



BCCWJコアデータにおける オノマトペ出現実態の分析

—現代オノマトペ実例辞書アプリ構築に向けて

内田 ゆず (北海学園大学)

高丸 圭一 (宇都宮共和大学)

乙武 北斗 (福岡大学)

木村 泰知 (小樽商科大学)

オノマトペを観察しています

(擬音語・擬態語)

今回は
お留守番…
しょんぼり



湯の川で
ほっこり



チャイチキ
がぶっ



内田

木村

乙武

高丸

函館ワイン
ぐびぐび



オノマトペ実例辞書を作りたい

■ たくさんの実例→新しい用法もわかる

雷がごろごろ鳴っている
夫が家でごろごろしている
メンバーがごろごろ変わる
...

新用法
変化するさま

■ 共起する表現→文脈と意味がわかる

雷ガ — ごろごろ — 鳴る

雷鳴の音

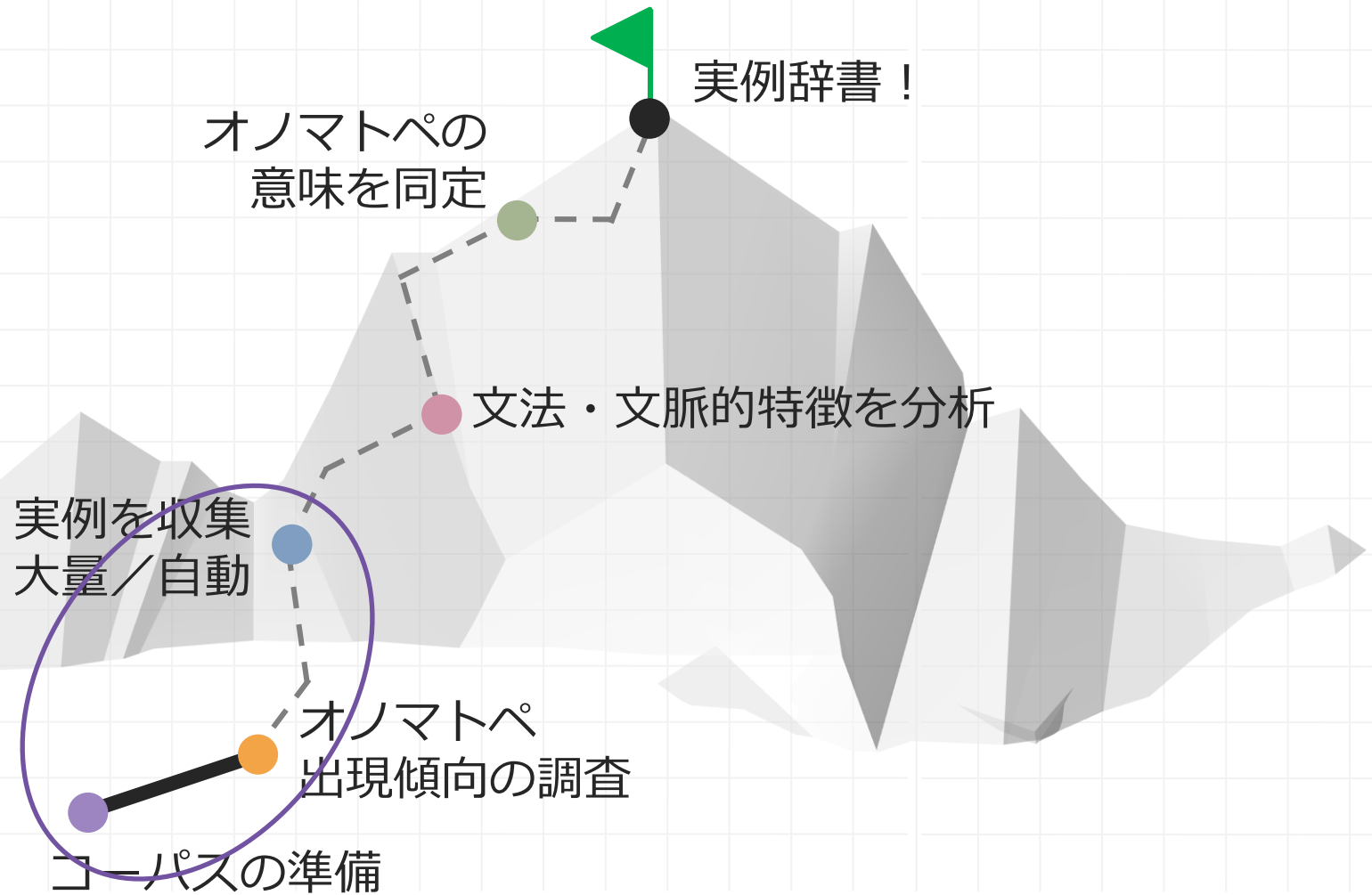
夫ガ
休日ニ
家デ

— ごろごろ —

する
寝る

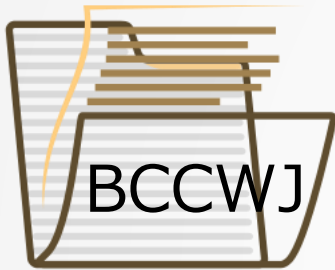
怠惰に
暮らすさま

行きつ戻りつ…



発表の概要

現代日本語書き言葉
均衡コーパス



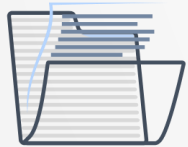
コアデータ

約100万語

形態素解析人手修正済み

① コアデータ中のすべてのオノマトペを分析

② ①に基づき抽出手法を構築（ベースライン）
→コアデータで抽出実験（クローズドテスト）



Yahoo! ブログデータ

約120万語

③ ②に基づき抽出手法を改良
→ブログデータで抽出実験（オープンテスト）

① コアデータ中の全オノマトペ分析

手順

1. ひらがな・カタカナのみからなる
2文字以上の形態素をコアデータからすべて抽出

198,829例

あいにく, これ, ドラクエ, すくすく, です, こと
ゲシュペンスト, んで, など, のぞく, おお, …

2. 人手で以下の項目をチェックする
 - A) オノマトペなのか？
 - B) オノマトペ辞典にある？
 - C) オノマトペの品詞は？

日本語オノマトペ辞典
4,506語



① コアデータ中の全オノマトペ分析

結果

A) オノマトペなのか？

分析対象	198,829
オノマトペ	2,048
非オノマトペ	195,927
感動詞	706
判断不能	148

574語のオノマトペが
2,048回出現

B) オノマトペ辞典にある？

辞典にある：493語

辞典にない：81語

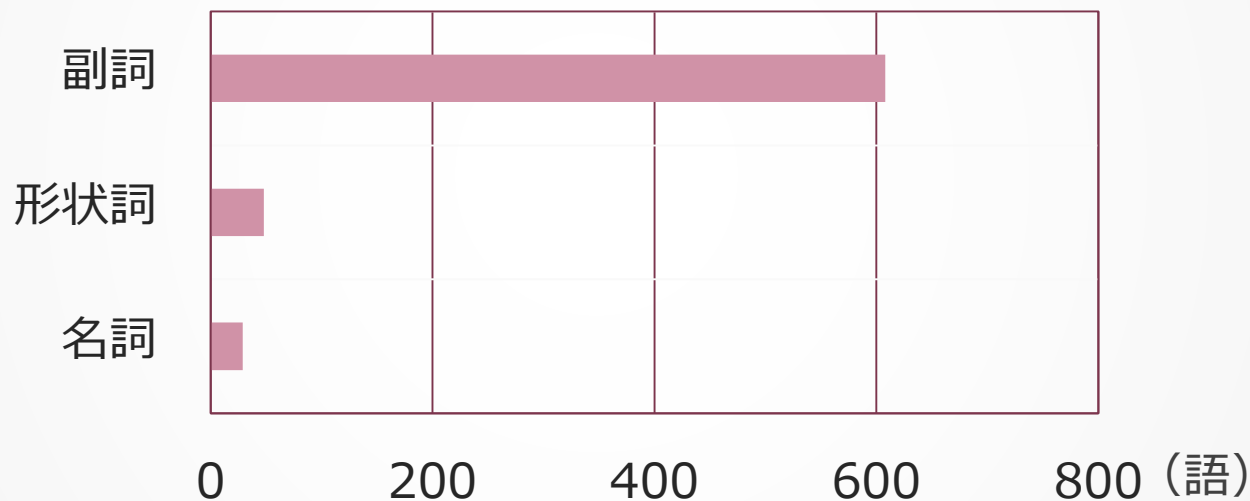
うがうが, てれん, ごふっ, ぱっきり,
ぴぽぴぽ, ぶふおっ, ぶんすか,
がたがたがた, たたったった,
ぱらぱらっ, めっちゃめちゃ, ふうふう

新語／繰り返し／促音／長音 …

① コアデータ中の全オノマトペ分析

結果

C) オノマトペの品詞は？



ほぼ副詞、形状詞と名詞に少し

オノマトペの自動抽出に品詞情報を利用できそう

② オノマトペ抽出実験（ベースライン）

目的

品詞情報のみを用いた単純な手法でどの程度の精度？

クローズドテストで検証

（辞典にないオノマトペはここでは無視）

手順

1. コアデータをMeCab（Unidic）で形態素解析
※形態素解析誤りが起きる
2. 以下の条件を両方満たす形態素をオノマトペと判断する

条件1：表層がオノマトペ辞典の見出し語と**完全一致**

条件2：品詞が**副詞**あるいは**形状詞**

② オノマトペ抽出実験（ベースライン）

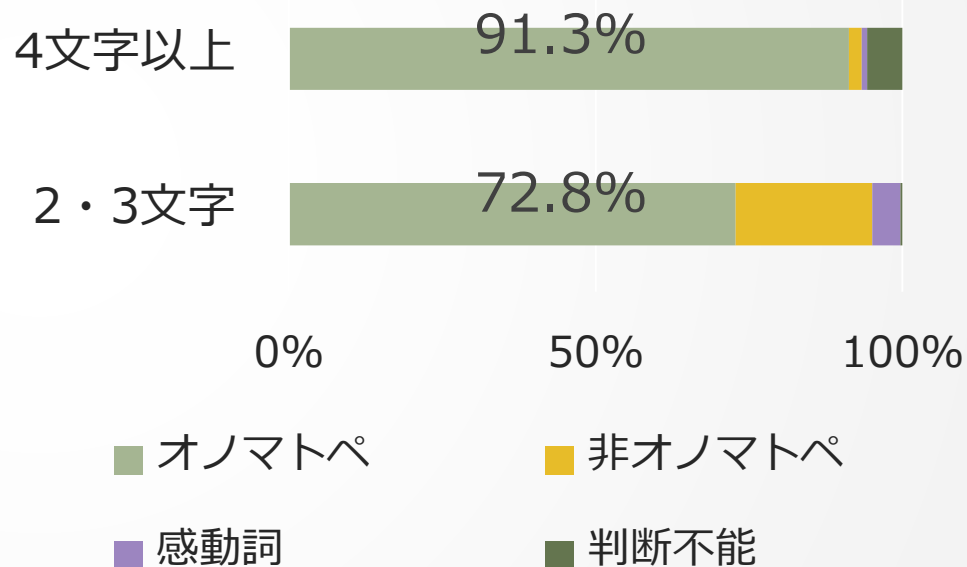
結果

全抽出数	2,076
オノマトペ	1,778
非オノマトペ	171
感動詞	42
判断不能	85

適合率：85.6%

再現率：86.8%

オノマトペの長さ と 適合率



短いオノマトペが悪い！！

② オノマトペ抽出実験（ベースライン）

エラー分析

2文字／長音／
カタカナ

育ち盛りの高校生，こーゆー添加物のこと…
一番目立っていたグレートデン。
どーでもイイ。

同音異義語

これが一般人のごくごく健全な感覚でしょう。
二十年も放置され，とうとう空家が一千戸に達した。
私にはたった一つだけ望みがあった。

助詞の影響

挽き出すときに，目がちゃっとひっかかるわけですね。
病気のペット（たとえばワンちゃんとしましょう）は…

これらの誤抽出を避けられるように手法を改良したい

③ オノマトペ抽出手法の改良

目的

ベースラインでの誤抽出、辞典にないオノマトペに対応
オープンテストで検証（ブログデータ）

改良の方針

- ◆ 辞典掲載語 + **副詞** / **形状詞** - **ストップワード**
 - 同音異義語、2文字の誤抽出
- ◆ 辞典掲載語で**名詞化**したもの
 - いらいら、どさくさ など
- ◆ 辞典掲載語 + **助詞** 「と」
- ◆ 辞典掲載語 + **促音** 「っ」
- ◆ 長音母音 → **長音** 記号（例：ぶうぶう → ぶーぶー）
- ◆ 繰り返しの**縮約**（例：がたがたがた → がたがた）

③ オノマトペ抽出手法の改良

結果

2・3文字のオノマトペのみ人手で分析

2文字	副/形状	名詞化	-促音	+促音	+と	長音/縮約	計
オノ	1689	0	583	0	0	5	2277
非オノ	2356	0	873	0	0	152	3381
判断不能	296	0	10	0	0	41	347
計	4341	0	1466	0	0	198	6005

2文字は品詞＋表層で抽出するのが難しい／長音も危険

3文字	副/形状	名詞化	-促音	+促音	+と	長音/縮約	計
オノ	5313	13	21	40	1921	165	7473
非オノ	942	19	72	8	122	24	1187
判断不能	11	0	0	0	16	0	27
計	6266	32	93	48	2059	189	8687

3文字ならばほどほどの精度

③ オノマトペ抽出手法の改良

結果

2文字オノマトペの実際の例

■ 副詞／形状詞-ストップワード

× これで夜の運転も安心さっ

○ ブロッコリーは小房に分けて、熱湯でサッとゆでる。

■ -促音

× どなたかコツ（コツツ）教えてください。

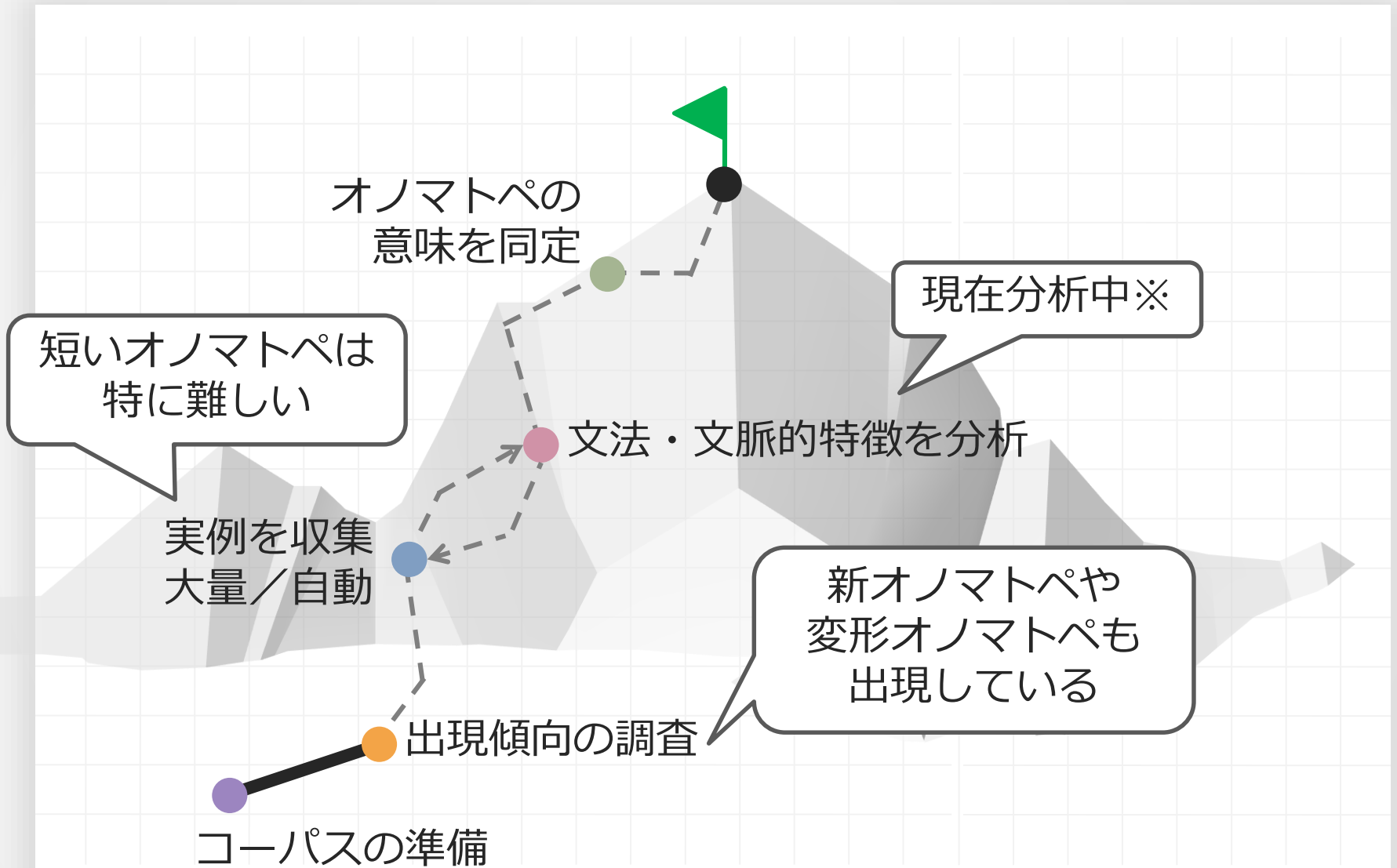
○ 見事全部が、キュ（きゅっ）、と仕上がりました。

■ 長音

× じい（じー）ちゃんとはあちゃんが病院に来てくれたよ☆

○ チビ娘達その後もじい（じー）と見つめて待つ*待つ

まとめ



※FSS2015 (9/2-4, 電気通信大学) で発表予定