ。	s191	0.46	0.07	0.00	0.00	9
o11	大衆が全てで感服に抗議を繰り返した。	s079	0.44	0.09	0.03	0.44	9
o01	娘が病院で医者に看護を聞いた。	s051	0.41	0.34	0.20	0.05	9
o02	熊がサーカスで自動車をトレーナーに奪った。	s144	0.41	0.38	0.12	0.10	5
o07	字のうまい青年が外国で生活の必需品を盗んだ。	s103	0.38	0.31	0.19	0.12	1
o04	船が船中庭で乗客を乗せた。	s099	0.27	0.42	0.21	0.10	1
o15	関係者がインターネットで個人情報を探知し知り合った。	s205	0.20	0.38	0.28	0.13	1
o28	青年が親友で救出時に胸に刺さった。	s101	0.19	0.51	0.21	0.08	9
o36	私が幼稚園でインフルエンザで学級を感染した。	s275	0.17	0.48	0.15	0.20	1
o13	シロヤが動物園で動物を盗んだ。	s157	0.10	0.42	0.29	0.20	8
o33	中学生が解題練習で英語に音読を習った。	s153	0.09	0.41	0.30	0.20	9
o30	新報さんが戦所で急に紙巻煙草を吸った。	s258	0.06	0.42	0.25	0.26	8
o19	高橋浩吉が胸で一目から十指を奪った。	s052	0.05	0.38	0.25	0.32	9
o14	部下が伝言棒で上司から噂を聞いた。	s204	0.04	0.41	0.29	0.23	9
o15	演劇が観劇で芝居に人工を演じた。	s054	0.04	0.21	0.39	0.36	8
o10	ランナーが路上で急病を克服し走り続けた。	s190	0.04	0.29	0.31	0.36	17
o12	幼童が親友で握手を握った。	s178	0.04	0.37	0.27	0.32	8
o18	字が論文で漢字を公開した。	s047	0.02	0.39	0.35	0.24	9
sum			17	9	5	5	270
count			36	36	36	36	36
average			0.48	0.26	0.13	0.13	7.50
max			0.92	0.51	0.39	0.46	17.00
min			0.02	0.06	0.00	0.01	1.00
stdev			0.32	0.14	0.12	0.14	2.96

Figure 1: 36 originals with attributes

The originals were clustered using DBSCAN, giving us 4 clusters. §2.2 deals with this result. The resulting clustering is laid out over a PC1-PC2 plane of PCA rotation and a Dim1-Dim2 plane of tSNE transformation. This gives us how the sentences are associated to each other.

2.2 Clustering of originals and all stimuli

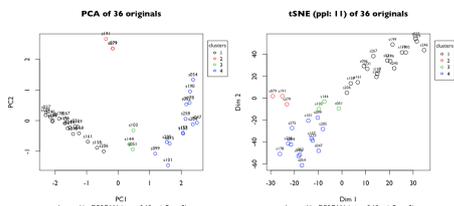


Figure 2: PCA and tSNE of 36 originals

2.2.1 Clustering originals only

DBSCAN was used to cluster the 36 originals to identify 4 clusters, which are shown in PCA and tSNE plots in Figure 2. tSNE in Figure 2 suggests the existence of a unified dimension along which acceptability can be measured. The deviance of Cluster 2 from it can be accounted for its special nature described below.

2.2.2 Clustering all 270 stimuli

For comparison, we present clustering of all stimuli including the 36 originals described above. Figure 3 gives PCA and tSNE plots of DBSCAN clustering. Note that the setting for eps and minPts parameters here is independent from those for the parameters for the DBSCAN clustering of the 36 originals.

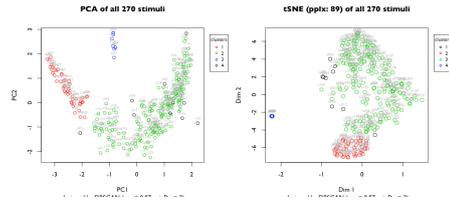


Figure 3: PCA and tSNE of all 270 stimuli

2.3 Layered Analysis

This clustering result seems natural in that clusterwise overplots in Figure 4 seem to capture both the similarities and dissimilarities quite well. Each cluster consists of instances with surprising similarities.

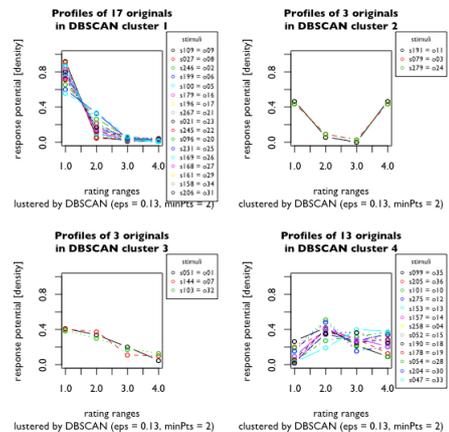


Figure 4: Clusterwise overplots of originals

Comparing with the data in Figure 1, we can safely conclude the following:

- (1) a. Cluster 1 identifies the group of “acceptable” stimuli, Cluster 4 the one of “unacceptable” stimuli, and Clusters 2 and 3 the ones of “ambiguous” stimuli, with different qualities.
- b. Cluster 3 consists of 3 instances, o03, o11 and o24, that have a “split” decision, in that raters are split into two groups of those who accept them fully and those who reject them fully.
- c. Cluster 4 consists of 3 instances, o01, o07 and o32, that have a nearly “uniform” distribution of (un)acceptability along the range, though with a slight inclination toward acceptability.

With these distinctions among the originals, we can now turn to effectively analyzing the effects of mutation to answer the research question: how random mutations to sentences affect their acceptabilities, and to what extent.

2.3.1 Analysis 1a: Mutations to originals in Cluster 4

We start our analysis with the one of the mutations to the originals to Cluster 4, a relatively large cluster, having 13

members which can be labelled as “unacceptable” stimuli. This is because the effects are simple and can be characterized most easily.

Effects of mutation are analyzed in three visualizations, as shown in Figure 5: i) overplot of response profiles (in the leftmost column); ii) proportional distribution of range values after reordering; and iii) layout of DBSCAN clusters in a rotated geometry G .

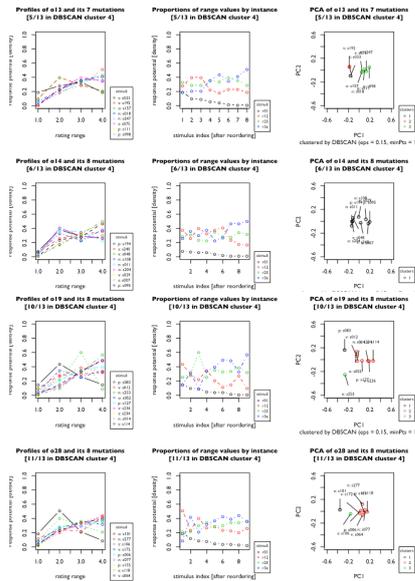


Figure 5: Mutations to originals to Cluster 4 [4 samples]

In each analysis, stimuli are ordered by their range values. So, their positions relative to 0 in the legend roughly indicate if mutations increase or decrease acceptabilities.

As suggested by the plots in Figure 5, virtually no mutations improve the original acceptabilities. In addition, the response profiles, i.e., shapes of distribution, stay unchanged for most cases. So, there seems to be a rule like “once (made) deviant, always deviant.” This finding is not truly surprising, let alone unexpected; yet, it deserves a mention because this is the first empirical confirmation of a truism on a large scale that has long awaited testing.

2.3.2 Analysis 1b: Mutations to originals in Cluster 3

Effects of mutations to cases in Cluster 3 are similar to the ones to stimuli in Cluster 4. Stimuli in Cluster 3 almost are almost never improved by any mutation, though, admittedly, we can hardly state generalizations safely because we have too few affected cases.

2.3.3 Analysis 1c: Mutations to originals in Cluster 2

Effects of mutations to cases in Cluster 2 are somewhat different from the ones to stimuli in Cluster 3. Unlike stimuli in Cluster 3, stimuli in Cluster 2 sometimes improve, though, again, we can hardly state generalizations safely

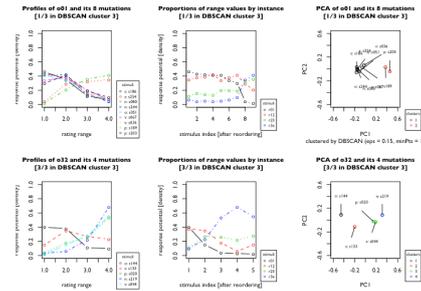


Figure 6: Mutations to originals to Cluster 3 [One case was dropped due to insufficient cases.]

due to having too few affected cases. It is interesting that o11 and o24 get more acceptable through certain mutations, though o03 does not possess such properties.

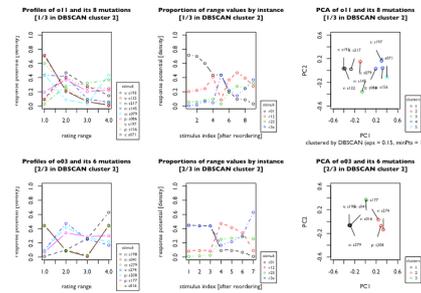


Figure 7: Mutations to originals to Cluster 2 [2 samples]

2.3.4 Analysis 1d: Mutations to originals in Cluster 1

Mutations to the originals in Clusters 2, 3 and 4 tell us virtually nothing interesting about how mutations affect acceptability, because we can hardly identify meaningful differences between mutations and their originals. This is not true of mutations to the originals in Cluster 1.

Mutation effects are different case by case, but generally speaking, we can observe the following:

- (2) a. Lexical changes (i.e., mutations to either n(ominals), v(erbals), or p(ositionals)) nearly always drop acceptability, while non-lexical changes (i.e., swapping) do not. Swapping often keeps the acceptability.
- b. It is not true, at least not observed, that n-changes reduce acceptability more than v- and p-changes, as far as degree of deviance is measured.
- c. Changes to p introduce only a slight amount of deviance, though it is unignorable.
- d. Effects of mutation are diverse, in that the number of clusters recognized is greater than the ones in other clusters, Clusters 2, 3 and 4.
- e. Mutations do not simply make sentences unacceptable, in that resulting profiles do not always look like those in the originals in Cluster

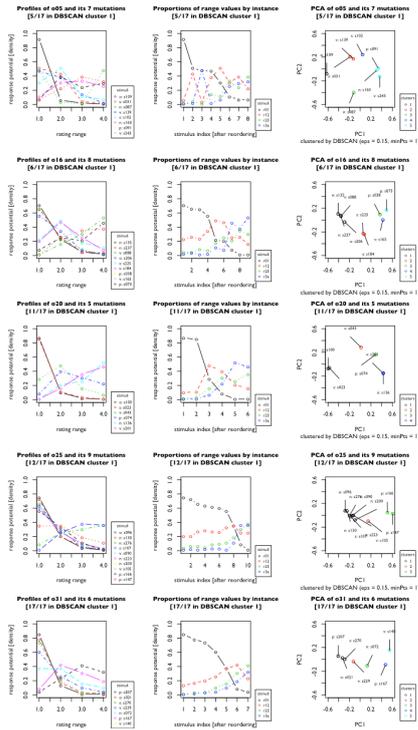


Figure 8: Mutations to originals to Cluster 1 [5 samples]

4. Rather, they look more like the originals in Cluster 3 or 2. So, the effects of mutation are more subtle than researchers have guessed.

With their relative importance, cases in Cluster 1 deserve further exploration. We do it by analyzing them with KL-divergence cross-table.

2.4 Analysis 2: KL-divergence for originals to Cluster 1

The diagrams in Figure 9 are heatmaps of the 9 samples, taken in Figure 8, allowing us to make the following observations:

- (3) a. p- and v-mutations tend to cluster together, but not always.
- b. swapping and n-mutations tend to lack systematic effects, though they often cause drastic deterioration.

2.5 Analysis 3: KL-divergence cross-table by mutation

Different types of mutation, i.e., p, v, n and s, have different magnitudes of effects. To estimate them, we can use KL-divergence heatmaps, as shown in Figure 10.

If the contrasts between “hot” and “cool” areas are sharp, the magnitude of effects can be judged to be large. If the

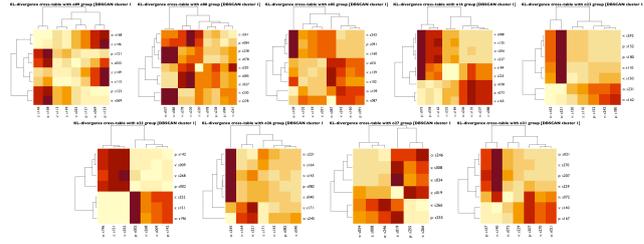


Figure 9: KL-div. heatmaps for Cluster 1 cases [o09, o08, o05, o16, o23, o22, o26, o27, o31]

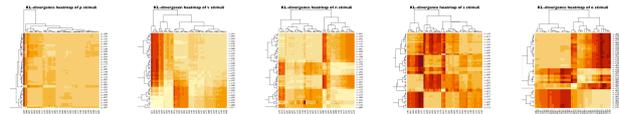


Figure 10: KL-div. heatmaps for mutation types [p, v, n, s, o]

degree contrasts are spread, effects are unsystematic. With this criterion, we can conclude that i) the effect of p mutations is least contrastive and most systematic; ii) that of v mutations is second least contrastive and rather systematic; iii) that of n mutations is second most contrastive and rather unsystematic; and iv) that of s mutations is most contrastive and most systematic. Note that the 36 originals are contrastive only because they are largely designed so.

3 Discussion and Conclusion

So, what are causes of reduced acceptabilities? With the results so far, we can now decide if we are able to answer this question. The most sincere answer to this would be: there is no simple answer to this question. What does this mean, then?

Simple dichotomy between “acceptable” and “unacceptable” kinds of sentences is too simple and unrealistic, though it is widely accepted and practiced in many areas of linguistics. What really matters about acceptabilities are response profiles (distributional pattern of responses) rather than categorical judgments, on the one hand, and regression to response profiles rather than binary classification (e.g., dichotomy = categorical judgment), on the other. This is a conclusion compatible with the results in [15].

What do linguists need to do with this reality, then? We would suggest that the goals of linguistics should be redefined. Redefined goals include the **construction of a mapping from sentences in textual form (e.g., values in s_{text} in Table 1) to response profiles (e.g., arrays of values [r01, r12, r23, r3x] in the table in Figure 1)**, yielding overplots in Figures 4. This can never be attained until large-scaled, bias-free databases like ARDJ are built for as many languages as possible. Until then, linguistics would never make itself an empirical science.

Acknowledgements

This research was supported by JSPS through Grant-in-Aid (16K13223).

We are grateful for Shunji Awazu (Jissen Women’s University), Asuka Terai (Future University of Hakodate) and Minoru Yamaizumi (Osaka University), who kindly helped our execution of Phase 1 of ARDJ Survey 2.

All statistic analyses were done on R version 3.6.3, a free statistical software developed by R Core Team (<https://www.r-project.org>).

References

- [1] Wayne Cowart. *Experimental Syntax: Applying Objective Methods to Sentence Judgments*. Sage Publishing, 1997.
- [2] Ira A. Noveck and Dan Sperber, editors. *Experimental Pragmatics*. Palgrave Macmillan, 2005.
- [3] Gary Dean Prideaux, Bruce L. Derwing, and William J. Baker. *Experimental Linguistics: Integration of Theories and Applications*. John Benjamins, 1979.
- [4] Carson T. Schütze. *The Empirical Base of Linguistics: Grammaticality Judgments and Linguistic Methodology*. Language Science Press/University of Chicago Press, 2016 [1996].
- [5] Jon Sprouse. Three open questions in experimental syntax. *Linguistics Vanguard*, Vol. 1, No. 1, pp. 89–100, 2015.
- [6] Jon Sprouse and Norbert Horstein. *Experimental Syntax and Island Effects*. Cambridge University Press, 2013.
- [7] Holly P. Branigan and Martin Pickering. An experimental approach to linguistic representation. *Behavioral and Brain Sciences*, Vol. 40, 2017.
- [8] Jonathan Baron. *Thinking and Deciding*. Cambridge University Press, 2000 [1988, 1994].
- [9] 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.
- [10] Joshua Klayman. Varieties of confirmation bias. *Psychology of Learning and Motivation*, Vol. 32, pp. 384–418, 1995.
- [11] Thomas Gilovich. *How We Know What Isn’t So: The Fallibility of Human Reason in Everyday Life*. Free Press., 1993. [(日本語訳) 人間, この信じやすきもの: 迷信・誤信はどうして生まれるか. 新曜社.]
- [12] David Robson. *The Intelligence Trap: Why Smart People Do Stupid Things and How to Make Wiser Decisions*. Hodder & Stoughton, 2019.
- [13] Stuart A. Vyse. *Believing in Magic: The Psychology of Superstition*. Oxford University Press, reprinted edition, 2000. [(日本語訳) 人はなぜ迷信を信じるのか: 思い込みの心理学. 朝日新聞社, 1999.]
- [14] 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.
- [15] Kow Kuroda, Hikaru Yokono, Keiga Abe, Tomoyuki Tsuchiya, Yoshihiko Aaso, Yuichiro Kobayashi, Toshiyuki Kanamaru, and Takumi Tagawa. Simulating acceptability judgments using ARDJ data. In *Proceedings of the 27th Annual Meeting of Association of NLP*, pp. 739–743, 2021.
- [16] 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.
- [17] 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.
- [18] 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, 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. 日本認知科学会第 36 回大会発表論文集, pp. 727–736, 2019.
- [19] 黒田航, 阿部慶賀, 栗津俊二, 寺井あすか, 土屋智行. ARDJ を基にした容認度評定値と反応時間の多変量解析と混合回帰分析. 認知科学会第 37 回大会発表論文集, pp. 919–928, 2020.
- [20] G. Evidence-based Medicine Group [Guyatt, J. Cairns, and D. et al.] Churchill. Evidence-based medicine: A new approach to teaching the practice of medicine. *Journal of the American Medical Association*, Vol. 268, No. 17, pp. 2420–5, 1992 (Nov 4).
- [21] David Isaacs and Dominic Fitzgerald. Seven alternatives to evidence based medicine. *The British Medical Journal*, Vol. 319, No. 7225, p. 1618, 1999.

A Data preparation

A.1 Construction of stimuli

To make the outcomes of ARDJ truly empirical and useful, we need to employ as unbiased stimuli as possible. The best way to implement it is to employ automatic generation rather than relying too much on human intuitions. Random mutation in DNA sequences serves as a model. The basic idea is the following:

(4) Steps of randomized generation with targets

- Step 1. Construct sentences, called “originals,” $O = \{o_1, o_2, \dots, o_n\}$ with or without deviance.
- Step 2. Introduce random mutations to instances of O . Let M denote the result.
- Step 3. Mix O and M and use its subsets for acceptability rating tasks.

More details of Step 2 are given in the following:

- (5) A) Random replacement of a lexical item under POS-identity (edit type: l(exical)); B) Random replacement of a postpositional case-marker (P) with another (edit type: p(ositional)); C) Random positional exchange of a given pair of NPs (or PPs) (edit type: s(wapping)).

provided that items for replacement are selected on a pre-defined measure such as distributional similarity.

The first and second authors developed Python scripts to implement mutations of A-, B- and C-classes. They are provided at <https://github.com/kow-k/Japanese-sentence-mutators>.

A.2 Surveys

By 2019, ARDJ completed two experiments. The first one, called “Survey 1,” was carried out in 2017. It was intended to be a pilot study with only a limited variety of responders (roughly 200 college students only) on 200 sentences for stimuli. Kuroda et al. [16] reported on this survey. The second experiment, called “Survey 2,” was carried out in 2018. This was the main study, with the stimulus set expanded to 300 with some overlap with Survey 1.

Survey 2 had two phases, Phases 1 and 2, and collected responses from 1,880 participants in total. Phase 1 was a small scale, paper-based survey, to which 201 participants (mostly all college or university students) contributed responses. This was comparable to the pilot study done in 2017. Phase 2 was a large scale web survey to which 1,679 participants contributed responses. They were significantly more varied in their attributes and we would safely state responses obtained were randomized enough. Survey 2 unified (s2u) data consists of unification of responses from Phase 1 and Phase 2. These data are freely available at <https://kow-k.github.io/Acceptability-Rating-Data-of-Japanese/>, but you need to register to use them.

A.3 Collection of ratings/responses

On acceptability rating, participants were asked to choose one of the four choices in (6).

- (6) **0.** 違和感がなく自然に理解できる文 [natural and easy to understand]; **1.** 違和感を感じるが理解可能な文 [more or less deviant but comprehensible]; **2.** 違和感を感じて理解困難な文 [deviant and difficult to understand]; **3.** 不自然な理解不能な文 [quite unnatural and incomprehensible]

Prefixes 0, 1, 2 and 3 are added to indicate the degrees of deviance, though they need not be on a single scale.

Outlier responders were filtered out using standard deviation ($0.6 < sd < 1.5$) and Mahalanobis distance (< 0.95). See Kuroda et al [17] for relevant details.

A.4 Standardizing responses

Note that gr0, ..., gr9 are different data sets, and cannot be directly compared. Comparison of them requires standardization. All groups were collapsed and responses were counted for each of the four rating ranges $r[0, 1)$, $r[1, 2)$, $r[2, 3)$, and $r[3, \infty)$.

Table 1 shows 10 samples of this process. These raw counts were then converted into proportions to item-wise sums. This gave us density array, $P = (p[0, 1), p[1, 2), p[2, 3), p[3, \infty))$, as exemplified in Table 2. Note that allowed response values were 0, 1, 2, and 3. These numbers are reinterpreted as ranges $r[0, 1)$, $r[1, 2)$, $r[2, 3)$, and $r[3, \infty)$, respectively, where $r[i, j)$ means a sum of response counts between i and j with i included and j excluded. The converted arrays of ranged response probabilities in this format are to be called “response potentials.” They are commensurable among groups with different sets of responders, and were used as effective encodings of the responses in a variety of analyses.

s.id	$r[0, 1)$	$r[1, 2)$	$r[2, 3)$	$r[3, \infty)$	sum
s021	128	25	5	2	160
s064	3	42	53	53	151
s128	5	50	32	60	147
s188	1	12	36	98	147
s231	119	27	10	4	160
s282.4	27	77	32	15	151

Table 1: Frequency table by ranges (6 samples)

s.id	$p[0, 1)$	$p[1, 2)$	$p[2, 3)$	$p[3, \infty)$	sum
s021	0.800	0.156	0.031	0.013	1.00
s064	0.020	0.278	0.351	0.351	1.00
s128	0.034	0.340	0.218	0.408	1.00
s188	0.007	0.082	0.245	0.667	1.00
s231	0.744	0.169	0.063	0.025	1.00
s282.4	0.179	0.510	0.212	0.099	1.00

Table 2: Density table by ranges (6 samples)