

オノマトペが修飾する 名詞の抽出と分析

～「一の」「一（と）した」を対象として～

乙武 北斗

(福岡大学)

内田 ゆず

(北海学園大学)

高丸 圭一

(宇都宮共和大学)

木村 泰知

(小樽商科大学)

第32回ファジィシステムシンポジウム (2016/08/31)



ono-collo.com



ono-collo.com

- **Onomatopoeia**
オノマトペ

- 擬音語・擬態語などと呼ばれる語群

- **Collocation**
コロケーション

- 語と語のつながり
- 一緒に使われる語の組み合わせ

語義とコロケーション

しっとり

- 語義① 湿気が軽く全体にいき渡っているさま
- 語義② 静かに落ち着いたさま
- 語義③ 女性の、静かな中にも色っぽさが漂っているさま



どんな意味？

ツヤツヤ・しっ通りの
髪を手に入れましたあ

語義曖昧性解消

どんな文脈？

しっとりとした



日本語学習者支援

目標は『オノコロ辞書』

- ・ 語義とコロケーションの実例を網羅したい

ぶらぶら

語義① 宙に釣り下がって揺れ動くさま

体の部位

足、腕、手

服飾品

バッグ、イヤリング

を

ぶらぶら

させる、振る

語義② あてもなく時間をかけて歩くさま

場所

広場、新宿、歓楽街、
スーパー、通学路

を

で

ぶらぶら

する、散歩する、歩く

オノマトペの典型的な使い方

- 副詞

雨で**しっとり**濡れる

係り先動詞で
語義がわかる

- サ変名詞

雰囲気**がしっとり**してる

コロケーションで
語義がわかる

- 連体詞

しっとりの髪
しっとりとした生活

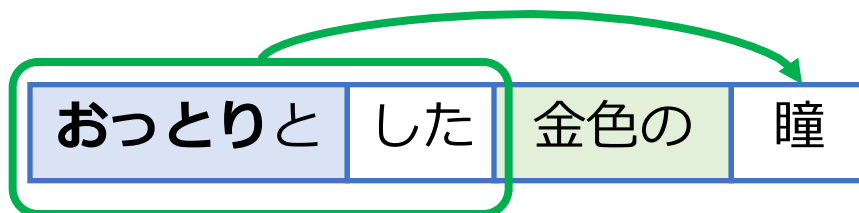
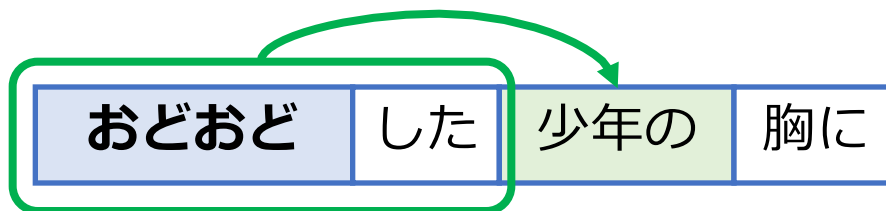
直後の名詞で
語義がわかる？

オノマトペが修飾する名詞

- ・オノマトペの直後の名詞で語義がわかる



- ・直後の名詞がオノマトペの修飾先である（？）



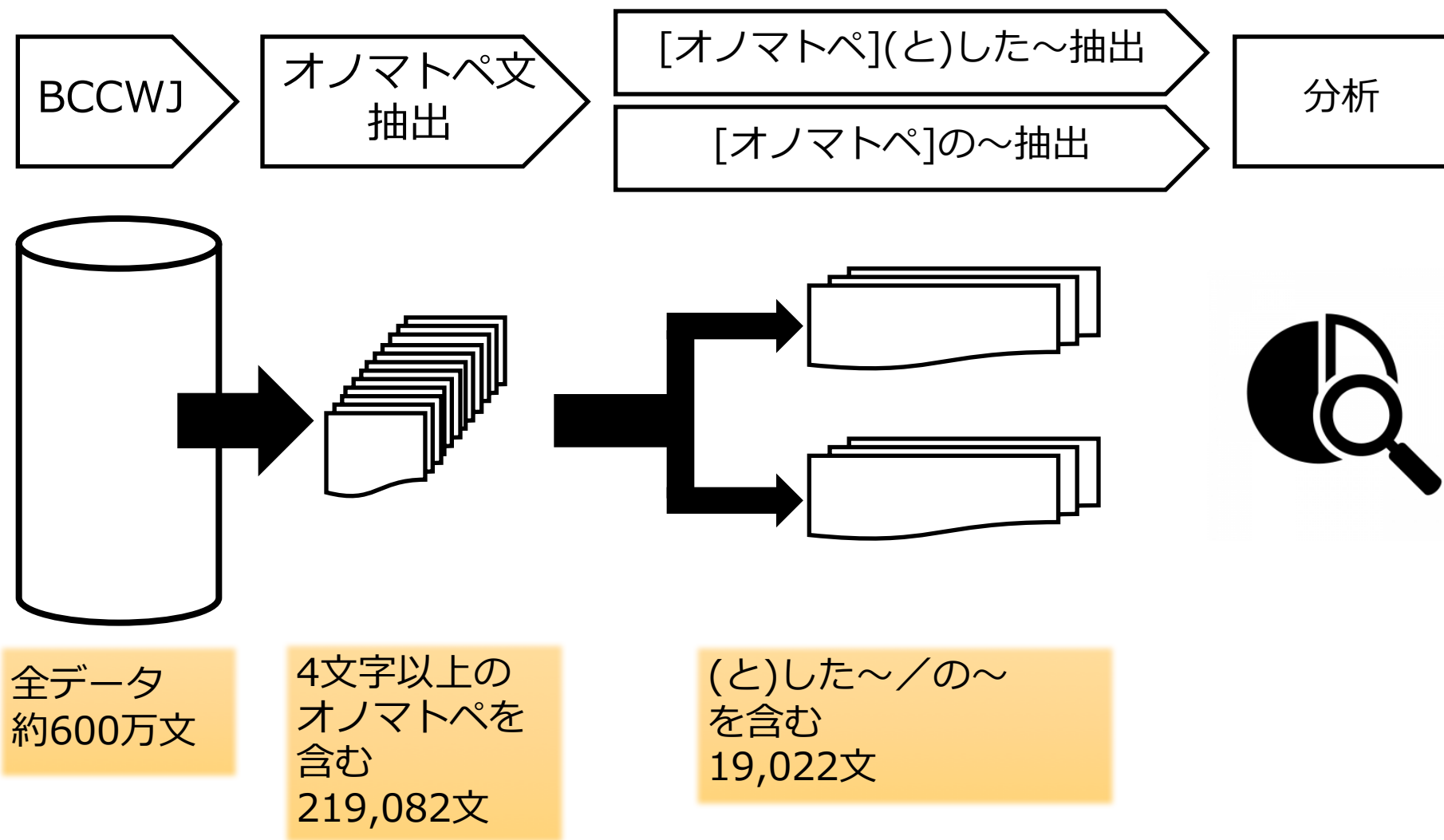
本分析の目的

- 【オノマトペ】（と）した
 - 【オノマトペ】の
- } の直後の名詞を抽出

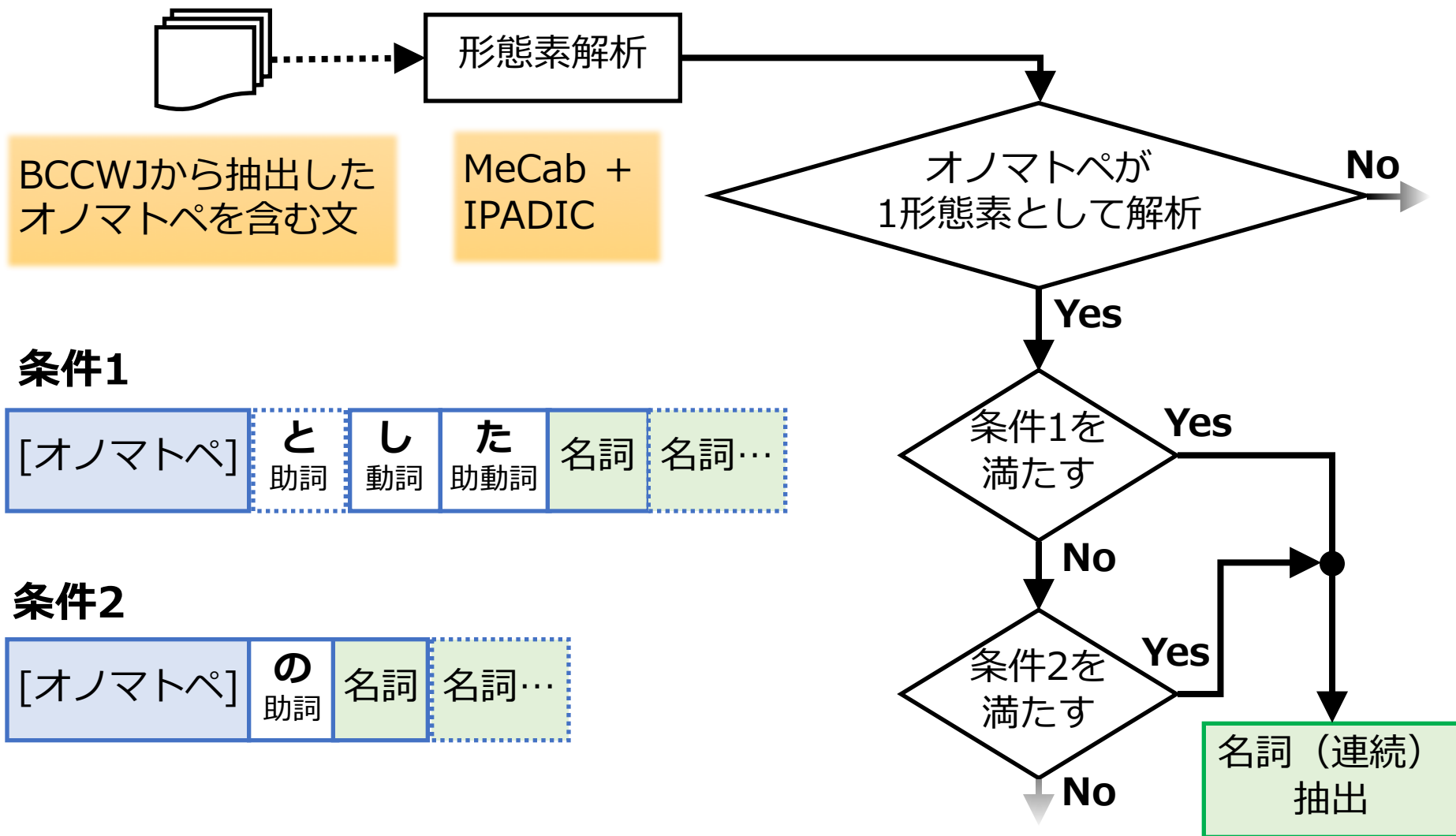
あっさり	した	デザイン
きょとんと	した	表情
ぎちぎちの	統制経済	

- オノマトペの修飾先として適切かどうか分析
 - 修飾先の自動推定手法実現への手がかりに
 - オノコロ辞書への登録

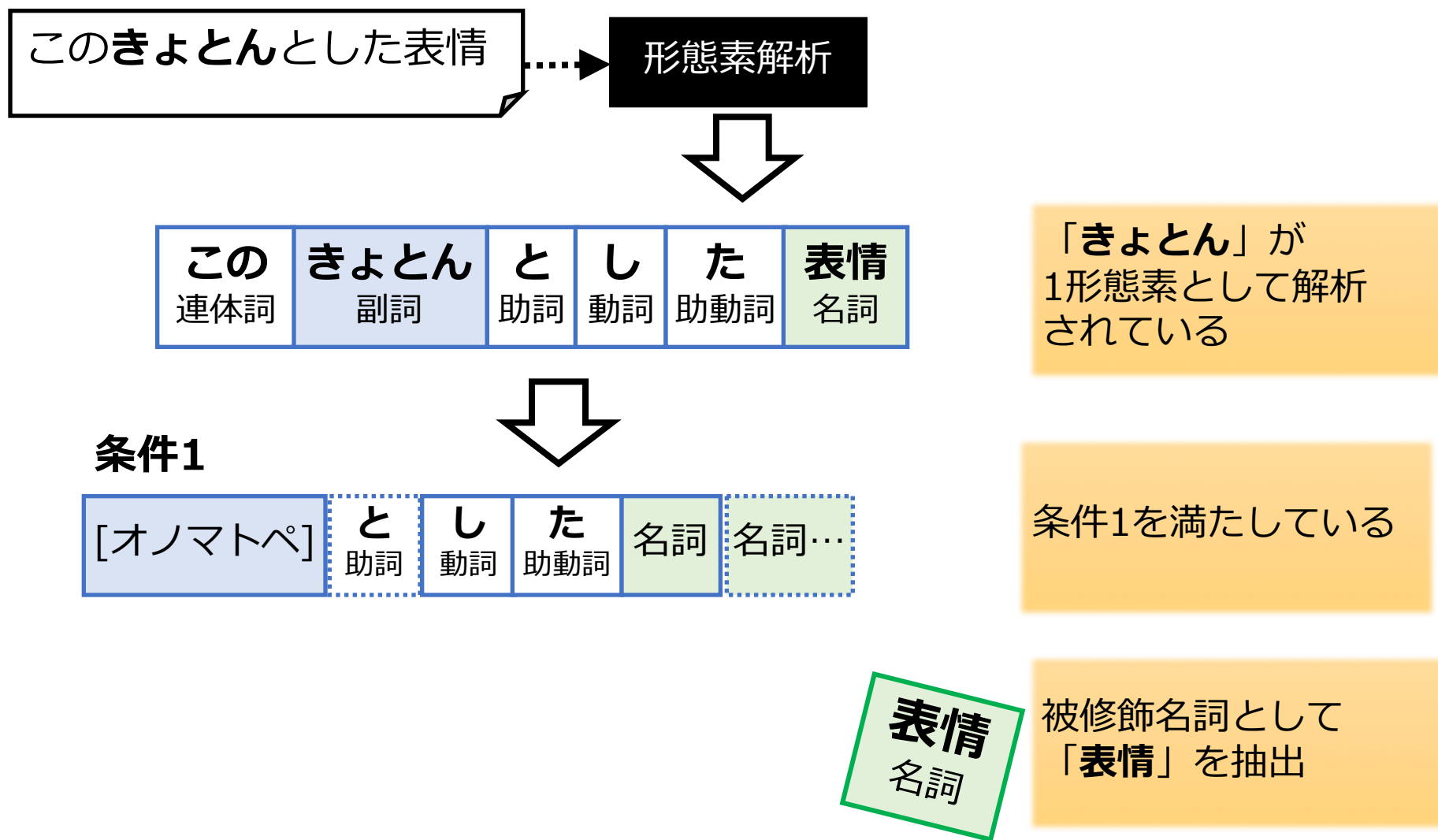
分析の手順



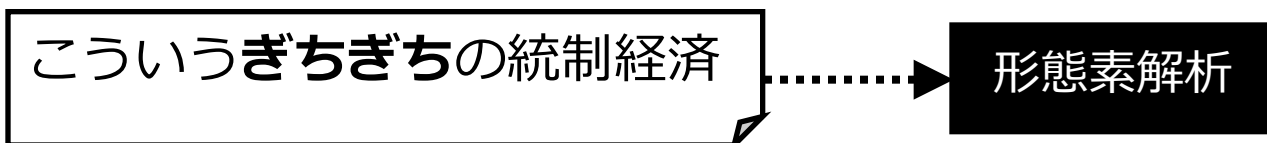
被修飾名詞の抽出方法



被修飾名詞の抽出例 (1)



被修飾名詞の抽出例 (2)



こういう	ぎちぎち	の	統制	経済
連体詞	副詞	助詞	名詞	名詞

「**ぎちぎち**」が
1形態素として解析
されている

条件2

[オノマトペ]	の	名詞	名詞…
	助詞	名詞	

条件2を満たしている
(名詞は2連続)

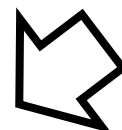


被修飾名詞として
「**統制経済**」を抽出

被修飾名詞が抽出できない例

あっけらかんとした明るい態度

形態素解析



あっけらかん	と	し	た	明るい	態度
副詞	助詞	動詞	助動詞	形容詞	名詞

「あっけらかん」が
1形態素として解析
されている



条件1

[オノマトペ]	と	し	た	名詞	名詞…
	助詞	動詞	助動詞	名詞	名詞…

条件1を満たさない
(直後が**形容詞**)

本手法では
何も抽出しない

抽出結果の評価のために

- 本手法を含む3パターンの出現頻度の調査

実際は (と)した～ も含む				ぎちぎちの 統制経済	ぴかぴかの 黒いかばん	実はがぶがぶ のみながら
① 本手法						
[オノマトペ] の 助詞 名詞 名詞…						
② 形態素一致						
[オノマトペ] の 助詞 任意の語						
③ 字面一致						
[オノマトペ] の 任意の文字列						

抽出結果

		(と) した	の	【合計】
①本手法	ぎちぎちの統制経済			
[オノマトペ]	の 助詞	名詞	名詞…	
		9,328	4,394	13,722
②形態素一致	ぴかぴかの黒いかばん			
[オノマトペ]	の 助詞	任意の語		
		11,691	4,880	16,571
③字面一致	実は がぶがぶ のみななら			
[オノマトペ]	の	任意の文字列		
		12,986	6,036	19,022

抽出結果

BCCWJの
オノマトペを含む文は
217,470

		(と) した	の	【合計】
①本手法	ぎちぎちの統制経済			
[オノマトペ]	の 助詞	名詞	名詞…	
		4.29%	2.02%	6.31%
②形態素一致	ぴかぴかの黒いかばん			
[オノマトペ]	の 助詞	任意の語		
		5.38%	2.24%	7.62%
③字面一致	実は がぶがぶ のみななら			
[オノマトペ]	の	任意の文字列		
		5.97%	2.78%	8.75%

抽出結果

②のみでカウントされた
2,849文には
形容詞で始まる名詞句が
含まれそう

		(と)した	の	合計
①本手法	ぎちぎちの統制経済			
[オノマトペ]	の 助詞	9,328	4,394	13,722
	名詞			
	名詞…			
②形態素一致	ぴかぴかの黒いかばん			
[オノマトペ]	の 助詞	11,691	4,880	16,571
	任意の語			
③字面一致	実はがぶがぶのみながら			
[オノマトペ]	の	12,986	6,036	19,022
	任意の文字列			

抽出結果

③のみでカウントされた
2,451文には
形態素解析誤りに起因する
抽出漏れが含まれそう

		(と) した	の	合計
①本手法	ぎちぎちの統制経済			
[オノマトペ]	の 助詞	名詞	名詞…	
		9,328	4,394	13,722
②形態素一致	ぴかぴかの黒いかばん			
[オノマトペ]	の 助詞	任意の語		
		11,691	4,880	16,571
③字面一致	実は がぶがぶ のみななら			
[オノマトペ]	の	任意の文字列		
		12,986	6,036	19,022

抽出結果

本手法の抽出結果が
すべて正しいとは限らない

おっとりと した 金色の 瞳

		(と) した	【合計】
①本手法	ぎちぎちの統制経済		
[オノマトペ]	の 助詞 名詞 名詞…	9,328	4,394
			13,722
②形態素一致	ぴかぴかの黒いかばん		
[オノマトペ]	の 助詞 任意の語	11,691	4,880
			16,571
③字面一致	実はがぶがぶのみななら		
[オノマトペ]	の 任意の文字列	12,986	6,036
			19,022

抽出結果の分析

		(と) した	の	【合計】
③字面一致	実は がぶがぶ のみななら			
[オノマトペ]	の 任意の文字列	12,986	6,036	19,022

- 分析対象 
- 分析者は著者ら4名
- 分析概要
 - 本手法の抽出結果がオノマトペの修飾先として適切か
 - そうでない場合、適切な修飾先はどこか

分析例（1）

自動抽出部分

ざらざらした舌で, ... ➡ 舌

うねうねとした急坂の
バリエーションが... ➡ 急坂

ざわざわとした。 ➡ (なし)

がつつした感じの
けだもののような**食欲** ➡ 感じ

分析者記入部分

係り先名詞 特記事項

○	(なし)
○	AのB Bは修飾先にならない
○	サ変
○	本当は「 食欲 」 がほしい

分析例 (2)

自動抽出部分

アッケラカンとした赤城
の性格から推察すると



赤城

アッサリとした線描も…



(なし)

しばらくじたばたした
あと, …



あと

ハロウィーンの夜, …



夜

分析者記入部分

係り先名詞 特記事項

性格

AのB

線描

解析ミス

(なし)

サ変

※

オノマトペじゃない

分析結果 -統計-

分析者による判定	細分類	件数	割合
抽出結果が適切	抽出結果がある	11,364	59.74%
	抽出結果がない	2,428	12.76%
抽出結果が誤り	適切な被修飾名詞が別に存在する	3,389	17.82%
	被修飾名詞は存在しない	764	4.02%
オノマトペではない		1,077	5.66%

分析結果 -統計-

分析者による判定	細分類	件数	割合
抽出結果が適切	抽出結果がある	11,364	59.74%
	抽出結果がない	2,428	12.76%
抽出結果が誤り	適切な被修飾名詞が別に存在する	3,389	17.82%
	被修飾名詞は存在しない	764	4.02%
オノマトペではない		1,077	5.66%

このうちの7%
オノマトペが正しく解析されず名詞を抽出できなかったことが
正しい出力結果につながったもの

翼	を	ふり	ふり	し	た	時
名詞	助詞	動詞	名詞	動詞	助動詞	名詞

分析結果 -統計-

分析者による判定	細分類	件数	割合
抽出結果が適切	抽出結果がある	11,364	59.74%
	抽出結果がない	2,428	12.76%
抽出結果が誤り	適切な被修飾名詞が別に存在する	3,389	17.82%
	被修飾名詞は存在しない	764	4.02%
オノマトペではない		1,077	5.66%

このうちの31%
オノマトペが正しく解析されず名詞を抽出できなかった事例

キッ
副詞

チリ
名詞

と
助詞

し
動詞

た
助動詞

手
名詞

分析結果 -統計-

分析者による判定	細分類	件数	割合
抽出結果が適切	抽出結果がある	11,364	59.74%
	抽出結果がない	2,428	12.76%
抽出結果が誤り	適切な被修飾名詞が別に存在する	3,389	17.82%
	被修飾名詞は存在しない	764	4.02%
オノマトペではない		1,077	5.66%

このうちの69%

形態素解析結果は正しいが、抽出結果は不適切
本手法の抽出方法自体を改善する必要あり

考察 -遠い修飾先-

抽出結果

久彦という人間の**ねっとり**とした、それでいて執拗に
言い寄ることを喜ぶ**性格**が感じ取られた



(なし)

うらやましいほど**つやつや**した開放的な新鮮な香気に
みちあふれた**くだもの**の生命が...



(なし)

- 修飾先が読点を越えた先であったり
何重にも渡る修飾を受けた先にある例
- 単純に形態素解析や係り受け解析結果から
適切な被連体修飾名詞を抽出するのは困難

考察 -形容詞・連体詞・接頭詞-

あつけらん 副詞	と 助詞	し 動詞	た 助動詞	明るい 形容詞	態度 名詞
うっとり 副詞	と 助詞	し 動詞	た 助動詞	その 連体詞	表情 名詞
カサカサ 名詞	の 助詞	お 接頭詞	肌 名詞		
のんびり 副詞	と 助詞	し 動詞	た 助動詞	ご 接頭詞	性格 名詞

- 抽出手法に形容詞や連体詞，接頭詞を許容する改善で対応可能

考察 -有用なコロケーションが別にある①-

抽出結果

クリーム <u>たっぷり</u> の <u>エクレア</u> は特に絶品	⇒	エクレア
本物 <u>そっくり</u> の <u>声</u> で鳴きます	⇒	声
新島の海面 <u>すれすれ</u> の <u>部分</u>	⇒	部分

- 【名詞＋オノマトペ】の形
 - オノマトペ直前の名詞が意味的な修飾先
- この形をとるオノマトペは種類が限られる傾向
 - たっぷり, そっくり, すれすれ, ぎりぎり...
- ただし例外もある

毎月たっぷりの額が送金できるように

考察 -有用なコロケーションが別にある②-

抽出結果

のどのイガイガの <u>改善</u> には	⇒	改善
太陽が <u>かんかん</u> の <u>時間帯</u> に水やりすると	⇒	時間帯

- オノマトペの前の文節が意味的な修飾先
- すべての例で前の文節が意味的な修飾先とは限らない
- 適切な被修飾名詞の抽出が難しい

考察 -有用なコロケーションが別にある③-

抽出結果

が つつ した <u>感じ</u> のけだもののような 食欲	⇒	感じ
グレイスが うんざり した <u>よう</u> な 声 で言った	⇒	よう

- 抽出結果は被修飾名詞として適切だが…
- オノマトペの有用なコロケーションが別に存在する形
- 「感じ」, 「よう」は汎用的に用いられる
 - コロケーションとしての特徴に乏しい

考察 - 並列関係 -

抽出結果

しっかりとした <u>基準</u> と行動を手に入れる	⇒	基準
カサカサの <u>粉</u> やひび割れがみられ	⇒	粉
しっかりした <u>味、香り</u> 、水色がきれいなお茶	⇒	味
しっかりした <u>打球感・打球音</u>	⇒	打球感

- 適切な被修飾名詞が複数あるような事例
- 「と」, 「や」, 読点による並列
 - 並列関係のすべての名詞が修飾先とは限らない

考察 - 判断が難しい連続名詞 -

抽出結果

しっとりとした花色選びもキーポイント



花色選び

「花色」のみが「しっとり」の修飾先？

しっかりとした花色選びが重要だ



花色選び

しっかりした選び方が重要
と解釈できる

考察 -形態素解析ミス-

抽出結果

カサカサ	の	白	つぽい	肌	⇒	白
名詞	助詞	名詞	名詞	名詞		

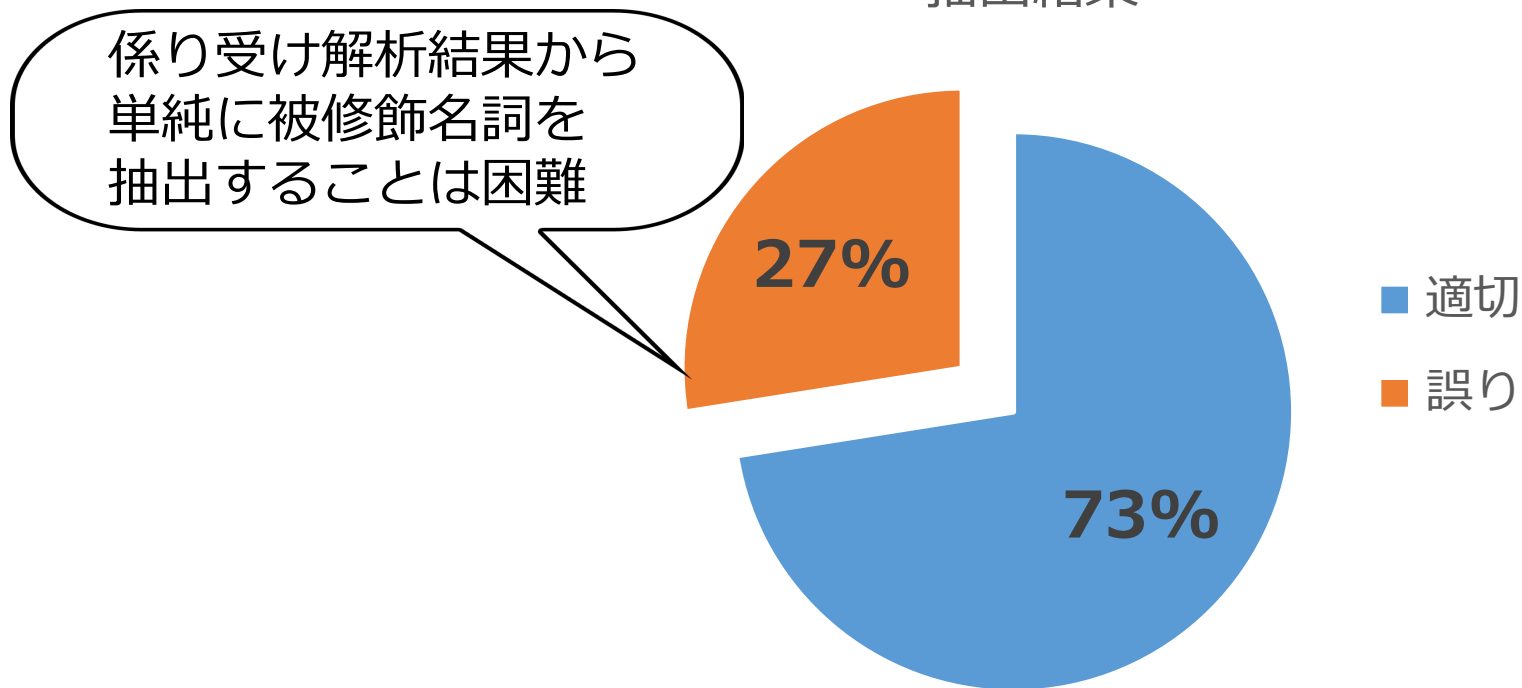
カツ	チリ	し	た	スーツ	姿	⇒	(なし)
副詞	名詞	動詞	助動詞	名詞	名詞		

するする	の	びるなめらかさ	⇒	びるなめらかさ
副詞	助詞	名詞		

- 抽出手法の前処理部分で起こる問題
- 形態素解析の辞書を拡張することでミスを減らせる可能性

まとめ

- 【オノマトペ】（と）した
 - 【オノマトペ】の
 - オノマトペの修飾先として適切か分析
- 抽出結果



次のステップ

出力結果は適切か？

ざらざらした舌で, ...	⇒ 舌	○
うねうねとした急坂のバリエーションが...	⇒ 急坂	○
アッケラカンとした赤城の性格から推察すると	⇒ 赤城	×
しばらく じたばた したあと, ...	⇒ あと	×

- 本手法をベースにオノマトペの修飾先として適切なものだけを抽出する改善
 - コロケーションを大量に集める
 - 本手法で抽出できないものを抽出する手がかりにしたい