

音素の類似性による対話型駄洒落の生成

WangYilin[†] 山本 岳洋[†] 大島 裕明[†]

[†] 兵庫県立大学 大学院情報科学研究科 〒651-2197 神戸市西区学園西町 8-2-1

E-mail: [†] ad24r004@guh.u-hyogo.ac.jp, t.yamamoto@sis.u-hyogo.ac.jp, ohshima@ai.u-hyogo.ac.jp

あらまし 本研究では、日本語における音韻の類似性に基づく言葉遊びの一形態である駄洒落に注目し、対話型駄洒落の生成手法を提案する。駄洒落とは、同一または非常に類似する音韻を持つ言葉を用いて遊ぶ言語的な遊戯であり、文中に音韻が類似する単語の集合が存在することによって特徴づけられる。本研究のアプローチは、ユーザから任意のテキスト（発話文）を受け取り、その中から特定の単語の音韻に類似する単語やフレーズを用いて、発話に対する面白い応答（返答文）を生成するものである。提案手法は、検索と生成、ランキングの3つのステップから構成されている。まず、検索のステップでは、音素の類似性に基づいて音韻が類似する単語やフレーズを検索する。次に、生成のステップでは、得られた単語やフレーズを返答文に含めるような制約を課したプロンプトを大規模言語モデルに入力することにより、対話型駄洒落の制約を満たした返答文の候補を生成する。最後のランキングのステップでは、生成された返答文の候補を対話の自然らしさでランキングし、もっとも自然な返答文を出力として採用する。評価実験では、人手による評価を通じて、生成された返答文が対話型駄洒落として適切かを評価した。

キーワード 駄洒落, 対話生成, 面白さ

1 はじめに

テクノロジーの急速な進展に伴い、スマートフォンや人工知能の普及が進んでいる。このような状況において、Siri や Alexa といった対話システムが登場し、ユーザは求める情報を容易に取得できるようになっている。しかし、これらの対話システムは主に回答の正確性と迅速性に焦点を当てており、対話そのものの楽しさを犠牲にしている場合が多い。

近年、社会環境の変化により独身者が増加し、より魅力的な会話体験に対する需要が高まっている。人々は、単に迅速かつ正確な応答を受け取るだけでは満足せず、対話システムとのやりとりを楽しみ、友人とのコミュニケーションに近いレベルを求めるようになってきている。このようなシステムは雑談対話システムと呼ばれ、特定の目的を持たず、娯楽性を重視した会話を楽しむことが可能である。しかし、雑談対話システムが人間の会話における自然さや面白さを再現することは依然として難しく、ユーザの期待に応える有意義なやりとりを実現することは困難である。

我々はこのような問題を解決するため、言葉遊び、特に駄洒落に焦点を当て、対話の面白さを向上させることを目的として研究を行っている。言葉遊びは、より広い文脈において、形態が類似しているが意味が異なる言語単位の組み合わせに基づく言語要素の操作として特徴づけられる。言葉遊びは、聴衆を楽しませたり、面白さを引き出したり、美的な楽しみを提供したりするなど、談話における機能によって分類されることが可能である [1]。本研究では、駄洒落という特定の言葉遊びを用いて対話における面白さを増強することを目指している。オックスフォード英語辞典 (OED) では、Pun を「2 つ以上の意味や関連性を暗示する言葉の使い方、または音韻が類似するが意味が

異なる 2 つ以上の単語の使用によって面白さを生み出すもの」と定義している。本研究でも、英語の Pun とほぼ同義の日本語の駄洒落に焦点を当てる [2]。

本研究では、一文で表現される一般的な駄洒落ではなく、対話型駄洒落に焦点を当てる。一般的な駄洒落は、以下に示す例のように、類似または同一の音韻の性質を共有しつつ、1 つの文の中で異なる意味を伝える 2 つの単語が含まれる。

「校長先生絶好調」¹

この例では、「校長」と「好調」の音韻が類似するため、駄洒落として捉えることができる。一方で、本研究で扱う対話型駄洒落は以下のように、対話における文脈を考慮した駄洒落である。

発話文 1: 「先生は校長に昇進したようです!」

返答文 1: 「おお、それは好調なニュースだね!」

発話文 2: 「荒川の女の子はどうですか?」²

返答文 2: 「あら、かわいいですよ。」

このように、発話文に含まれる単語と音韻が類似する単語やフレーズが発話文に対する返答文に含まれている。本研究では、このような駄洒落を「対話型駄洒落」と定義し、その生成と評価を行うことを目的とする。このような対話型駄洒落を生成することができれば、雑談対話システムが人間のように自然で面白い会話を再現できると考えた。

提案手法は、音韻が類似する単語やフレーズの検索、返答文の生成、そして返答文のランキングという 3 つのステップで構成される。音韻が類似する単語やフレーズの検索では、まず辞書を用いて音韻が類似する単語やフレーズを抽出し、意味的類

1: <https://ja.wikipedia.org/wiki/駄洒落> より引用

2: <https://dajare.jp/works/89096/> に掲載されている例に基づいて著者が作成した。

似度や発話文における自然さなどで返答文に含めるべきものをランキングする。返答文の生成では、得られた単語やフレーズをもとに、大規模言語モデル（LLM）を用いて、発話文に対する返答文の候補を生成する。最後に、生成された返答文の候補を発話文との関連性および対話としての自然さの観点からランキングする。評価実験では、人手による評価を用い、生成された対話型駄洒落の有効性を評価する。

本稿の構成は以下の通りである。まず、2 節では駄洒落の生成と評価に関する関連研究を述べる。3 節では本研究で扱う対話型駄洒落および問題定義、対話型駄洒落の生成に関する提案手法について述べる。4 節では評価実験と結果について述べ、最後に本稿をまとめる。

2 関連研究

これまで、駄洒落に関するさまざまな研究が行われている。例として、音韻が類似する語の検索に関する研究や、駄洒落の生成、そして生成された駄洒落や面白さの研究などがあげられる。

2.1 音韻が類似する語の検索に関する研究

近年、音韻が類似する語の検索は、主に駄洒落やなぞなぞといった言葉遊びの文脈で注目されている。これまでの研究の多くは、音素列の類似性に基づく音韻が類似する語の検索を行っていた。音素とは音声上の最小単位となる音韻の単位を指す³。たとえば、前田らは音素間の距離表を作成し、2 つの単語間の音素の類似性を求めることで、音韻が類似する語を求める手法を提案している [3]。また、島谷らはカタカナやローマ字の編集距離といった情報の組み合わせに基づいて単語の音韻間の類似性を求める手法を提案している [4]。また、これらの研究とは異なるアプローチとして、子音間と母音間の類似性を定義し、単語間の子音間および母音間の編集距離に基づいて単語の音韻間の類似性を求める手法も提案されている⁴。

本研究では、音韻が類似する語を検索するために、編集距離の算出方法を適用する。従来の手法とは異なり、本稿では 2 つの単語の音素列全体の編集距離を計算するのではなく、子音および半母音部分と母音部分の編集距離をそれぞれ独立に算出する。このアプローチにより、完全に同一の母音を有しつつ、子音および半母音が非常に類似している単語を音韻が類似する語として特定することが可能となる。

2.2 駄洒落生成に関する研究

駄洒落を生成する手法がこれまでに提案されている。たとえば、キムらは、長い文字列を持つ単語の類音語を置換し、テンプレートに基づいて駄洒落なぞなぞを生成する手法を提案している [5]。後藤らは、画像情報を基に適切なプロンプトを提示し、GPT を用いた併置型駄洒落文の生成手法を提案している [6]。南らは、振り仮名注釈付きコーパスを活用して言語モデルに日

本語の読み方を学習させ、その後、学習済みモデルを駄洒落とその言い換え文のペアを用いてファインチューニングすることで、駄洒落を自動生成する手法を提案している [7]。

英語における Pun 生成の研究は、日本語における研究とは異なる焦点を持っている。具体的には、同音異義語のペアを利用して、複数の意味を持つ一文を生成することが含まれる。この文脈において、一般的なテキストコーパスを用いて条件付きニューラル言語モデルを訓練し、言語モデルから Pun を生成する手法が提案されている [8]。また、意外性をもたらすことによって面白さの感覚を高めることを目的とした、ニューラルネットワークに基づく Pun 生成に関する研究も存在する [9]。さらに、同音異義語のペアは、質問と回答の形式での返答文においても多くの研究で利用されている。Binsted らは、Pun 生成の初期プロトタイプである JAPE⁵を開発した [10]。また、音声学および意味論に関する言語リソースを活用し、テンプレートに基づく Pun の抽出と生成に関する研究も行われている [11]。

日本語においてこれらの研究は非対話的な駄洒落の生成だが、本研究は対話型駄洒落の生成に取り組んでいる点で異なる。対話型駄洒落に関する定義は存在する [12] が、生成に関する研究は著者らの知る限りでは未だ存在していない。対話型駄洒落は非対話的な駄洒落とは異なり、言葉の適切な使用と対話全体の流暢性を考慮する必要がある。これは、対話型駄洒落を考察する上で重要な技術的課題となる。

2.3 駄洒落の評価に関する研究

駄洒落の面白さに関する評価手法の研究は複数存在する。キムらは、人間が創作した駄洒落なぞなぞと手製の辞書を用いて生成された駄洒落なぞなぞに対して、電子メールを通じた主観的評価実験を実施した。この研究は、辞書が駄洒落の面白さに一定の影響を及ぼすことを示唆している [5]。さらに、谷津らは、駄洒落の面白さの評価を、入力データの 2 値分類タスクとして捉え、「面白い」と感じる駄洒落と「面白くない」と感じる駄洒落の 2 つのクラスを設定し、分類器を用いて「面白い」駄洒落の検出手法を提案した [13]。花房らは、駄洒落単体と駄洒落を含む対話の面白さを評価することにより、対話文脈における面白さの変化を分析する手法を提案し、話者の反応について考察を行っている [14]。

3 対話型駄洒落の生成

3.1 対話型駄洒落の定義

本研究における対話型駄洒落は、音韻が同一または類似するが、意味が異なる 2 つの要素が、異なる 2 つの文に存在する。2 つの要素は、一方が発話文、もう一方が発話文に対する返答文に配置される。1 節であげた 2 つの例は、どちらも対話型駄洒落の例である。

発話文とは、対話において話者が発する文を指し、本研究における入力となる。また、**返答文**とは発話文に対する返答を指し、本研究の出力となる。本研究では、1 節であげた例におけ

3 : <https://ja.wikipedia.org/wiki/音素>

4 : <https://zenn.dev/fura2/articles/c97c8a59f3cc19> 日本語の響きの近さを測る

5 : Joke Analysis and Production Engine

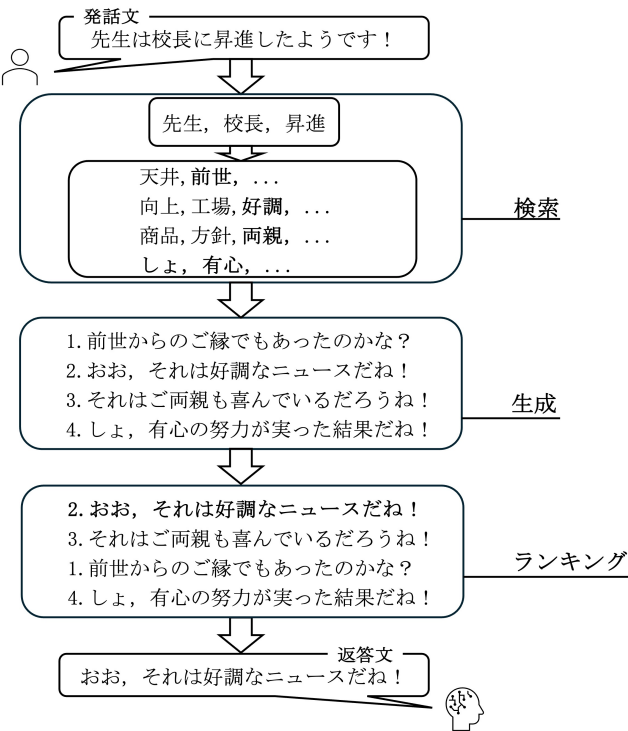


図 1 提案手法の流れ.

る「校長」や「荒川」のように、発話文にて音韻が同一または類似する要素が含まれる要素のことを駄洒落対象語と呼ぶ。また、駄洒落対象語と音韻が類似する要素のことを音類似フレーズと呼ぶ。たとえば、「荒川」という駄洒落対象語に対する音類似フレーズとしては「新川（あらかわ）」「唐皮（からかわ）」「あら、かわいい」などがあげられる。

3.2 提案手法の概要

本研究では対話型駄洒落の生成について取り組む。図 1 に提案手法の流れを示す。提案手法は、音類似フレーズの検索、返答文候補の生成、返答文のランキングという 3 つのステップで構成されている。

まず、音類似フレーズの検索では、与えられた発話文から駄洒落対象語を抽出し、駄洒落対象語に対する音類似フレーズを検索する。このとき、音類似フレーズを単語の出現頻度に基づいてフィルタリングした後、元の発話文との関連性や、音類似フレーズの発話文中での自然さなどによるランキングを行い音類似フレーズの候補を選択する。

次に、得られた音類似フレーズの候補に対して大規模言語モデル（LLM）を用いて返答文の候補を生成する。このとき、プロンプトに返答文に音類似フレーズを含めるような制約を記述することで、対話型駄洒落の制約を満たすような出力を行う。

最後に、生成された返答文の候補に対して、発話文と生成された返答文全体を自然さの観点からランキングし、最適な返答文を選択しユーザに返す。

以降、それぞれのステップについて詳しく説明する。

表 1 モーラの構成要素.

子音	k, g, s, z, c, t, d, n, h, b, p, m, r, v
半母音	j, w
母音	a, i, u, e, o
特殊音素	N, T, R

3.3 音素の類似性による音類似フレーズの検索

本節では、音素の類似性に基づいた音類似フレーズの検索手法を説明する。このステップの目標は、与えられた発話文（例：先生は校長に昇進したようです！）から、駄洒落対象語（例：「先生」、「校長」、「昇進」）を抽出し、返答文に含めるべき駄洒落対象語と音韻が類似する単語やフレーズ（例：「前世」、「好調」、「しょ、有心」）を出力することである。

以降、まず、音韻の類似性を計算するために必要となるモーラ の概念について述べる。次に、発話文からの駄洒落対象語の抽出について述べ、各駄洒落対象語に対して音類似フレーズの候補を取得する手法について述べる。その後、音類似フレーズの候補を単語の出現頻度によってフィルタリングする方法について述べ、最後に、フィルタリングされた候補から返答文に含めるべき音類似フレーズをランキングするための 2 つの手法について述べる。

3.3.1 モーラ

モーラとは、音韻論における音節より小さい単位であり、音節量を測るのに用いられる [15]。例えば、「カ」と「キャ」はいずれも 1 モーラに分類される。2 つの単語の音韻の類似性を評価するためには、モーラ間の類似性を分析することが不可欠である。そのため、単語のカタカナ表記をモーラ列に分割する必要がある。具体的には、「ガッキューシンブン」という単語のカタカナ表記は、「ガ、ツ、キュ、ー、シ、ン、ブ、ン」というモーラ列に分割される。各モーラはカンマによって区分されている。モーラの構造が一定であることを踏まえ、正規表現アプローチを用いてカタカナ列を分割することが可能である⁶。

また、既存の研究によると、1 モーラは子音、半母音、母音の 3 つの要素で構成される [16]。表 1 に示すように、子音音素は 14 種類、半母音音素は 2 種類、母音音素は 5 種類、特殊音素は 3 種類あり、モーラはこれらの要素によって構成される。したがって、モーラ間の類似性を測るために、表 2 に示されているようにモーラと英文字の変換表を作成した。例えば、モーラ「カ」は「k_a」に、モーラ「キャ」は「kja」に変換される。音素が存在しない場合は、「_」で示す。以降、このようにして変換された英文字列を「音素列」と呼ぶ。

表 2 に示されている以外にも、3 種類の特種音素、すなわち長音（ー）、撥音（ン）、促音（ッ）が存在する。これらは、半母音や母音を含まない 3 モーラとしても認識される。日本語の文脈において、「ー」は通常、前のモーラの音節を延長する長音を示す。本研究におけるモーラ変換プロセスでは、「ー」は前のモーラの母音に置き換えられる。具体的には、「ハ、ー」は「ハ、

6 : <https://qiita.com/shimajiroxyz/items/a133d990df2bc3affc12> 日本語（カタカナ）をモーラ単位で分かち書き【Python】

表 2 変 換 表.

	a	i	u	e	o	ja	ju	je	jo	wa	wi	we	wo
k	カ	キ	ク	ケ	コ	キャ	キュ		キョ	クア	クイ	クエ	クオ
g	ガ	ギ	グ	ゲ	ゴ	ギャ	ギユ		ギョ	グア			
s	サ	シ	ス	セ	ソ	シャ	シュ	シェ	ショ	グア		スイ	
z	ザ	ジ	ズ	ゼ	ゾ	ジャ	ジュ	ジェ	ジョ	グア		ズイ	
c	ツア	チ	ツ	ツエ	ツォ	チャ	チュ	チェ	チョ	グア			
t	タ	ティ	トゥ	テ	ト		テュ			グア			
d	ダ	ディ	ドゥ	デ	ド		デュ			グア			
n	ナ	ニ	ヌ	ネ	ノ	ニャ	ニユ			グア			
h	ハ	ヒ	フ	ヘ	ホ	ヒャ	ヒュ	ヒエ	ヒョ	ファ	フィ	フェ	フォ
b	バ	ビ	ブ	ベ	ボ	ビャ	ビュ		ビョ	グア			
p	パ	ピ	プ	ペ	ポ	ピャ	ピュ		ピョ	グア			
m	マ	ミ	ム	メ	モ	ミャ	ミュ		ミョ	グア			
r	ラ	リ	ル	レ	ロ	リャ	リュ		リョ	グア			
v	ヴァ	ヴィ	ヴ	ヴェ	ヴォ								
—	ア	イ	ウ	エ	オ	ヤ	ユ	イエ	ヨ	ワ	ウィ	ウェ	ウォ

ア」と表記される。

3.3.2 発話文からの駄洒落対象語の抽出

音類似フレーズの検索を行うためには、発話文から駄洒落対象語を抽出する必要がある。本研究では、発話文に含まれる名詞、動詞、形容詞、形容動詞のうち、モーラ数が3以上のものを全て駄洒落対象語として抽出した。形態素解析は MeCab [17] を用いた。

後述する音類似フレーズ抽出のため、得られた駄洒落対象語を表 2 を参照して選定した単語から音素列に変換しておく。例えば、発話文「先生は校長に昇進したようです！」から、駄洒落対象語「先生」、「校長」、「昇進」を抽出し、それに対応する音素列（「s_e,N___s_e,___i」, 「k_o,___u,cjo,___u」, 「sjo,___u,s_i,N___」）に変換する。

3.3.3 音類似フレーズの候補の取得

本研究では、音類似フレーズを取得するため、既存の辞書に登録された単語を利用した。本研究では、UniDic 辞書 [18] および現代日本語書き言葉均衡コーパス (BCCWJ) 短単位語彙表を使用した。UniDic 辞書には、英単語や絵文字を含む 878,229 語の日本語が収録されている。BCCWJ は、現代日本語の書き言葉の全体像を把握するために構築されたコーパスであり、BCCWJ 短単位語彙表 [19] は、BCCWJ 全体および BCCWJ を構成する各レジスターおよびコアデータについて頻度 1 までの見出し語を収録している。また、BCCWJ 短単位語彙表には単語、読み、出現頻度などの情報が含まれている。

本研究では、これらの辞書を用いて、以下に示す 4 通りのアプローチで音類似フレーズの候補を取得する。

1. 同音異義語

同音異義語とは、音韻が同一であるが意味が異なる 2 つの単語を指す。駄洒落対象語と音素列が一致する単語を辞書から見つけ、音類似フレーズとして取得する。たとえば、「校長」という駄洒落対象語に対して、「好調」という音類似フレーズを同音異義語として取得することができる。

2. 類音語

類音語とは、音韻が類似するが同一ではない 2 つの単語を指す。今回は、2 つの単語の音素列において母音列が完全に一致し、子音列の編集距離が 1 である単語を類音語として定義した。単語間の音韻の類似性を算出するためには、音素列間の編集距離を計算する必要がある。日本語においては半母音が 2 つしか存在せず、通常は音韻に対する影響が小さいため、本論文では子音と半母音をまとめて子音として扱うことにした。

たとえば、駄洒落対象語である「歯医者 (h_a,___i,sja)」と辞書内の「会社 (k_a,___i,sja)」との音韻の類似性を計算する場合、子音列 (h_,___,sj) と (k_,___,sj), および母音列 (a,i,a) と (a,i,a) の編集距離をそれぞれ計算する。計算結果は、母音列間の編集距離は 0 (完全一致)、子音列間の編集距離は 1 であり、類音語の定義に合致するため、「会社」を駄洒落対象語「歯医者」の類音語として取得することができる。

3. 駄洒落対象語を含む単語

駄洒落対象語を含む単語とは、駄洒落対象語の音素列を含む単語を指す。例えば、駄洒落対象語「ナイス」を含む単語として「内水 (ナイスイ)」が挙げられる。このように、駄洒落対象語の音素列を含む単語を取得することで、駄洒落対象語の音素列で始まる「内水 (ナイスイ)」, 駄洒落対象語の音素列で終わる「タナイス」、さらに駄洒落対象語の音素列が途中に出現する「グナリスト」といった 3 種類の単語を取得することができる。

4. 駄洒落対象語を含むフレーズ

これまでに説明した手法は、全て単語を音類似フレーズとして取得する手法であった。ここで説明する駄洒落対象語を含むフレーズとは、駄洒落対象語の音素列を含むフレーズを指す。対象とするフレーズは、読点で区切られた 2 つの単語を組み合わせたものである。例えば、「はい、社長」という 2 つの単語の組み合わせだったテキストが 1 つのフレーズとして該当する。既存の対話型駄洒落データを分析した結果、「感動詞、名詞」「感動詞、形容詞」「感動詞、形容動詞」という構造を持つ組み合わせが対話型駄洒落に出現することがあることがわかった。そこで、単純な単語だけではなく、こうした構造を持つフレーズを

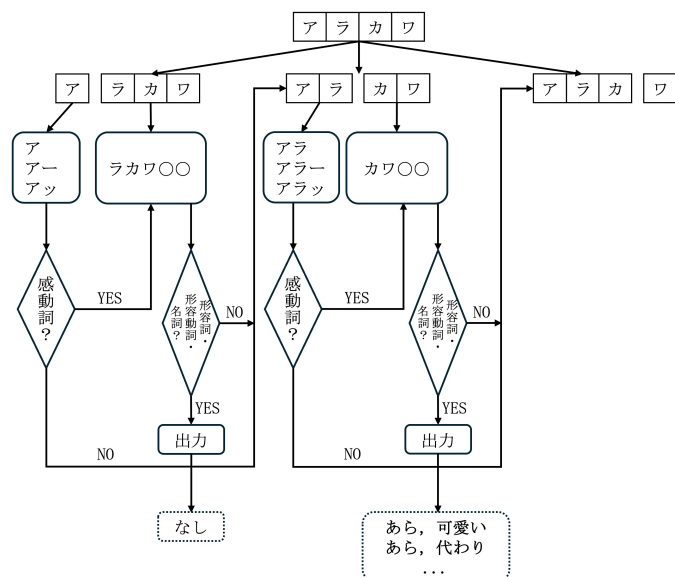


図2 フレーズの検索手法。

音類似フレーズとして取得することで、より多様な返答文が生成できると考えた。

駄洒落対象語を含むフレーズを取得するための手法を以下に提案する。まず、駄洒落対象語のモーラ列を2つのセグメントに分割し、各セグメントには最低1つのモーラを含めることとする。さらに、日本語の感動詞の多くは「ッ」や「ー」で終わるという特性を考慮し、感動詞を網羅的に特定するために、最初のセグメントに「ッ」と「ー」を追加する。最初のセグメントに対応する単語に感動詞が見つかった場合、次のセグメントで始まるモーラ列に対応する単語を辞書から取得する。得られた結果を基に、名詞、形容詞、形容動詞の単語を特定し、カンマで区切って最初の感動詞と結合する。逆に、感動詞が見つからなかった場合は、異なるセグメンテーション手法を用いて辞書の検索を継続する。このアプローチにより、図2に示される検索手法を通じて、駄洒落対象語を含むフレーズを取得することが可能となる。この手法により、たとえば、「荒川」という駄洒落対象語に対して、「あら、可愛い」や「あら、代わり」といったフレーズを取得することができる。

3.3.4 出現頻度による音類似フレーズのフィルタリング

先述した方法で取得された音類似フレーズの候補は、辞書中の単語から取得しているため、一般に用いられない単語も候補として取得される。そのような一般的に用いられない単語を返答文に含めしまうと、多くの人には伝わらない対話型駄洒落になってしまうと考えられる。

そこで、多くの人が理解できるような返答文を生成するために、前述の BCCWJ 短単位語彙表を用いて音類似フレーズのフィルタリングを行う。具体的には、BCCWJ 短単位語彙表における出現頻度に基づき、出現頻度が高い単語（同音異義語、類音語および駄洒落対象語を含む単語）とフレーズ（駄洒落対象語を含むフレーズ）を1つの駄洒落対象語あたりそれぞれ20個の計40個を取得する。このとき、駄洒落対象語を含むフレーズは、感動詞とその後の単語の2つの単語から構成される

ため、後者の単語（名詞、形容詞、形容動詞）の出現頻度に基づいてフィルタリングを行う。例えば、「あら、かわいい」というフレーズの場合、後半の「かわいい」に着目する。

このフィルタリングにより、1つの駄洒落対象語あたり40個の音類似フレーズを取得する。

3.3.5 音類似フレーズのランキング手法

最後に、取得した音類似フレーズの候補から、返答文に含めべき音類似フレーズをランキングする。このために、本研究では、類似度ランキング法と自然さランキング法という2つの手法を提案する。

1. 類似度ランキング法

著者らが人手で対話型駄洒落を分析したところ、対話型駄洒落は、以下の例に示されるように、返答文中の音類似フレーズ「あら、かわいい」が発話文中の駄洒落対象語「女の子」と意味的に類似しているケースがみられた。

発話文: 「荒川の女の子はどうですか？」

返答文: 「あら、かわいいですよ。」

そこで、本研究で提案する類似度ランキング法は、駄洒落対象語と音類似フレーズの意味的類似度に基づいて音類似フレーズをランキングする。いま、発話文に含まれる駄洒落対象語の集合を $W = \{w_1, w_2, \dots, w_n\}$ とする。ここで、 n は駄洒落対象語の数である。また、先述した方法で候補として得られた音類似フレーズの集合を $P = \{p_1, p_2, \dots, p_m\}$ とする。 m は候補として抽出した音類似フレーズの個数であり、本研究では $m = 40 \times n$ である。ここで、音類似フレーズの要素 p_i は、 p_i が駄洒落対象語を含むフレーズの場合は感動詞を除いた単語の部分を表す。

駄洒落対象語と音類似フレーズの意味的類似度は、既存の埋め込みモデルでそれぞれの単語の埋め込み表現を求め、コサイン類似度を求める。音類似フレーズ $p_j \in P$ と駄洒落対象語集合 W との類似度を以下の式に従って求める。

$$p_j = \max_{w_i \in W} \text{sim}(\mathbf{p}_j, \mathbf{w}_i), \quad j = 1, 2, \dots, m \quad (1)$$

ここで、 \mathbf{p}_j と \mathbf{w}_i は p_j と w_i の埋め込み表現をそれぞれ表す。本研究では、埋め込みモデルとして intfloat/multilingual-e5-large⁷を用いた。式(1)の値の降順に上位10件の音類似フレーズを次節で述べる生成に用いる。

2. 自然さランキング法

対話型駄洒落の中には、以下の例のような、返答文に含まれる音類似フレーズ「五千元」を、発話文中の駄洒落対象語である「ご声援」に置き換えても、発話文として自然な文となるケースが見られた。

発話文: 「ご声援をいただいてありがたかったね！」⁸

返答文: 「えっ!? 五千元をもらったって!？」

そこで、本研究では、発話文中の駄洒落対象語を音類似フレーズに置き換え、置き換え後の文の自然さを測ることで音類

7: <https://huggingface.co/intfloat/multilingual-e5-large>

8: <https://ameblo.jp/syunkon/entry-12835773522.html> に掲載されている例に基づいて著者が作成した。

似フレーズをランキングする自然さランキング法を提案する。本研究では、文の自然さを測る指標としてパープレキシティを用いる。パープレキシティは言語モデルが与えられたテキストをどれだけうまく予測できるかを表す指標である。いま、あるモデルのパラメータを θ とするとき、このモデルがテキスト $T = t_1 t_2 \dots t_N$ を出力する確率は以下のように表される。

$$p(T; \theta) = \prod_{i=1}^N p(t_i \mid t_1, t_2, \dots, t_{i-1}; \theta)$$

(2)

このとき、このテキスト T に対するパープレキシティは以下のよう求められる。

$$\text{Perplexity}(T; \theta) = \exp \left(-\frac{1}{N} \log p(T; \theta) \right)$$

(3)

本研究では、このパープレキシティを、文章の自然さを表す指標として用いる。パープレキシティが小さいほど、モデルがテキストを出力する確率は高くなるため、そのテキストは一般的であり自然な文章だと捉えることができる。本研究では、パープレキシティの計算には rinna/japanese-gpt2-small⁹ [20] [21] を用いた。このパープレキシティを用いて、発話文中の駄洒落対象語を音類似フレーズを置き換えた文の自然さを求める。いま、音類似フレーズ p_j が駄洒落対象語 w_i から抽出されたとする。このとき、発話文中の駄洒落対象語 w_i を p_j に置き換えた新たな文 T_j を作成する。そして、 T_j のパープレキシティを式 (3) に従って求め、パープレキシティが小さな値となった音類似フレーズ上位 10 件を次節の返答文の生成に用いる。

3.4 音類似フレーズを用いた返答文の生成

3.3 節の手法により、返答文に含めるべき音類似フレーズの集合が得られる。本研究では、音類似フレーズを含んだ返答文を生成するように指示したプロンプトを大規模言語モデル (LLM) に入力することで、対話型駄洒落としての制約を満たした返答文を生成する。

本研究で作成したプロンプトを図 3 に示す。作成したプロンプトの特徴について述べる。1 点目は、返答文として含めるべき音類似フレーズを明示的にプロンプトに記載している点である。これにより、LLM が音類似フレーズを含んだ返答文を生成することが期待される。2 点目は、発話文に含まれる単語は返答文には含めないようにしている点である。これは、発話文と返答文に同一の単語や表現が含まれることで、異なる意味を持つ単語を含めるという駄洒落としての性質が満たされなくなることを避けるためである。3 点目は、返答文として短く簡潔なものを出力するように指示している点である。これは、既存の駄洒落を著者らが分析した結果、駄洒落の多くは比較的短いテキストであったためである。このプロンプトを用いて、10 個の音類似フレーズに対してそれぞれ 1 件の返答文を生成することで、10 件の返答文候補を得る。

次の条件を満たす返答文を生成してください。

- 条件
1. 発話文S、駄洒落対象語W と 音類似フレーズP が与えられます。
 2. 生成する返答文には 音類似フレーズP を必ず含めてください。
 3. 発話文S に含まれる単語や表現は、できるだけ返答文に含めないようにしてください。
 4. 駄洒落対象語W は生成する返答文に表示されてはならない。
 5. 生成する返答文の長さは短く簡潔なものにしてください。

入力例1
発話文S: 「先生は校長に昇進したようです!」
駄洒落対象語W: 「校長」
音類似フレーズP: 「好調」

理想的な出力1
「おお、それは好調なニュースだね!」

入力例2
発話文S: 「この毛布、すごく温かいんだよ! 冬にぴったり!」
駄洒落対象語W: 「毛布」
音類似フレーズP: 「もう、不思議」

理想的な出力2
「もう、不思議なくらい温かくて、手放せなくなりそうだね!」

入力
発話文S: 「{発話文}」
駄洒落対象語W: 「{駄洒落対象語}」
音類似フレーズP: 「{音類似フレーズ}」

図 3 返答文生成に用いたプロンプト。

3.5 返答文のランキング

最後に、得られた返答文候補をランキングすることで、対話型駄洒落として適切な返答文を選択する。我々は、対話型駄洒落として適切な返答文は、入力の発話文に対して自然な返答となっており、多くの人にとって理解しやすい返答が相応しいと考えた。

そこで、本研究では発話文と生成された返答文を結合した文章に対してパープレキシティを求め、パープレキシティが最も小さい返答文を最終的な出力結果として出力する。

4 評価実験

本節では、3 節で提案した手法の有効性を検証するための実験とその結果について述べる。

4.1 データ

提案手法の有効性を検証するため、対話型駄洒落の評価用のデータを構築した。具体的には、既存の駄洒落サイトを参考に対話型駄洒落を作成したものに加え著者らが独自に対話型駄洒落を作成した。既存の駄洒落サイトとしては、DajareStation¹⁰ およびダジャレナビ¹¹を参考にした。本研究では、合計 50 件の対話型駄洒落（つまり、発話文と返答文のペア）を作成し評価に用いた。

4.2 評価に用いた手法

本研究では、音類似フレーズのランキング手法として 2 種類

9 : <https://huggingface.co/rinna/japanese-gpt2-small>

10 : <https://dajare.jp/>

11 : <https://dajarenavi.net/>

次の条件を満たす返答文を生成してください。	
条件	
1. 発話文S が与えられます。	
2. 生成する返答文には、発話文S に含まれる単語と音韻が類似する単語を用いて駄洒落を作成してください。	
3. 発話文S に含まれる単語や表現は、できるだけ返答文に含めないようにしてください。	
4. 生成する返答文の長さは短く簡潔なものにしてください。	
5. 駄洒落に使用した音韻が類似する単語を明確に示してください。	
入力例	
発話文S: 「先生は校長に昇進したようです！」	
理想的な出力	
「おお、それは好調なニュースだね！」	
分析	
「校長」と「好調」は音韻が類似しています。	
入力	
発話文S: 「{発話文}」	

図 4 ベースラインの返答文生成に用いたプロンプト。

を提案した。そこで、その 2 種類のランキング手法の性能の違いを調べる。また、提案手法に用いる大規模言語モデルを変えることによって生成結果の質が変わるのかを検証するため、本研究では 2 つの大規模言語モデルを用いて実験を行う。最終的に、本研究では以下の 6 つの手法を比較した。

- 人手で作成した理想例: 4.1 節で用意した返答文を出力する手法。人手で作成した対話型駄洒落がどの程度の質を持っているのかを評価するために用いる。
- GPT-4o (ベースライン): 音類似フレーズをプロンプトに含めずに、図 4 に示されたプロンプトに基づいて、GPT-4o モデルで返答文を生成する手法。提案手法のベースラインとして用いる。
- GPT-4o (類似度ランキング法): 3.3 節で提案した、類似度ランキング法により音類似フレーズを選択し、GPT-4o を用いて返答文を生成する手法。
- GPT-4o (自然さランキング法): 3.3 節で提案した、自然さランキング法により音類似フレーズを選択し、GPT-4o を用いて返答文を生成する手法。
- Llama-3.1-Swallow (類似度ランキング法): 3.3 節で提案した、類似度ランキング法により音類似フレーズを選択し、Llama-3.1-Swallow を用いて返答文を生成する手法。
- Llama-3.1-Swallow (自然さランキング法): 3.3 節で提案した、自然さランキング法により音類似フレーズを選択し、Llama-3.1-Swallow を用いて返答文を生成する手法。

実験に用いた GPT-4o¹² のバージョンは 2024-08-06 である。また、Llama-3.1-Swallow は Llama-3.1-Swallow-8B-Instruct-v0.3¹³ を用いた。GPT-4o は、OpenAI が 2024 年 5 月に発表したもので、テキスト、音声、画像、動画を統合的に処理できる次世代の AI モデルである。一方、Llama-3.1-Swallow モデルは、2024 年 12 月に更新されたモデルで、日本語の高い言語理解・生成・対話能力を発揮する大規模言語モデルである [22] [23] [24]。返答文の多様性を確保するために、GPT-4o モ

表 3 各評価項目に対する κ 係数の結果。

	駄洒落の認識	自然さ	面白さ (3 段階)	面白さ (2 段階)
κ 係数	0.48	0.44	0.34	0.45

デルと Llama-3.1-Swallow モデルの temperature パラメータを 1 に設定した。また、Llama-3.1-Swallow モデルにおいては top_p を 0.9、max_tokens を 128 に設定した。

4.3 評価項目

生成された返答文の質を評価するため、以下の 3 つの基準で評価者に各手法が生成した返答文を評価してもらった。

- 対話型駄洒落だと感じられるか (駄洒落の認識): 発話文と返答文が対話形式の駄洒落として成立しているかを評価する。発話文に含まれる単語と音韻が類似する単語やフレーズが返答文に含まれると評価者が認識すれば 1 を、そうでなければ 0 と評価してもらった。
- 自然な対話だと感じられるか (自然さ): 発話文と返答文のつながりがスムーズで、論理的に自然であるかどうかを評価する。駄洒落でなくとも、返答文が対話として自然であれば 1 を、そうでなければ 0 と評価してもらった。
- 面白い対話だと感じられるか (面白さ): 返答文がどの程度面白さを持っているかを評価する。読んだ瞬間に面白さを感じた場合は 1 を、読んだ瞬間でなくでも、少し考えた後に面白さを感じた場合は 0.5 を、面白さを感じない場合は 0 を評価してもらった。

2 名の評価者にこれら 3 つの項目で返答文を評価してもらった。評価には 2 名の評価者のスコアの算術平均を用いた。また、上記 3 つの項目に加え、評価者の評価を基に、対話型駄洒落としての面白さについても評価した。具体的には、評価者が「駄洒落の認識」について 0 をつけた場合は「面白さ」のスコアを 0 とし、「駄洒落の認識」について 1 をつけた場合は「面白さ」のスコアをそのまま用いるという計算式で新たに「対話型駄洒落としての面白さ」を評価項目として用意した。

2 名の評価者の評価がどの程度一致するのかを検証するため、 κ 係数を求めた。表 3 はそれぞれの評価項目における κ 係数の結果である。ここで、面白さ (3 段階) とは、面白さを 0, 0.5, 1 の 3 段階で評価してもらった際の係数、面白さ (2 段階) とは面白さが 0.5 と 1 の評価された場合は両者とも 1 と扱った際の係数である。表 3 より、「駄洒落の認識」と「自然さ」については κ 係数が 0.4 を超えており中程度の一致が見てとれる。一方、「面白さ (3 段階)」については 0.34 と「駄洒落の認識」と「自然さ」に比べて一致度が低くなっているものの、面白さを 2 段階として扱った場合は κ 係数が 0.45 となっておりこちらも中程度の一致が見てとれる。

4.4 実験結果

表 4 にそれぞれの手法における各評価項目の評価結果を示す。また、表 5 に、それぞれの手法の生成結果の例を示している。これらの結果は Llama-3.1-Swallow (類似度ランキング法) を

12 : <https://openai.com/index/hello-gpt-4o/>
13 : <https://swallow-llm.github.io/llama3.1-swallow.ja.html>

除き、2名の評価者とも「駄洒落の認識」に1を評価し、かつ、「面白さ」に1を評価した、すなわち面白い対話型駄洒落として評価された例である。

4.4.1 音類似フレーズのランキング手法の比較

音類似フレーズのランキング手法として、類似度ランキング法と自然さランキング法を提案した。「駄洒落の認識」の評価項目をみると、自然さランキング法によって生成された返答文が対話型駄洒落として認識される可能性が高いことがわかる。また、面白さについても、自然さランキング法の方が類似度ランキング法よりも高いスコアとなっていることがわかる。

4.4.2 大規模言語モデルの比較

GPT-4oモデルとLlama-3.1-Swallowモデルの性能を比較した結果、類似度ランキング法に基づく評価では、GPT-4oモデルが3つの基準すべてにおいてLlama-3.1-Swallowモデルを上回る結果が得られた。一方、自然さランキング法による評価では、自然さ以外の2つの基準においてLlama-3.1-SwallowモデルがGPT-4oモデルを上回る結果が見られた。両手法の結果を総合的に考察すると、Llama-3.1-Swallowモデルによって生成された結果は駄洒落に認識されやすい一方で、GPT-4oモデルによって生成された結果はより自然で論理的であることが示唆される。また、面白さの表現に関しては、両モデル間に顕著な差異は見られなかった。

4.4.3 提案手法とベースラインの比較

本論文では、提案手法とGPT-4o（ベースライン）によって得られた結果の比較分析を行う。まず、提案手法は音韻が類似する音類似フレーズを検索し、それに基づいて返答文を生成することにより、対話型駄洒落を生成するものである。実験結果に基づくと、生成モデルのいずれにおいても、GPT-4o（ベースライン）によるスコアと比較して、提案手法による対話型駄洒落の識別度は0.3以上高いことが示された。また、自然さスコアに関しては、Llama-3.1-Swallow（自然さランキング法）を除くすべてのモデルが0.1以上のスコアを示した。これらの基準においてベースラインよりも提案手法が優れていると評価できる。

しかし、面白さに関する評価（3段階および2段階）では、提案手法はGPT-4o（ベースライン）に劣る結果となった。この「面白さ」は対話型駄洒落として認識されずに面白いと評価された返答文も存在する。そのため、対話型駄洒落として認識され、かつ面白いと評価された際にスコアを獲得する「対話型駄洒落としての面白さ」で提案手法とGPT-4o（ベースライン）を比較すると、GPT-4o（自然さランキング法）およびLlama-3.1-Swallow（自然さランキング法）の結果がGPT-4o（ベースライン）よりもやや良好となる結果となったが、両者の差は顕著ではないことが確認された。

4.5 考察

まず、2つの音類似フレーズのランキング手法による実験結果の分析から、単語選択の違いが対話の面白さに影響を及ぼすことが明らかになった。自然さランキング法がより面白い結果を生む理由は、単語の音韻が非常に類似するため、同一の文に

おいて曖昧さや誤解が生じる可能性が高く、これに基づく返答がより面白くなるからである。また、GPT-4o（ベースライン）は提案手法に比べて、面白さの観点においてほぼ常に優れていることが示された。具体的な例の分析を通じて、面白いと評価された返答文に含まれる単語やフレーズの多くは、検索結果には現れないことが判明した。例えば、表5でGPT-4o（ベースライン）による生成した返答文では「焦点」に対応する「商店」という音類似フレーズが含まれている。この音類似フレーズは検索可能であるものの、発話文との関連性が低く、発話文に置換した際にも不自然であるため、最終的に10個の音類似フレーズには含まれず、返答文生成には使用されなかった。このことから、適切な単語の選択が面白さの程度に一定の影響を与えると考えられ、特に意外性における面白さにおいてその影響が顕著であることが示唆される。

さらに、生成された対話の自然さとその面白さの間には明確な正の相関関係が存在しないことが明らかになった。つまり、対話が不自然であっても、面白いと感じられる場合があるということを示している。例えば、表5でGPT-4o（類似度ランキング法）による生成した結果は両者の評価者によって不自然であると判断されたが、両者ともに面白いと感じた。また、Llama-3.1-Swallow（類似度ランキング法）による生成した結果は両者の評価者によって自然であると判断されたが、両者ともに面白くないと感じた。このことから、著者らは対話における意外性が、自然さよりも面白さの変化をより適切に反映していることが多いと考えられる。

また、生成モデルにおいては音韻の類似性に関する理解が不十分であり、面白さの理解に関しても表現が適切ではない。本研究において生成モデルを用いて得られた結果は、3つの基準全てにおいて、人手で作成した理想例の結果と顕著な相違が見られた。特に、「駄洒落の認識」の項目に関しては、Llama-3.1-Swallow（自然さランキング法）を用いた場合、0.84という結果を示し、理想例の結果0.93に接近したものの、依然として改善の余地が残されている。他の基準に関しては、生成モデルの結果は理想例と比較して大幅に低い数値を示している。特に、「対話型駄洒落としての面白さ」に関する指標において、理想例の0.32と生成モデルの最高結果である0.10との間には、依然として顕著な差異が存在している。しかしながら、生成モデルの結果は、文脈の意外性が面白さに対して肯定的な影響を及ぼす可能性があることを示唆している。また、発話文の表現を返答文に含めないようにという指示が出されているにもかかわらず、発話文の表現や言葉を含む生成結果が依然として多く存在している。

5 まとめと今後の課題

本研究においては、雑談対話システムが自然な会話や面白さを再現する際の課題に焦点を当て、日本語における音韻の類似性に基づく言葉遊びの一形態である駄洒落に特に注目し、対話型駄洒落の生成手法を提案した。提案手法は、検索と生成、ランキングの3つのステップから構成されており、対話型駄洒落の生成を行なった。検索のステップでは、音素の類似性に基づ

表 4 各手法の評価結果.

	駄洒落の認識	自然さ	面白さ (3 段階)	面白さ (2 段階)	対話型駄洒落としての面白さ (3 段階)
人手で作成した理想例	0.93	0.66	0.34	0.50	0.32
GPT-4o (ベースライン)	0.44	0.30	0.10	0.17	0.09
GPT-4o (類似度ランキング法)	0.72	0.48	0.07	0.12	0.07
GPT-4o (自然さランキング法)	0.73	0.38	0.09	0.15	0.09
Llama-3.1-Swallow (類似度ランキング法)	0.72	0.40	0.05	0.09	0.05
Llama-3.1-Swallow (自然さランキング法)	0.84	0.28	0.10	0.16	0.10

表 5 生成結果の例.

	発話文	生成した返答文
GPT-4o (ベースライン)	今日の会議の <u>焦点</u> は何ですか？	焦点は「 <u>商店</u> 」でセール中ですよ！
GPT-4o (類似度ランキング法)	<u>配送</u> ですか？	はい、操作はお任せください！
GPT-4o (自然さランキング法)	今日の天気、すごくいいね！	今日の空はペンキで塗ったみたいにきれいだね！
Llama-3.1-Swallow (類似度ランキング法)	この店舗の改装が始まるんだ。	お店の改造、楽しみだね！
Llama-3.1-Swallow (自然さランキング法)	<u>オオカミ</u> が大好きだ。一緒に描こう。	おー、 <u>紙芝居</u> でオオカミを描くのはいいアイデアだね！

いて音韻が類似する単語やフレーズを探索するための検索システムを開発した。生成のステップでは、適切なプロンプトを提示し、GPT-4o モデルと Llama-3.1-Swallow モデルを用いて自然な返答文を生成した。ランキングのステップでは、生成された返答文を対話の自然らしさでランキングし、もっとも自然な返答文を採用した。

本研究では評価のために 50 件の対話型駄洒落を用意したが、より大規模な評価を行うためには評価用データをさらに作成する必要がある。また、今回の実験では返答文のランキング手法の効果については未検証であり、今後ランキング手法の影響について評価する予定である。また、本稿では生成段階における面白さのヒントを提示していないため、生成段階の内容についてもさらなる研究と改善が求められる。加えて、今後は意外性に基づく面白さに焦点を当てつつ、面白さに影響を与える他の要因についても検討し、雑談対話の面白さを向上させることを目指している。

謝辞

本研究は JSPS 科学研究費助成事業 JP24K03228, JP22H03905, JP21H03554, JP21H03775 による助成を受けたものです。ここに記して謝意を表します。

文献

[1] Verena Thaler. Varieties of wordplay. In *Sebastian Knospe, Alexander Onysko & Maik Goth (eds.), Crossing Languages to Play with Words: Multidisciplinary Perspectives*, pp. 47–62. De Gruyter, 2016.

[2] 大竹孝司. ‘pun’ にない「駄洒落」の柔軟性: 日本語の言葉遊びからの証拠 (<特集> 最適性理論の実験検証と実験音声学の理論整備). 音声研究, Vol. 14, No. 1, pp. 76–85, 2010.

[3] 前田実香, 鬼沢武久. 単語の関連性とおもしろさを取り入れたなぞかけ生成. 感性工学研究論文集, Vol. 5, No. 3, pp. 17–22, 2005.

[4] 島谷二郎, 中村泰. Soramimic-限定された単語による空耳日本語文自動生成システムの開発. 研究報告ヒューマンコンピュータ

インタラクション (HCI), Vol. 2018-HCI-176, No. 5, pp. 1–6, 2018.

[5] ビンステッドキム, 滝澤修. 日本語駄洒落なぞなぞ生成システム “BOKE”. 人工知能, Vol. 13, No. 6, pp. 920–927, 1998.

[6] 後藤颯志, 浅野歴, 森田武史. シーングラフと GPT に基づく画像に関連する併置型駄洒落生成. 人工知能学会第二種研究会資料, Vol. 2023, No. SWO-062, p. 04, 2024.

[7] 南智仁, 清雄一, 田原康之, 大須賀昭彦. 日本語の仮名表記を学習した言語モデルを用いた文章の言い換えによる駄洒落の生成. 第 38 回人工知能学会全国大会, 2G5–GS–6–03, pp. 1–4, 2024.

[8] Zhiwei Yu, Jiwei Tan, and Xiaojun Wan. A neural approach to pun generation. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 1650–1660, 2018.

[9] He He, Nanyun Peng, and Percy Liang. Pun generation with surprise. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pp. 1734–1744, 2019.

[10] Kim Binsted and Graeme Ritchie. A symbolic description of punning riddles and its computer implementation. *arXiv preprint arXiv:cmp-lg/9406021*, 1994.

[11] Bryan Anthony Hong and Ethel Ong. Automatically extracting word relationships as templates for pun generation. In *Proceedings of the Workshop on Computational Approaches to Linguistic Creativity*, pp. 24–31, 2009.

[12] Meri Giorgadze. Linguistic features of pun, its typology and classification. *European Scientific Journal, ESJ*, Vol. 10, No. 10, 2014.

[13] 谷津元樹, 荒木健治. 駄洒落の面白さにおける要因の分析. ファジィシステムシンポジウム講演論文集, Vol. 32, pp. 237–242, 2016.

[14] 花房竜馬, 荒木健治. 対話文脈付き駄洒落データベース構築のための駄洒落を含む対話の面白さの評価及び分析. 知能と情報 (日本知能情報ファジィ学会誌), Vol. 35, No. 3, pp. 683–692, 2023.

[15] Haruo Kubozono. On the universality of mora and syllable (features on theories of syllable and mora). *Journal of the Phonetic Society of Japan*, Vol. 2, No. 1, pp. 5–15, 1998.

[16] 秋永一枝. 日本語の音節 (拍) は幾つか. 講座日本語教育, 第 5 分冊, pp. 11–21, 1969.

[17] Taku Kudo. Mecab : Yet another part-of-speech and morphological analyzer. 2005.

[18] 伝康晴, 小木曾智信, 小椋秀樹, 山田篤, 峯松信明, 内元清貴, 小

磯花絵. コーパス日本語学のための言語資源: 形態素解析用電子化辞書の開発とその応用. 日本語科学, Vol. 22, pp. 101–123, 2007.

- [19] 国立国語研究所コーパス開発センター. 『現代日本語書き言葉均衡コーパス』短単位語彙表 (version 1.0). <https://repository.ninjal.ac.jp/records/3234>.
- [20] Tianyu Zhao and Kei Sawada. rinna/japanese-gpt2-small.
- [21] Kei Sawada, Tianyu Zhao, Makoto Shing, Kentaro Mitsui, Akio Kaga, Yukiya Hono, Toshiaki Wakatsuki, and Koh Mitsuda. Release of pre-trained models for the Japanese language. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation*, pp. 13898–13905, 2024.
- [22] Aaron Grattafiori, Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Alex Vaughan, Amy Yang, Angela Fan, and et al. The Llama 3 herd of models, 2024.
- [23] Kazuki Fujii, Taishi Nakamura, Mengsay Loem, Hiroki Iida, Masanari Ohi, Kakeru Hattori, Hirai Shota, Sakae Mizuki, Rio Yokota, and Naoaki Okazaki. Continual pre-training for cross-lingual LLM adaptation: Enhancing Japanese language capabilities. In *Proceedings of the First Conference on Language Modeling*, 2024.
- [24] Naoaki Okazaki, Kakeru Hattori, Hirai Shota, Hiroki Iida, Masanari Ohi, Kazuki Fujii, Taishi Nakamura, Mengsay Loem, Rio Yokota, and Sakae Mizuki. Building a large japanese web corpus for large language models. In *Proceedings of the First Conference on Language Modeling*, 2024.