

形態素解析手法と通俗的単語群に基づく 類音文変換システム

金久保 正明^{1,a)}

受付日 2012年11月1日, 採録日 2013年4月5日

概要: 近年, いわゆる「ことば工学」の一環としてシステムに駄洒落を自動生成させる試みが盛んになり, 駄洒落に基づくなぞなぞの生成等, 様々な応用のほか, 面白い駄洒落を自動的に絞り込む方法等も模索されている. 一方, 形態素解析では入力文に対して複数の形態素解析結果が得られることがあり, このとき意図されていなかった解析結果は駄洒落として見なすことができる. また, 駄洒落と見なせる複数の文の間では, 共通する読みの平仮名文字列における双方の単語区切り位置が食い違う場合もあり, これらは意外性の高い駄洒落になると思われる. そこで, 本論文では形態素解析を類音に拡張し, 文節レベルの変換が可能なシステムを提案する. 駄洒落の面白さを高めるため, 通俗的な名詞群を基本体言とし, 他の単語もそれらと接続する可能性の高いもので揃えた. 区切り位置の異なる文が生成されやすくするため, 短い単語を増やし, 特に文頭, 文末に来る読み1字の単語を多く登録するような工夫をしている. 詳細な試験評価により, 一定レベルの意外性, 面白さを有する変換文の生成が確認された.

キーワード: 言語感覚, 言葉遊び, 駄洒落, 慣用句, 形態素解析

Similar Sounds Sentences Generator Based on Morphological Analysis Manner and Low Class Words

MASAAKI KANAKUBO^{1,a)}

Received: November 1, 2012, Accepted: April 5, 2013

Abstract: Two different sentences which resemble each other in pronunciation are humorous. Japanese language morphological analysis systems can make two different sentences which have pronunciations which are in step with each other. This paper proposes the generation system of sentences which sound like entered sentences by morphological analysis manner. The proposed system has something low class word database, to generate funny sentences. This paper shows subjects experiments in order to confirm the validity of the proposed system.

Keywords: language sense, wordplay, pun, idiomatic phrase, morphological analysis

1. はじめに

近年, いわゆる「ことば工学」の研究がさかんになっている [1], [2], [3]. 様々な「言葉遊び」等がシステムで自動生成されているが, なかでも読みの一致または類似する2語を探せば生成できる駄洒落は自動生成に適していると考

えられ, 駄洒落に基づくなぞなぞ生成等の様々な応用システムも開発されている [4], [5], [6]. なぞなぞ以外にも, 駄洒落は「北海道, でっかいどう」等のキャッチコピーに使われる場合も多い. また, 駄洒落等の音の類似は誤聴をもたらすが, 誤聴は「空耳アワー」[7]等のテレビ番組コーナーになったり, 一般からの誤聴の投稿で構成する人気サイト [8]等もある. 駄洒落やそれに基づく誤聴は, 多くの人気を集めうる面白さがあるため番組やCMに使われると考えられ, 単なる言葉遊びの枠を超えて経済的な価値を有し

¹ 静岡理科大学
Sizuoka Institute of Science and Technology, Fukuroi,
Shizuoka 437-8555, Japan

a) kanakubo@cs.sist.ac.jp

ているといっても過言ではない。システムによる自動生成の研究も意味のあるものと考えられる。

一方で、形態素解析では入力文に対して複数の形態素解析結果が得られることがあり、このとき意図されていなかった解析結果は駄洒落と見なすことができる。入力文字変換を使った駄洒落の作り方を解説しているサイト [9] も存在する。また、形態素解析により発生する駄洒落は、複数の句にまたがる長いものが生成可能で、たとえば「裏庭には鶏→裏庭に埴輪取り」のように、単語間の区切りの位置が食い違う場合も多い。これは単なる単語の置き換えによる駄洒落よりは人間にとっては発想しにくいと考えられ、その分、意外性に基づく面白さを誘発する可能性が高いと思われる。また、単語区切りが変わることで、意味的なずれを生じる効果も期待できる。駄洒落の面白さの要因を様々に分析した文献 [10] でも、品詞間の区切り位置の食い違いについては指摘されていない。しかし、形態素解析により発生する駄洒落は、当然、読みが完全に一致するものに限定される。

一般に、面白いとされる駄洒落や誤聴には類音によるものも多く、完全一致よりはるかに多様な駄洒落を形成しうる。そこで本論文では、形態素解析の方法を用いて、類音による生成も可能なシステムを提案する。提案システムの入力は、ある文の読みとしての「かな文字列」で、その文の単語区切り位置の情報も入力され、完全一致または類音の読みを持つ違う文章を可能な限り出力する。生成される文は、システムに登録された単語から構成され、決められた品詞間接続規則に基づいて生成されていく。生成された様々な文に対し、単語区切り位置の違い具合、音の一致度、印象的な単語を含むか、の3項目から評価し、得点の高いものから表示する。一般的に駄洒落の面白さの要因として、音的なずれと意味的なずれの2つがあげられるが、提案システムでは、類音により音的なずれを、形態素解析結果の単語区切りが変わることで意味的なずれを表現しているといえる。

提案システムでは、単語間の区切り位置の食い違いが多く発生するように比較的短い言葉のみを登録している。また、生成文をできるだけ面白いものにするため、「性・食・物欲・遊び」等の人間の俗的、動物側面に関係する体言を集め、基本体言とした。また、他の単語も基本体言を中心に固め、体言は基本体言と接続可能な単語群に絞り、用言や副詞等も既存の言語資源を利用して、基本体言と接続する可能性の高いものに絞る等して違和感の少ない文生成を試みている。また、「あっ」等の文頭に来る語、終助詞等の文末に来る語で、印象を強める働きがあり、かつ読みが平仮名1字のものを多く登録し、柔軟な文単位の類音変換が発生しやすいように工夫する等している。

以降、2章で提案システムの概要を述べる。3章では、生成された文章の数や発音の類似度、チューリングテスト、

感性的な評価の結果を述べる。4章で結論と今後の課題についてまとめる。

2. 提案システムの概要

2.1 提案システムの処理の流れ

図1に、提案システムの処理の流れを示す。平仮名の入力文字列の先頭から1~5文字を取り、類音変換し、その類音を読みにつなぐ単語のうち、前の単語の品詞に接続可能な品詞のものを追加する。この処理を通常の形態素解析と同様に繰り返す。処理可能な文字列がなくなったら終了、得られた文字列の得点順に出力する。

2.2 単語の選定基準

「笑い」の本質を考察したベルグソンの著書 [11] では、「人間の中に動物的・機械的なものを見出した時」に笑いが生ずると指摘されている。また北垣 [12], [13] は、通俗的概念が笑いの源泉であるとし、具体的に「排泄・性・食・支配・金銭等」の項目をあげている。これはベルグソンも指摘する人間の動物側面にほかならない。提案システムでは、出力文をなるべく面白くするため、また生成の容易さと多様性の確保のため、単語データベース（通俗的単語群）を以下の方針に従って作成した。

- 人間の動物側面に関係の深い体言（以下、基本体言）を集める。
- イメージを描きやすくするため、基本体言は具体物に限定する。
- 他の体言、用言、副詞等も基本体言と共に起しうるものに限定する。

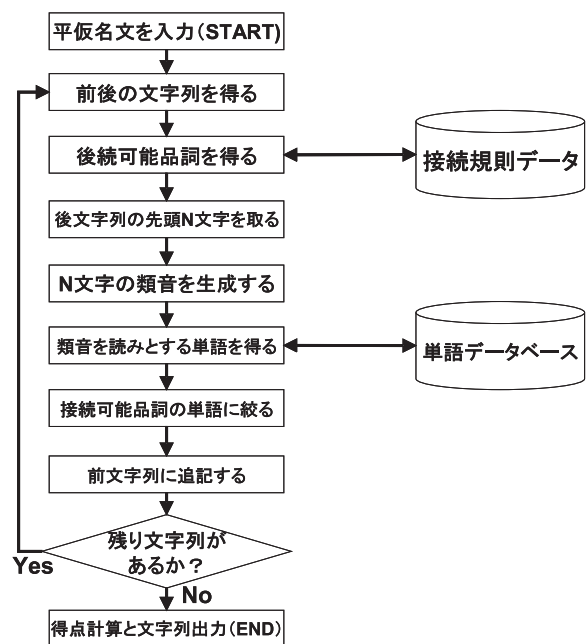


図1 提案システムの処理の流れ

Fig. 1 Flowchart of the proposed system.

表 1 体言の選定基準

Table 1 Criterion of selecting nouns.

体言グループ	Wikipedia のカテゴリ	総取得単語数	具体例
性別を表す人間	「男性」「女性」	142	少年, 長女
服装	「服飾」「衣類」	127	背広, 浴衣
動物	「哺乳類」	66	猫, ラクダ
食物	「食品」	589	寿司, キャベツ
贅沢品	「装身具」「室内装飾」	83	勲章, ダイヤ
玩具	「玩具」	63	ビー玉, めいぐるみ
職業で表す人間	「職業一覧」	228	教師, パイロット
体の一部	「髪型」「顔」「化粧」	175	顔, 虫歯
色彩	「色名一覧」	141	青, 黄色
日用品	「日用品」「家具」	304	コップ, 布団
建物	「建築物」	175	劇場, 自宅
屋外の名称	「景観」「地形」	167	ビーチ, 牧場
地名	「国名一覧」	261	日本, 静岡
時間の名称	「暦」「記念日」「年中行事」	150	水曜, 昨夜

- 変換を細かく多様化するため, 読みの短い単語で揃える (原則, 平仮名 4 文字以下).

以下, 品詞別に具体的な選定基準を述べる.

2.2.1 体言の選定基準

国研教育基本語彙 [14] (6,104 語) および Wikipedia 日本語版 [15] の見出し語から, 条件に合う単語を手で収集した. Wikipedia は, 辞書等よりも一般的でくれた若者言葉, 比較的新しい単語等も幅広く網羅されているためである. まず, 体言のグループを決め, Wikipedia では対応するカテゴリおよび下位カテゴリの見出し語から収集した. これらを表 1 に示す.

上記の動物のカテゴリの中で, 「排泄」関連は単語数が少ないうえ, どう考えても下品な出力結果になるため除外した. また「性」についても露骨な表現は猥褻となるため使用できない. 性的イメージを持つ単語として, 「性別で表した人間」および「服装」を設定した. 「支配・金銭」に関係の深い具体物として, 権威や物欲, 虚栄心と結び付く「贅沢な品物」をあげた. これは, 世界高級品協会選定のブランドトップ 100 [16] から, 著名なブランド名も追加登録した. 「玩具」は, 動物的な欲求が多い子供を象徴する概念として追加した. さらに, 動物自体も単語群に追加した. ベルグソンの指摘で「動物に面白さを感じるのは擬人化による」があるため, 比較的人間と似た形状の哺乳類に限定した.

基本体言とは別に, 基本体言と直接接続, あるいは「の」「と」等の助詞を挟んで共起可能な体言 (以下, 一般体言) を設定し, 同様に収集した. このうち「体の一部」は, 眼鏡等の顔等に付着する物も含めている. 「建物」は人間が, 「屋外」は人間または哺乳類が存在しうる場所とした. 地名は国名のほか, 都道府県名, 県庁所在地名等とした. 「時間の名称」は, 朝・先週等の時間帯を示す名詞である.

2.2.2 用言の選定基準

動詞・形容詞・形容動詞 (用言) は, 京都大学の黒橋・

表 2 用言収集に用いた体言

Table 2 Nouns used for selecting declinable words.

基本体言グループ	用言収集に用いた検索用体言	用言の例
性別を表す人間	男性・女性	来る, 思う
服装	服	着る, 選ぶ
動物	動物	座る, 鳴く
食物	食べ物	食べる, 好む
贅沢品	ブランド・宝石・宝物 指輪・ネックレス	高い, 良い
玩具	おもちゃ	買う, 触る

河原研究室が提供する京都大学格フレーム [17] から収集した. 用言は活用語尾が固定的なため, 長い音の固定を抑えるため, 語幹の読みが平仮名 2 文字以内の短いものに限定した. 同研究室の格フレーム検索 [18] の「名詞から検索」を利用し, 表 2 に示す基本体言グループごとの上位概念で検索し, 頻度 20 以上を持つ用言を接続に使われる格助詞とともに登録した.

用言の総数は 721 語となった. 「人間」カテゴリでは「男性」「女性」の双方の検索で, 意外と出現する用言の差はほとんどなかった. 「人間が」「人間に」等の格関係に統一して該当する用言を登録した. これらの検索語は試行錯誤で比較的输出結果が多いことを確認してから用いている. また, 「こわい→こええ」といった形容詞のくだけた読みは, 別の単語として登録した. なお, 各用言について, 収集した基本体言カテゴリとの接続関係を確認し, 「が・に・を・で」の主要格助詞について, 接続関係が不足していると判断された 140 の用言に対し, 計 268 の接続関係を人手で追加した.

2.2.3 副詞・他の品詞の選定基準

京大格フレーム検索では, 用言終止形を入力して検索すると, 「修飾」という見出しで共起する副詞が表示される. その中から頻度 100 以上のものを, その用言に接続可能な副詞として登録した. ただし, 頻度 100 以上のものがない場合は, 最も高い頻度の副詞を 1 つだけ登録した. 「全然」

「大して」のようにほとんど否定助動詞としか共起しないものは音型が固定するため除外した。「簡単だ」のように、形容動詞語幹+断定助動詞の形で出力されるものは、「簡単に」のように連用形に直して副詞として登録した。副詞の総数は、445語である。

連体詞は国研基本語彙から収集、それぞれ基本体言の各グループに接続しうるかを人手で決め、21語を登録した。用言活用語尾（五十音各行の五段活用語尾，上一段および下一段活用語尾，カ変およびサ変活用語尾）および助動詞は、中学生用の国文法解説サイト [19] にあげられているものを登録した。助詞は表 3 に示すもののみとした。類音変換を作りやすくする観点から 1 文字のもので揃え、2 文字のものは印象を強める強調的な意味を持つものだけとした。強調でも「すら」「さえ」「しか」は否定助動詞と共起するケースが多いため除外した。

話言葉では「えー」等が文頭に、「な」等の終助詞（変形も含む）が文末によくあり、これらは、入力文字列を左右に少しスライドし、多様な類音変換を可能にするため、多く登録した。「で,」「でも」のような接続詞も、前の文が書

かれていない以上、順接・逆接の矛盾は生じないため、文頭語として登録した。国研基本語彙の接続詞、感動詞のほか、平仮名 2 字の組合せすべてについて該当するかを人手で検討した。登録数は文頭 117 語、文末 71 語である。

2.3 品詞間の接続規則

出力する類音変換後の文は、図 2 に示す品詞間の接続規則に基づくものに限定した。括弧はなくても良い箇所、+ は接続、・はいずれかが付きうるものを示す。「体言+(間の助詞)+体言」の体言は、最低 1 つが基本体言であることを条件とした。格助詞等は「服を着る→服着る」のように省略しても意味が通ずる場合が多く、省略可とした。

2.3.1 複雑な文型の排除

用言は体言と接続する可能性のみで選ばれているため、用言連体形+体言+用言の形は、用言どうしの組合せに違和感を生ずるおそれがある。体言+助詞+体言+用言の形も、用言が先の体言と接続する可能性があるか分からない。体言+助詞+体言+助詞+体言等も、最初と最後の体言の接続する可能性が分からないため、これらの文型は除外している。助動詞どうしの接続も音型が固定するため除外した。大幅に文型を制限したが、これでも平仮名 20 文字程度の入力文には対応でき、十分な長さと思われる。

2.3.2 体言どうしの接続規則

表 4 および表 5 に、提案システムで用いた体言間の接続規則を示す。縦に並ぶカテゴリの単語が左に来て、横に

表 3 登録した助詞
Table 3 Selected particles.

助詞の位置による種類	登録した助詞
体言と用言の間 (格助詞等)	が・は・を・に・へ・で・と・も
体言と用言の間 (強調的)	こそ・まで・だけ
体言と体言の間	の・と・か・や・より・が・は

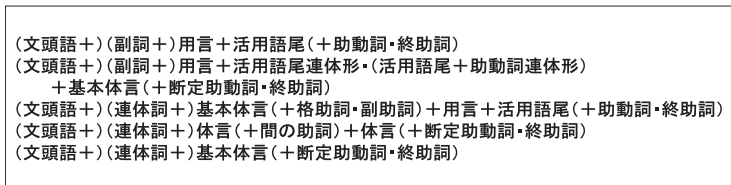


図 2 品詞間接続規則

Fig. 2 Juncture rules of the proposed system.

表 4 体言+基本体言の接続規則

Table 4 Juncture rules from nouns to fundamental nouns.

	人間	服装	動物	食物	贅沢品	玩具
人間	とやよりか	との	と	との	との	との
服装	との+	とやよりか	×	×	と	×
動物	と	×	とやよりか	×	×	×
食物	と	×	×	とやよりか	×	×
贅沢品	と	と	×	×	とやよりか	と
玩具	と	×	×	×	と	とやよりか
職業	とのやよりか	との	と	との	との	との
体一部	×	×	×	×	×	×
色彩	×	の+	×	×	×	の
日用品	と	と	×	×	×	×
建物	の	×	×	×	×	×
屋外	の	×	の	×	×	×
地名	の+	×	×	の+	×	×
時間	×	の	×	の	×	×

表 5 基本体言+一般体言の接続規則

Table 5 Juncture rules from fundamental nouns to the other nouns.

	職業	体一部	色彩	日用品	建物	屋外	地名	時間
人間	とやよりかはが	の	×	との	×	×	×	×
服装	との	×	はが	と	×	×	×	×
動物	と	の	×	×	×	×	×	×
食物	と	×	×	×	×	×	×	×
贅沢品	と	×	×	×	×	×	×	×
玩具	と	×	はが	×	×	×	×	×

並ぶカテゴリの単語が右に来る場合に、間に入りうる助詞をつなげて示したものである。+は間に助詞を挟まない直接接続の可能性を示す。×はいずれの接続も不可であることを示す。

並置を示す「と・や」、比較・選択を示す「より・か」は、同じカテゴリの基本体言どうし、「人間」と「職業」の間には使えるものとした。また並置の「と」は、具体物である体言と使用者になりうる「人間・職業」の間に入りうるものとした。他に括弧内の共通点により、並置に違和感が少ないと考えられる「服装」と「贅沢品」(美しさ)、「贅沢品」と「玩具」(遊びの要素)、「服装」と「日用品」(生活用品)の間も可能とした。

助詞「の」は所有の意味では、「人間・職業」+の+「人間と動物以外の基本体言と日用品」、「人間・動物」+の+「体一部」は違和感の少ない組合せとして可能とした。「～である (be 動詞的)」の意味では、「職業・服装」+の+「人間」、「服装」+の+「職業」、「色彩」+の+「人工物である基本体言」を可能とした。「～にある」の意味では、「建物」+の+「人間」、「屋外」+の+「人間・動物」、「地名」+の+「人間・食物」を可能とした。また、「朝の食事」等の使い方から、「時間」+の+「服装・食物」は可能とした。

主格を示す「は・が」は、「人間」+は(が)+「職業」、「服装・玩具」+は(が)+「色彩」を認めた。体言間の直接接続は、結び付きが強いだけに下手な組合せは違和感が大きいおそれもある。服装+人間, 色彩+服装, 地名+人間, 地名+食物のみを可能とした。それぞれ、「ジーパン刑事」「赤服」「パリ娘」「宇都宮餃子」等の言い回しはよく見かけるためである。

これら体言間の接続規則は集合どうしの規則なので、個々のケースでは違和感が残ることもありうる。生成された類音変換の評価において、違和感(意外性)についても検証することとした。

2.4 類音変換の定義

まったく同じ発音による言い換えのほか、類音による言い換えを行う。類音とは似た発音で、「か→た」のような同じ母音(以下、同母音)を持つ文字等が代表的である。田辺 [20] は、駄洒落を構成しうる類音として同母音の他、同

表 6 類音種別ごとの類似度

Table 6 Assigned degree of similarity of each sound-alike types.

類音種別	ランク	類似する単語例
近い同母音	A	かた がた
他の同母音	B	さぬき たぬき
同行音	C	かち けち
きょ等→近母音	A	きょう ぎょう
きょ等→他母音	C	そうとう そうぞう
い・う→ん (逆も)	B	こいん こんい
同字・あ行→音引き (逆も)	A	はあと は一と
同字省略	B	いいえ いえ
近母音省略	C	つづき つき
他母音省略	C	たかい たい
どえ→で等	C	しずおか しぞか
い・う・ん・ー・っ省略	B	まっと まと
同字挿入	C	かお かかお
近母音挿入	C	さえ さざえ
他母音挿入	C	はと はあと
い・う・ん・ー・っ挿入	B	まと まんと
で→どえ等	C	です どえす

行音(し→す, のような子音共通), 濁音(と→ど, 等), 追加(つぶら→つぶらや等の同母音等の1字追加), 交換(連接する2文字の入れ換え, さらわれる→さわられる等)をあげている。また、「テーブル起こし」を業とする方の作成したサイトに「聞き間違いをしやすい言葉一覧」[21]があり、非常に豊富な例があげられているため、その登場頻度等も考慮した。これらを参考に、提案システムでは、表 6 に示すものを類音とした。また、類音の中でも違和感の程度に差があることを考慮し3段階に分類した(A, B, Cの順に類似度が高い)。ただし、この分類の妥当性を検証するため、被験者試験において、生成された類音文字列が確かに元の文字列と類似する発音を持つと思うかを尋ねることとした。

なお、ここで「きょ等」とあるのは、小さい「ゃ・ゆ・よ」等からなる2文字で、母音は小さい字に従う(きょ, の場合は「お」)。同母音の変換対象には、これらの文字も含めている。近い同母音とは、「か」と「が」のように、濁音・半濁音のあるなしのみが違いの場合であり、その他の同母音とは区別した。同字・あ行→音引き, は同じ字が2つ続いた場合, ある字とその母音(あ行)が続いた場合に後の文字を音引き(ー)に変換することを示す。同字・近母音・

他母音省略は、同様に同じ字や近い母音、他の母音が連続した場合に後の字を省略することを指す。挿入はある字の後に同じ字等を入れることを指す。なお、これらの小さい平仮名、「い・う・ん・ー」を「省略向き文字」として、単語選定の際の文字数上限（4文字）はこれらを含む場合は5文字とした。

田辺論文にあげられているもの以外は、「聞き間違いをしやすい言葉一覧」で登場頻度が高かったものである。また、くだけた表現として「です→どえす」のような言い回しも追加した。これは de → doe のように、ローマ字表記すると1字挿入（逆は省略）に該当し、類音の一種と考えられる。「交換」は田辺論文であげられた例は、同母音の変換であり、そうでない入れ換えは違和感が高く、聞き間違い例のサイトでも見られなかったため除外した。

2.5 類音文字列への変換処理

単語ごとの区切り情報が付された入力文字列が、平仮名表記で与えられる。これを受けて、提案システムは以下の処理を行う。

- (1) 入力文字列の残りの先頭から1~5文字を順に取り、
- (2) 以下の処理を行う。入力文字列の残りが4文字以下である場合は、その長さまでの処理を行う。入力文字列が空になったら、探索終了となる。
- (2) 切り出した文字列が1文字の場合は、読みが完全に一致する読み1文字の単語のみを接続候補とする。2文字以上の場合、類音変換の定義であげた方法に従い、同文字数・省略・挿入に基づくあらゆる類音変換を行い、その読みを持つ単語はすべて接続候補とする。他に読みが完全に一致する単語も候補とする。なお、違和感を抑えるため、図3に示すように、切り出した文字列の中で変換処理を行うのは1カ所に限定した。挿入処理は、切り出した文字列が5文字の場合に行わない。図3の灰色は変換・挿入箇所、省略処理の黒は2文字から生成された1文字を示す。これらの探索で、接続候補の単語がなかった場合は、(1)に戻る。
- (3) (2)で得られた接続候補のうち、品詞接続規則に合致する単語群のすべてについて、それぞれ変換文字列の後に付け、切り出した文字列を切った入力文字列を生成して(1)に戻す。品詞接続規則によって、合致する単語がなかった場合は変換文字列の生成を行わずに(1)に戻す。

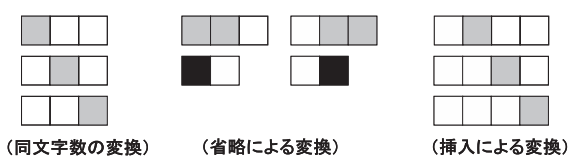


図3 類音文字列への変換位置

Fig. 3 Transformation position for similar sounds.

- (4) (1)~(3)の処理を再帰的に繰り返し、すべての枝についての処理が終われば全体の探索を終了する。ただし、文の途中で終わっているものは、生成文字列として扱わない。

- (5) 終了後、各生成類音文字列に対し、後述する得点を算出し、得点順に表示する。

2.6 変換文字列の得点

変換文字列に得点を付け、できばえをある程度は自動評価することを試みた。得点の要素として、類音変換の類似度、入力文字列と比較した単語の切り位置の違い、基本体言その他の印象を強める単語の数、の3要素の重み付き線形和で評価する。

2.6.1 類音変換箇所の類似度

平仮名の入力文字列と変換文字列の読みを、先頭から1字ずつ比較する。同音、類音の種類ごとにそれぞれ与えた得点に従って、得点の総和を求め、総和を文字列長で割った値を類似度の得点とする。また、類音が連続すると違和感が高まるため、1字の場合よりも低い得点とする（具体的な得点パラメータは3章で述べる）。

2.6.2 単語の区切り位置の違いによる得点

たとえば、「生木を裂く→寝巻きを裂く」といった類音変換の場合、先頭の単語が「なまき」→「ねまき」と変化しただけであり、比較的連想が容易な駄洒落であると考えられる。しかし、「真っ赤+な+嘘」→「あっ+カワウソ」という変換では、単語区切りの位置が変わっている。生成された単語が元の文章では埋もれている状態で、連想は比較的難しくなると考えられる。さらに、意味的に大きく変える効果があると考えられるため、提案システムでは、このような変換に高い得点を与えることとした。図4の左側は区切り位置が食い違う場合、右側は一致している場合である。新たな文字列を生成する過程で、元の文字列の入力平仮名列が、どこで区切られたかを記録しておく。仮に、元の文字列を縮めた類音変換が行われ、3文字が2文字になった場合（「カカオ」→「顔」等）でも、元の文字列の方で3文字の位置に生成された方の区切り位置を記録する。生成後、入力文字列と生成文字列の双方から相手も同じ位置で区切られているかを見て、異なっていれば、得点を1ポイント付ける（左の例では3点となる）。

提案システムで、変換の対象となる入力文字列は、日本

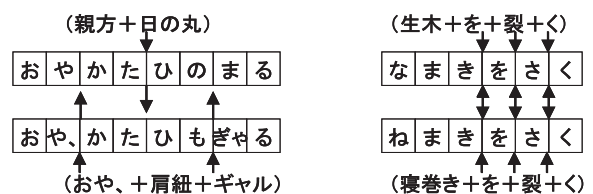


図4 単語区切り位置の違い

Fig. 4 The difference of compartmental locations of words.

表 7 提案システムに登録した強調的な副詞

Table 7 Emphatic adverbs stored in the proposed system.

いっぱい・ほとんど・必ず・何時も・ずっと・大変
断然・やたら・兎に角・特に・結構・あまり
かなり・ドンドン・急に・随分・一気に・益々
なかなか・グッと・隅々・皆・まるで・一斉に・常々
全然・全く・大分・こんなに・絶対・きっと・常に
多々・しばしば・度々・山程・非常に・もちろん
何時でも・とても・大いに・とっても・減法・一切
あんまり・よっぽど・何より・余程・何とも・いやに
実に・余りに・一番・全部・皆に・最も・遥かに
たっぷり・やけに・グングン・一層・一度に・すっぱり
思いっきり・たらふく・ギッシリ・ズラリ・俄かに
一躍・ビッシリ・グンと・続々・完全に・メッキリ
ガンガン・すこぶる・すっかり・ガラッと・見る見る
大勢・ガブガブ・忽ち・活発に・どうしても
どっぷり・思いきり・盛んに・猛烈に・みだりに
バンバン・バシバシ・ボンボン・豊に

表 8 体言・副詞以外の強調語

Table 8 Other emphatic words.

品詞	単語群
助動詞	う・よう・まい (意志), たい・たがる (希望)
副助詞	こそ・だけ
文頭語 (あ系)	あ, ・あつ・あー・ああ, ・あら, ・あれ,
文頭語 (う系)	う, ・うっ・うむ, ・うん,
文頭語 (え系)	え, ・えっ・えー・ええ,
文頭語 (お系)	お, ・おっ・おー・おお, ・おや,
文頭語 (へほ系)	へー・へえ, ・ほー・ほう, ・ほお,
文頭語 (ま系)	まっ・まー・まあ,
終助詞	ぞ・ぜ・ね・わ・よ・さ・な
「ぞ」の派生	ぞっ・ぞー・ぞう・ぞお
「ぜ」の派生	ぜっ・ぜー・ぜえ
「ね」の派生	ねっ・ねー・ねえ・よね・のね
「わ」の派生	わっ・わー・わあ・わね・わよ
「よ」の派生	よっ・よー・よお・のよ
「さ」の派生	さっ・さー・さあ・のさ・ってさ
「な」の派生	なっ・なー・なあ・のさ・ってな
他の文末語	のみ・のだ・べき・べし・とも

語の平仮名で表記できるものであれば何でもよい。ただし、前段で述べた「単語の区切り位置の違いによる得点」を算出するため、入力文字列と同時に、その単語ごとの区切りも入力する必要がある。たとえば、「まっかなうそ」と入力するだけでなく、「まっか*な*うそ」等と、ユーザは区切り情報を入力する必要がある。

2.6.3 印象を強める単語による得点

提案システムでは、面白さに結び付く体言グループとして基本体言を選んでいるので、変換文字列に含まれる数を得点とした。一般体言でも「体の一部」は動物的・性的なイメージもあるので得点とした。このほか、何らかの「強調」的な意味を持つ語は、変換文字列の印象を高めると考えられるため得点に加えた。具体的には、「意志」「希望」を示す助動詞、「こそ・だけ」等の限定（取立て）を示す副助詞、「ぞ・よ」等の強い気持ちを示す終助詞、「おお」等の強い気持ちを示す文頭の語、程度副詞等に得点を与え

た。体言以外で得点を与えられた強調語をすべて表 7 および表 8 に示す。

程度副詞には「必ず」「急に」のように、確実性・変化の速度を示すものも含まれている。また、文頭語の「あ、」等は表記上は読点が入っているが、読みは「あ」等と読点のない状態で登録されている。

3. 試験評価

提案システムの有効性を確かめる試験評価では、入力文字列のジャンルとして、慣用句を選定した。平仮名にした慣用句と、単語の切り位置を入力し、システムにより生成された類音文字列の評価を行った。評価は、システム生成の文字列について、システム自体により高得点の付いたもの、システムでは比較的低得点だったが、人手で選別したもの、また人間が作成したものについて、各 20 ずつ比較した。

3.1 類音文字列の評価関数

2.5 節で述べた評価の 3 要素の重み付けは以下で述べるようにした。印象を強める単語の数は、出現個数を 4.55 で割った値を得点とした。4.55 は、今回の実験に用いた入力文字列の平均単語数である。単語の切り位置の違いは 1 カ所 0.25 点として、出現数に応じて加算する。これは、各単語数 -1 の平均が切り位置の数であることから、2 つの文字列の全部が食い違った場合におおよそ 2 点となるように設定したものである（切り位置の違いを重視して他のパラメータより重みを倍とした）。類音度の評価は、文字ごとに類音度に応じて加算し、最後に文字数で割って平均をとった。具体的には、完全一致は 1 点、2 章の表 6 に示す A ランクは 0 点、B ランクは -1 点、C ランクは -2 点とした。また類音が 2 文字続いた際に、さらに 1 点を減点した。類音度が低いと急激に違和感が高まることを考慮して、大幅に減点することとした。

3.2 入力文字列のジャンル

入力文字列のジャンルとして、試験評価では「慣用句」を選んだ。たとえば固有名詞や外来語等では意外性のある変換結果が得られるかもしれない。しかし、一般的な単語の連続とは合わないことが危惧された。その点、慣用句は、体の一部を示す語等の比較的簡単な単語で構成されている場合が多く、1 文の長さも 10 文字以内の短いものが多いことから、提案システムによる変換が容易であると考えられたためである。

また、慣用句は国語の問題にもなる等、知っていることが教養であるとして、権威付けられている側面がある。それをまったく異なる言葉で言い換えることによって、権威を裏返す面白さが期待できると思われた。具体的には、市販の辞典 [22] に収められている 1,563 の慣用句のすべてを

表 9 システム選出上位 20 の類音文字列

Table 9 Best 20 similar sounds sentences chosen by the proposed system.

元の慣用句 (読み)	類音文字列	得点 (×10)
それ見た事か (それみたことか)	それに蛸とか	19.34
真っ赤な嘘 (まっかなうそ)	まっカワウソ	18.56
おくびにも出さない (おくびにもださない)	おう, ピキニもダサイ	18.01
雨後の筍 (うごのたけのこ)	う, この筍	17.97
親方日の丸 (おやかたひのまる)	おや, 肩ひもギャル	17.84
十把一からげ (じゅっぱひとからげ)	うっパイと唐揚げ	17.7
ウンともスンとも (うんともすんとも)	うん, 通すとも	17.2
足が速い (あしがはやい)	あ, 滋賀野菜	16.9
芋の子を洗うよう (いものこをあらうよう)	干物こそ有るよう	16.62
あの手この手 (あの手この手)	あの猫の手	16.06
こと此処に至って (ことここにいたって)	コートこそ似たって	15.9
足が出る (あしがでる)	あ, 鹿です	15.9
心が通う (こころがかよう)	このロバかヒョウ	15.82
あつと言う間に (あつというまに)	あ, 問う兄	15.82
思いも寄らない (おもいもよらない)	お, 桃より梨	15.65
無い袖は振れない (ないそではふれない)	はい, 袖は紅	15.25
痛し痒し (いたしかゆし)	イタチか牛	15.23
お目に掛る (おめにかかる)	お, 飯か粥	15.23
今か今かと (いまかいまかと)	居まいかと	15.2
蜘蛛の子を散らすよう (くものこをちらすよう)	獣こそ焦らすぞう	15.09

対象とした。

3.3 慣用句の類音文字列生成率

提案システムは、平仮名の入力文字列の先頭から、類音の範囲で一致する読みを持つ単語で置き換えていくため、登録単語数が増えれば、文の最後までの変換成功率（生成率）も高まると考えられる。しかし、慣用句によっては、元々変換不可能なものがあるかもしれない。そこで、類音文字列が存在することが明らかな慣用句群を定義し、それに対する提案システムの生成率を調べた。まず、1,563の慣用句の単語に基づく読みの区切り位置情報を用いて、平仮名2文字以上（1文字では助詞が多くなるため）で出現頻度の高い読みの区切りを抽出した。多い順から「はな」「むね」「はら」「くち」等、上位20の読みを持つ305の慣用句を抽出した。あらかじめ「はな→花, 鼻, etc.」等の各読みの類音から得られる単語をピックアップしておき、各慣用句が類音文字列を持つかを検討したところ、251の慣用句が該当した。このうち、提案システムの登録単語により類音文字列が生成できた慣用句は101（40.2%）となった。

3.4 生成された類音文字列

1,563のうち476の慣用句が、文末まで類音変換ができた。それ以外は途中で接続可能な該当単語がなくなり、失敗している。生成率は約30.5%となった。前節の40.2%より少ないのは、元々類音変換不可能な慣用句も含まれているためと考えられる。また、生成された類音文字列の総数は3,119で、生成された慣用句1つあたり、平均6.55の文字列が生成されたことになる。表9に、システムの評価に

基づき、高得点を得た上位20までを示す。ただし、同じ慣用句に対する変換グループからは、最高点だけを選出してある。

また、これとは別に、システムの評価ではあまり上位にいかなかったが、この中では比較的よくできたと思われる20の類音文字列を人手で選んだ結果を表10に示す。得点は、システムによる評価である。

3.5 被験者による評価

被験者は、大学2~4年生30人である。前節にあげた計40の類音文字列と人間作成の20の類音文字列を混ぜ、被験者ごとにランダムな順番で印刷したアンケート用紙（1人あたりは半数の30問とした）を配り、以下の項目について、4. そう思う, 3. ややそう思う, 2. あまりそう思わない, 1. そう思わない, の4段階評価で答えてもらった。制限時間は特に設けなかった。また、被験者には人間が作ったものが3分の1混ざっていると伝え、それぞれの類音文字列について、システムと人間のいずれが作ったと思うかを問うチューリングテストも実施した。

- ユーモアがある。
- 楽しさがある。
- (総合的に) 面白さがある。
- 意外性がある。
- 確かに発音が似ている。

表11に、混ぜ込んだ人間作成の類音文字列を示す。これらは、誤聴の投稿で構成する有名なサイト[8]等から、慣用句に相当するものを選んだものである。

表 10 人手で選出した 20 の類音文字列

Table 10 Best 20 similar sounds sentences chosen by human.

元の慣用句 (読み)	類音文字列	得点 (×10)
芋の子を洗うよう (いものこをあらうよう)	着物こそ洗うよう	15.4
真っ赤な嘘 (まっかなうそ)	あ, カワウソ	15.23
顔に書いてある (かおにかいてある)	なお, 二階である	13.75
烈火の如く (れっかのごとく)	ねっ彼女とか	12.2
下にも置かない (したにもおかない)	下着も置かない	12.2
愛想を尽かす (あいそうをつかす)	あ, 衣装を溶かす	11.9
お株を奪う (おかぶをうばう)	おかずを奪う	11.84
看板が泣く (かんばんがなく)	カバが鳴く	11.48
難色を示す (なんしょくをしめす)	南極女子です	10.95
心が躍る (こころがおどる)	ところが踊る	9.64
影を落とす (かげをおとす)	金を落とす	8.86
対岸の火事 (たいがんのかじ)	台湾の菓子	7.91
事なきを得る (ことなきをえる)	この柿を得る	7.55
拍車が掛かる (はくしゃがかかる)	学者かギャル	6.36
見当を付ける (けんとうをつける)	弁当を作る	5.95
膝を交える (ひざをまじえる)	ビザを混ぜる	5.53
業を煮やす (ごうをにやす)	ゾウを逃がす	5.53
自棄のやんばち (やけのやんばち)	夜明けのカンパチ	5.05
押しの一手 (おしのいって)	落ちないって	5.0
勿怪の幸い (もっけのさいわい)	おっ, 貶さない	5.0

表 11 人間が作成した類音文字列

Table 11 Human-made similar sounds sentences.

元の慣用句 (読み)	類音文字列
神のみぞ知る (かみのみぞしる)	蟹の味噌汁
豚に真珠 (ぶたにしんじゆ)	豚と心中
九死に一生を得る (きゅうしにいっしょうをえる)	急死に一生終える
押っ取り刀で来る (おっとりがたなでくる)	夫にガタが来る
夢多かりし頃 (ゆめおおかりしころ)	夢を借りし頃
灯台下暗し (とうだいまくらし)	大正デモクラシー
場末の酒場 (ばすえのさかば)	松江の酒場
若気の至り (わかげのいたり)	わき毛の痛み
善は急げ (ぜんはいそげ)	電話, 急げ
石橋を叩いて渡る (いしばしをたたいてわたる)	石橋を叩いて割る
正直に言う (しょうじきにいう)	掃除機に言う
私的な会話 (してきなかいわ)	素敵な会話
腐っても鯛 (くさってもたい)	クワって重たい
憎まれっ子世にはばかる (にくまれっこよにはばかる)	肉コロッケ菌に挟まる
輪になって踊る (わになっておどる)	ワニになって踊る
雨降って地固まる (あめふってじかたまる)	豆食って痔止まる
寝込みに来る (ねこみにくる)	猫見に来る
人っ子一人いない (ひとっこひとりいない)	一人っ子一人以内
ほとぼりが冷める (ほとぼりがさめる)	外堀から攻める
我が意を得たり (わがいをえたり)	童は見たり

3.6 チューリングテストの結果

システム生成の類音文字列に対し、システム生成と正解した率は全体で 56.5%だった。一方、人間作成の類音文字列に対する正解率は 65%となった。システム生成のものは 50%台で、人間作成と見分けがつきにくかったと考えられる。人間作成のものを 65%の確率で正解した被験者達が見分けられなかったということから、提案システムでは明らかにシステム生成らしいぎこちない作品はあまり生成されなかったものと考えられる。

3.7 読みの類似度の調査結果

システムが生成した計 40 の類音文字列について、被験者が「確かに発音が似ている」の設問に対してそう思うか否かを 4 段階評価した結果を集計した。肯定的評価 (4. 3.) の数と、否定的評価 (2. 1.) の数の総計をそれぞれ比較した。ただし、4. と 1. はそれぞれ強い評価なので数を 2 倍にした。また、得点の総和を標本数とする母比率の検定 (Z 検定) も行った。肯定的評価の回答総数は 549、否定的評価は 242 で、Z 値は 10.9 となった。Z 値は標準正規分布に従うため、危険率 1%でも十分に有意差がある。提案シ

表 12 感性評価の比較

Table 12 Comparison of the “kansei” estimations.

	ユーモア			楽しさ			(総合的な)面白さ			意外性		
	肯定数	否定数	Z 値	肯定数	否定数	Z 値	肯定数	否定数	Z 値	肯定数	否定数	Z 値
1 群	222	140	4.31	199	170	1.51	204	168	1.87	248	136	5.72
2 群	262	112	7.76	238	130	5.63	246	134	5.75	247	130	6.03
3 群	296	117	8.80	271	116	7.88	279	119	8.02	243	155	4.41

表 13 Z 値が高かった類音文字列

Table 13 Similar sounds sentences with high Z score.

類音文字列	評価項目	肯定スコア	否定スコア	Z 値
足が速い→あ, 滋賀野菜 (1 群)	ユーモア	20	1	4.15
親方日の丸→おや, 肩ひもギャル (1 群)	ユーモア	22	0	4.69
(同)	楽しさ	21	0	4.58
(同)	意外性	20	1	4.15
拍車が掛かる→学者かギャル (2 群)	ユーモア	20	1	4.15
(同)	(総合的)面白さ	21	1	4.26
お株を奪う→おかずを奪う (2 群)	ユーモア	18	1	3.9
(同)	楽しさ	21	0	4.58
灯台下暗し→大正デモクラシー (3 群)	ユーモア	23	1	4.49
(同)	楽しさ	18	1	3.9
(同)	(総合的)面白さ	22	1	4.38
九死に一生を得る→急死に一生終える (3 群)	ユーモア	20	0	4.47
雨降って地固まる→豆食って痔止まる (3 群)	ユーモア	23	0	4.8
(同)	楽しさ	23	0	4.8
(同)	(総合的)面白さ	22	0	4.69
豚に真珠→豚と心中 (3 群)	ユーモア	23	0	4.8
(同)	楽しさ	20	1	4.15
(同)	(総合的)面白さ	22	0	4.69
正直に言う→掃除機に言う (3 群)	ユーモア	21	2	3.96
輪になって踊る→ワニになって踊る (3 群)	ユーモア	21	1	4.26
(同)	楽しさ	19	0	4.36
(同)	(総合的)面白さ	19	1	4.02
神のみぞ知る→蟹の味噌汁 (3 群)	ユーモア	24	0	4.9
(同)	楽しさ	24	0	4.9
(同)	(総合的)面白さ	23	0	4.8
押っ取り刀で来る→夫にガタが来る (3 群)	楽しさ	18	2	3.58
(同)	(総合的)面白さ	18	2	3.58
憎まれっ子世にはびこる→肉コロッケ菌に挟まる (3 群)	ユーモア	20	3	3.54
(同)	楽しさ	22	1	4.38

システムは、発音が似ていると感じられる類音文字列を生成できたといえる。

3.8 感性評価の比較

システムが生成しシステム自体が選出した類音文字列 (1 群), システムが生成し人手で選出した類音文字列 (2 群), 人間が制作した類音文字列 (3 群) の種類別に, 5 項目の感性評価項目について 4 段階評価の数の総和を求め, 前節と同様に Z 検定を行った結果を表 12 に示す。各項目のカンマで区切った数字は左から, 肯定的評価の総数, 否定的評価の総数, Z 値を示している。

意外性を除くすべての項目において, 人間制作, システム生成人手選別, システム生成システム選別の順で成績が良い傾向が見られる。しかし, システム生成システム選別でも, 「ユーモアがある」「意外性がある」では危険率 1%で

の有意差を検出している。システム生成人手選別では, すべての項目で危険率 1%での有意差が検出された。全体的に, システム生成システム選別で評価が比較的良かった理由として, これらの類音文字列の中での相対評価になってしまったことも 1 つの要因と考えられる。

表 13 に, 個別の類音文字列の各評価項目において, Z 値が比較的高かった (3.5 以上) のものをまとめた。

やはり人間制作 (3 群) のものが多いが, システム生成 (1 群, 2 群) もそれに匹敵する高評価を得たものもあったことが分かる。

以上のことから, システム生成の類音文字列においても, 「ユーモアがある」「意外性がある」等の評価において, 半数以上の被験者から肯定的評価が得られることが確認された。また, なかには, 人間制作のものに匹敵する高評価を得られるものもあることが明らかになった。

表 14 SD 法による評価を行った類音文字列

Table 14 Similar sounds sentences estimated by SD method.

A 群	B 群	C 群
灯台下暗し→大正デモクラシー	愛想を尽かす→あ、衣装を溶かす	心が通う→ところが火曜
九死に一生を得る→急死に一生終える	油が乗る→あ、豚が乗る	出来ない相談→で、嫌いそう
雨降って地固まる→豆食って痔止まる	それ見た事か→それに蛸とか	肌に合う→ただ、似合う
正直に言う→掃除機に言う	真っ赤な嘘→あっカワウソ	名を上げる→なお、揚げる
輪になって踊る→ワニになって踊る	ウンともスンとも→うん、通すとも	手も足も出ない→でも、蟻も出ない
神のみぞ知る→蟹の味噌汁	雨後の筍→う、この筍	申し分が無い→そう、新聞が無い
押っ取り刀で来る→夫にガタが来る	あの手この手→あの猫の手	肩の荷を降ろす→また、糊を降ろす
憎まれっ子世にはびこる→肉コロッケ歯に挟まる	下にも置かない→下着も置かない	対岸の火災→海岸の蚊かい？
D 群	E 群	
顎が落ちる→アナゴが落ちる	隔世の感→学生の缶	
虫の息→牛の息	当を得る→蝶を得る	
至難の業→次男の痣	琴線に触れる→便箋に触れる	
箔を付ける→白衣を着ける	見当を付ける→電燈を着ける	
一言も無い→毒も無い	腰が据わる→講師が座る	
対岸の火事→台湾の菓子	拍車が掛かる→作者が躍る	
埒もない→ランチも無い	足を運ぶ→和紙を運ぶ	
バトンを渡す→ヴィトンを渡す	幕が下りる→客が下りる	

人間制作の類音文字列が、「蟹の味噌汁」「夫にガタが来る」等、日常生活において十分にありうることを述べているのに対し、システム生成の類音文字列は「南極女子」「夜明けのカンパチ」等、なぜこのような組合せとなるのか不思議に感じられるものも出力されている。それぞれ、「南極に興味を持つ女子」「夜明けに食べる寿司」等の解釈も可能であるが、唐突な印象は否めない。品詞間接続規則で、可能なものに絞るようにしたため、解釈できない異常な結合は見られなかったが、やはり多少の不自然さは残る結果となった。このような違いが、意外性ではシステムの方が高かったが、ユーモア・面白さで人間制作に及ばなかった原因の1つであると考えられる。

提案システムで採用した単語グループ間の接合規則以上の細かい意味による接続規則を取り込み、上記のような違和感の解消を図ること、さらに単語のカテゴリ、属する単語数を増やして多様な類音文字列の生成を可能にすること、等が課題となる。

3.9 SD 法による提案手法の有効性の検証

提案システムでは、生成された類音変換文に通俗的単語群や強調語（以下、基本体言等）が含まれていること、および形態素解析結果の単語区切りをまたぐ変換になっていることに得点を与えていることが特徴である。この有効性を検証するために、基本体言等を含まない類音変換文、単語区切りのまたぎがない類音変換文との印象の比較を行った。人間制作（A 群）、基本体言等を含み単語区切りのまたぎがあるシステム生成（B 群）、単語区切りのまたぎはあるが基本体言等がないシステム生成（C 群）、基本体言等はあるがまたぎがないシステム生成（D 群）、双方ともないシステム生成（E 群）に分け、表 14 に示す各 8 個の変換例についての印象評価を行った。

これらの変換例は、人間作成は表 13 にあげられたものを選び、システム生成は類音度による影響を受けないように、類音度得点が比較的高く生成数も多かった類音度 6.5～8.5 点の範囲から、総合得点の高い順に選んだ。なお、基本体言等を含まない一般的な単語のみで構成される変換文字列の生成のため、「性別で示す人間→職業で示す人間」「哺乳類→鳥類・魚類・昆虫」「贅沢品・玩具→家電・事務用品」に同数ずつ置き換えた単語データベースを用いた。

10 人の大学生を被験者として、SD 法（意味微分法）による印象評価を行った。7 段階の評価で、左側の形容詞への得点を 6～0 点とし、類音文字列の同じグループに対するすべての評価（8 文字列×10 人＝80 評価）の平均値を結んだグラフを図 5 に示す。10 種類の評価尺度は、「親しみがある」「面白い」「楽しい」「ユーモラス」「勢いがある」等は、基本単語等の効果測定用に設定したものである。また「変な」「分かり難い」「元と遠い」「意外な」「発音違う」等の項目は、単語区切りのまたぎの効果測定のために設定した。

5 群の変換文字列グループのすべての 2 つの群の組について、尺度ごとに平均値の差の検定（t 検定）を行った。以下、両側検定で危険率 1% 以下の有意差が出た組をあげる。「親しみ」では、A 群と B 群がそれぞれ C、E 群より高いとの結果になった（基本体言等を含む D 群とは差がつかなかった）。「面白さ」でも、A 群と B 群がそれぞれ C、D、E 群よりも高かった。しかし、「ユーモラス」では、A 群が C、D、E 群に差をつけたのに対し、B 群は有意差が出なかった（危険率 5% 以下では、C 群に対してのみ有意差が得られた）。A 群（人間制作）と B 群（システム生成で基本体言等と区切りのまたぎを有する）は、これらの項目で他より有意に高い結果が出た（A 群と B 群間では有意差は出なかった）。しかし、「ユーモラス」では人間制作の方

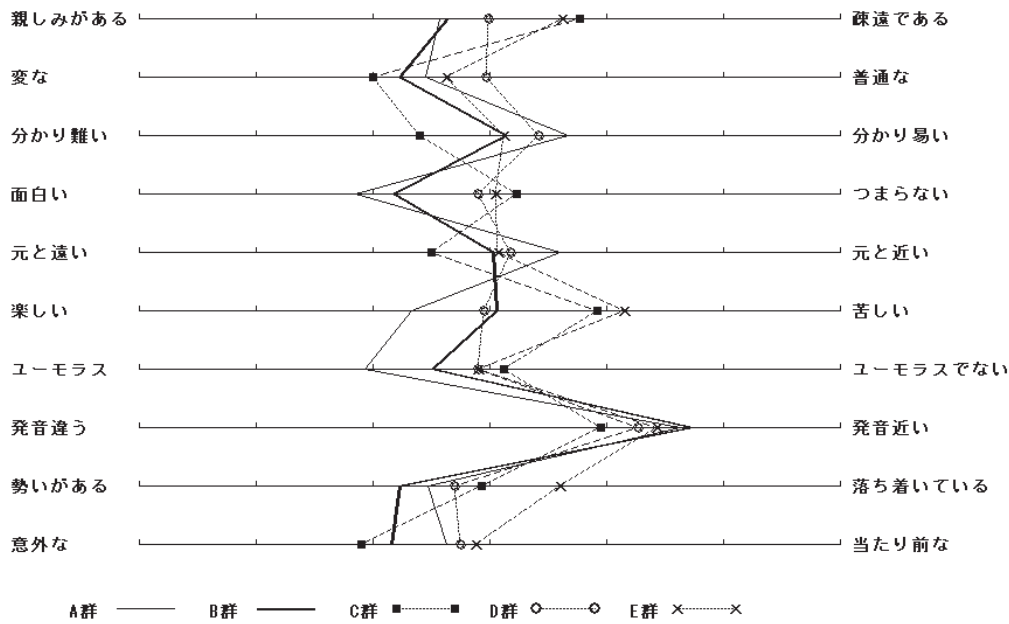


図 5 SD 法による評価結果

Fig. 5 Comparison of the SD method estimations.

表 15 類音による生成文の増加

Table 15 Increase in the number of generated sentences by similar sounds.

類音のグループ	生成文の数	全体の割合 (%)
完全一致	5	0.0016
同母音	1,010	32.4
同行音	1,654	53.0
い・う→ん (逆も)+同字・あ行→音引き (逆も)	1,681	53.9
同字の省略と挿入	1,697	54.4
同母音の省略と挿入	2,685	86.1
い・う・ん・ーっの省略と挿入	3,042	97.5
で→どえ等 (逆も)	3,119	100.0

がシステム生成を明らかに上回っていた。

C 群 (単語またぎはあるが基本体言等はない) に着目すると、「分かり難い」で A 群と D 群に対し有意に高かった (5%以下ではすべての群に有意差がある)。「意外な」で A, D, E 群に、「変な」で D 群より有意に高かった。単語のまたぎを有する B 群も、「意外な」では E 群に対し (5%以下では C 群にも) 有意差をつけた。単語のまたぎは、分かり難い違和感に加え、特に意外性の原因となっていることを示唆する結果となった。

D 群 (基本体言等はあるが単語またぎはない) については、「楽しい」で C, E 群より有意に高かった。「勢いがある」で E 群より有意に高かった。基本体言等を有さない C, E 群よりは「楽しさ」で上回ったが、「面白さ」「ユーモラス」等では差がつかず、基本体言等単独ではこれらを得るのは難しく、単語またぎと両方有する B 群で初めて「面白さ」について効果が得られていたことが分かった。「勢い」で E 群に対し有意差があったのは、基本体言等に「お」等の強調語が含まれているためと考えられる。

以上の結果から、基本体言等と単語またぎの両方を有す

る群では「面白さ」を得ることができ、基本体言等単独でも「楽しさ」は得られること、単語またぎにより「意外性」を得られること等の可能性が示されたといえる。

なお、因子分析では、「親しみ」「面白い」「楽しい」「ユーモア」で因子負荷量の絶対値 0.8 を超えた第 1 因子、「変な」「分かり難い」で因子負荷量の絶対値 0.6 を超えた第 2 因子が抽出された (両方で累積寄与率は 69.7%)。前者は「面白さ・楽しさ」を、後者は「難解・違和感」を表す因子と考えられる。

3.10 類音による変換可能性の向上度

提案システムでは、類音の部分が多いほど変換文の得点は下げているが、類音変換の導入により、変換可能性を高めている。様々な類音により、どの程度生成変換文が増えたかを検証した。類音をタイプの似ているグループにまとめ、1,563 の慣用句から生成された 3,119 の類音文字列が、どのように増えたかを表 15 にまとめた。

完全一致で生成された変換文は 5 つしかなかった。表の数は、ある類音グループを追加して全体でいくつになった

かを示している。同母音への変換（「きょ」等の変換も含む）を追加すると、全体の3割を超え、同行音への変換の追加で半数を超えている。同字の省略・挿入等はあまり多くなく、同母音の省略と挿入で再び大幅に増えている。「い・う・ん・ー・っの省略と挿入」も比較的多かった。同母音をはじめとする類音の追加により、完全一致のみより飛躍的に変換可能性が高まったことが分かる。

3.11 提案システムの応用に関する考察

駄洒落はCMのキャッチコピー、商品名、雑誌記事のタイトル等、意外と多く使われている。たとえば商品名では「ナイシトル（小林製薬。内脂肪をとる、からの駄洒落）」「アラウノ（パナソニック。便器）」「茶ップリン（駿河屋忠兵衛。お茶のプリン）」等があげられ、CMコピーでも「汗を掻いたらアセロラレモン（飲料）」「北海道、でっかいどう（旅行）」等、数多く見ることができる。また、たとえばシェイクスピアのような高度な文学作品でも駄洒落は多用されていることが知られている。

提案システムは慣用句に限らず、任意の平仮名の文字列をすべて対象とできるため、たとえば商品やサービスに関連する文字列の類音変換の結果から、上記のようなCMコピーや商品名を考えることが可能となる。このような発想支援システムとしてのアプリケーションがまず考えられる。また、関連付けたい複数の単語（「茶」と「プリン」等）を入力すると、それらを可能なすべての順列で並べ替えた文字列を作った後、類音変換処理を行う等の前処理の追加で、さらに幅広い変換を可能にすると思われる。駄洒落は日記等の文章に彩を添える効果もあるため、駄洒落化した文を入力する文章作成支援システムとしてのアプリケーションも考えられる。

4. おわりに

平仮名の入力文を形態素解析と同様に単語群に変換するが、その過程で「誤聴を楽しむ」といった言葉遊びになりうる類音変換を行うシステムを開発し、その効果を確認した。登録単語群には、人間の俗的・動物的側面に関係の深い、性別・衣服・食物・装飾品等に係る体言を基本に据え、他の品詞群もそれらと接続可能な単語群に限定し、印象の強い感動詞等も増やして面白さを高める効果を狙っている。類音文字列が柔軟に生成されやすくするため、文頭・文末に来る単語等も多く用意し、比較的短い単語のみを登録した。また、類音の種類も、同母音・同行音に限らず、特定の文字の省略や挿入、置き換え等を入れて多彩になるよう工夫した。生成された類音文字列は、人間には一般に発想しにくいと思われる元の文との単語区切り位置の違いを重視したシステムによる評価がされ、意外性を持つ変換結果を選別できるようになっている。

市販の辞典に収められた1,563の慣用句を入力文字列と

した実験では、約30%の類音変換成功率となり、システムが上位に選出した20の類音変換例、システムが生成し人手で選んだ20の類音変換例に対する被験者の感性評価では、有意にユーモア、面白さ、意外性を備える類音変換文字列も多数得られたことが確認された。また、人間が作成した慣用句の類音変換例と比較しても同等の評価を得るシステム生成例もあり、人間作品を混ぜたチューリングテストにおいても、区別のつかない例が多かったと結論することができた。提案システムが生成する類音文字列は、元の入力文と比較して「確かに発音が似ている」と感じられるものが多かったことも確認された。SD法による印象評価でも、提案手法の特徴である基本単語群は「楽しさ」を、単語区切り位置の違いは「意外性」を、両方あわせ持つ場合は「面白さ」を有意に高めることが確認された。

今後の課題としては、提案システムで採用した単語グループ間の接続規則以上の細かい意味による接続規則を取り込み、残ってしまう違和感の解消を図ること、さらに単語のカテゴリ、属する単語数を増やして多様な類音文字列の生成を可能にすること、等があげられる。

参考文献

- [1] 松澤和光, 堀 浩一, 金杉友子, 阿部明典: ことば工学入門, 人工知能学会誌, Vol.15, No.3, pp.446-455 (2000).
- [2] 金杉友子, 松澤和光, 笠原 要: アバウト推論の「言葉遊び」への適用, 信学技報, NLC96-31 (1996).
- [3] 滝澤 修, 柳田益造, 伊藤 昭, 井佐原均: 日本語修辞表現の工学的解析—駄洒落・アイロニー・トートロジー, 信学技報, TL97-2 (1997).
- [4] 松澤和光, 金杉友子, 阿部明典: コンピュータ上の言語感覚実現に向けて: B 級機関, 人工知能学会全国大会論文集, Vol.12, pp.685-686 (1998).
- [5] キム ビンステッド, 滝澤 修: 日本語駄洒落なぞなぞ生成システム“BOKE”, 人工知能学会誌, Vol.13, No.6, pp.920-927 (1998).
- [6] ヨーナス・シューベルグ, 荒木健治: 日本語を対象とした謎掛けの自動生成, 情報処理学会研究報告, 2007-NL-178, pp.91-95 (2007).
- [7] タモリ倶楽部, 入手先
<<http://www.tv-asahi.co.jp/tamoriclub/index.html>>.
- [8] 誤聴の部屋, 入手先
<<http://www.sutv.zaq.ne.jp/shirokuma/gocho.html>>.
- [9] 駄洒落の作り方講座, 入手先
<<http://ameblo.jp/totherma/entry-10344318836.html>>.
- [10] 滝澤 修, 柳田益造: 駄洒落処理の工学的実現について, 人工知能学会言語・音声理解と対話処理研究会, SIG-SLUD-9202-5, pp.37-46 (1992).
- [11] ベルグソン (著), 林 達夫 (訳): 笑い, 岩波文庫 (1976).
- [12] 北垣郁雄: おかしの誘発にかかわるファジィ論的定式化と通俗性/劣弱性, 日本ファジィ学会誌, Vol.2, No.1, pp.100-104 (1990).
- [13] 北垣郁雄: おかしの誘発モデルに基づいた「笑うコンピュータ」の開発事例, 日本知能情報ファジィ学会誌, Vol.15, No.5, pp.577-583 (2003).
- [14] 国立国語研究所 (編): 教育基本語彙の基本的研究—教育基本語彙データベースの作成, 明治書院 (2001).
- [15] Wikipedia 日本語版, 入手先
<<http://ja.wikipedia.org/wiki/>>.

- [16] 世界高級品協会ブランドトップ 100, 入手先
(<http://prw.kyodonews.jp/opn/release/201201201897/>).
- [17] 河原大輔, 黒橋禎夫: 高性能計算環境を用いた Web から
の大規模格フレーム構築, 情報処理学会自然言語処理研
究会, 171-12, pp.67-73 (2006).
- [18] 京都大学格フレーム検索,
入手先 (<http://reed.kuee.kyoto-u.ac.jp/cf-search/>).
- [19] High School 国語 文法, 入手先
(<http://www.hello-school.net/harojapa000top.htm>).
- [20] 田辺公一朗: 駄洒落のコンピュータによる処理—駄洒落
生成システムの基本設計, 産能大学紀要, Vol.26, No.1,
pp.65-74 (2005).
- [21] 聞き間違いをしやすい言葉一覧,
入手先 ([http://www006.upp.so-net.ne.jp/
okosoken/magirawashii.html](http://www006.upp.so-net.ne.jp/okosoken/magirawashii.html)).
- [22] 米川明彦, 大谷伊都子 (編): 日本語慣用句辞典, 東京堂
出版 (2005).



金久保 正明 (正会員)

平成 15 年慶應義塾大学大学院後期博士課程修了。平成 16 年より東京工科大学コンピュータサイエンス学部助手。平成 18 年より静岡理工科大学理工学部情報システム学科講師。平成 22 年より同大学総合情報学部人間情報デザイン学科准教授。ことば工学, 発想支援システムに関する研究に従事。博士 (工学)。平成 16 年感性工学会論文賞。電気学会, 日本知能情報ファジィ学会各会員。