

## 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

## Pun Processing in a computer A Primary Design for a Pun Generator

田辺 公一郎  
Koichiro Tanabe

#### Abstract

Most of Japanese puns are made from a sentence by substituting one new word for a word in it. In many cases, several words are found as a substitute for one. In this paper, an idea to select one word from others is proposed. Based on it, a primary design for a pun generator is proposed.

#### 1. はじめに

ユーモアは、人間の知的活動の一つである。このため、人工知能の分野においては、ユーモアの生成・理解、および、その計算機による処理に関して、多くの研究が行われている。ユーモアの一つに駄洒落がある。駄洒落は、「類似した発音をもつ複数の語を組み合わせて曖昧性をもたせることによって、ユーモラスな効果を与える言語表現」と大まかに定義される [ビンステッド98]。そして、人工知能学会のことは工学研究会 [須永00] においても、感性・創造性といった、従来型の自然言語処理とは異なるスタンスからの研究対象の一つに挙げられており [金杉99B], [松澤00]。B級機関と呼ばれる駄洒落生成システムも開発されている [金杉99A]。

[ピンステッド98]では、駄洒落に関して次のように述べている：

駄洒落は、言語構造が比較的単純・明解で、その理解・生成において世界知識が概ね不要と考えられ、定式化がしやすいため、コンピュータで扱うことが比較的容易な対象と考えられる。また、豊富なサンプルがあり、文学等の分野での考察がある程度なされていることから、コンピュータ処理の結果に対する評価がやりやすい特徴もある。

[田辺03]における分析でも、駄洒落の類似した発音をもつ語の組合せにはいくつかの明確なパターンが存在する。そこで、本論文では、駄洒落のコンピュータによる処理を目標とし、駄洒落生成システムの基本設計を行う。以下、2.で駄洒落生成のパターンを示し、3.で駄洒落生成システムの基本的な機能と処理の概要について述べる。

## 2. 駄洒落生成のパターン

駄洒落の生成については、ある単語を、発音が同じである別の単語、或は、発音が類似している別の単語で置き換える、という手法が提案されている [ピンステッド98]、B級機関でも同様の手法が採用されている [金杉99A] [榎本99] の駄洒落について分析を行った結果、やはり同様の手法によるものが大多数であった [田辺03]。本章では、その概略を示す。

### 同音異義語による置換

ある言葉を同一の発音の別の言葉で置き換えることで駄洒落を生成する。

《例》 見よ、事務で鍛えたこの体

これは、「ジム (gym)」を「事務」で置き換えている。

### 類似音語による置換

ある言葉を音が類似した別の言葉で置き換えることで駄洒落を生成する。これには次の種類がある。

#### (2・1) 同母音

ある言葉を、その読みの1音が異なる行の同じ母音の音となっている言葉で置き換えることで駄洒落を生成する。

《例》 今アンカーにタヌキが渡されました

これは、「タスキ」を「タヌキ」で置き換えている。音に注目すると、「す」が「ぬ」に変わっている。五十音図では、「す」はさ行で「う」を母音とし、「ぬ」はな行で「う」を母音とする。すなわち、いずれも、行は異なるが母音は同じである。

(2・2) 同行音

ある言葉を、その中の1語が同じ行の前後の1語となっている言葉で置き換えることで駄洒落を生成する。

《例》 俺の料理にケツつける気か

これは、「けち」を「けつ」で置き換えている。音に注目すると、五十音図では、「つ」はた行の「ち」の次の語である。

(2・3) 濁音

ある言葉を、その読みの1音が濁音あるいは平音となっている言葉で置き換えることで駄洒落を生成する。

《例》 お願い、本堂のこと教えて 寺の秘密

これは「ほんとう」を「ほんどう」で置き換えている。音に注目すると、「と」が「ど」に変わっている。

(2・4) 追加

ある言葉を、その読みに1語が追加された読みの言葉で置き換えることで駄洒落を生成する。

《例》 円谷な瞳 ウルトラマン

これは、「つぶら」を「つぶらや」、すなわち、元の単語に1文字追加された言葉で置き換えている。

(2・5) 交換

ある言葉を、その読みの中の2語を入れ替えた読みとなっている言葉で置き換えることで駄洒落を生成する。

《例》 波にさわられる

これは、「さわられる」を「らわが」で置き換えている。読みに注目すると、「らわ」が「わら」に変わっている。すなわち、読みの2語が交換されている。

(2・6) 発音

ある英単語をその発音(読み)に相当する日本語で置き換えることによって駄洒落を生成する。

《例》 トム? そうや 関西育ちの外人

これは、「ソーヤ」を「そうや」で置き換えている。

(2・7) 組み合わせ

(2・1) から (2・7) を組み合わせで駄洒落を生成する。

《例》 古き、よき、死体 ミイラ

ズラかぶってますけど、まあちょっと上がって下さいな

前者は「じだい」を「したい」で置き換えていて、2語が平音になっている。これは、(2・3)の組み合わせである。また、後者は「ちらかって」を「ずらかぶって」で置き換えていて、(2・2)(2・1)(2・3)(2・4)の組み合わせである。

以上が駄洒落の主なパターンである。[榎本99]の駄洒落では、この他に、ある言葉中の文字の形が類似した別の文字による置換(例 しばらくお持ち下さい)文節の変更(例 木の幹のまま ログハウス生活)といったパターンもあるが、少数でもあり、システムの構築を優先させるため、初期システムでは対象外とする。

また、B級機関では、ある文の文末に似た音の語を無理矢理入れる、という手法も用いられている[金杉99A]

《例》 一日の長 一日のチョーヨンピル

ただし、[榎本99]ではこの手法はほとんど見られないので、本論文では対象外とする。

### 3. 駄洒落生成システム

本章では、2.で示したパターンに基づいて駄洒落を生成するシステムの機能を決め、処理について考察し、基本設計を行う。

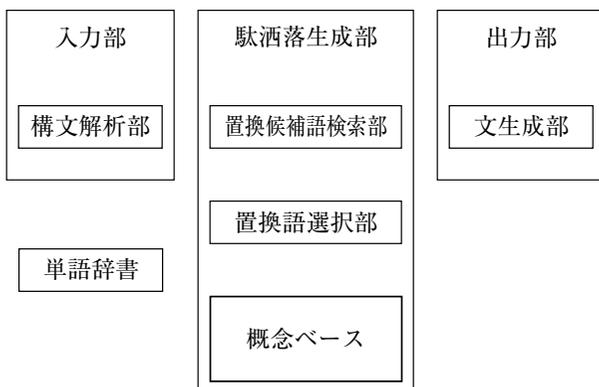
#### 3.1 システムの機能と基本構成

本論文で設計する駄洒落生成システムは、ある文を元にして、駄洒落を作成し、その駄洒落を表示する。元になる文は、外部から入力するものとする。従って、システムの基本機能は次のようになる。

- ( ) 入力 ... 文
- ( ) 処理 ... 駄洒落の生成、すなわち、入力された文を駄洒落に変換する
- ( ) 出力 ... 処理結果、すなわち、生成された駄洒落

駄洒落は、2.で示したパターンに基づいて、入力された文中のある言葉を別の言葉で置き換えることにより生成する。そのためには、文を構成している言葉(単語)が区別されなければならない。これは( )で行うものとする。( )では、単語を文に戻して表示する。( )の処理は次のステップから構成される。

- ( ) 置換対象となる単語を選ぶ
- ( ) 単語を置き換えるための同音異義語および類似音語を見つける
- ( ) 見つかった同音異義語または類似音語の中から1つを選ぶ



このため、システムは、( )から( )の機能をもつ3つのモジュールから構成される。そして、各モジュールは個々の処理を受け持つサブモジュールと、同音異義語や類似音語を検索するための単語帳(辞書)および概念ベースから構成される。システムの構成を図1に示す。

図1 システムの構成

「入力部」は、駄洒落の元となる文が入力され、「構文解析部」でその文を単語に分解する。それらの単語のいくつかに対して、「駄洒落生成部」の「置換候補語検索部」で同音異義語および類似音語を検索する。「単語辞書」はこれらのサブモジュールにおいて使用される。「置換語選択部」は、見つかった同音異義語および類似音語の中から1つを選択する。その際に、「概念ベース」を利用する。「出力部」は、「文生成部」で元の文の単語、および、同音異義語または類似音語を結合して文を作り、それを表示する。この処理の流れの概略を図2に示す。

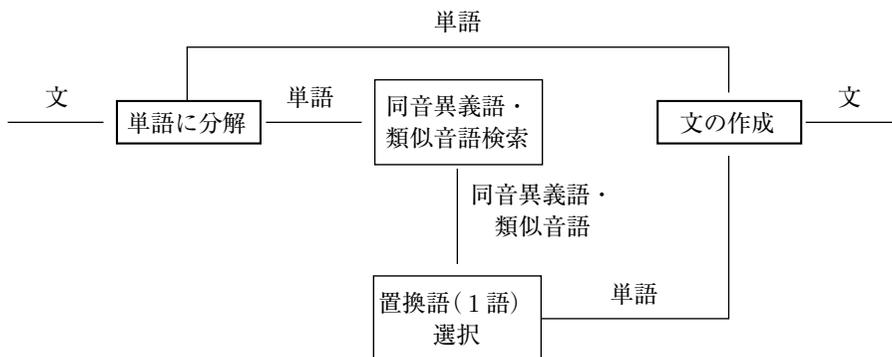


図2 処理の流れ

### 3.2 処理の概要

「構文解析部」における処理は、自然言語処理の構文解析の手法 [ 田中90 ] を応用することで実現できる。また、「文生成部」における、単語を文へ戻す操作は、文字列の演算により行われる。本節では、「駄洒落生成部」における処理について考察する。

「駄洒落生成部」が処理の対象とする単語、すなわち、置換対象となる単語であるが、[ 榎本99 ] からランダムに抽出した駄洒落について、置き換えの対象となった単語の品詞を調べると、表1のような結果が得られた。従って、この順位で同音異義語および類似音語を探す対象とすればよい。ただし、システムの構築を優先させるため、初期システムでは名詞だけを対象とすることにする。

品 詞	出現率
名 詞	68%
動 詞	11%
形容動詞	6%
形容詞	3%
副 詞	2%
その他	12%

表1 駄洒落において置換対象となった品詞

「置換候補語検索部」における処理は、同音異義語および類似音語を探すことである。同音異義語を探すには、読み（音）で引く（検索する）ことのできる辞書（単語帳）を用意しておけば、同じ読み（音）をもつ単語を選びだすことができる。

類似音語を探すには、まず、単語の読み（音）を2章で示した（2・1）から（2・7）のパターンに基づいて変更する。これは、（2・1）から（2・7）の各パターンをルール化し、それらを適用することにより行う。次に、得られた読みで辞書（単語帳）を引いて（検索して）単語を選びだす。すなわち、得られた読みに相当する単語が辞書にあれば、それを選ぶのである。

従って、「置換候補語検索部」は単語辞書を検索するサブモジュールと、（2・1）から（2・7）の各パターンに基づくルールを知識ベースにもつ推論モジュールから構成される。これを図3に示す。

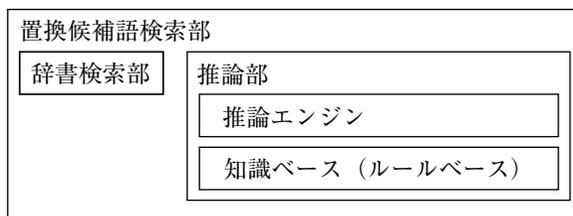


図3 「置換候補語検索部」の構成

「置換語選択部」では、「置換候補語検索部」で選びだされた複数の同音異義語または類似音語から1つを選択する。それを行うために、[田辺03]では、

特定の観点から見てもっとも差が大きいとみなされ、

かつ、より一般的である単語を選ぶ

という規則を提案した。この規則をシステム化するために、概念ベース [グエン00],[佐藤00],[松澤00]を応用する。

概念ベースは、ことばが表す意味 (= 概念) をデータベース化したものである。各概念は、その特徴を表す属性と、属性がその概念でどの程度重要かを表わす0から1の数値 (重要度) の対の集合で表わされる。例えば、「馬」という概念は

“馬” = ( “動物” : 0.5, “家畜” : 0.2, ... )

などと表現される [松澤00]。そして、

概念と属性から構成される行列を初期の概念ベースとする。そのイメージを図4に示す [グエン00]。属性の重要度を用いて、概念と概念の類似度を計算する [グエン00],[佐藤00]。概念間の類似度は、話題や文脈等の状況で変化する。例えば、動物の話なら、「馬」は「豚」に似ているが、乗物としてなら「自動車」により似ている [松澤00]。従って、話題や文脈に応じて、類似度の計算に用いる属性を変えるのである。

		属性				
		動物	家畜		移動	
概念	馬	0.5	0.2	...	0.1	
	乗物	0.1	0	...	0.4	

図4 概念ベースのイメージ

概念ベースは、本来、類似する概念を見つけるために提案されたものである。しかし、これを逆に使うことで、差が大きい概念を見つけることができる。類似度の低い概念を探せばよいのである。従って、「置換語選択部」での処理は次のように行われる。

文を構成する単語 (のいくつか) に共通する属性 (概念) を「観点」とする。

同じ「観点」をもつ同音異義語および類似音語を選ぶ。

選びだされた同音異義語および類似音語の中から、類似度が最も低いものを採用する。

この処理を実現するためには、概念ベースが鍵となる。概念ベースが、同音異義語または類似音語の選択において、中心的役割を担うからである。従って、本システムでは、このような処理が実行できるように、概念ベースを構築する、すなわち、属性を決め、重要度を設

定する必要がある。

以上の処理概要に従って、文「見よ、ジムで鍛えたこの体」を例にとって、駄洒落を作ってみる。

- ( ) 単語に分解「見よ・、・ジム・で・鍛えた・この・体」
- ( )
  - ( ) 同音異義語および類似音語の検索対象(名詞)「ジム」,「体」
  - ( ) 同音異義語および類似音語の検索
    - 「ジム」の同音異義語: 「事務」,「時務」,「寺務」
    - 「ジム」の類似音語
      - (2・1)のパターン: 「義務」,「字句」,「磁区」,「軸」,「JIS」,「自負」,「慈父」,
      - (2・3),(2・1)のパターン: 「医務」
    - 「体」の類似音語
      - (2・1)のパターン: 「サラダ」,「カナダ」
  - ( ) 「ジム」,「体」に共通する属性
    - 「運動」,「筋肉」,「過負荷」,「重労働」等
  - ( ) 「過負荷」,「重労働」を観点とする単語
    - 「事務」,「時務」,「寺務」,「医務」,「義務」
  - ( ) 類似度が最も低い単語「事務」
- ( ) 単語の置き換え「見よ・、・事務・で・鍛えた・この・体」  
文字列の加算「見よ、事務で鍛えたこの体」

もちろん、( ) , は、概念ベースがそのように構成されているものと仮定しての話である。言い換えれば、概念ベースの構成が適切であれば、駄洒落がうまく生成されることが期待できる。このため、先にも述べたが、本システムの構築においては、適切な概念ベースの構築が鍵となる。

しかし、概念ベースは近年になって提案されたものであるため、その設計・開発に関しては、いわゆる定番となる手法はまだない。目的別に設計・開発が行われているのが現状である。そこで、本システムでも、概念ベースは、プロトタイプングを通じて、設計・修正を繰り返しながら構築して行く予定である。

#### 4. まとめ—詳細設計および構築に向けて

本論文では、駄洒落生成システムの構築を目的として、システムの基本設計を行った。そして、システムの基本的な機能を定義し、処理の概要について考察した。

本論文での駄洒落生成の手法は、同音異義語あるいは類似音語による置換によるものである。これまでに作成された駄洒落生成システムの多くは、同音異義語による置換で駄洒落を生成するものである。類似音語による置換を取り入れているシステムは少数であり、それらにおいても明確な手法は提案されていない。本論文では、類似音語による置換の手法として、7つのパターンを提案した。これらにより、駄洒落の生成を、より柔軟で明確に行えることが予想される。

置換に用いる単語の選択では、概念ベースを応用する。概念ベースは、類似する概念を見つけるために提案されたもので、本論文ではその逆の使用法、すなわち、差が大きい(類似していない)概念を見つけるために使うことを提案した。

今後の設計・開発の過程においては、概念ベースが鍵となる。概念ベースが、同音異義語または類似音語の選択において、中心的役割を担うからである。しかし、概念ベースの設計・開発に関しては、先にも述べたが、構造化手法やオブジェクト指向といったような、主流となっている手法はない。目的別に設計・開発が行われているのが現状である。従って、今後の設計・開発においても、駄洒落向け概念ベースを設計・開発する必要がある。それには、まず、駄洒落に向けた概念と属性を選び、それらの重要度を設定しなければならない。そして、それらから構成される行列を基に、駄洒落向けにカスタマイズして、概念ベースを設計するのである。この、駄洒落向け概念ベースの設計・開発は、今後の最優先課題とする。

また、本論文では、文字の形が類似した別の文字による置換・文節の変更・文の文末に似た音の語を無理矢理入れる等の駄洒落生成の手法は取り入れなかった。これらも今後の課題とする。

## 参考文献

- [グエン00] グエン, 石川, 笠原: 「ベクトル表現された概念に対する類似度計算法」, 人工知能学会第4回ことば工学研究会資料, SIG-LSE-A001, pp.49-55 (2000)
- [ピNSTEDD98] ピNSTEDD, 滝澤: 「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)
- [安藤00] 安藤, 岡本, 石崎: 「連想属性における感性語の特徴と定量的記述」, 人工知能学会第5回ことば工学研究会資料, <http://ultimavi.arc.net.my/banana/Workshop/Program/5.html> (2000)
- [榎本99] 榎本 (監修): 御教訓大言海, PARCO 出版 (1999)
- [金杉99A] 金杉: 「B級機関 ~ コンピュータにおける「言語感覚」の目覚め~」, 人工知能学会第1回ことば工学研究会資料, SIG-LSE-9901-P5 (1999)
- [金杉99B] 金杉, 松澤: 「ことば遊びの世界」, 人工知能学会第3回ことば工学研究会資料, <http://ultimavi.arc.net.my/banana/Workshop/Program/3.html> (1999)
- [佐藤00] 佐藤, 藤本: 「綺麗だったら美しい! ~ 概念ベースによる意味照合」, 人工知能学会第4回ことば工学研究会資料, <http://ultimavi.arc.net.my/banana/Workshop/Program/4.html> (2000)
- [須永00] 須永, 松浦, 堀, 松澤, 阿部: 「座談会: ことば工学のすすめ」, 人工知能学会誌, Vol.15, No.3, pp.456-467 (2000)
- [田中90] 田中 (主任): 「解析手法」, 人工知能学会 (編) 人工知能ハンドブック, 自然言語編, 3章, pp.226-235 (1990)
- [田辺03] 田辺: 「駄洒落の形式化」, 産能大学紀要, 第24巻, 第1号, pp.213-222 (2003)
- [松澤00] 松澤, 堀, 金杉, 阿部: 「ことば工学入門」, 人工知能学会誌, Vol.15, No.3, pp.446-455 (2000)