

面白いアナグラムとはどんなアナグラムか

土井 遥^{1†} 山本 優衣奈^{2†} 佐藤 理史³

¹ 名古屋市立向陽高等学校

² 三重県立桑名高等学校

³ 名古屋大学大学院工学研究科

ssato@nuee.nagoya-u.ac.jp

概要

アナグラムとは、「言葉の綴りの順番を変えて別の語や文を作る遊び」である。本稿では、表記と読みの組からなる入力に対して、アナグラム候補を順位付きで出力するシステムを示す。さらに、システムで生成した大量のアナグラム候補の中から「面白い」と思えるアナグラムを収集・分類し、「面白いアナグラムとはどんなアナグラムか」という問いに対して、ひとつの仮説を提示する。

1 はじめに

アナグラムとは、「言葉の綴りの順番を変えて別の語や文を作る遊び [1]」である。日本語のアナグラムは、読み(ひらがな文字列)の順番を変えて作るのが一般的であり、たとえば、「時計/とけい」から「毛糸/けいと」を作ることが、これに相当する。

ひらがな文字列の並べ替えは、機械的に可能であり、容易である。たとえば、7文字のひらがな文字列の並べ替えは、 $7! = 5040$ 個存在する。しかしながら、その中で日本語として意味をなすものは限られており、「面白い」ものとなると、さらに限られている。それゆえ、面白いアナグラムを作るとは知的なパズルであり、ある種の創作とみなすことができる。

鈴木ら [2] は、あらかじめ用意した文節集合に含まれる文節をつなげて、アナグラムを生成する方法を提案・実装した。しかし、この生成法には意味的適格性の判定が陽に含まれておらず、意味の通るアナグラムは出力のほんの一部に限られた。後続研究 [3] では、係り受けペアの一般化による意味的適格性判定が試みられ、一定の効果が見られたが、根本的な解決には至らなかった。

本研究では、「面白いアナグラムを作る」ことを

† Equal contribution

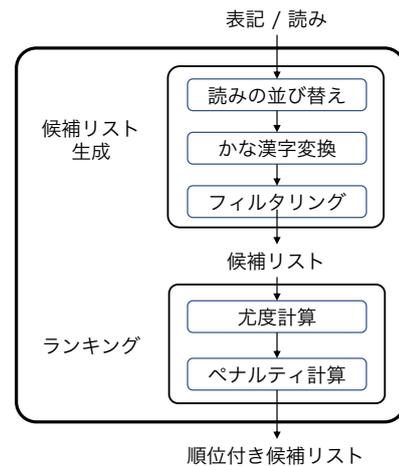


図1 アナグラム生成システム

研究目標として掲げ、有望な候補を順位付きで出力するシステムの実現を試みる。さらに、システムで生成した大量のアナグラム候補の中から「面白い」と思えるアナグラムを収集・分類し、「面白いアナグラムとはどんなアナグラムか」という問いに対して、ひとつの仮説を提示する。

2 アナグラム生成システム

図1に、今回実装したアナグラム生成システムの概要を示す。生成手順は、おおきく、(1) 候補リスト作成と、(2) ランキングの2つのステップから構成される。

2.1 候補リスト生成

候補リスト生成では、与えられた入力から、アナグラムの候補リストを作成する。入力は、表記(日本語文字列)とその読み(ひらがな文字列¹⁾)の組である。なお、入力それ自身は、生成する候補リストに含められない。

1) 本研究における「ひらがな」は、JISコード(ISO-2022-JP)の0x2421から0x2473までの83文字のうち、小書の「わ」と「ゐ・ゑ」の3文字を除いた80文字である。

2.1.1 読みの並び替え

候補リスト作成の最初のステップは、読み(ひらがな列)の並び替えを求めるステップである。たとえば、入力「名古屋大学 / なごやだいがく」は、読みの長さが7で、すべての文字が異なるので、 $7! = 5,040$ 個の並び替えが得られる。

このステップでは、ひらがな列として適切なもののみ生成する。具体的には、以下の制約を課す。

1. 拗音「ゃ・ゅ・ょ」は、ア行を除くイ段の文字「き・し・ち・に・ひ・み・り・ぎ・じ・ぢ・び・び」の直後の文字としてのみ採用する。
2. 撥音「ん」、促音「っ」、小書き文字「ぁ・ぃ・ぅ・ぇ・ぉ」は、並び替え文字列の先頭の文字として採用しない。

このため、「東京大学 / とうきょうだいがく」の読みは9文字であるが、「う」が2回現れ、「ゃ」は「き」の直後にしか配置できないため、生成される並び替えは9!ではなく、 $8! / 2 = 20,160$ 個となる。

2.1.2 かな漢字変換

次のステップでは、かな漢字変換システムを利用して、得られたそれぞれのひらがな列を日本語文字列(かな漢字文字列)に変換する。かな漢字変換システムには、MeCab[4]とMeCab-skkserv[5]のかな漢字変換辞書を利用した。

一般に、かな漢字変換では、複数の変換候補が考えられるが、ここでは、それぞれのひらがな列に対して、変換結果をひとつだけ生成する。

2.1.3 フィルタリング

候補生成の最後のステップでは、以下に示す2つのフィルタを順に適用し、候補を絞り込む。

読みの不一致フィルタ かな漢字変換によって得られた日本語文字列の標準的な読みが、変換前のひらがな文字列と一致しない場合、読みアナグラムとしての完全性に欠ける。そこで、日本語文字列の読みを形態素解析システム MeCab (IPA 辞書) を用いて取得し、かな漢字変換前のひらがな列と一致しない場合は、その候補を削除する。このフィルタにより、入力「名古屋大学 / なごやだいがく」に対する候補は5,040個から4,184個に絞られる。

文字 n-gram フィルタ 文字 n-gram に基づいて、日本語文字列としての可能性が著しく低いものを候補から削除する。具体的には、国立国語研究所が公

開している日本語の文字 n-gram データ [6] から、文字 2-gram は頻度が101回以上の1,384,419種類、文字 3-gram は頻度が11回以上の39,670,672種類を採用し、このリストに存在しない文字 2-gram および文字 3-gram を含む候補をリストから削除する。このフィルタによって、「名古屋大学 / なごやだいがく」では、4,184個の候補が387個に絞られる。

2.2 ランキング

ランキングでは、候補リスト生成で得られたそれぞれの候補にスコアを付与し、スコア順にソートした候補リストを出力する。

2.2.1 スコアの定義

入力 I に対する候補 C のスコアを、以下の式で定義する。

$$\text{score}(C, I) = \text{GPT-2}(C.h) - \text{penalty}(C, I) \quad (1)$$

ここで、 I と C は、いずれも表記(日本語文字列)と読み(ひらがな文字列)の組であり、 $\text{GPT-2}(C.h)$ は、候補の表記 $C.h$ に対する GPT-2 [7] の尤度を表す。 $\text{penalty}(C, I)$ は、候補と入力の類似性に基づくペナルティを表す。

2.2.2 尤度計算

候補表記の、日本語として妥当性(文法的適格性と意味的整合性)を GPT-2 の尤度によって見積もる。GPT-2 の実装としては、公開されている日本語用の実装 [8] の medium モデルを利用した。

2.2.3 候補と入力の類似性ペナルティ

尤度のみでランキングすると、ほとんどの場合、入力自身が最上位の候補となる。さらに、「名古屋大学」に対して「大学名古屋」や「名古屋大工が」のように入力の一部を色濃く残した候補が上位に来ることが多い。これらの候補の順位を下げるため、入力との類似性に基づくペナルティを導入した。

ペナルティの計算は、次のように行う。ペナルティの値は、候補 C と入力 I の表記の類似度 sim_h と読みの類似度 sim_y の和に10を掛けた値と定義する。

$$\text{penalty}(C, I) = 10 * (\text{sim}_h(C, I) + \text{sim}_y(C, I)) \quad (2)$$

表記の類似度は見た目の類似度、読みの類似度は音の類似度に相当し、見た目と音の両方を考慮することを意図している。

表1 「名古屋大学/なごやだいがく」に対する上位10件

順位	スコア	表記	読み
1	-27.053	長い訳語だ	ながいやくごだ
2	-27.232	訳語がないだ	やくごがないだ
3	-27.538	名護大学や	なごだいがくや
4	-27.857	嫌だな語学	いやだなごがく
5	-28.397	大納屋がごく	だいなやがごく
6	-28.609	嫌な語学だ	いやなごがくだ
7	-31.420	大学名古屋	だいがくなごや
8	-31.811	語学嫌だな	ごがくいやだな
9	-31.878	名子が代役	なごがだいやく
10	-31.945	第五夜がなく	だいがやがなく

2つの文字列 X と Y の類似度は、両者に共通する文字 n -gram の割合として計算する。

$$\text{sim}(n, X, Y) = \frac{\text{dup}(n\text{-gram}(X), n\text{-gram}(Y))}{\min(|n\text{-gram}(X)|, |n\text{-gram}(Y)|)} \quad (3)$$

ここで、 $n\text{-gram}(X)$ は文字列 X に含まれる文字 n -gram のリストを、 $|\cdot|$ はリストの要素数を表す。関数 $\text{dup}(A, B)$ は、2つのリスト A, B に共通して現れる要素の個数を表す。ただし、 A または B に同一要素が複数存在し、かつ、それらが A と B に共通して現れる場合、共通して現れる個数としては、少ない方の個数を採用する。たとえば、 A に「なご」が2回現れ、 B に1回現れる場合は、「なご」が共通して現れる個数としては1を採用する。このような計算により、上記の sim の値域は $[0, 1]$ となる。

表記の類似度は、文字 3-gram と文字 2-gram の類似度の平均値とする。

$$\text{sim}_h(C, I) = \frac{1}{2} (\text{sim}(3, C.h, I.h) + \text{sim}(2, C.h, I.h)) \quad (4)$$

一方、読みの類似度は、文字 4-gram と文字 3-gram の類似度の平均値とする。

$$\text{sim}_y(C, I) = \frac{1}{2} (\text{sim}(4, C.y, I.y) + \text{sim}(3, C.y, I.y)) \quad (5)$$

ここで、 h は表記文字列を、 y は読み文字列を表す。

このように、表記と読みで n の値を変更するのは、一般に、読み文字列の方が長くなるからである。なお、式 (2) の定数 10、式 (4) の定数 3 と 2、式 (5) の定数 4 と 3 は、実際のデータを観察して定めた。

3 出力例

入力「名古屋大学/なごやだいがく」に対する出力の上位10件を、表1に示す。なお、表記と読みに含まれる区切りは、かな漢字変換の際に得られる単語区切りである。出力された候補の中では、「嫌だな語学/いやだなごがく」(4位)、「嫌な語学だ/いやなごがくだ」(6位)、「語学嫌だな/ごがくいやだな」(8位)が面白い。

4 面白いアナグラムとは

機械的に生成したアナグラム候補は、おおきく次のように分類できる。

1. アナグラムとは認められないもの(日本語として不自然なもの)
2. アナグラムと認められるもの(日本語として自然なもの)
 - (a) 特に面白くないアナグラム
 - (b) 面白いアナグラム

今回採用したアナグラム生成アルゴリズムでは、日本語文字列としての可能性が著しく低いものを文字 n -gram フィルターで排除するが、日本語文字列らしさは、GPT-2 の尤度としてスコア化しているだけである。そのため、ランキングの下位には、日本語として不自然なものが多数含まれる。

日本語として自然な語・句・文は、アナグラムとして認定できる。しかしながら、その中で「面白い」と思えるアナグラムは限られている。

面白いアナグラムとはどのようなアナグラムか。これを調査するために、普通名詞 6,750 語に対して、網羅的にアナグラム候補を生成した。入力として採用した語は、MeCab 用の IPA 辞書に収録されている、読みの長さが6文字から9文字の普通名詞全てである。こうして得られた出力を調査(目視)して、著者らが面白いと思うアナグラムを収集し、それらの分類を試みた。以下では、その分類を示し、付録にその実例を示す。なお、これらの分類は排他的分類ではない。複数の分類に属するアナグラムも存在し、それらは面白さの度合いが高い。

4.1 単独で面白いアナグラム

単独で面白いアナグラムとは、アナグラム自身に面白さを感じるものである。言い換えるならば、元となった語(入力)が何だったかに依存しない面白さである。

文形式となるアナグラム 完結した文 [付録 (1)]、あるいは、それに準ずる形式のアナグラムは、面白いと感じることが多い。文は、ある事態を表す形式であり、読み手は、その事態を想起することが容易である。端的に言えば、具体的なイメージが湧きやすい。そのイメージに我々は反応し、面白いと感じるのではないだろうか。

意外性がある複合語となるアナグラム 意外性がある複合語となる場合 [付録 (2)] も面白い。これらの複合語は実際には存在しないが、あってもおかしくないという意味で、我々の好奇心を刺激するのであろう。

ナンセンス系アナグラム 単体で面白いアナグラムの分類のひとつとして、「ナンセンス系」と名付けたアナグラムがある [付録 (3)]。このタイプのアナグラムでは、思いもよらなかった意外な組み合わせに、想像力が刺激されるのだろう。

オノマトペを含むアナグラム オノマトペを含むアナグラム [付録 (4)] が生成されることは比較的少ない。つまり、希少性がある。オノマトペには臨場感があり、イメージを想起させる効果がある。そのような効果も、面白さに寄与しているのだろう。

評価や感情を表す形容詞を含むアナグラム 評価や感情を表す形容詞を含むアナグラム [付録 (5)] は、面白と感じることが多い。これは、面白さの評価が、感覚・感情に基づく評価であることが関係しているのかもしれない。

4.2 入力との関連で面白いアナグラム

入力との関連で面白いアナグラムとは、「この入力からこんなアナグラムが作れるのか」という驚き、あるいは、巧みさに面白さを感じるアナグラムである。

入力と関連するアナグラム アナグラムが入力と何らかの意味で関連する場合 [付録 (6)]、面白と感じることが多い。人間がアナグラムを作る場合、入力と関連するアナグラムを作ることは、かなり難しい。そのため、制作者のスキルの高さ(巧みさ)を評価するのだろう。

入力とアナグラムがひと続きの文となる その特殊な場合として、入力とアナグラムで、ひと続きの文を構成できる場合 [付録 (7)] がある。

ダジャレ系アナグラム ダジャレ系アナグラム [付録 (8)] とは、元の語と音は似ているのに、意味的にはまったく関係ないアナグラムである。そのギャップに面白さを感じる。

入力とギャップがあるアナグラム 入力との意味的ギャップが面白いアナグラムもある [付録 (9)]。何だかよくわからないところが、逆に面白いのだろう。

4.3 面白さを感じる要因

以上、我々が面白いと感じるアナグラムを見てきた。これらを整理すると、我々が面白いと感じるアナグラムは、おおよそ次のように総括できよう。

1. 具体的なイメージを想起しやすいアナグラム
2. 驚きや意外性があるアナグラム、想像力や好奇心を喚起するアナグラム
3. 入力と強く関連するアナグラム、入力との落差がはげしいアナグラム

この総括からわかるように、面白さの要因はひとつではない。いくつかの要因が組み合わせると、面白さが際立ってくる。つまり、面白さを捉えるためには、複数の軸を考える必要がある。

今回得られたアナグラムの中で、最も面白と判断したもののひとつは、次のアナグラムである。

他力本願/たりきほんがん
→ 本気が足りん/ほんきがたりん
【他力本願、本気が足りん】

このアナグラムは、アナグラムが文形式となり、入力とアナグラムがひと続きの意味の通る文となる。さらに、「足りん」と話し言葉的になって臨場感を醸し出し、面白さを際立たせている。

5 おわりに

本稿では、入力に対するアナグラムの候補を順位付きで出力するシステムを示すとともに、システムが生成した候補から「面白いアナグラム」を収集・分類し、面白さを醸し出す要因を分析した。

本研究により「面白いアナグラムを作る」という目標は、かなり達成できたと考える。生成した面白いアナグラムの一部を付録に示したが、これ以外にも数百個の面白いアナグラムが得られている。次の目標は、「面白いアナグラムを選ぶ」能力の実現であるが、本稿で示した「面白いアナグラム」の整理と分類は、その足掛かりになろう。

謝辞

本稿の内容は、名大 MIRAI GSC (<http://www.iar.nagoya-u.ac.jp/miraigsc/>) の第 2 ステージで行った研究が元となっている。

参考文献

- [1] 新村出（編）. 広辞苑 第七版. 岩波書店, 2018.
- [2] 鈴木啓輔, 佐藤理史, 駒谷和範. 文節データベースを用いた日本語アナグラムの自動生成. 第 10 回情報科学技術フォーラム (FIT-2011), RF-009, 第 2 分冊, pp. 97–102, 2011.
- [3] 鈴木啓輔, 佐藤理史, 駒谷和範. アナグラム生成における文節列の意味的適格性の判定法の検討. 言語処理学会第 18 回年次大会発表論文集, pp. 1308–1311, 2012.
- [4] MeCab: Yet Another Part-of-Speech and Morphological Analyzer, (2021-12 閲覧). <https://taku910.github.io/mecab/>.
- [5] mecab-skkserv, (2021-12 閲覧). <http://chasen.org/%7Etaku/software/mecab-skkserv/>.
- [6] GSK2020-C 「国語研日本語ウェブコーパス」 n-gram データ・頻度表, (2021-12 閲覧). <https://www.gsk.or.jp/catalog/gsk2020-c/>.
- [7] openai/gpt-2, (2021-12 閲覧). <https://github.com/openai/gpt-2>.
- [8] gpt2-japanese, (2021-12 閲覧). <https://github.com/tanreinama/gpt2-japanese>.

付録 面白いアナグラムの具体例

文形式となるアナグラム

- (1) a. 頭でっかち / あたまでっかち
→ ちまたで悪化 / ちまたであっか
- b. 功労賞 / こうろうしょう
→ 売ろう コショウ / うろう こしょう
- c. 生殺与奪 / せいさつよだつ
→ いつ左折だよ / いつさせつだよ
- d. 連尺商い / れんじゃくあきない
→ あれ禁句じゃない / あれきんくじゃない
- e. セラミックエンジン / せらみっくえんじん
→ えっ自民落選 / えっじみんらくせん

意外性がある複合語となるアナグラム

- (2) a. サウンドトラック / さうんどらっく
→ 皿うどん 特区 / さらうどん とっく
- b. 緞帳役者 / どんちょうやくしゃ
→ やんちゃ 食堂 / やんちゃしょくどう
- c. 相対性理論 / そうたいせいりろん
→ 総理 生態論 / そうりせいいたいろん

ナンセンス系アナグラム

- (3) a. 一日千秋 / いちじつせんしゅう
→ セイウチ 出陣 / せいうちしゅつじん
- b. 般若心経 / ほんにゃんしんぎょう
→ 社運は人魚 / しゃうんはにんぎょ
- c. 蜀江の錦 / しょっこうのにしき
→ キノコに失笑 / きのこにしっしょう
- d. 蝸牛角上 / かぎゅうかくじょう
→ 食うか 授業か / くうかじゅぎょうか
- e. 反対称律 / はんたいしゅうりつ
→ 両親はタイツ / りょうしんはたいつ

オノマトペを含むアナグラム

- (4) a. 一宿一飯 / いっしゅくいっばん
→ 行く一瞬パツ / いくいっしゅんぱつ
- b. 一昨昨週 / いっさくさくしゅう
→ サクサク一周 / さくさくいっしゅう
- c. 胸突き八丁 / むなつきはっちょう
→ 千夏は今日ムッ / ちなつはきょうむつ
- d. ニッポンチャレンジ / にっぽんちゃれんじ
→ レンジにちゃっぼん / れんじにちゃっぼん
- e. 似た者夫婦 / にたものふうふ
→ うふふのモニタ / うふふのみにた

評価や感情を表す形容詞を含むアナグラム

- (5) a. 第一人称 / だいいちにんしょう
→ 超いい男子に / ちょういいだんしに
- b. 骨粗鬆症 / こつそしょうしょう
→ 少々そこつ / しょうしょうそこつ
- c. ジャイアントパンダ / じゃいあんたぼんだ
→ ジャパンだと安易 / じゃぼんだとあんい
- d. 風林火山 / ふうりんかざん
→ うんざり花粉 / うんざりかふん
- e. 乾坤一擲 / けんこんいつてき
→ インコって危険 / いんこってきけん

入力と関連するアナグラム

- (6) a. 総合大学 / そうごうだいがく
→ そういう 語学だ / そういうごがくだ
(「大学」と「語学」)
- b. 知識工学 / ちしきこうがく
→ 講師が鬼畜 / こうしがきちく
(「知識工学」と「講師」)
- c. 最年少 / さいねんしょう
→ うん最初ね / うんさいしょね
(「最年少」と「最初」)
- d. 名題役者 / なだいやくしゃ
→ 医者役だな / いしゃやくだな
(「役者」と「医者役」)
- e. 方角違い / ほうがちがひ
→ 北緯が違ふ / ぼくいがちがう
(「方角」と「北緯」)

入力とアナグラムがひと続きの文となる

- (7) a. 一日中 / いちにちじゅう
→ 知事に注意 / ちじにちゅうい
【一日中、知事に注意】
- b. 百科事典 / ひゃっかじてん
→ 光ってんじゃ / ひかってんじゃ
【百科事典、光ってんじゃ】
- c. 相思相愛 / そうしそうあい
→ 試合早々 / しあいそうそう
【試合早々、相思相愛】
- d. 健康上 / けんこうじょう
→ 好条件 / こうじょうけん
【健康上、好条件】
- e. 器械体操 / きかいたいそう
→ 競う大会 / きそうたいかい
【器械体操、競う大会】

ダジャレ系アナグラム

- (8) a. 香典返し / こうでんがえし
→ 公園で餓死 / こうえんでがし
- b. 亜成層圏 / あせいそうけん
→ 麻生政権 / あそうせいけん
- c. 山河襟帯 / さんかきんたい
→ サンタ解禁 / さんたかいきん
- d. 猿の腰掛 / さるのこしかけ
→ 酒の コルシカ / さけのこるしか
- e. 交感神経 / こうかんしんけい
→ 更新関係 / こうしんかんけい

入力とギャップがあるアナグラム

- (9) a. 天地神明 / てんちしんめい
→ メンチ進呈 / めんちしんてい
- b. 全知全能 / ぜんちぜんろう
→ 膳所の運賃 / ぜぜのうんちん
- c. 柴枯盛衰 / さいこせいすい
→ せこい水泳 / せこいすいえい
- d. 猿の腰掛 / さるのこしかけ
→ 今朝の コルシカ / けさのこるしか
- e. 酒池肉林 / しゅちにくりん
→ チリに君主 / ちりにくんしゅ