

オノマトペ関連語辞書の 逐次的精製とその利用に関する研究

総合情報学研究科
知識情報学専攻

インタラクションデザインの理論と実践

13M7105

上間 大生

目次

1	序論	1
1.1	本研究の背景	1
1.2	オノマトペの工学的利用に関する研究	1
1.3	解くべき課題	3
2	関連研究	5
2.1	オノマトペ辞書に関する研究	5
2.2	オノマトペに関する書籍	7
2.3	先行研究の問題点	8
2.4	本研究の位置付け	8
3	提案手法	9
3.1	提案	9
3.2	デザイン指針	10
4	実装	15
4.1	オノマトペ関連語対応関係の取得	15
4.2	オノマトペ関連語の抽出	19
4.3	オノマトペ関連語辞書の参照	21
5	オノマトペ類似性の逐次的収集	27
5.1	関連語によるオノマトペ類似性	27
5.2	オノマトペ類似性データの精錬	31
6	検証	32
6.1	オノマトペ関連語辞書の対応評価	32
6.2	Twitter を用いた結果との比較	35
6.3	関連語テンプレートによる類似オノマトペ精錬の検証	38
7	オノマトペ関連語辞書の利用	40
7.1	関連語辞書の閲覧方法	40
7.2	オノマトペ類似関係の精錬	40
8	議論	43
8.1	提案手法の有効性	43
8.2	提案手法の問題点と解決策	43
8.3	関連語用例文を用いたオノマトペの取得	44
9	結論	46

A	付録	48
A.1	6 章で行った関連語テンプレートによる精錬の検証結果	48

1 序論

本章では、本研究の実施に至った背景を説明し、対象とする課題を明確にする。

1.1 本研究の背景

近年、様々な分野においてオノマトペへの関心が高まっている。オノマトペとは擬音語・擬態語のことである。擬音語とは「さらさら」、「ざあざあ」、「わんわん」といったような実際の音をまねて言葉とした語であり、擬態語とは「にやにや」、「ふらふら」、「ゆったり」のように視覚・触覚など聴覚以外の感覚印象を言語音で表現した語である [14]。日本語オノマトペは、主に口語でのコミュニケーションに用いられており、4500 語以上あるとされている [12]。オノマトペはその親しみやすさから食品の食感やイメージを伝える手段として商品名 (e.g., 「もちもちパン」¹ 「ふわふわスフレ」² 「ふんわり鏡月」³ など) にも利用されており、多くの日本人にとって身近な表現であることが伺い知れる。またオノマトペは、人が持つ感覚的な概念をアウトプットする手段として横断的な利用を目的に注目を集めており、研究の分野においてもオノマトペを専門的な知識や操作を必要とするコンピュータの入力に用いることで、より人が持つ感覚に近い操作を可能にするツールとして、オノマトペをコンピュータの入力に利用する研究が進められている。

以下ではそのようなオノマトペを活用した工学的研究を紹介する。

1.2 オノマトペの工学的利用に関する研究

コンピュータシステムの入力手段にオノマトペを利用することで、ユーザがシステムを利用した際のシステムユーザの利便性を向上させるアプローチについていくつかの研究が行われている。

小松らが開発した「もやもやドロ잉」は、画像編集ソフトウェアにおけるエフェクトの選択時にオノマトペの直感的な印象を利用することで、システムの対象ユーザとなるエフェクトについて詳しくない初心者が画像編集を行う際の支援を目指してシステムが構築されている [11]。このシステムでは、ユーザがオノマトペに込めたイメージを画像中の所望の対象にエフェクトとして反映させることができるドロ잉ツールとなっている (図 1.1 参照)。写真の編集に使うエフェクトの印象を、画像処理の専門的知識 (この場合、エフェクト名) を持たない人が表現することを可能にするため、オノマトペを利用してそれを実現している (図 1.2 参照) これはオノマトペが持つ、曖昧な感覚を言葉として表現することができる点を活用した研究であるといえる。

Kanwipa らは料理レシピを食に関するオノマトペをクエリに用いて検索することができるレシピ検索システムに関する研究を行っている [2]。この研究では、日本において料理や味覚をオノマトペを用いて表現することが多いという点に着目し、料理レシピページから料理に関する用語 (材料, 調理法, 食感) とオノマトペの共起度を収集することで料理とオノマトペの対応を

¹<http://www.showa-sangyo.co.jp/special/oyatsu/mochimochi.html>

²<https://www.yamazakipan.co.jp/product/04/souffle.html>

³<http://www.suntory.co.jp/sho-chu/kyougetsu/funwari.html>



図 1.1: もやもやドローイング (文献 [11] より図引用)

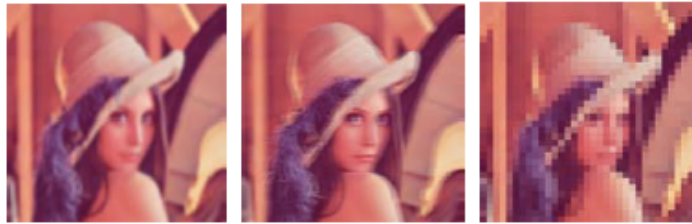


図 1.2: オノマトペによる画像エフェクトの変化 (文献 [11] より図引用)



図 1.3: オノマトペロリ

調べる取り組みが行われている。

その結果をもとに構築されたシステム「オノマトペロリ」では、検索クエリとして与えられるオノマトペを関連の強い他の語に置き換えて検索(オノマトペが「ピリッ」だった場合、共起度の強かった「唐辛子」や「山椒」)する手法と、レシピとオノマトペの関連度を用いてレシピ検索を行う手法の2種類の検索アプローチでレシピを検索することができる(図 1.3)。

また、システムの入力にオノマトペを利用するだけでなく、システムの出力に用いた研究も存在する。

寺島らはデジタルコミック制作における動的表現の手段の1つとして、デジタルコミックのコマ内に書き文字として描かれたオノマトペである音喩に着目し、効果線を用いて音喩の活字を動かす表現を生成するシステム(以下、キネティックオノマトペシステムと記す)を考案した[9]。デジタルコミック制作者を対象ユーザとして、ユーザの「効果線を引く」という動作を入力に用いて、動く音喩を生成するシステムを提案している(図 1.4)。このシステムではユー

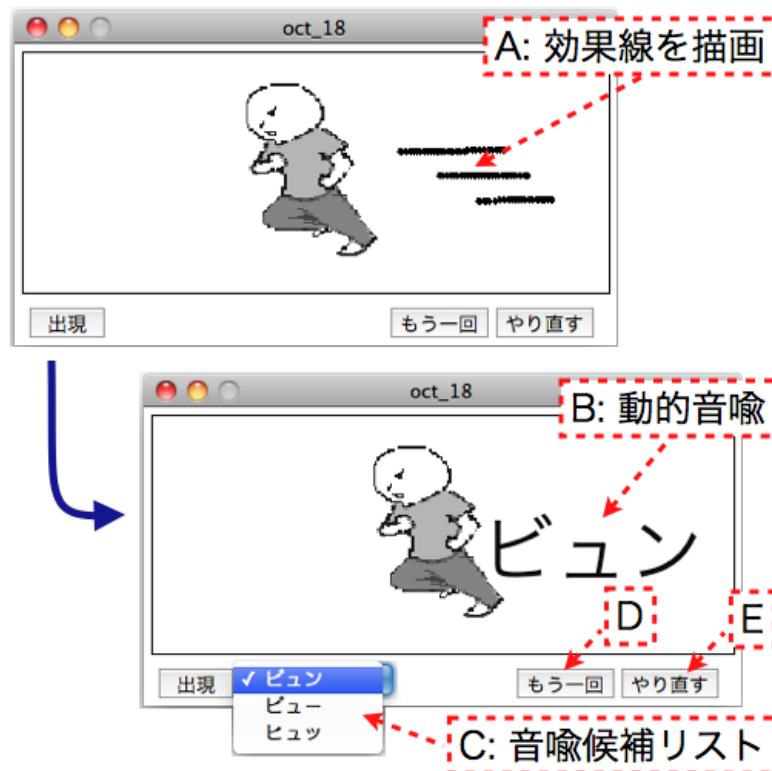


図 1.4: 寺島システム (文献 [9] より図引用)

ザがシステム内に効果線を描画する (図 1.4 中 A) と、その効果線の線種と描画速度に基づいて、関連する音喩を出現させ描画した線の形と描画速度に沿って動くようになっている (図 1.4 中 B)。

効果線の線種と描画速度の組み合わせが複数の音喩と重複している場合は、効果線描画後に音喩候補リストが出現し、そのリストの中からユーザが表示させたい音喩を選択することができる (図 1.4 中 C)。このシステムを実装するためには効果線の種類と描画速度の組み合わせにオノマトペを対応付ける必要があるため、システム構築前に事前被験者アンケートを実施し、それらが人々の間でどのように対応付けられているのか、またその対応づけは被験者間で共通しているのかを明らかにしている。

1.3 解くべき課題

このように画像エフェクトの効果とオノマトペ、料理とオノマトペ、効果線とオノマトペといった多様な利用方法を目的として様々なオノマトペを用いたシステムが提案されている。

しかし、これらオノマトペを用いるシステムには共通の問題点が存在する。それは、システム構築の際、製作者の主観のみでオノマトペとシステムの動作を対応させた場合、ユーザが意図する入出力オノマトペと動作の対応が必ずしも一致するとはいえないという点である。システム制作者が考えるオノマトペの利用と、ユーザが考えるオノマトペのイメージが異なった場合、ユーザがオノマトペを入力しても意図した操作をできない可能性がある。

この原因として、オノマトペの多義性が考えられる。

内田らによれば、日本語オノマトペは複数の意味を持つことが多く、オノマトペ単体での意味の分類は難しいとされている [7]。例えば日本語オノマトペに「ちゅーちゅー」というものがある。これはネズミの鳴き声の擬音語として絵本などでよく用いられている。しかしこの「ちゅーちゅー」には「吸う音」を表す意味も存在する。よって「ちゅーちゅー吸う」と「ちゅーちゅー鳴く」では同じ「ちゅーちゅー」を用いていても文の意味が異なってしまう。このような理由から、システム内でオノマトペを利用する場合、人によってオノマトペとその入出力に対する動作の対応イメージが異なる場合があると考えられる。このことを考慮に入れ、入出力に用いるオノマトペとシステム内の動作が一致しているかどうかを確かめるために、多くの研究ではシステム構築に際して被験者アンケートが行われている。

しかし、事前に被験者アンケートを行うことでアンケート作成や実施に時間を要するため、このことがオノマトペを工学的に利用した研究の妨げになっている可能性がある。このような問題はオノマトペを用いる研究において共通の課題であるため、オノマトペの内容を示す辞書があれば効率的に研究を行うことに繋がるのではないかと考えられる。

また、システム制作後に被験者アンケートを行った結果、システム内で用いられているオノマトペと動作の対応が被験者間で大きく異なった場合にはシステムを作り直す必要があり、無駄な時間を要してしまうという問題も発生する可能性がある。

そこでオノマトペを利用したシステムを構築するために、オノマトペをどのように利用することが適切なのか、どのオノマトペがどんな「動き」や「状況変化」と対応しているのかを調べる事ができる辞書が必要になると考える。オノマトペがどのような概念と対応しているのかを予め調べることで、システムに用いるオノマトペを適切に選択することができ、製作者とシステムユーザの考えるオノマトペのイメージの違いが生じる問題を解消できるのではないかと考えた。

2 関連研究

本章ではオノマトペ辞書に関する研究について述べる。筆者はオノマトペ辞書に関する研究をオノマトペ用例収集に関する研究と、オノマトペ辞書生成に関する研究の大きく 2 つに分類できると考えている。ここでは筆者が考える 2 つの分類に分けて先行研究を紹介する。

2.1 オノマトペ辞書に関する研究

2.1.1 オノマトペ用例収集に関する研究

オノマトペにどのような使われ方があるのかを調べるために、オノマトペを含む用例文を参照する方法が挙げられる。しかし、オノマトペはその全てが一般的な日本語辞典に紹介されていない。そこで Web 等を用いてオノマトペ用例文を収集する研究が行われている。

本節では、これまでに行われたオノマトペ用例文収集に関する研究を紹介する。

オノマトペは新語・造語が多いと言われており [4]、この観点から多種多様な情報が更新・追加される環境として Web を利用したコーパス¹の収集手法が数多く考案されている。

Web を利用したオノマトペ用例文の収集に関する研究として、平田らによる Twitter を用いてオノマトペ記述データを収集するシステムが挙げられる [13]。この研究では Twitter を用いてオノマトペ用例文を収集するために、お題を Tweet してフォロワーに答えてもらうという形式をとっている。

具体的には、システムが日本語で特定の表現をオノマトペを含む用例文として投稿するようフォロワーに示唆する Tweet を自動的に一定時間おきに投稿する。そして、フォロワーからお題の回答となるオノマトペを含む文章をリプライで受け取ることでオノマトペ記述データを収集する取り組みがされている。オノマトペは口語表現でよく用いられるという点において、人にオノマトペを含む文章を記述してもらうというこの手法は、オノマトペの使い方を集める手法としてふさわしいと言える。また SNS である Twitter を使うことで広範囲に渡って用例文を収集できるという利点がある。

しかしこの手法では、提示されたお題に対するリプライから用例文を収集する手法であるため、事前にフォロワーを集めておく必要があり、かつフォロワーがお題に沿った文章を必ずツイートしてくれるとは限らない。そのため、多くの用例文を収集することは難しい。

それに対して、オノマトペを含む Tweet を要求する形ではなく、Twitter に投稿された全 Tweet からオノマトペを含む文章を抽出することで、多くのオノマトペ用例文を収集しようとした研究も存在する。上間らは Twitter を用いてオノマトペを取得し、その用例文を用いてオノマトペと単語の共起関係からオノマトペとどのような概念・事象・対象が対応しているのかを調べる研究を行った [6]。

この研究では電子書籍の特徴である動的コンテンツの利用が可能であるという点に着目し、デジタル絵本における動的表現付与の入力手法にオノマトペを用いたインタラクティブオノマトペ絵本を提案している (図 2.2 参照)。

そのようなオノマトペを入力手段として用いたシステム実装の問題として、オノマトペとシステム内の動作の対応をどのように対応付けることが適切であるか判断が難しいことを挙げて

¹ コンピュータ上で扱うことができる大規模な言語データのこと。

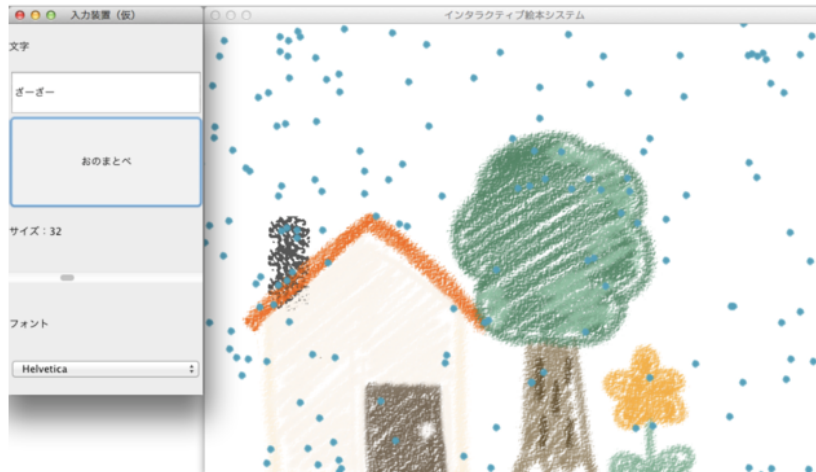


図 2.1: オノマトペ絵本 (文献 [17] より図引用)

	A	B	C	D	E	F	G
1	ちよこちよこ	ふらふら	うるうる	ぶらぶら	ばたばた	さっさ	ととと
2	する	する	する	する	する	さる	する
3	てる	てる	てる	てる	てる	する	てる
4	ある	なる	いる	行く	いる	寝る	く
5	なる	いる	できる	くる	なる	いる	いる
6	いる	くる	なる	いる	くる	てる	おおう
7	やる	行く	くる	ぶる	ある	帰る	やる
8	くる	れる	行く	ある	行く	ほる	なる
9	見る	寝る	ある	なる	く	やる	くる
10	思う	思う	れる	帰る	やる	くる	ある
11	れる	ある	思う	思う	れる	終わる	れる
12	行く	すぎる	見る	終わる	ちゃう	行く	とる
13	出る	帰る	買う	見る	思う	く	みる
14	みる	飲む	みる	買う	できる	なる	行く
15	いく	歩く	やる	れる	いく	れる	いう
16	言う	やる	帰る	食べる	すぎる	せる	おる
17	買う	見る	言う	みる	せる	入る	思う
18	くれる	く	すぎる	いく	終わる	ある	いく
19	く	言う	ちゃう	く	寝る	いく	見る
20	できる	ちゃう	探す	歩く	帰る	言う	せる
21	ちゃう	食べる	いう	やる	見る	いう	買う
22	寝る	いく	来る	すぎる	しまう	見る	ちる
23	せる	いう	く	言う	いう	くれる	寝る
24	いう	みる	いく	ちる	忘れる	思う	言う
25	来る	しまう	しまう	来る	言う	すぎる	すぎる
26	入る	終わる	くれる	いう	くれる	来る	終わる
27	食べる	くれる	わかる	ちゃう	来る	みる	ふる

図 2.2: 「歩く、走る」に関するオノマトペの関連語知識ベース

いる。そしてこの研究では、Twitter を用いて「歩く、走る」に関するオノマトペを含む Tweet を収集しその Tweet 内単語の共起回数からオノマトペが対応している概念・事象・対象などを調べることでオノマトペ関連語知識ベースを制作し、システム内の動作とオノマトペを対応付ける手法を提案している (図 2.1 参照)。

しかし、この手法には非常にノイズが多く、関連語とは思えない語が共起語リスト共起回数順の上位に出現していた。その原因として、最大文字制限 140 字で気軽に記述される Tweet には話し言葉が多く、単なる単語の羅列や主語述語が存在しない文などが含まれており、非文になっているものも多かったことが挙げられる。

Twitter と同じように自由記述文を集める手段として、ブログを用いる方法も考えられる。ブログは制限文字数が大きく、より自由に記述することができるため Twitter と比べると非文が少ない可能性がある。

内田らはブログ記事からのオノマトペ用例文の自動抽出手法を行っている [8]。この研究では日本語オノマトペの大規模なコーパス構築を目指して、大量のデータを低コストで入手できる点、現代日本語の用法を反映している点、新聞等のコーパスに比べて感情表現が含まれやすい点などに着目して、ブログを用いて用例文を自動収集する手法を提案している。提案手法では 299 語のオノマトペをクエリとして用例文を収集しており、41315 件分のブログ記事に適用した結果、15437 文の用例文を抽出している。オノマトペを含む文内における、オノマトペの後続要素と係り受け分析を行った結果をもとに、オノマトペを含む文の特徴を割り出し、オノマトペ用例文自動収集適合率 96.2% という高い結果を得ている。

また、大量の文書集合や語彙資源を必要としない点で高い適合率を達成したことがこの研究の利点だといえる。

2.1.2 オノマトペの辞書生成に関する研究

次にオノマトペの辞書生成を目的とした研究について述べる。

奥村らは、Web 上のオノマトペが含まれる文章を使ってオノマトペの概念を自動で収集し、記録する手法を提案している。この研究では、機械学習を用いて生成したオノマトペ候補語を、Web で検索することで用例文を収集してオノマトペの含まれるテキストコーパスを集め、用例文を抽出することでオノマトペ概念辞書を生成している [4]。

この研究の着眼点として、数多い日本語オノマトペに対応するために Web からオノマトペの含まれる文章をテキストコーパスとして収集している点が挙げられる。前述したように、日本語オノマトペは約 4500 語あるとされており、それら全てのオノマトペの関連語を収集するためには多くのオノマトペを含むコーパスが必要になると考えられる。また新しいオノマトペやオノマトペの新しい使い方が生まれる可能性があるため、それらに対応していくためには日々情報量が増加する Web 上のテキストコーパスを利用することが好ましいといえる。またオノマトペ用例文を収集するために、機械学習を用いてオノマトペの候補になり得る語、オノマトペ候補語の生成を行っている。これはオノマトペのコンピュータ上にオノマトペとして扱える断定的な文字列が存在しないために用いられた手法であるが、新語・造語が多いとされるオノマトペの未知語を探し出す手法として適切な方法だといえる。

2.2 オノマトペに関する書籍

オノマトペとその内容を調べる方法として、オノマトペの内容を示した日本語オノマトペ辞典 [12] が出版されている。この辞書には日本語オノマトペが約 4500 語記載されており、各オノマトペが擬音語であるか、擬態語であるかの分類や、どのような音・様子を表すオノマトペであるかを紹介している。また、2470 語のオノマトペを自然・人間・事物に分類しており、各オノマトペごとに意味・用法の違いを詳細に紹介している。

しかし、各オノマトペがどのようなものの動きや状況と対応しているのか、全ての関連する語が書かれているわけではない。また新しいオノマトペが生まれた時、オノマトペの新しい使い方が生まれたときに意味を調べることができない。

そのような関連語からオノマトペを検索する手法として、インターネットサービスとして goo

が提供しているオノマトペディア²[16]がある。この辞書は検索したいオノマトペをクエリとしてテキストフォームに入力することで、そのオノマトペを含む用例文を閲覧することができる。またオノマトペでない語句を検索し、関係するオノマトペの用例文を閲覧することもできる。

しかし、どの語句がオノマトペと関係しているのか一覧することはできず、関連語ごとの関連の度合いなどを知ることはできない。

2.3 先行研究の問題点

オノマトペが示す内容を明らかにするために、これまでに日本語オノマトペの意味や用法を明らかにするために多くの用例文収集研究が行われてきた。

しかし、オノマトペは多義語が多いという特徴を持ち、オノマトペの意味を調べるために収集された用例文は複数の文章に様々な語句を含むことが考えられるため、多種多様な文章からオノマトペの意味を判断することは簡単ではないといえる。また一つのオノマトペを含む文章が複数ある場合、用例文を参照しただけでは検索者が調べるオノマトペがどのような意味を持つか判断できない可能性がある。

例えば「うろうろ」というオノマトペは「歩く」、「走る」という概念に分類されている [12]。しかし、実際には歩くことを表す「うろうろする」という表現もあり、この場合「する」という動詞と「うろうろ」というオノマトペは関係していると言える。

既存の用例文辞書を閲覧するだけでは、「歩く」という動きを表す単語と関連するオノマトペである「うろうろ」が関係しているということを文章から読み取り、判断しなければならない。

またオノマトペを用いたシステム構築など、オノマトペとどのような事象が対応しているのか調べるために使う際、自分が用いたいオノマトペと概念が一般的に使われるものかどうか調べるために、該当する用例文を探さなければならない。

2.4 本研究の位置付け

以上をふまえ、用例文から語句を抽出し、単語単位でのオノマトペとの語句の関係を明らかにしておくべきではないかと考えた。オノマトペと対応する事象や状況の違いを、関連する単語として対応付けておくことができれば、オノマトペがどのような事象を表しているのかの判断が容易に行えるようになると考えている。

人は普段見聞きするオノマトペから、その対象や状況の違いを連想する。その連想される対象を示す単語がオノマトペの関連語ではないかと考えられる。このオノマトペの関連語を取得する手段として、人が書いたオノマトペが使われている文章から収集できるのではないかと考えた。

本研究ではオノマトペとそのオノマトペが含まれる文章の共起語との関係を明らかにし、オノマトペと対応する関係語として記録しておく「オノマトペ関連語辞書」の構築を目指す。

²<http://dictionary.goo.ne.jp/onomatopedia/>

3 提案手法

本章では提案手法について述べる。

3.1 提案

3.1.1 オノマトペ関連語辞書

我々は2章で述べたような問題に対応するため、オノマトペの関連語を探し出し、その対応関係を記すオノマトペ関連語辞書の構築を目指す。また 4500 語にも及ぶとされているオノマトペを網羅するために、オノマトペ関連語の自動収集に適した手法について検討を行う。オノマトペ関連語辞書を整備することで、オノマトペの工学的利用を円滑に行えるようになると考えている。

本来、オノマトペの「意味」を記述するデータベースが存在すれば、問題は解決するよう思える。しかし、上記したようにオノマトペそのものが曖昧な表現であり、明確な「意味」を定義することは難しい。日本語オノマトペ辞典 [12] ではオノマトペの意味・用法について解説されており、その内容を説明するために用例を用いている。しかし意味や用法がわかるが、オノマトペを用いる事象や使い方を全て知ることはできない。

そこで我々のアプローチとして、機械的に判別できる単語単位での収集、記録の対象として「関連語」を対応づけることとした。オノマトペと対応する関連語を記録しておくことで、関連語からオノマトペを調べ出すことができるようになる。また関連語との対応付けにより、オノマトペと共によく用いられる事象やその対象となる概念を関連語から知ることができるようになると考えている。

例えば、「うろうろ」というオノマトペは「歩く」という概念に分類されている [12]。しかし、現状では「歩く」という語句からどのようなオノマトペがその概念に該当するか、どのオノマトペが対応しているかを明らかにするためには被験者アンケートで確かめる方法以外はない。また、「うろうろ」がどのような事象を示しているのかを調べ、「歩く」という単語がそれに当たると断定するに至る場合にも同じことが言える。これがオノマトペの工学的利用の妨げになっていると考えられる。そこで「うろうろ」「歩く」というようにオノマトペと関連語の対応を記録した辞書があれば、オノマトペから関連語、関連語からオノマトペを調べ出すことが出来る。ただし、このオノマトペ関連語辞書を手作業で日本語オノマトペ 4500 語全てに対応させることは困難である。そこで、提案するオノマトペ関連語辞書を自動で生成する手法について検討を行う。

3.1.2 オノマトペの文脈

オノマトペが使われる文にはオノマトペと他の単語を含めた文脈が存在すると考えられる。

例えば「どんぶらこ」というオノマトペがある。このオノマトペを聞くと、おそらく多くの人が川上から桃が流れてくる様子を想像すると思われる。このオノマトペは有名な童話「桃太郎」で用いられており、「川上からどんぶらこ、どんぶらこと大きな桃が流れてきました。」といったように「桃が川を下る」様子を表す際に用いられるオノマトペである。

しかし、このオノマトペが他の文脈で用いられることはほとんどなく、「どんぶらこと小枝が流れる」といったように対象を桃から他にものに変えて使うことや、「桃がどんぶらこと落ち

る」といったようにその行為を変えて用いることもない。どんぶらこは「桃がどんぶらこと流れる」という文脈で用いるのが適切だといえる。

このように「桃」が“流れる”という“対象”と“行為”の組み合わせがオノマトペのりよう適切性に影響すると考えられる。

本研究において取得したいと考えているのは、上記したオノマトペとその対象及び、行為との関連であり、オノマトペと関連語を対応づけて記録するオノマトペ関連語辞書の構築を目標としている。その一環として、まずどのような単語がオノマトペ用例文における共起単語とその共起頻度の観点から関連語の取得を試み、オノマトペと関連語の一対一の対応づけを行う。

3.1.3 オノマトペの類義語・対義語

オノマトペは事象のニュアンスを伝える手段として使われることがあり、このとき使い分けられるオノマトペは類義語、対義語として分類することができると考えられる。そこで本研究で構築したオノマトペ関連語を基に、オノマトペが持つ細かなニュアンスの違いを関連語の観点から知ることができないかと考えた。

例えば、オノマトペ関連語辞書内のあるオノマトペの関連語に「雨」という単語があったとすると、そのオノマトペは「雨」と関連するオノマトペである可能性がある。そのとき、他のオノマトペに「雨」が関連語として含まれていたとすると、それぞれ「雨」と共起したそれらオノマトペはなんらかの概念で類似している可能性がある。このようなオノマトペ同士の関係を関連語の関係から調査し、提案するオノマトペ関連語辞書の関連度指標として利用できないか検討する。

3.2 デザイン指針

本節では提案するオノマトペ関連語辞書の生成手法について概説し、生成における各段階ごとに手法と考慮すべき点について解説する。

本研究ではオノマトペから関連語の検索、関連語からオノマトペの検索を行うことができるオノマトペ関連語辞書の構築を目指す。また、オノマトペと関連語の関連度を共起頻度で表すことで、各オノマトペと関連語の関連度合いを共起頻度から知ることができると考えている。これにより、どのような単語がオノマトペとより関連度が強いかわ知ることができ、オノマトペを用いたシステム製作時に使用するオノマトペと対象の関連を決める際に役立つと考える。

アプローチとしては、文章内に含まれるオノマトペと共起する他の単語の頻度情報を取り、共起頻度の高い単語がオノマトペとの関係性を持っているのかを調べる。そして、辞書として記録しておくことことで、オノマトペから関係語を調べられる仕組みづくりを行う。

簡単に手順を説明する(図 3.1 参照)。

まずオノマトペが使われている文章を使うため、各オノマトペごとに用例文を収集し用意する。次にオノマトペが使われている文を抽出し、その共起語を取得する。共起語は共起回数を取得しておき、そのオノマトペが使われた文章数で割ることで共起頻度を求める。その後、オノマトペごとに各共起語とその共起頻度を記録しておく。

本研究ではこの共起頻度をオノマトペ関連語辞書で扱う「関連度」として扱う。

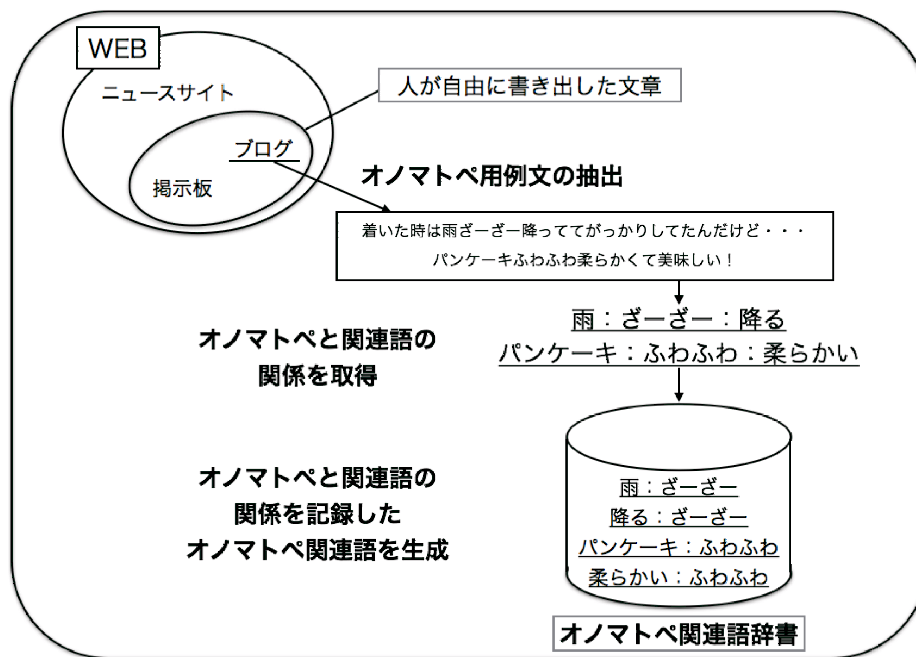


図 3.1: オノマトペ関連語辞書の生成

3.2.1 コーパス

まずはじめに、オノマトペ用例文を収集する必要がある。利用するテキストコーパスは、4500語以上あるとされている日本語オノマトペに対応することを考慮すると、量の多いものを用意する必要があるといえる。そこでオノマトペを含むテキストコーパスを入手するために、他の先行研究に倣って Web を利用することとした。

次に、Web 上のどのような文章にオノマトペが含まれているのか予測調査を行った。その結果、大手新聞社によって運営されているサイトやその他のニュースサイトなど、機関を通じて公的に発信されている情報にはオノマトペがほとんど用いられていないことが分かった。これは、オノマトペが口語でのコミュニケーションに用いられやすい曖昧な表現であることが原因だと考えられる。

その一方で個人が自由に記述するブログなどではよく利用されていることが判明した。

そこで我々は、オノマトペは人が話す際に利用されるという点に注目して、人が略式なコミュニケーションを行うために書きだした文章のうち、オノマトペの含まれるものを調べることで、オノマトペと他の単語の関係性を調べることができると推測した。

これまでにミニブログサービスである Twitter を用いて用例文を収集し、Tweet 内共起語の共起回数を取得することでオノマトペの関連語が取得できないか調査を行ってきた [6]。

Twitter は利用者数が多い点、口語文が多く用いられている点が特徴である。そのため、新語オノマトペや既存オノマトペの新しい使われ方も発見できるのではないかと考えた。また Twitter の特徴として、一度に投稿できる文章の文字数が最大 140 字に制限されていることが他の一般的なブログサービスと異なっていることが挙げられる。ユーザは制限文字数内で表現したい内容を書かなければならないため、1 つの tweet に含まれるの内容が複数になることが少ない傾向がある。例えば「歩く」に関するオノマトペが文章内に含まれていれば、「誰が歩く



図 3.2: アメーバブログの検索結果

のか」「どこを歩くのか」といったオノマトペの対象となる単語が含まれるはずである。そのため、オノマトペが文章内に含まれていればそのオノマトペに関する語句が含まれている可能性が高く、オノマトペの関連語を抽出できると考えた。この研究の結果、共起頻度の上位に関連語が出現した。

しかし、コーパスとして利用した Twitter には最大 140 字の字数制限があるためにユーザが書いた文章は要約されている場合が多いことから文章構造として正しくないものも多かった。その上 Tweet 内共起を調べたために関連していないであろう語句も共起回数の上位に出現してしまい、関連語の算出結果にノイズが非常に多かったという点が問題として挙げられている。

また、この研究では「歩く、走る」に関するオノマトペ 56 語の収集しかしておらず、利用できるオノマトペは 39 語と非常に少ないため、様々なオノマトペを用いたシステムで用いるには網羅性が低い。オノマトペ知識ベースも共起語の共起回数で順をつけてはいるが、オノマトペごとにコーパスが異なるため、共起回数を指標として使うことはできない。

このような問題点を解決するために、今回は用例文の収集対象としてブログを用いることとする。インターネット上で広く普及しているブログもユーザ数が多く、個人が自由に扱うことができるため口語表現が用いられる場合も多い。また、ブログには字数制限が無い文章量が多く、長文の記述を取得する事ができる。そのためより正確な構造の文章を取得できると考えられる。

収集対象としてアメーバブログ¹を用いることとした。(図.3.2) アメーバブログの利用者数は 2014 年 9 月時点で 4000 万人を突破しており²、累計記事投稿数は 19 億を超えるとされている。このことから多くのユーザが様々なテーマでブログを書いていると推測でき、オノマトペを含む文章も多いと予想される。

以上の理由から、アメーバブログを対象としてオノマトペ用例文の収集を行う。

¹<http://www.ameba.jp>

²<https://www.cyberagent.co.jp/news/press/detail/id=9281&season=2014&category=ameba>

3.2.2 オノマトペと関連語の判定

本稿では、オノマトペとその関連語の取得を行う。関連語が関連していると判断する基準として、文内共起を用いることとした。その理由は、オノマトペが用いられている文に共起している単語はそのオノマトペと関係している可能性が高いと考えたからである。

例としてオノマトペ「ぎーぎー」を用いて解説する。

「ぎーぎー」は強い雨が降るときに用いられるオノマトペであり、日本語オノマトペ辞典には「雨がげげしく降る音」として記載されている [12]。例えば、この「ぎーぎー」というオノマトペを含む「雨がぎーぎー降る」という文章が収集できたとすると、この文内に共起した「雨」と「降る」を関連語だと考えられる。この文を「雨, が, りーりー, 降る」といったように形態素解析を行い、単語の切り分けをすることができる。今回はこれらの単語の内、単語単体で意味を持つ「名詞、動詞、形容詞、副詞」の 4 種の品詞について抽出を行い、その共起頻度を求めることとする。

3.2.3 オノマトペの類似関係の取得

生成したオノマトペ関連語辞書はオノマトペをクエリとしてその関連語を、また何らかの語句をクエリとしてそれと関連するオノマトペを検索することができる。ある特定語を検索に掛けた際の検索結果として、その語と共起していたオノマトペが複数出現した場合、それらのオノマトペは検索語を通じた類似語である可能性があると考えられる。

例えば「雨がぎーぎー降る」という表現と「雨がしとしと降る」という表現がある場合、前者は「強い雨」、後者は「弱い雨」を表現している。この時「ぎーぎー」と「しとしと」というオノマトペはどちらも「雨」という名詞や、「降る」という動詞に関するオノマトペだと言える。加えてこれらのオノマトペは「雨が降る」という概念において、強度の違いを表す類似語であると言える。

このように、生成したオノマトペ関連語辞書に含まれる名詞や動詞などの語句を元に、オノマトペ類似関係が取得できるのではないかと考えた。

以下にそのような関連語の抽出に関する研究を紹介する。

渡部らは関連語が登録されたシソーラスを人手で構築することが非常にコストがかかる問題であることを背景として、Web を大規模なテキストコーパスとみなし、検索エンジンを用いて関連語を抽出する手法を提案している [10]。この研究では、関連語をある語に対して関係の強い語と定義し、関連性のタイプによって分類することができるとしている。その上で、関連語を得たい語 (対象語) の文章内における出現パターンを取得し、そのパターンを検索にかけ、検索エンジンの検索結果として提示されるスニペットをもとに対象語と同じパターンで共起する単語を取得し、関連語を抽出している。

対象語の出現パターンを取得する点において、筆者らが取り組むオノマトペ関連語からオノマトペ類似語を取得する手法と類似しており、オノマトペにおいてもオノマトペと関連語の出現パターンを利用して関連オノマトペの取得ができる可能性がある。

また、渡部らは関連語の例として Miller らによる名詞同士の関係を挙げ、同義、大義、上位-下位、全体-部分という 4 つの関係の定義を紹介している [3]。名詞の場合はこのような関係性で関連語が存在している。オノマトペも関連語の対応を明らかにできれば、今後オノマトペの利用において役立つのではないかと考える。

加藤らは Web 上にある文章から語の共起を用いてその類似関係を取得する研究を行っている [15]。この研究では単語 A, B, C が在り、単語 D を求めるとき、単語 A, B を含む文章で共起頻度が高かった単語が、単語 C を含む文章で出現するとき、単語 B の類似語 D を取得できるのではないかという仮定を検証している。本研究においてもオノマトペ関連語の関係から類似オノマトペの取得を行うことができるのか検証するため、参考とする。また、この研究では Web から文章を取得する際に検索エンジンの検索結果から得られるタイトルとスニペットをもちいて、単語用例文を取得している。この手法であれば直接用例文が含まれる個別の Web ページを参照する必要がないため非常に効率的であり、本研究を進める上でも参考とする。

4 実装

本章では実装について各段階ごとに詳細を述べる。

提案システムでは大きく分けて 3 つのステップでオノマトペと関連語の対応関係を取得する。

まず 1 つ目が Web からオノマトペを含む文章のコーパスを収集する段階である。2 段階目に収集したコーパスからオノマトペ用例文のみを取得する。3 段階目にオノマトペ用例文から関連語の抽出を行う。

4.1 オノマトペ関連語対応関係の取得

4.1.1 オノマトペ用例文収集

本提案手法ではオノマトペをもとに文章検索を行い、用例文を取得する。そのためクエリとなる電子的に扱うことができるオノマトペ文字列が必要となる。

多くのオノマトペに対応するためにはオノマトペを含む文章を自動で取得できる仕組みが必要となり、本研究においても検討を行ったが、今回は人手で用意することとした。オノマトペ文字列の自動取得については次節で詳しく述べる。

今回は、日本語オノマトペ辞典 [12] にある意味分類別さくいん欄に記載された語のうち p 33 から p 55 に記載されている、大きく 5 つに分類された 1537 語のオノマトペを人手でテキストファイルに打ち込むことでオノマトペ文字列を用意した (表 4.1 参照)。この 1537 語は分類ごとのオノマトペ語数を合計した値であり、分類中のオノマトペに一部に重複を含むことがわかった。この重複を除くと 1188 語であった。

提案システムではオノマトペ文字列ごとにオノマトペ用例文を収集し、生成したコーパスをオノマトペの名前を付けたファイルで管理する。その時点で重複名のファイルは上書きされるため、重複したオノマトペは消える。システムでは、改行区切りでオノマトペを記述したテキストファイルを読み込み、そのテキストファイルから取得したオノマトペ文字列をクエリとして用例文の収集を行うよう実装した。

ブログ記事からオノマトペを含む文章を抽出する方法として、アメーバブログ検索結果のページに含まれるスニペット¹を用いることとした。予備調査を行ったところ、アメーバブログの検索結果にスニペットが表示されており (図.3.2)、このスニペット内にオノマトペを含む一文が表示されていることがほとんどだったため、このスニペットを取得し、オノマトペ用例文の抽出を行うこととした。

アメーバブログの Web サイトにはブログのキーワード検索があり、このページの検索 URL を編集し検索したいクエリを指定することで検索結果のページを直接取得することができる。このとき、API などを用いることなくブログの検索ページの URL をクエリとしてブログの検索を行うことができる。また一度の検索につき 100 件でしか結果を取得できないが、検索回数の制限は存在しない。そのため辞書生成のために多くの語句を検索にかけることができ、また簡易な方法で文章の検索を行うことができる。このような理由からアメーバブログ検索 URL を用いて検索結果を取得することとした。

¹検索エンジンの検索結果を表示する際に、検索にマッチした結果の一部を要約して表示するもの

表 4.1: オノマトペ辞典における分類

大分類	小分類	オノマトペ語数
温度に関する	暑い、暖かい	24
	寒い冷たい	16
火・土に関する	火、燃える、燃やす	19
	土、土砂、岩石	20
感情・感覚に関する	慌てる、もがく、落ち着かない	39
	元気がない	29
	痛む	43
	泣く	47
	驚く	38
	怒る、不機嫌、無愛想	40
	思う、感じる	20
	躊躇う、怯む	10
	笑う	75
	喜ぶ	11
水・液体に関する	流れる、垂れる、注ぐ	24
	波立つ、泡立つ、沈む	23
	濡れる	19
	湿る、にじむ	17
	水滴、滴る、落ちる、跳ねる	22
性格・性質に関する	元気な	17
	平気、平然	33
体格・姿に関する	太った、頑健な	30
	痩せた	9
・	・	・
・	・	・

・ ・	・ ・	・ ・
天気に関する	雨、雪、氷 風、吹く 照る、晴れる	27 32 15
動作・状態に関する	歩く、走る 吐く、もどす 働かない 震える 言う、話す 見る、見える、睨む 眠る、寝る 飲む、酔う 起きる、立つ 騒ぐ 咳をする、むせる 食べる、噛む、舐める 疲れる	58 11 15 34 82 52 26 67 21 32 16 71 13
動き・変化に関するオノマトペ	進む、運ぶ 落ちる 折れる ゆれる 切る、切れる 破る、破れる 貼る、くっつく 刺す 曲がる、しわがよる	45 72 33 49 52 18 21 28 22
	計	1537 語

一つのオノマトペあたりの結果取得数は、アメーバブログの一度の検索につき取得できる結果数が最大 100 件までであるため、100 件取得している。しかし、オノマトペを検索にかけ取得できた結果数が 100 件に満たなかったものも存在した。オノマトペごとのコーパス数が大きく異なってしまうと偏りがでてしまい、語句の共起頻度を算出するにあたり望ましくない。また、文章数が少ない場合には共起したものが一組であったとしても共起頻度は高くなってしまふ。そこで、本研究では取得オノマトペ用例文の数が 20 以下だったものは結果を算出しないこととした。

尚、コーパスの取得期間は 2014 年 12 月から 2015 年 1 月であった。

4.1.2 クエリとなるオノマトペについて

クエリとなるオノマトペの設定方法について、4500 語以上あるとされているオノマトペに対応するため、オノマトペ自体の自動収集が行えないか検討した。

オノマトペ用例文の自動収集手法における先行研究として、奥村らが行ったオノマトペ概念辞書の自動生成に関する研究を挙げる。この研究では日本語オノマトペの文字列による型分類に着目し、その型に当てはまるオノマトペ候補語を生成し、その語句をもとに用例文を自動収集を行っている。

オノマトペには一定のパターンがあるものが多く、XYXY 型 (e.g., “がたがた”, “もぐもぐ”)、AつB り型 (e.g., “がっかり”, “ひょっこり”) などの型に分類することができる。このパターンにそって平仮名を組み合わせることでオノマトペ候補語を生成し、候補語をクエリとして検索エンジンを用いて Web 上のテキストを検索し、用例文を取得した [4]。この研究の利点として、事前にクエリとなるオノマトペ文字列を自動生成することで用意しているため、オノマトペ文字列を用意する必要がない点が挙げられる。

しかし、この手法では型に合わせてひらがなの組み合わせを行うため数が膨大になることが予想され、オノマトペでない語も全て検索にかけることになるため膨大な時間を要する。また現在は検索エンジンの検索 API に制限を設けている場合が多く、制限内で非常に多くのパターンを全て検索にかけると多くの時間を要するため、短時間でオノマトペを含む文章を多く集める目的には適していない。ブログを用いてオノマトペ用例文を収集する本提案手法においては検索対象となる Web サイトサーバへの負担が大きくなり現実的でない。

そこで、本研究の提案手法であるオノマトペをクエリとして用例文を収集し、関連語を取得するという手法に着目し、一部のオノマトペに共起した関連語を元に、その関連語を含む用例文からオノマトペを取得できるのではないかと考え、予備調査を行った。

調査には日本語オノマトペ辞典の意味分類別さくいんに記載された、天気に関するオノマトペ欄の「雨、降る」という分類語句を用いた。調査方法は、クエリとして分類語句「雨、降る」(オノマトペの関連語)を用いて、アメーバブログの検索エンジンに「雨」と「降る」を入力して and 検索を行い、オノマトペが含まれる文章を収集できるかを調べる。調査はウェブブラウザを用いてアメーバブログの検索欄に関連語入力して検索し、検索結果の一部を参照してオノマトペが含まれるかを目視で確認した。調査は 2014 年 11 月に行った。

50 件程度のスニペットを調査した結果、検索結果に「雨、降る」に関するオノマトペを見つけることはできなかった。逆の手順として、「ぎーぎー」というオノマトペを検索した際に「雨」や「降る」を含めた文章を取得することは可能であるため、全検索結果の中にオノマトペ

が全く入っていないということはないと考えられる。しかし、オノマトペを指定しない状況でオノマトペを含む文章を探し出すことは難しく、また、そこにオノマトペが入っていたとしてもオノマトペ文字列をオノマトペとして認識することができない。

本研究ではオノマトペの関連語を取得するため、オノマトペの含まれる用例文が取得できていなければならない。今回はシステムのユーザが任意のオノマトペを入力し、そのオノマトペに関する関連語が即座に取得できるような仕組みづくりを行うことで、初期オノマトペを用意する必要がなくなるため、今回はオノマトペを人手で用意することとした。

4.2 オノマトペ関連語の抽出

次にオノマトペごとに取得できた用例文集合を元に関連語の抽出、共起頻度の算出を行う。

語の共起頻度を関連度とするため、文章取得数が極端に少なかったものに関しては共起していると言えない可能性がある。そのため、今回は閾値を 20 とし、スニペット数が 20 を上回る場合のみ、共起頻度を算出することとした。

4.2.1 一文判定

先に取得したオノマトペ用例文集合から関連語の取得を行うが、収集したスニペットから出来る限りオノマトペに係る単語のみを含む文を切り出す必要がある。実際に取得できたスニペットを用いて説明する。

オノマトペ「ざーざー」取得で取得出来たスニペットの例を以下に示す。

「ざあざあ」を含むスニペット集合

週末はまたもや紅葉見て来ました。今度は高尾山!高尾山着いた時は雨ざーざー降っててがっかりしてたんだけど、山頂着いた瞬間にキレイに晴れた!天気が良いと紅葉の赤が映えるなー(^~^) ケーブルカー使ったけど少し登山もしたので、運動出来...

...ったとよ!そして この夜の打ち上げで食べた甘いお醤油のお刺身がちょ〜美味しかったとよ♪ホテルに戻り一泊した次の朝、ようやくあの方が遅れて九州に到着したとよ。雨雲さんが。ざーざー降りなんとよ!リーダーが予定していた観光コース...

... たい!!ってグズグズ...体力消耗してないから寝れなくてグズグズ...眠いなら寝なさいって怒ったらこの顔(笑)寝れないもんって甘えちゃん雨もやんだから着替えて出ようとしたらまたざーざー降りだしてめっちゃ怒ってました夕方にやっと買い...

上記のスニペット集合のうち、一つ目のスニペットを見ると、

週末はまたもや紅葉見て来ました。今度は高尾山!高尾山着いた時は雨ざーざー降ってて
 がっかりしてたんだけど、山頂着いた瞬間にキレイに晴れた!天気が良いと紅葉の赤が映え
 るなー(^~^) ケーブルカー使ったけど少し登山もしたので、運動出来...

オノマトペ「ざーざー」を含む一文は下線部となり、この一文を取り出す必要がある。

そこで特定の記号表現を使い、スニペットの中からオノマトペを含む文だけを取り出せない
 かと考えた。オノマトペを中心としてその前後のどこかに記号が含まれていた場合にその記号
 で文を切り、オノマトペを含む文を取り出すこととした。

使った記号は“。 , ? , ? , ! , ! , . , 「 , 」 , 半角スペース” (“ ” と , を除く) とした。

オノマトペを含む文を切り出したものを以下に示す。

「ざーざー」用例文集合

高尾山着いた時は雨ざーざー降っててがっかりしてたんだけど、山頂着いた瞬間にキレイ
 に晴れた

ざーざー降りなんとよ

眠いなら寝なさいって怒ったらこの顔(笑) 寝れないもんって甘えちゃん雨もやんだから着
 替えて出ようとしたらまたざーざー降りだしてめっちゃ怒ってました夕方にやっと買い

このとき、3つ目の文を見ると“(笑)”という表現があり、文が切れている事がわかる。この
 ことから括弧記号である“(”と “) ”も判定に使用すると複文を分けることができると考えら
 れるが、“(”と “) ”は文内で使われることも多いため、今回は除外した。

また“、 ”であるが、他の用例文を見てみると“、 ”で文が切れているその前後でオノマトペ
 と関連しているであろう語が分かれている場合も多くあった。そのため“、 ”も除外すること
 とした。

このようにオノマトペを含む一文を切り出して生成した用例文集合を元に、単語の共起頻度
 を取得する。

4.2.2 共起頻度

次に共起語の共起頻度を算出する手法について説明する。まず4.2.1節で生成したオノマト
 ペ用例文を形態素解析器 MeCab [1] にかかけ、分かち書きを行う。このとき、オノマトペ文字列
 「ざーざー」が「ざー、ざー」のように2語に切り分けられてしまうことが判明した。そこで、
 4.1.2節で述べた検索クエリとなるオノマトペのリストを元に MeCab ユーザ辞書にオノマトペ
 文字列を登録しておき、分かち書きに用いることとした。

「ざーざー」を含む用例文の分かち書き

高尾山, 着いた, 時, は, 雨, ざーざー, 降っ, て, て, がっかり, し, て, た, ん, だ, けど, , , 山頂, 着いた, 瞬間, に, キレイ, に, 晴れた

分かち書きを行った後、単語ごとに形態素解析を行い、品詞が「名詞、動詞、形容詞、副詞」に当たる単語を原型で抽出する。

用例文の単語 (原型)

高尾山, 着く, 時, は, 雨, ざーざー, 降る, てる, て, がっかり, する, てる, た, ん, だ, けど, , , 山頂, 着く, た, 瞬間, に, キレイ, に, 晴れる, た

各文に出現した単語を抜き出し、オノマトペごとの用例文集合に何回出現しているかをカウントする。

これら一連の処理をオノマトペ用例文集合の一行毎に行い、共起語ごとの共起回数を用例文数で割ることで共起頻度を算出する。

出力形式はオノマトペ毎に共起語、出現頻度、品詞を共起頻度降順にリスト化し、CSV ファイルで出力することとした。

4.3 オノマトペ関連語辞書の参照

4.3.1 各オノマトペの関連語参照

生成したオノマトペ関連語辞書の一例として、オノマトペ「ざーざー」の関連語辞書を示す(表 4.2, 4.3, 4.4, 4.5)。名詞、動詞については語数が多いため上位 30 語を、形容詞と副詞は単語数が少なかったため全てを表示している。

本論における表としては品詞ごとに分けて表示しているが、実際のデータは一つの csv ファイルに上から名詞、動詞、形容詞、副詞の順で記録している。また、各品詞毎に共起頻度順でソートして記録しており、各オノマトペごとに全関連語を収録している。

オノマトペ「ざーざー」の関連語は「雨」や「降る」が考えられ、オノマトペ辞典にも「天気に関するオノマトペ」の内「雨、降る」に分類されるオノマトペとして収録されている。

提案手法で生成したオノマトペ関連語辞書の共起名詞の内、「ざーざー」ともっとも共起頻度が高かった名詞は「雨」であり、その共起頻度は 0.788 であった。2 位の「日」は共起頻度が 0.141 であり、その差は 0.647 と大きく開いている。このことからオノマトペ「ざーざー」と「雨」は文内で特に多く共起していると言え、関連語が取得できたといえる。

次にもっとも共起頻度が高かった単語は「する」であり共起頻度は 0.282 であったが、2 位には「降る」が出現しており、その共起頻度は 0.270 である。その差は 0.011 であり頻度に大きな開きはなく、上位に関連語が出現していると言える。

このように共起語の共起頻度を指標として、オノマトペと関連語を参照することができる。

表 4.2: 「ぎーぎー」に関する名詞 (上位 30 語) 表 4.3: 「ぎーぎー」に関する動詞 (上位 30 語)

共起名詞	共起頻度	品詞	共起動詞	共起頻度	品詞
雨	0.788235294118	名詞	する	0.282352941176	動詞
日	0.141176470588	名詞	降る	0.270588235294	動詞
今日	0.141176470588	名詞	てる	0.223529411765	動詞
中	0.117647058824	名詞	なる	0.164705882353	動詞
朝	0.0941176470588	名詞	いる	0.105882352941	動詞
私	0.0941176470588	名詞	降りる	0.0823529411765	動詞
笑	0.0823529411765	名詞	出る	0.0705882352941	動詞
月	0.0705882352941	名詞	行く	0.0705882352941	動詞
外	0.0705882352941	名詞	やる	0.0588235294118	動詞
前	0.0588235294118	名詞	れる	0.0588235294118	動詞
こと	0.0588235294118	名詞	ある	0.0588235294118	動詞
土曜日	0.0470588235294	名詞	くる	0.0470588235294	動詞
降り	0.0470588235294	名詞	寝る	0.0470588235294	動詞
みんな	0.0470588235294	名詞	着く	0.0470588235294	動詞
昨日	0.0470588235294	名詞	来る	0.0352941176471	動詞
みなさん	0.0470588235294	名詞	言う	0.0352941176471	動詞
ぶり	0.0470588235294	名詞	思う	0.0352941176471	動詞
明日	0.0470588235294	名詞	滑る	0.0235294117647	動詞
引っ越し	0.0352941176471	名詞	やむ	0.0235294117647	動詞
年	0.0352941176471	名詞	くれる	0.0235294117647	動詞
時間	0.0352941176471	名詞	曇る	0.0235294117647	動詞
よう	0.0352941176471	名詞	おる	0.0235294117647	動詞
今	0.0352941176471	名詞	終わる	0.0235294117647	動詞
ライブ	0.0352941176471	名詞	すぎる	0.0235294117647	動詞
感じ	0.0352941176471	名詞	折り畳む	0.0235294117647	動詞
傘	0.0352941176471	名詞	いく	0.0235294117647	動詞
店内	0.0352941176471	名詞	持つ	0.0235294117647	動詞
韓国	0.0235294117647	名詞	見る	0.0235294117647	動詞
店長	0.0235294117647	名詞	怒る	0.0235294117647	動詞
出勤	0.0235294117647	名詞	食べる	0.0235294117647	動詞
:	:	:	:	:	:

表 4.4: 「ぎーぎー」に関する形容詞 (上位 30 語) 表 4.5: 「ぎーぎー」に関する副詞 (上位 30 語)

共起名詞	共起頻度	品詞	共起動詞	共起頻度	品詞
楽しい	0.0588235294118	形容詞	しとしと	0.0588235294118	副詞
ない	0.0588235294118	形容詞	ちょっと	0.0235294117647	副詞
寒い	0.0470588235294	形容詞	なかなか	0.0235294117647	副詞
赤い	0.0352941176471	形容詞	まだ	0.0235294117647	副詞
眠い	0.0235294117647	形容詞	全然	0.0235294117647	副詞
強い	0.0235294117647	形容詞	ちょうど	0.0235294117647	副詞
いい	0.0235294117647	形容詞	グズグズ	0.0235294117647	副詞
悪い	0.0235294117647	形容詞	ぽつん	0.0235294117647	副詞
良い	0.0235294117647	形容詞	ずっと	0.0235294117647	副詞
すごい	0.0235294117647	形容詞			

4.3.2 単語をクエリとした関連オノマトペ検索

関連語 1 語による検索

またオノマトペ関連語辞書では、オノマトペではない単語、関連語を元に検索を行うことができる。例として、「雨」と「降る」をそれぞれクエリとして検索し、共起したオノマトペを図 4.1, 4.2 に示す。単語による検索結果の順は、単語とオノマトペの共起頻度順になっている。

「雨」の結果を見ると「ぎーぎー」や「しとしと」、「どしゃどしゃ」「ざんざん」などの雨の様子を表すようなオノマトペが結果の上位に出現していることが分かる。また「降る」の場合も同様に「しとしと」「どしゃどしゃ」「ざんざん」「ぎーぎー」と降る様子を表すオノマトペが上位に出現している。オノマトペ関連語辞書のユーザはこのように、オノマトペとの対応が知りたい単語をクエリとして検索することで、関連語から関連オノマトペを調べることが可能となる。

またこの結果を見ると、「雨」や「降る」に関係して「ぎーぎー」や「しとしと」など「雨が降る」様子の程度のニュアンスを表すような、意味が似たオノマトペが取得できていることが分かる。

一方で、下位に表示されたオノマトペを見ると、まったく異なるオノマトペが含まれてもいた。「雨」の検索結果下位を見ると「かんかん」「うろうろ」「ぐにゃ」などが出現しており、一方「降る」では「もくもく」「うらら」「ぽかぽか」など関係ないであろうオノマトペが出現している。

しかしながら上記した通り、「雨」「うろうろ」各語の関連オノマトペ上位には同じオノマトペが出現している。このことから、それら 2 語が同時に使われた文章において共起しているオノマトペが上位に出現したと考えられる。

そこで、単語による検索では 2 語検索を行えるよう実装を行った。

関連語 2 語による検索

以下に「雨、降る」の 2 語を用いて検索を行った結果を示す (図 4.3)。

ここで問題になるのが、この結果は本研究で用いたオノマトペ用例文の一文内において 2 語

```

-----雨-----
ざーざー 0.788235294118
しとしと 0.778947368421
どしゃどしゃ 0.6
ざんざん 0.510204081633
じとじと 0.466666666667
ざんざ 0.395833333333
ぼつぼつ 0.242424242424
ぐちょぐちょ 0.20202020202
ぐしょぐしょ 0.175257731959
じめじめ 0.161290322581
ばらばら 0.151515151515
どんより 0.135416666667
ざぶざぶ 0.134146341463
びちょびちょ 0.131313131313
べちょべちょ 0.123711340206
びしゃびしゃ 0.12
びちゃびちゃ 0.118279569892
びしょびしょ 0.103092783505
べちゃべちゃ 0.1
ずくずく 0.0845070422535
むわっ 0.0752688172043
ぶらぶら 0.0721649484536
ばらっ 0.0707070707071
びゅー 0.0666666666667
びゅーびゅー 0.063829787234
ひっ 0.0631578947368
ぼたぼた 0.06
こんこん 0.0597014925373
ぼたり 0.0561797752809
じゃーじゃー 0.0537634408602
ばらり 0.0531914893617
じみじみ 0.0520833333333
えっちらおっちら 0.0515463917526
びっちより 0.0510204081633
びちゃびちゃ 0.0510204081633
こくん 0.0510204081633
むしむし 0.0487804878049
びんしゃん 0.046511627907
しなしな 0.0425531914894
ぼとり 0.0422535211268
ぐっしょり 0.0421052631579
びゅーびゅー 0.0421052631579
びゅんびゅん 0.0412371134021
ぐにゃ 0.0408163265306
うろろろ 0.040404040404
かんかん 0.0379746835443

```

図 4.1: クエリ「雨」による検索の結果

```

-----降る-----
しとしと 0.421052631579
どしゃどしゃ 0.357142857143
ざんざん 0.30612244898
ざーざー 0.270588235294
ばらばら 0.262626262626
ざんざ 0.239583333333
どさどさ 0.2125
ぼたぼた 0.18
ぼつぼつ 0.171717171717
どかどか 0.15306122449
こんこん 0.119402985075
どんより 0.114583333333
ちらちら 0.11
じとじと 0.0933333333333
ちらほら 0.0909090909091
わっさわっさ 0.0877192982456
こんもり 0.0808080808081
びゅーびゅー 0.0736842105263
ちらりちらり 0.0729166666667
びゅんびゅん 0.0721649484536
びちゃびちゃ 0.0714285714286
さくっ 0.0666666666667
はらはら 0.0612244897959
さぶさぶ 0.0606060606061
ばらっ 0.0606060606061
びしゃびしゃ 0.06
べちゃべちゃ 0.06
どさっ 0.0571428571429
ざぶざぶ 0.0487804878049
しゃんしゃん 0.046511627907
ぶつぶつ 0.0454545454545
じめじめ 0.0430107526882
あたふた 0.0412371134021
べちょべちょ 0.0412371134021
びしょびしょ 0.0412371134021
ずんずん 0.0412371134021
ふー 0.0384615384615
びくびく 0.032967032967
しゃーしゃー 0.0326086956522
じゃーじゃー 0.0322580645161
むわっ 0.0322580645161
びゅーびゅー 0.031914893617
どぼどぼ 0.0315789473684
ぽかぽか 0.03125
うらら 0.03125
ぐしょぐしょ 0.0309278350515
もくもく 0.0309278350515

```

図 4.2: クエリ「降る」による検索の結果

```

-----雨-----
ざーざー 0.788235294118
しとしと 0.778947368421
どしゃどしゃ 0.6
ざんざん 0.510204081633
じとじと 0.466666666667
ざんざ 0.395833333333
ぼつぼつ 0.242424242424
ぐちょぐちょ 0.20202020202
ぐしょぐしょ 0.175257731959
じめじめ 0.161290322581
ばらばら 0.151515151515
どんより 0.135416666667
ざぶざぶ 0.134146341463
びちょびちょ 0.131313131313
べちょべちょ 0.123711340206
びしゃびしゃ 0.12
びちゃびちゃ 0.118279569892
びしょびしょ 0.103092783505
べちゃべちゃ 0.1
ずくずく 0.0845070422535
むわっ 0.0752688172043
ぶらぶら 0.0721649484536
ばらっ 0.0707070707071
びゅー 0.0666666666667
びゅーびゅー 0.063829787234
ひっ 0.0631578947368
ぼたぼた 0.06
こんこん 0.0597014925373
ぼたり 0.0561797752809
じゃーじゃー 0.0537634408602
ばらり 0.0531914893617
じみじみ 0.0520833333333
えっちらおっちら 0.0515463917526
びっちょり 0.0510204081633
びちゃびちゃ 0.0510204081633
こくん 0.0510204081633
むしむし 0.0487804878049
びんしゃん 0.046511627907
しなしな 0.0425531914894
ぼとり 0.0422535211268
ぐっしょり 0.0421052631579
びゅーびゅー 0.0421052631579
びゅんびゅん 0.0412371134021
ぐにゃ 0.0408163265306
うろうろ 0.040404040404
かんかん 0.0379746835443

```

図 4.3: クエリ「雨, 降る」による 2 検索の結果

が同時に共起した結果ではないという点である。

提案手法ではオノマトペごとのオノマトペ用例文集合を一つのコーパスとし、それに含まれる単語の共起回数を数えることでオノマトペとの共起頻度を算出している。そのため、各用例文の一文内でオノマトペ以外のどの単語とどの単語が共起しているかの関係は考慮していない。

図 4.3 に示した 2 語検索は、第 4.3.1 節で示したオノマトペと関連語辞書 (表 4.2, 4.3, 4.4, 4.5) 内に検索にかけた 2 語が両方含まれている場合、結果を表示している。

以上、オノマトペを含む用例文を収集し、オノマトペとその関連語とその共起頻度を記録したオノマトペ関連語辞書を生成方法を述べた。オノマトペ関連語辞書に登録されたオノマトペは 1188 語であった。またオノマトペ関連語辞書の参照について、その実装と参照方法について述べ、取得できた関連語にどのようなものがあつたかについて述べた。

5 オノマトペ類似性の逐次的収集

5.1 関連語によるオノマトペ類似性

実装したオノマトペ関連語辞書では、オノマトペ用例文に含まれた単語全てを抜き出し、オノマトペと各共起語の一対一対応で辞書を生成している。そして4では例として「雨」「降る」という単語それぞれを検索にかけ、結果を示した。

そして、それぞれの語の関連オノマトペ検索結果に同様に含まれるオノマトペ (e.g., “しとしと”, “ぎーぎー”) と、そうでないもの (e.g., “ぐっしょり”, “どさどさ”) があることがわかった。

前者のオノマトペの場合、その表現例としては「雨がしとしと」「しとしと降る」のようにどちらでも使うことができる。一方で後者は「雨でぐっしょり」というような表現はあるが「ぐっしょり降る」という表現はない。また「雨がどさどさ」という表現はないが「どさどさ降る」という使い方は適切であると考えられる。

「雨」「降る」のどちらの結果にも出現したオノマトペに関しては、第3.1.2節で述べたように、オノマトペには特定の単語との組み合わせが存在し、「雨－オノマトペ－降る」のような「名詞, オノマトペ, 動詞」の三つ組になっている場合がある。

またそのような「名詞－オノマトペ－動詞」の組が存在する時、オノマトペだけが他のオノマトペと入れ替えることが出来る場合がある。この時、名詞と動詞の間に入るいくつかのオノマトペは、その前後に位置する関連名詞と関連動詞の組で関連した類似語である可能性がある。

例えば、名詞「雨」と動詞「降る」の単語の組があるとき、「ぎーぎー」や「しとしと」、「ぼつぽつ」など複数のオノマトペがこの組に当てはまる。

このとき、名詞と動詞の組はオノマトペを挟むテンプレートとなる (以下、関連語テンプレート)。このオノマトペの対象 (名詞) と行為 (動詞) テンプレートを発見することで、類似するオノマトペを判断できるのではないかと考えた。

そこで「名詞－オノマトペ－動詞」の形で記述されたオノマトペ用例文にどのようなものがあり、どのような組があるのか調査を行った。

「ぎーぎー」を含む文

今日は雨が“ぎーぎー”降ってて外に出られなかった

このようなオノマトペ用例文があるとき、オノマトペより前に出現している名詞「今日」「雨」とオノマトペより後ろに出現している動詞「降る」「出る」などを全て取得し、

「今日－ぎーぎー－降る」

「今日－ぎーぎー－出る」

「雨－ぎーぎー－降る」

「雨－ぎーぎー－出る」

このように「名詞－オノマトペ－動詞」を対応付け、これらを関連語テンプレートとした。その後これらのような組み合わせがオノマトペごとの用例文集合に何回出現しているかを算出した。

その結果、取得できた組み合わせは膨大な数に上り、また回数の上位には「今日－オノマトペ－いる」「日－オノマトペ－する」など、意味を判断することが困難な組み合わせが多く出現したため、関連語テンプレートを発見するに至らなかった。

そこで、オノマトペ用例文の内、オノマトペと関連する名詞、動詞がどのように使われているのか、助詞なども含めて詳細に調査し、文内でオノマトペが使われるルールを探すこととした。

「ざーざー」を含む文

今日は 雨 が “ざーざー” 降って て外に出られなかった

その結果、上記のように「名詞, 助詞(が), オノマトペ, 直後の動詞」の順でオノマトペと関連語が用いられている文章が多かった。

そこで、

「名詞 が オノマトペ 動詞」

の型に当てはまる文を探し、その名詞と動詞を抽出することで、オノマトペを挟むテンプレートを抽出し、そのテンプレートがオノマトペごとのスニペット内に何回出現しているか、出現回数を数えた。

しかし、出現回数が少ないものが多く、一度しか出現していないテンプレートがほとんどであった。また出現数が同じものも多く、どのテンプレートが特徴的であるか判断できないため、今回は 5 回以上出現した組み合わせのみを関連語テンプレートとして扱うこととした(表 5.1 参照)。

その結果 5 回以上出現した関連語テンプレートは 15 組であった。この内、「の－オノマトペ－起きる」という組み合わせが出現したが、「の」は名詞としてどのような意味をなすか判断できないため除外した。

このように、いくつかのオノマトペで関連語テンプレートを発見することができた。

しかし、これらの組み合わせに入るオノマトペは多いとは言えず、また上記以外の他のオノマトペも入りうることが考えられる。

そこで、第 4.3.2 節で紹介した関連語 2 語検索の結果を利用できるのではないかと考えた。

オノマトペ関連語辞書の 2 語検索は上記したように用例文中の「名詞－動詞」の関係を取得しているわけではないが、オノマトペを元に関連した「名詞」「動詞」をそれぞれ複数個取得出来ている。よって、擬似的に「名詞－オノマトペ－動詞」関係を作り出すことができる。

この関係を用例文から直接取得したオノマトペが入りうる「名詞－動詞」の組み合わせをもとにオノマトペ関連語辞書で 2 語検索を行い、合わせて提示することで、関連語テンプレートを元にした類似オノマトペ取得の逐次的収集が行えるのではないかと仮定した。

表 5.1: 用例文から抽出した関連語テンプレート

名詞	オノマトペ	動詞
胃	むかむか	する
雨	どしゃどしゃ しとしと ざんざ ざんざん ざーざー	降る
関係	ぎく	する
胸	むかむか	する
心	ほかほか ざわざわ	する
心臓	ばくばく	する
雪	どさどさ ぼたぼた	降る
雪	どさどさ	落ちる
頭	くらくら	する
疲れ	どっと	出る
鼻	むずむず ひくひく	する
風	びゅーびゅー びゅんびゅん びゅーびゅー そよそよ	吹く
目	ちかちか しょぼしょぼ ぎょろぎょろ ぎょろっ ぱちくり	する
Tシャツ	じっとり	する
(除外) の	むっくり	起きる

表 5.2: 関連語テンプレート雨、降るに含まれていたオノマトペ

名詞	オノマトペ	動詞
雨	しとしと ざーざー ざんざ ざんざん どしやどしや	降る

表 5.3: 関連語テンプレート雨、降るをもとに関連語辞書 2 語検索を行った結果

名詞	オノマトペ	動詞
雨	しとしと ざーざー ざんざん ざんざ じとじと どしやどしや ぽつぽつ ぱらぱら どんより ぽたぽた	降る

例として、用例文内に 5 回以上出現していた関連語テンプレート「雨 – 降る」に含まれていたオノマトペと、その関連語テンプレート「雨」「降る」を用いてオノマトペ関連語辞書に 2 語検索をかけた結果のうち、上位 10 件の組をそれぞれ提示する (表 5.2, 5.3 参照)。

用例文から直接見つけた「雨 – オノマトペ – 降る」の関係は 5 組であり (表 5.2 参照)、オノマトペ関連語辞書から「雨」と「降る」に関係するオノマトペは 10 組であった (表 5.3 参照)。

用例文から抽出した関連語テンプレートに加え、関連語辞書を用いることで足りない類似オノマトペを追加することができる。

このことから、用例文から抽出した関連語テンプレートと、そのテンプレートを用いてオノマトペ関連語から 2 語検索を行った結果得られた疑似関連語テンプレートを合わせることで、類似するオノマトペの逐次的収集が行えると言える (図 5.1 参照)。

しかし、これらを合わせた結果の中にはその組み合わせとして適切といえないノイズが含まれている。

そこで、我々はこの関係性の絞り込みをもとに、人によるフィードバックを行うことで、関連語テンプレートを用いたオノマトペ類似性データの精錬を行えないかと考えた。

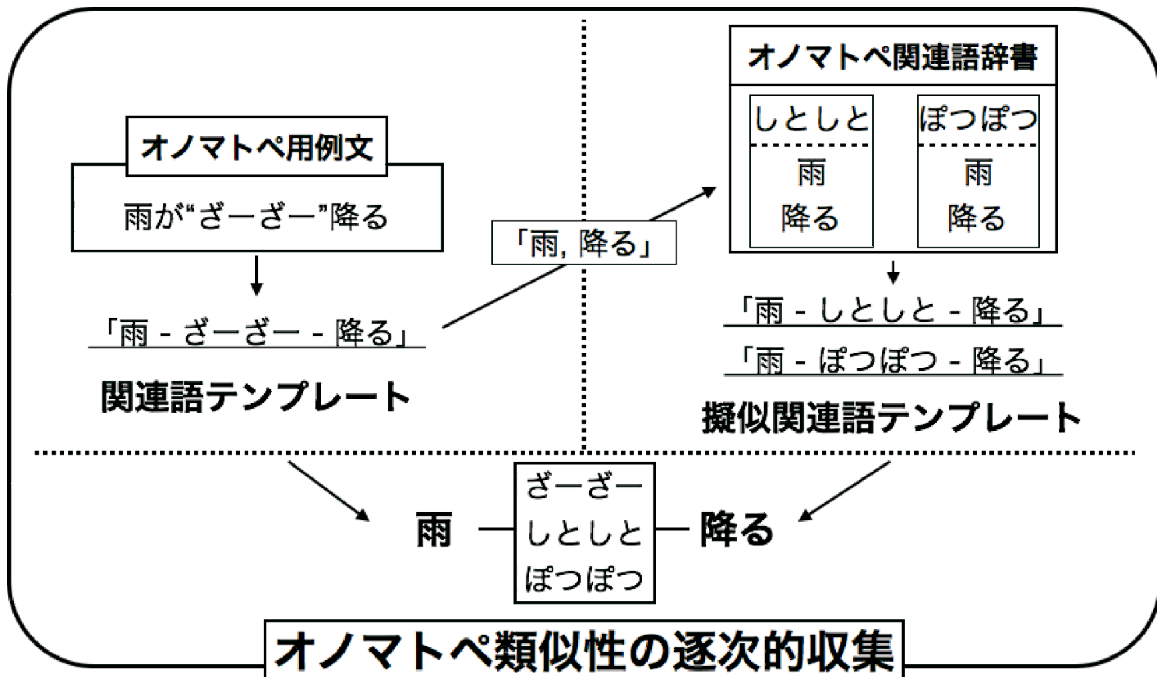


図 5.1: オノマトペ類似語の逐次的収集

5.2 オノマトペ類似性データの精練

ここまで機械による自動判断でオノマトペ関連語辞書生成、関連語によるオノマトペ検索、関連語テンプレートによる類似オノマトペの絞り込みを行ってきた。しかし、オノマトペとその関連語には正解データがなく、機械のみによる完全な自動判断を行うことはできない。そこで統計的数値によるある程度の絞り込みをもとに、最終判断として「人」を用いることを考えている。

そのような人による判断を元にデータの精練を行う手法としてレリバンスフィードバック、適合フィードバックという概念が存在する。

レリバンスフィードバックとは J.J.Rocchio が提唱した概念であり、文書検索において初期検索結果が検索者の情報要求に沿っているかどうかを評価し、その評価をもとに検索クエリを修正して次の検索を行う手法である。この仕組みをもとに繰り返し検索とクエリ修正を行うことで情報要求に沿った結果が提示されるようになる [5]。

本研究は文書検索を行うものではないが、このレリバンスフィードバックを参考にオノマトペと関連語テンプレートの組を提示し、人が見て自分の感覚に沿う組み合わせであるかどうかを判断し、その結果をもとにオノマトペと関連語テンプレートの組み合わせに重み付けすることで精練を行うことができるのではないかと考えた。

そこで、本当に人がオノマトペと関連語テンプレートの組み合わせを見て、組み合わせに違和感を覚えるものを判断できるのかどうか、アンケート形式で検証した。

検証の詳細については 6 章で述べる。

6 検証

本章では、前章で提案した手法について検証を行う。

6.1 オノマトペ関連語辞書の対応評価

提案手法ではオノマトペ文字列をもとにオノマトペ用例文を収集し、集めた用例文から関連語を取得している。実装を行ったオノマトペ関連語の自動収集システムを動作させ、関連語とされる単語の取得が行えているのか、検証を行う。

6.1.1 検証手続き

本研究では日本語オノマトペ辞典 [12] から意味分類別さくいんに記述されているオノマトペ 1537 語を利用し、オノマトペと関連語の対応を取得した。分析にあたり、日本語オノマトペ辞典の分類語句を関連語と仮定し、取得された関連語の中にこれらの分類に用いられる語句が含まれているかを検証する。含まれていれば関連語を取得できたことが確認できる。

分析に用いるデータであるが、量が非常に多いため一部に限り行う。今回は動きに関するオノマトペに分類された「歩く、走る」に関するオノマトペ 56 語を分析対象とする。

提案手法で取得した「歩く、走る」に関するオノマトペ 56 語の関連語を参照し、動詞「歩く」「走る」がそれぞれ含まれているのかを確認する。

尚、第 4.1.1 節で解説したように、関連語取得時に共起の前提としてオノマトペ用例文取得数が 20 以下だったものについては関連語の取得を行っておらず、関連語辞書内に結果が存在しない。「歩く、走る」に関するオノマトペのうち“かっぽかっぽ”、“しゃりりしゃりり”、“たかたか”、“だだーっ”、“どっしどっし”、“とっとっ”のオノマトペ 5 語は用例文数が 20 文に満たなかったため、分析から除外することとし、残る 51 語での検証を行う。

検証の方法は、第 4.3.2 節で紹介した単語によるオノマトペ検索を用いて、「歩く」「走る」をそれぞれ検索にかけ、出現したオノマトペがオノマトペ辞典の 51 語にどれだけ該当するのかを調べるものである。

オノマトペ辞典の「歩く、走る」に関するオノマトペの分類では「歩く」に関するものと「走る」に関するオノマトペを一纏まりで紹介しているため、分類の 2 語をそれぞれ検索し、合わせて評価することとした。

6.1.2 検証の結果

検証結果を表 6.1 に示す。「歩く」「走る」の結果、出現したオノマトペに丸をつけている。分析から除外した取得用例文数が 20 に見たなかったオノマトペは「除外 (20 以下)」、関連語の取得は行えていたが「歩く」「走る」が共起していなかったものは「共起なし」と表示している。

2 語それぞれの検証結果、検証に用いた 51 語中 36 語が関連語として取得できていた。一方で取得できていなかったオノマトペは 15 語であった。

6.1.3 新たな関連語とオノマトペの対応

検証のために「歩く」「走る」でオノマトペを検索した結果、提示された関連オノマトペの中にオノマトペ辞典の分類に記載されていないが、「歩く」「走る」と関係すると思われるオノマ

表 6.1: 「歩く、走る」に関するオノマトペの関連語検証

オノマトペ	歩く	走る	用例文数	共起していない
うろうろ	○			
えっちらおっちら	○			
かつぽかつぽ			除外 (20 以下)	
さっさ	○			
さっさっ				共起なし
しゃなりしゃなり	○			
しゃらりしゃらり			除外 (20 以下)	
すたこら	○	○		
すたこらさっさ	○	○		
すたすた	○			
せかせか	○			
たー				共起なし
たかたか			除外 (20 以下)	
たじたじ		○		
たたーっ	○	○		
だだーっ			除外 (20 以下)	
たたたた		○		
だだだだ		○		
たったかたったか	○	○		
たったっ	○	○		
たどたど		○		
ちょこちょこ				共起なし
てくてく	○			
てけてけ	○	○		
とーん				共起なし
とことこ		○		
どたばた				共起なし
・	・	・	・	・
・	・	・	・	・

・ ・	・ ・	・ ・	・ ・	・ ・
どっしどっし			除外 (20 以下)	
とっと				共起なし
とっとこととこ	○	○		
とっとっ			除外 (20 以下)	
ととと				共起なし
とぼとぼ	○			
のこのこ		○		
のさのさ				共起なし
のしのし	○			
のそのそ	○	○		
のっしのっし	○	○		
のらりのらり				共起なし
のろのろ	○	○		
のろりのろり	○	○		
ばかばか		○		共起なし
ばたばた				共起なし
ばたばた				共起なし
ぱっぱ				共起なし
ぱっぱか				
ぴたぴた				共起なし
ひよこひよこ	○	○		
ふらふら	○			
ぶらぶら	○			
よたよた	○	○		
よちよち	○	○		
よぼよぼ	○			
よろよろ	○			
らったった	○			
わたわた				共起なし

表 6.2: 「歩く」「走る」に関係するであろうオノマトペ

「歩く」オノマトペ	「走る」オノマトペ
ちゃかちゃか	
どかどか	とつとことつとこ
のっそり	どたどた
ひたひた	ちょろちょろ
ぽてぽて	

トペが出現していた (表 6.2)。

これらのオノマトペは日本語オノマトペ辞典の「歩く、走る」オノマトペ 56 語の分類に含まれていないオノマトペである。オノマトペにはその表現に明確な答えがないが、これらのオノマトペは「歩く」や「走る」行為を説明する際に適切なオノマトペであると考えられる。

例えば「ちゃかちゃか歩く」という表現は「早歩き」を表していると思われる。また「どかどか歩く」は「大きな足音をたてて歩く」表現であり、足音が立っていない場合は「横柄な態度」という印象も受ける。「のっそり歩く」は「ゆっくり身体が重たそう」という表現に思える。「ひたひた歩く」は裸足などで足音を立てずに歩くか、裸足で冷たい場所 (例えば冬の風呂場) などを歩く時に用いることができると考えられる。一方、「とつとことつとこ」はそのオノマトペを聞いただけで走っていることが分かる。また「どたどた走る」は「大きな足音を立てて走る」様子であり、「ちょろちょろ走る」は「小さな子供が走り回る」様な様子を表す時に用いられる。

このことから、日本語オノマトペ辞典には記載されていなかったが、これまで明らかにされていなかったオノマトペと行為の対応を取得できていると考えられる。

6.2 Twitter を用いた結果との比較

Twitter をコーパスとして Tweet 内共起を用いて関連語の取得を行った関連語知識ベースの結果と今回行った結果の比較を行う。表 6.3 に Twitter で行った共起回数をもとに示した関連語知識ベース「ぶらぶら」の結果上位 21 件、表 6.4 に今回生成したオノマトペ関連語辞書「ぶらぶら」の結果上位 21 件を示す。

手続きについて述べる。

関連語の評価には明確な解が存在しない。そのため、結果の中に含まれた関連語のうち、尤もらしいと感じるものに下線を引いた (図 6.3, 6.4 参照)。その後、それが関連語といえるのかどうか、取得先の用例文をもとに考察する。

Twitter をコーパスとした関連語知識ベースとオノマトペ関連語辞書のうち、共に出現した関連語として「街」や「駅」がある。これらの関連語は場所を表す関連語といえ、「街をぶらぶら」「駅をぶらぶら」などの使い方が考えられる (図 6.3, 6.4 参照)。

本研究で取得した「ぶらぶら」のスニペットから抽出したオノマトペを含む一文のうち、「街」を含む文を確認したところ、

「ぶらぶら」と「街」が含まれる用例文

水族館出たあとはぶらぶら 街 を探検しながら色々お話した

インドの時とは違ってなんだか wifi が強いぞ!! 街 をぶらぶら～ひたすら街ブラだけ～街も綺麗やし、野菜も綺麗

のように「ぶらぶら」する場所を表している語として「街」が含まれており、「街をぶらぶら」と用いられていることが分かる。

次に、アメーバブログを用いた今回の結果においては関連語知識ベースには含まれていない「人」や「散歩」「ショッピング」などの語句も出現していた。

「人」という単語が関連語として取得されていた理由は、対象(行為者)がぶらぶらすること示す結果だと考えることができ、オノマトペ「ぶらぶら」においては「人がぶらぶら」とはいうが「犬がぶらぶら」と使うことはない。

スニペットから抽出したオノマトペを含む一文を確認したところ、

「ぶらぶら」と「人」が含まれる用例文

夕涼みの 人 がぶらぶらしています

ホームセンターの他にもいろんなお店が入ってる場所だったので、みゆぼんと二 人 でぶらぶらしてたら

前者は「ぶらぶら」する対象として「人」が用いられていることが分かる。後者は人数を数える単位として「人」が使われているが、文内において「ぶらぶら」する「人」の人数を表すために「人」が用いられているため、「人」が関連語だといえる。

「散歩」を含む文は、

「ぶらぶら」と「散歩」が含まれる用例文

12月26日、義姐のマンションに止めてもらい、朝から近所をぶらぶらお散歩しました

僕は、今日は、何気なく、ぶらぶらと近所を散歩していましたさて、今日は、こんな曲聞いてみました見せちゃいますよー

「ぶらぶらと散歩」のように用いられており、「ショッピング」を含む文は、

「ぶらぶら」と「ショッピング」が含まれる用例文

ぶらぶらショッピングしたり、買い食いしたり

今日は娘とお台場デートお台場でぶらぶらウィンドウショッピング中に axes

表 6.3: 「ぶらぶら」と共起語 (名詞) の
共起回数:Twitter を用いた場合

共起名詞	共起回数
ん	1146
今日	1108
の	1043
笑	716
朝	677
さん	668
日	599
こと	484
今	619
<u>街</u>	582
<u>駅</u>	579
中	527
そう	515
何	504
<u>買い物</u>	494
ちゃん	474
私	456
バイト	428
あと	408
前	401
暇	398
:	:

表 6.4: 「ぶらぶら」と共起語 (名詞) の
頻度:Ameblo を用いた場合

共起名詞	共起頻度
今日	0.19
時間	0.15
日	0.1
中	0.1
<u>街</u>	0.09
<u>人</u>	0.09
笑	0.09
あと	0.09
<u>散歩</u>	0.08
の	0.08
<u>店</u>	0.07
結局	0.06
<u>駅</u>	0.06
方	0.06
後	0.06
仕事	0.06
こと	0.05
それ	0.05
ランチ	0.05
<u>ショッピング</u>	0.05
その後	0.05
:	:

「ぶらぶらショッピング」と用いられている。これらの結果から、「ぶらぶら」するのは「人」であり、「ぶらぶら」する対象は「場所」であるといえる。

また、取得できた関連語が、取得先となっている用例文において非文ではなく、主述関係のある文章構造で用いられていることが分かる。

Twitter の結果と比較して、今回行ったアメーバブログを用いた結果の方が上位のノイズが減っているように思われる。また上位に出現した関連語のうち尤もらしい単語も数が多かった。

用例文を確認した結果、オノマトペと関連語が関連した状態で共起していることがわかり、今回の提案手法において Twitter とアメーバブログでコーパスが異なること、スニペット内からオノマトペを含む一文を切り出しておいたことがノイズ低減の要因と考えられる。

6.3 関連語テンプレートによる類似オノマトペ精練の検証

章で関連語テンプレートによって、類似オノマトペを判断できるのではないかという仮説を元に、用例文から「名詞－オノマトペ－動詞」の型に当たる組み合わせを取得し、またその取得した組み合わせの中に含まれる「名詞－動詞」組をオノマトペ関連語辞書の 2 語検索にかけ擬似的に「名詞－オノマトペ－動詞」の組み合わせを作り出し、それらの結果をあわせることでオノマトペの類似関係を逐次的に収集できることが分かった。

しかし、取得したオノマトペ類似関係データにはノイズが含まれており、また「名詞－オノマトペ－動詞」関係のうち「名詞－オノマトペ」は関係するが「動詞」は用法として間違っているような複雑な関係のものを含むことがわかった。

そこでそのようなノイズを人の判断に委ねて、データに重み付けをすることで尤もらしいオノマトペ類似関係を取得できないかと考えた。

本節では人が「名詞－複数オノマトペ－動詞」の組み合わせをみて、その組み合わせが尤もらしいかどうかの判断を下すことができるのかどうかを確かめた検証について述べる。

6.3.1 検証手続き

検証は被験者を使ったアンケート形式で行う。

被験者に第 6.3 節で用例文から取得した「名詞－オノマトペ－動詞」関係になったもの(一例として図 5.1, 5.2, 5.3 参照)と、その関係から「名詞－動詞」を抜き出し、オノマトペ関連語辞書 2 語検索にかけ取得できた擬似「名刺－オノマトペ－動詞」関係の両方をリストにして提示し、「オノマトペと関連語の組み合わせが正しい、普段耳にする、日常会話で使う」と感じた場合にチェックを入れてもらうことで精練を行ってもらった。

被験者は、23 歳から 25 歳の男性 5 名、女性 1 名の計 6 名であった。使用した関連語テンプレートは 14 組(図 5.1 参照)であり、用例文から抽出した関連語テンプレートとオノマトペの組み合わせは 29 組、オノマトペ関連語辞書 2 語検索から得られた擬似関連語テンプレートとオノマトペの組み合わせは 145 組であった。

実験を行う際には、例として実験に含まない「人－うろうろ－歩く」というテンプレートを使って解説を行い、「人がうろうろと歩く」のように、単語の間に助詞である「が、の、に、を」や「と」などを入れて考えてもらっても構わない、と指示した。また、組み合わせの中には用例文から抽出した関連語テンプレートと関連語辞書から生成した擬似関連語テンプレートの組が重複する行もあったが、一度目の回答と同じようにチェックを入れてもらうこととした。よって、後述する検証結果の表 A.14 内では重複組は最初に出現した一組だけ掲示することとする。

6.3.2 検証結果

検証結果の一例として「胃－オノマトペ－する」の結果を示す(表 A.14 参照)。「胃－むかむか－する」は用例文から見つかった関連語テンプレートと、関連語辞書 2 語検索結果の両方に出現していたものであり、それ以外の 9 組は 2 語検索からのみ取得できた組み合わせである。

検証の結果、被験者らはオノマトペと関連語テンプレートの組み合わせをみて、尤もらしいものとそうでないものを判断できていた。また、多くの実験参加者の間でオノマトペと関連語テンプレートの組み合わせを見て共通認識があることが分かった。

表 6.5: 「胃 – オノマトペ – する」の検証結果

組み合わせ	参加者 A	参加者 B	参加者 C	参加者 D	参加者 E	参加者 F	合計
胃 むかむか する	○	○	○	○	○	○	6
胃 きりぎり する	○	○	○	○	○	○	6
胃 がっちり する							0
胃 ずきんずきん する			○	○			2
胃 じくじく する							0
胃 しん する							0
胃 でこぼこ する							0
胃 たた一つ する							0
胃 げぼげぼ する							0
胃 がぶがぶ する							0

上記したように用例文から取得できた「名詞 – オノマトペ – 動詞」の組み合わせは一番上の「胃 – むかむか – する」であったが、オノマトペ関連語辞書で「胃」「する」の 2 語を検索にかけ得られたもののうち「胃 – きりぎり – する」と「胃 – ずきんずきん – する」の 2 組が尤もらしいと判断されていた。

「胃 – むかむか – する」は食べ過ぎや飲み過ぎなどで胃がもたれた状態であることを示す表現であり、「胃 – きりぎり – する」は胃が痛む状態を表す表現であると思われる。また参加者 C, D の二人だけが尤もらしいと判断した「胃 – ずきんずきん – する」であるが、「ずきんずきん」とは痛みを表すときに用いられるオノマトペであり、胃が痛む状態を表現していると判断したものと思われる。しかし、他の参加者は「ずきんずきん」は「頭」が痛む際に用いる表現であり、「胃」が痛むときにはそぐわないと判断したために結果が別れてしまったと考えられる。

このように参加者によって多少の揺れは存在するが、よく使う、よく耳にする組み合わせに関しては共通認識があり、全参加者が同じような傾向で尤もらしいと判断できていることが分かる。この結果をレリバンスフィードバックすることで、関連語テンプレートをういたオノマトペの類似性の表示順位などに重みをつけ、次のユーザがみるときには尤もらしいものが上位に提示されるようになる。

これらのことから、オノマトペ関連語辞書 2 語検索を使った擬似関連語テンプレートの中にもその組み合わせとして尤もらしいものが存在し、用例文から取得できなかった関連語テンプレートをういた類似オノマトペを増やすことができるといえる。

この関連語テンプレートによるオノマトペ類似性をレリバンスフィードバックする機能をオノマトペ関連語辞書インタフェースに付与することで、逐次的な精錬を行うことができると考える。

7 オノマトペ関連語辞書の利用

本章では提案手法により生成したオノマトペ関連語辞書の利用方法について述べる

7.1 関連語辞書の閲覧方法

ユーザはオノマトペとの対応を知りたい語句をオノマトペ辞書で検索をかける。その結果、その語句が関連語として含まれるオノマトペを閲覧することができる。このとき、検索クエリを関連語を含むオノマトペ全てが表示される。オノマトペが複数に渡る場合、それらオノマトペの関連語を見比べることで、自分が使用したいオノマトペの選択を行えるようになると思う。

単純な検索をかけることで関連語が含まれているオノマトペを調べることができる。また、共起頻度順に見ることができるため、共によく用いられているオノマトペと関連語の組を上から順に見ただけで発見することができる。

例として関連研究で述べた「オノマトペロリ」[2] で用いられている「辛い」という単語をもとにオノマトペ関連語辞典に検索をかけた結果を示す(図 7.1)。

すると一番に「ぴりっ」、二番目に「ぴりり」というオノマトペが出現していることが分かる。本研究においてはレシピをもちいてオノマトペ関連語を取得したわけではないが、ブログをもちいて網羅的に関連語を取得した結果として料理に関するオノマトペとその関連語の関係性を取得できていると言える。

このようにオノマトペを用いたシステムの製作者がユーザである場合、オノマトペとシステム内に用いる特定の状態やオノマトペとそれが関係する対象を単語から調べることができる。

また、関連語から検索を行う場合、どのオノマトペとどの程度関連しているか、共起頻度を用いて分析することができる。例えば、頻度 1 位の「ぴりっ」は共起頻度が 0.153 であるが 2 位「ぴりり」は 0.103、3 位の「だくだく」は 0.063 であることからその頻度には大きな差があり、もっともよく用いられているのはオノマトペ「ぴりっ」であると推測される。

加えて、関連語の検索結果が一つのオノマトペではないため、関連語共起の観点から類似するオノマトペを知ることでもできる。例えば「辛い表現として「ぴりっ」を使いたい」が他には「オノマトペも知りたい」ユーザがいたとする。そこでオノマトペ関連語辞書で「辛い」をキーワードに検索を行い、結果を参照する。そうすると「ぴりっ」の他に「ぴりり」や「ひりひり」「ひーひー」などがあつたことを知ることができる。もちろんノイズも含まれているが、ユーザ自身がこの検索結果を見て文字情報や音象徴からある程度の判断を行うことは可能であると考えられる。

7.2 オノマトペ類似関係の精練

生成したオノマトペ関連語辞書をもとに単語による検索を行うことで、類似するオノマトペを閲覧することができる。このとき、第 6.3 節で示すような名詞と動詞の関連語テンプレートと、オノマトペ関連語辞書に対して 2 語で検索にかけた結果から算出した関連語の組み合わせテンプレートをもとに、類似するオノマトペを閲覧する事ができる。しかし、5 章で述べたように組み合わせテンプレートにはノイズが多く存在する。

そこで、オノマトペ関連語辞書の利用の一部分として、レリバンスフィードバックを用いて

```

-----辛い-----
びりっ 0.15306122449
びりり 0.103448275862
だくだく 0.063829787234
びりり 0.0612244897959
どくどく 0.0459770114943
ひりひり 0.0444444444444
ぶすっ 0.0365853658537
ぜーぜー 0.031914893617
どぼどぼ 0.0315789473684
ひーひー 0.0315789473684
ちろっ 0.0315789473684
しょぼしょぼ 0.03125
がぼがぼ 0.0307692307692
ぐちょぐちょ 0.030303030303
ずきんずきん 0.0253164556962
いがいが 0.025
ぞぞっ 0.025
びきびき 0.0246913580247
ごぼごぼ 0.0243902439024
しょぼん 0.0229885057471
うはうは 0.0227272727273
しん 0.0224719101124
ぎゅー 0.0224719101124
むしゃくしゃ 0.0224719101124
ぎくん 0.0222222222222
にっ 0.0222222222222
ばたばた 0.0217391304348
でこぼこ 0.0215053763441
とろっ 0.0215053763441
からから 0.0215053763441
さめざめ 0.0210526315789
しこしこ 0.0210526315789
ぐっすり 0.0208333333333
こりこり 0.0208333333333
とっぶり 0.0208333333333
くよくよ 0.0208333333333
ぐちゅぐちゅ 0.0208333333333
とろとろ 0.020618556701
べろっ 0.020618556701
ぼってり 0.020618556701
ごちょごちょ 0.0204081632653
どっぶり 0.0204081632653
のそのそ 0.0204081632653
わっ 0.0204081632653
こくこく 0.020202020202
ぐっ 0.020202020202
じりじり 0.02

```

図 7.1: クエリ「辛い」による検索の結果

このノイズを取り除く精錬機能を付与することを考えている [5]。

人の判断をもとに精錬を行うインタフェースの開発を行うことで、オノマトペと関連語テンプレートに重み付けを行い、オノマトペ類似度データの精錬が進むと考えている。

その際、多くの人に利用してもらえる様なインタフェースを構築することが重要になると考えられる。

以上の様に、利用者が知りたいオノマトペを入力することで関連語辞書が逐次的に生成され、その結果をもとに新たなユーザに利用してもらい、オノマトペ間の類似度をレリバンスフィードバックを用いて精錬する辞書の構築ができる。

現時点では Python で構築したプログラムを実行する際に引数としてクエリを与える形で検索を行うが、今後 Web で利用できるインタフェースの実装を目指す。

8 議論

本章では、6章で述べた実験の結果に基づき、提案手法の考察を行う。

8.1 提案手法の有効性

これまでオノマトペが指し示す内容を調べるためにはオノマトペ辞書から調べる、アンケートをとるなどといった方法しかなかった。しかし、多くの文章をもとに生成したオノマトペ関連語辞書を用いることで、オノマトペがどのような単語と関係しているかを知ることができる。また、関連語テンプレートをを用いて導き出したオノマトペ類似度をもとに、類似するオノマトペや使われ方の似たオノマトペを参照できる。

オノマトペを用いたシステム構築の際、このオノマトペ関連語辞書を参照し、指標として用いることで、多くの人がどのようにオノマトペを使っているかを簡単に確認することができる。システム構築の指標として用いることができる。また、構築したシステムの利用ユーザがシステム内で利用するオノマトペと、それと対応した動作の対応への戸惑いを軽減できるのではないかと考えられる。

8.2 提案手法の問題点と解決策

8.2.1 取得できないオノマトペ

本提案手法を用いてオノマトペとその関連語を取得する場合、コーパスを変えれば取得できないオノマトペ関連語は存在しない。しかし、短すぎるオノマトペはオノマトペでない表現にその文字列が含まれているがために誤って取得してしまう場合があることが判明した。例として歩く、走る動作を表すオノマトペとして日本語オノマトペ辞典に分類されているオノマトペ「たーっ」を挙げる。

このオノマトペをもとに収集したコーパスを確認したところ、「やったー」「よかったー」などのオノマトペでない語句がその殆どであり、「たーっ」というオノマトペ表現を含む文章は見られなかった。このことから、他の仮名記述にまぎれてしまう可能性のあるオノマトペの場合、本提案手法は有効では無い。これは Mecab が形態素解析に失敗していることが原因である。

この問題を解決する方法として、取得に失敗したパターンを集めておくことで解決できる可能性がある。具体的には取得した短いオノマトペ用例文のオノマトペの前後にある文字を見て、オノマトペでないものを辞書登録しておくことで解決できると考える。

また、他の方法として機械学習等を用いることで文字列を判断できれば、オノマトペかそうでないかを自動的に判断することができる可能性がある。

8.2.2 取得できない語句

今回は用例文収集にブログを用いたが、専門的な用語などを取得することは難しい。その結果、もやもやドロイングに用いられるような画像処理エフェクト名など多くの人に知られていない語句とオノマトペの対応を取得できなかった。

この問題はコーパスの選択に左右されるが、口語表現によく用いられるオノマトペと一般的でない専門用語が一文内に共起する形で記述された文章を用意する必要があり、そのようなこ

とは稀であるため、提案手法での関係性取得は困難だと言える。

しかし、専門的な用語を含む文章にも様々な単語が用いられており、提案したオノマトペ関連語に記載された関連語を通じて、オノマトペがよく用いられる文章とそうでない文章の何らかの繋がりを見つけることができる可能性はある。

また、専門用語との対応付けは別の手法で行い、類似オノマトペ間での差 (e.g., ざあざあ: 激しい, しとしと: 激しくない) のような情報が取得できれば、オノマトペ同士での程度の差を示すことができると考えられる。この程度の差を指標として用いることで、オノマトペと対応させたい表現の差を対応付けすることができる可能性がある。

8.3 関連語用例文を用いたオノマトペの取得

関連語テンプレートをもとに文章を収集することでオノマトペを含む文が収集できる可能性はある (第 4.1.2 節参照)。しかし、含まれるオノマトペの数が非常に少ないことが予想されるため、高精度なオノマトペ文字列の自動判断を行う必要があると考えられる。この技術に関しては第 4.1.2 節で解説したように、奥村らが行ったオノマトペ概念辞書生成においてオノマトペの認識に関する研究成果を報告している [4]。

本研究はオノマトペとその関連語の関係を明らかにすることを目的としており、オノマトペ文字列の自動認識に重点を置いていないため、今回は人手でオノマトペを用意して検索クエリとして用いた。しかし、オノマトペ関連語辞書を新語オノマトペの関連語を取得する、網羅性を向上させるためにオノマトペ文字列の自動認識は必要な技術であるといえる。

オノマトペ文字列の自動認識は本研究においても今後検討していくべき課題である。

8.3.1 新たなオノマトペ文字列の取得における問題点

前節で述べたとおり、関連語テンプレートをもとに用例文を収集すればオノマトペ文字列を取得できる可能性がある。しかし、それを用いて未入力のオノマトペ文字列を獲得するためには、本研究のように初期検索の段階で分類されたオノマトペのうちの全てを利用する場合、困難となる。なぜならその分類をもとに類似した全ての関連オノマトペを既に入力してしまった状態となり、それらのオノマトペの関連語をもとに用例文を収集しても、関連語によって類似したオノマトペを取得してしまうことが多いからである。

本研究では第 4.1.2 節でも述べたように、オノマトペ用例文の収集によって取得できたオノマトペの関連語をもとに関連語の用例文を収集し、それに含まれるオノマトペを収集しようと試みた。しかし、オノマトペ文字列を取得出来たとしても、関連語を介して関係したオノマトペが出現する可能性が高いため、すでに初期段階で用意したオノマトペを取得してしまう場合がほとんどである。よってオノマトペの網羅性を向上できない。

関連語テンプレートを用いて新しいオノマトペ文字列を入手したい場合、初期段階で用意するオノマトペは、その分類が異なるものを用意することが好ましいと考えられる。

8.3.2 分類において見落とされたオノマトペ

第 6.1.3 節で解説したように、オノマトペ辞典の分類に含まれていないが、本来分類されるべきオノマトペを発見することができた。これは「文内共起」の観点からオノマトペ関連語が取得出来ていることを明らかに示すものであると考えられる。

これまでに確実な分類が行われていなかったり、見落とされていたオノマトペの分類に関して、オノマトペ関連語辞書を用いることで関連語とオノマトペの関係の発見につながったと考えられる。

8.3.3 関連語群によるオノマトペのクラスタリング

本研究では行わなかったが、オノマトペにおける関連語群から単語集合ベクトルを算出し、クラスタリングを行うことも可能であると考ええる。そうすることで、関連語同士の関係からオノマトペの類似語や対義語の判断を行うことができる。

しかし、オノマトペによっては多義語の存在があるため、関連語辞書内に複数の文脈を含む単語が含まれている場合もある。そのとき、単語集合ベクトル算出において、通常の記事ベクトルと同じような結果を得られるかどうか定かではなく、その複数の文脈から生成した関連語辞書をもとに単語集合ベクトルを算出したもの同士を比較し類似度を算出した場合に、果たして類似オノマトペとされるもの同士がベクトルの観点から類似するに至るかは疑問が残る。

これを踏まえると、オノマトペ多義語の観点から一語のオノマトペでも用例文の意味をもとに関連語辞書を生成し分ける必要があり、意味毎に単語集合ベクトルを算出しなければならないと言える。

9 結論

本稿ではオノマトペを利用した研究において、オノマトペとそのオノマトペが示す動作がどのように対応しているのかを調べる必要がある問題点について述べ、オノマトペの工学的研究での円滑な利用を目的として、オノマトペとその“対象”と“行為”を明らかにするためのオノマトペ関連語辞書を提案した。

そのオノマトペ関連語辞書の生成手法としてブログを利用してオノマトペ用例文を収集することでオノマトペと関連語の共起頻度を調べる手法を提案し、実装と分析を行った。

提案したオノマトペ関連語辞書はオノマトペとその関連語を対応付けて記録し、オノマトペの関連語を閲覧すること、単語をクエリとして単語に関連するオノマトペを検索することができる。

また、類似するオノマトペや細かいニュアンスを伝えるオノマトペの関係性を明らかにすることを目指して、オノマトペ用例文には、オノマトペとその他の単語を含めた文脈が存在することについて考察し、オノマトペと関連語にテンプレートになりうる関係性があるのではないかと仮定した。その仮定をもとに関連語とオノマトペの関係について調査した結果、関連語テンプレートが存在し、また関連語テンプレートに複数のオノマトペが入り、特定の関連語テンプレートに入る複数のオノマトペは類似関係にある可能性があることが確認された。

オノマトペの類似性を示すことができるオノマトペ関連語辞書生成の一環として、どのようなオノマトペと関連語テンプレートの組み合わせが存在するか調査し、事前に用例文から生成したオノマトペ関連語辞書のデータを用いることで類似オノマトペの逐次的取得を実現した。

その際、問題であったノイズを除去する手法として、レリバンスフィードバックに着目し、オノマトペ関連語テンプレートの尤もらしさを人が判断できるのかを検証した。

検証の結果、人が判断できることがわかり、関連語テンプレートを用いたオノマトペ類似性のノイズをレリバンスフィードバックを用いて精錬する手法を提案した。

今後は、オノマトペ関連語辞書を簡易に利用できるインタフェースの開発を行い、レリバンスフィードバックを用いた精錬機能を実装する。本研究では、これまでに類似オノマトペの精錬にレリバンスフィードバックを用いることができることを検証したが、元データに対するフィードバックにおいてどのように重み付けをすべきか、また重み付けをしたデータをどのように提示すべきか、類似するオノマトペ同士のニュアンスの違いを明らかにする手法などについては検討できていない。

また、ブログではないコーパスや、用例文による関連語抽出手法の精度を向上させることで、関連語の確定やノイズ除去、対応語数を増やすことが可能になると考えられるため、この点については今後検討する必要がある。

謝辞

本研究を纏めるにあたり、関西大学総合情報学部総合情報学研究科の松下光範教授にご指導ご鞭撻を賜りました。研究活動を通して、論理の組み立て方や、研究への取り組み方など多くのことを学ぶことができました。また、学会や研究会への参加を後押ししていただき、沢山の貴重な経験ができました。ここに感謝の意を表します。

提案システムの実装にあたり、関西大学大学総合情報学部の朝田君には多大なるご協力をいただきました。また技術的な側面だけでなく、研究そのものについての意見も多く貰い、朝田君なくして結果を出すことはできなかったと思います。御礼申し上げます。

その他、研究を進めるにあたり、様々な議論を交わし助言を下さいました松下ゼミの方々、また参考にした論文の著者の方々に心より感謝致します。

最後に、ここまで支えてくれた父、母、励ましてくれた弟に感謝いたします。

A 付録

A.1 6 章で行った関連語テンプレートによる精練の検証結果

表 A.1: 「胃 – オノマトペ – する」の検証結果

組み合わせ	参加者 A	参加者 B	参加者 C	参加者 D	参加者 E	参加者 F	合計
胃 むかむか する	○	○	○	○	○	○	6
胃 きりきり する	○	○	○	○	○	○	6
胃 がっちり する							0
胃 ずきんずきん する			○	○			2
胃 じくじく する							0
胃 しん する							0
胃 でこぼこ する							0
胃 たた一つ する							0
胃 げほげほ する							0
胃 がぶがぶ する							0

表 A.2: 「雨 – オノマトペ – 降る」の検証結果

組み合わせ	参加者 A	参加者 B	参加者 C	参加者 D	参加者 E	参加者 F	合計
雨 どしゃどしゃ 降る			○	○	○		3
雨 しとしと 降る	○	○	○	○		○	5
雨 ざんざ 降る			○				1
雨 ざんざん 降る							0
雨 ざーざー 降る	○	○	○	○	○	○	6
雨 ぱらぱら 降る	○	○	○	○	○	○	6
雨 じとじと 降る		○					1
雨 ぽつぽつ 降る	○	○	○	○	○	○	6
雨 どんより 降る							0
雨 ぼたぼた 降る			○				1

表 A.3: 「関係 – オノマトペ – する」の検証結果

組み合わせ	参加者 A	参加者 B	参加者 C	参加者 D	参加者 E	参加者 F	合計
関係 ぎく する							0
関係 ぎくしゃく する	○	○	○	○	○	○	6
関係 むしゃくしゃ する							0
関係 つんけん する							0
関係 ごたごた する	○	○	○	○		○	5
関係 うずうず する							0
関係 いらいら する			○				1
関係 ごつごつ する							0
関係 がっくり する							0
関係 おっとり する							0

表 A.4: 「胸 – オノマトペ – する」の検証結果

組み合わせ	参加者 A	参加者 B	参加者 C	参加者 D	参加者 E	参加者 F	合計
胸 むかむか する	○	○	○	○	○	○	6
胸 ぶすぶす する							0
胸 ずきり する			○				1
胸 ぐさぐさ する		○	○				2
胸 ざわざわ する	○	○	○	○	○	○	6
胸 ずきずき する	○	○	○	○		○	5
胸 きゅー する		○	○				2
胸 する							0
胸 する							0
胸 する							0

表 A.5: 「心 – オノマトペ – する」の検証結果

組み合わせ	参加者 A	参加者 B	参加者 C	参加者 D	参加者 E	参加者 F	合計
心 ほかほか する	○	○	○	○	○		5
心 ざわざわ する		○	○	○	○		4
心 ばくばく する			○				1
心 ぼろぼろ する							0
心 こせこせ する							0
心 ずたずた する		○	○				2
心 ほんわか する	○	○	○	○		○	5
心 ぽっかり する		○	○		○		3
心 くよくよ する			○				1
心 ほわっ する	○	○					2

表 A.6: 「心臓 – オノマトペ – する」の検証結果

組み合わせ	参加者 A	参加者 B	参加者 C	参加者 D	参加者 E	参加者 F	合計
心臓 ばくばく する	○	○	○	○	○	○	6
心臓 どくどく する		○	○				2
心臓 ときとき する	○	○	○	○	○	○	6
心臓 むずむず する							0
心臓 じゅくじゅく する							0
心臓 どつきどつき する		○	○				2
心臓 ぽきぽき する							0
心臓 ごぼごぼ する							0
心臓 とくどく する			○				1
心臓 ぴこぴこ する							0

表 A.7: 「雪 – オノマトペ – 降る」の検証結果

組み合わせ	参加者 A	参加者 B	参加者 C	参加者 D	参加者 E	参加者 F	合計
雪 どさどさ 降る		○					1
雪 ぼたぼた 降る		○	○	○			3
雪 しとしと 降る	○		○				2
雪 こんもり 降る		○	○				2
雪 ぱらぱら 降る	○	○	○	○	○	○	6
雪 こんこん 降る	○	○		○			3
雪 ちらほら 降る	○	○	○		○		4
雪 どうかどか 降る		○					1
雪 ざんざん 降る							0
雪 ちらちら 降る	○	○	○	○	○	○	6

表 A.8: 「雪 – オノマトペ – 落ちる」の検証結果

組み合わせ	参加者 A	参加者 B	参加者 C	参加者 D	参加者 E	参加者 F	合計
雪 どさどさ 落ちる		○	○		○	○	4
雪 ぼたぼた 落ちる			○			○	2
雪 こんもり 落ちる							0
雪 はらはら 落ちる	○	○					2
雪 ぽとん 落ちる	○	○	○				3
雪 ぼたぼた 落ちる	○		○				2
雪 ほろっ 落ちる							0
雪 ずぶずぶ 落ちる							0
雪 ぱらぱら 落ちる	○	○	○	○			4
雪 ぽとぽと 落ちる		○	○	○		○	4

表 A.9: 「頭 – オノマトペ – する」の検証結果

組み合わせ	参加者 A	参加者 B	参加者 C	参加者 D	参加者 E	参加者 F	合計
頭 くらくら する	○	○	○	○	○	○	6
頭 ずきんずきん する	○	○	○	○		○	5
頭 ずきずき する	○	○	○	○	○	○	6
頭 ごちゃごちゃ する	○	○	○	○	○	○	6
頭 ぐらぐら する	○	○	○	○	○		5
頭 ぐりぐり する		○					1
頭 くしゃくしゃ する		○	○				2
頭 ぽんぽん する		○	○	○	○		4
頭 ぽやぽや する			○				1
頭 ぽけっ する	○	○	○	○			4

表 A.10: 「疲れ – オノマトペ – 出る」の検証結果

組み合わせ	参加者 A	参加者 B	参加者 C	参加者 D	参加者 E	参加者 F	合計
疲れ どっと 出る	○	○	○	○	○	○	6
疲れ どーっ 出る		○	○		○		3
疲れ どぼどぼ 出る							0
疲れ ちょろちょろ 出る							0
疲れ どっ 出る	○	○	○		○		4
疲れ げっそり 出る							0
疲れ ぎくしゃく 出る							0
疲れ こぼこぼ 出る							0
疲れ ぼんやり 出る							0
疲れ たらたら 出る		○					1

表 A.11: 「鼻 – オノマトペ – する」の検証結果

組み合わせ	参加者 A	参加者 B	参加者 C	参加者 D	参加者 E	参加者 F	合計
鼻 むずむず する	○	○	○	○	○	○	6
鼻 ひくひく する	○	○	○	○	○	○	6
鼻 ぐずぐず する		○	○	○	○		4
鼻 むずむず する	○	○	○	○	○	○	6
鼻 ひりひり する	○	○	○			○	4
鼻 つんつん する		○	○				2
鼻 ぐちゅぐちゅ する	○		○	○			3
鼻 ずーずー する			○		○		2
鼻 すすっ する					○		1
鼻 ずるずる する		○	○			○	3
鼻 ぐりぐり する		○					1

表 A.12: 「風－オノマトペ－吹く」の検証結果

組み合わせ	参加者 A	参加者 B	参加者 C	参加者 D	参加者 E	参加者 F	合計
風 びゅーびゅー 吹く	○	○	○	○	○	○	6
風 びゅんびゅん 吹く		○	○	○	○		4
風 びゅーびゅー 吹く	○	○	○	○	○		5
風 そよそよ 吹く		○	○		○	○	4
風 びゅー 吹く	○	○	○		○		4
風 さわさわ 吹く							0
風 びゅー 吹く	○	○	○		○		4
風 ひゅーひゅー 吹く	○	○	○	○	○	○	6
風 びゅんびゅん 吹く		○	○		○		3
風 ゆらりゆらり 吹く					○		1

表 A.13: 「目－オノマトペ－する」の検証結果

組み合わせ	参加者 A	参加者 B	参加者 C	参加者 D	参加者 E	参加者 F	合計
目 ちかちか する	○	○	○	○	○	○	6
目 しょぼしょぼ する	○	○	○	○	○	○	6
目 ぎょろぎょろ する	○	○	○	○	○	○	6
目 ぎょろっ する	○	○	○		○		4
目 ぱちくり する	○	○	○	○		○	5
目 ぱしぱし する		○	○			○	3
目 ぱちくり する	○	○	○	○		○	5
目 ぱっちり する	○	○	○		○	○	5
目 がくり する							0
目 うるるん する		○	○				2
目 ぴくぴく する	○	○	○			○	4

表 A.14: 「T シャツ – オノマトペ – 濡れる」の検証結果

組み合わせ	参加者 A	参加者 B	参加者 C	参加者 D	参加者 E	参加者 F	合計
T シャツ じっとり 濡れる	○	○	○	○		○	5
T シャツ ぐっしょり 濡れる	○	○	○	○		○	5
T シャツ びしゃびしゃ 濡れる	○	○	○				3
T シャツ ぐしょぐしょ 濡れる	○	○	○		○	○	5
T シャツ べちょべちょ 濡れる	○	○	○		○		4

参考文献

- [1] Kudo, T.: MeCab : Yet Another Part-of-Speech and Morphological Analyzer, <http://mecab.sourceforge.net/> (2005).
- [2] Lertsumruaypun, K., Watanabe, C., Lertsumruaypun, K., Watanabe, C. and Nakamura, S.: Onomatoperori : Recipe Recommendation System Using Onomatopoeic Words, *IPSSJ SIG Technical Report*, Vol. 2009-DD-73, No. 6, p. 7 (2009).
- [3] Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D. and Miller, K. J.: Introduction to wordnet: An on-line lexical database*, *International journal of lexicography*, Vol. 3, No. 4, pp. 235–244 (1990).
- [4] Okumura, A., Saito, S. and Okumura, M.: Automatic construction of a Japanese onomatopoeia dictionary using text data on the WWW, *Joho Shori Gakkai Kenkyu Hokoku*, Vol. 2003, No. 23, pp. 63–70 (2003).
- [5] Rocchio, J. J.: Relevance feedback in information retrieval, *The Smart retrieval system - experiments in automatic document processing* (Salton, G.(ed.)), Englewood Cliffs, NJ: Prentice-Hall, pp. 313–323 (1971).
- [6] Uema, H., Hashui, D. and Matsushita, M.: Constructing a Knowledge Base of Onomatopoeia-Action Relations for an Interactive Illustrated Storybook., *The Second Asian Conference on Information Systems(ACIS2013)*, Vol. 6, pp. 1–6 (2013).
- [7] 内田ゆず, 荒木健治, 米山淳: ブログ記事から抽出したオノマトペの多義性について, 日本知能情報ファジィ学会 ファジィ システム シンポジウム 講演論文集, Vol. 27, pp. 193–193 (2011).
- [8] 内田ゆず, 荒木健治, 米山淳: ブログ記事からのオノマトペ用例文の自動抽出手法, *Journal of Japan Society for Fuzzy Theory and Intelligent Informatics*, Vol. 24, No. 3, pp. 811–820 (2012).
- [9] 寺島亜耶香, 上間大生, 松下光範: 効果線の描画に着目した動的音喩の付与手法, 第 26 回人工知能学会全国大会論文集, No. 1M2-OS-8b-3, 人工知能学会, pp. 1–4 (2012).
- [10] 渡部啓吾, Danushka, B., 松尾豊, 石塚満: 検索エンジンを用いた関連語の自動抽出, 知能と情報 : 日本知能情報ファジィ学会誌 : journal of Japan Society for Fuzzy Theory and Intelligent Informatics, Vol. 23, No. 5, pp. 739–748 (2011).
- [11] 寺島宏紀, 小松孝徳: もやもやドロ잉: オノマトペの印象をエフェクトとして反映するドロ잉ツールの開発, 第 26 回人工知能学会全国大会論文集, No. 1M2-OS-8b-2, 人工知能学会, pp. 1–4 (2012).
- [12] 森田康夫: 日本語オノマトペ辞典, 小学館, 第 1 版 edition (2007).

- [13] 平田佐智子, 澤井大樹, 藤井弘樹, 喜多伸一: Twitter を用いたオノマトペ記述データの収集システム, 第 25 回人工知能学会全国大会論文集, No. 1C2-OS4b-9, 人工知能学会, pp. 1–3 (2011).
- [14] 新村出: 広辞苑, 株式会社岩波書店, 第六版 edition (2008).
- [15] 加藤誠, 大島裕明, 小山聡, 田中克己: 語の共起を用いた Web からの類似関係検索 (ウェブ情報とデータベースに関して (ポスター講演)), 電子情報通信学会技術研究報告. DE, データ工学, Vol. 108, No. 329, pp. 9–10 (2008).
- [16] 浅賀千里, Yusuf, M., 渡辺知恵美: オノマトペ用例辞典における用例を意味により分類するためのクラスタリング手法の諸検討, 日本データベース学会 letters, Vol. 6, No. 2, pp. 45–48 (2007).
- [17] 上間大生, 蓮井大樹, 松下光範: オノマトペを入力手段としたインタラクティブ絵本システムの提案, No. D-09, EntertainmentComputing, pp. 1–3 (2012).