

オノマトペの語義決定に寄与するコロケーションの分析

藤田実智斗^{*1} 内田ゆず^{*2} 荒木健治^{*3}

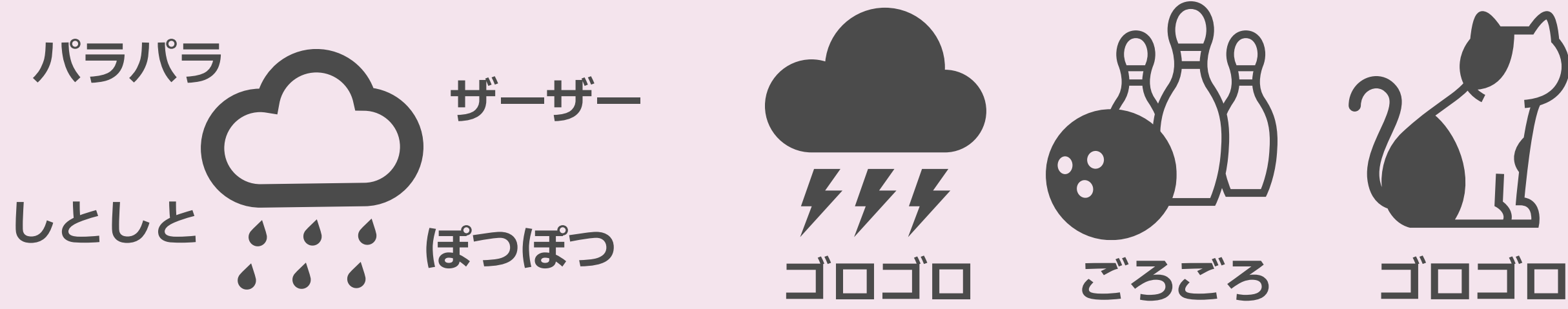
^{*1} 北海道大学 大学院情報科学院

^{*2} 北海学園大学

^{*3} 北海道大学 大学院情報科学研究院

研究背景

日本語オノマトペは種類が多く、その多くは複数の語義を持つ。



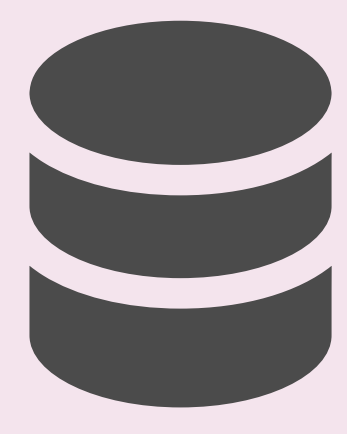
雨の降り方も様々

同じオノマトペでも
いろんな意味がある

日本語母語話者はこうした語義の判別を感覚的に行うため、その判別基準を日本語学習者が学ぶことは難しい。

そこで、オノマトペの使い分けが分かりやすくなるようなオノマトペコロケーションデータベースの構築をしたい！

天気が悪くなってきて、
ゴロゴロ言い始めた



語義①
雷が鳴る音。また、雷が鳴るような音。

共起頻度が高い語
雷 天気 鳴る 言う . . .

本発表では、オノマトペコロケーションデータベース構築の自動化に向けて、語義とコロケーションの関係を分析する。

語義に関係している語はどれ？

語義決定に寄与する語（寄与コロケーション）を抽出して語義とコロケーションの関係を分析！

『現代日本語書き言葉均衡コーパス（BCCWJ）』上で頻度が高い10種類の多義オノマトペを対象とする。

対象オノマトペを含む各400文について、3人の評価者が語義と語義を決める際に手がかりとなる文節をマークする。

文節	評価者1	評価者2	評価者3	合計値	「ごろごろ」の語義① 「雷が鳴る音。また、雷が鳴るような音。」の例
その	0	0	0	0	
音は、	1	0	0	1	
雷が	1	1	0	2	
ゴロゴロ	-	-	-	-	
鳴るみたいに	1	0	1	2	
遠くまで	0	0	0	0	
聞こえるんだ。	0	0	0	0	

2人以上がマークした文節の
先頭の単語を分析対象語として
抽出する

「ごろごろ」の語義を決めている語
(寄与コロケーション)

雷 鳴る

「ごろごろ」の抽出例（一部抜粋）

語義番号	定義	寄与コロケーション	
		名詞	動詞
語義①	雷が鳴る音。また、雷が鳴るような音。	音, 雷, 雨, おなか, 轟き, 声, 喉, 天気, 猫, 腹, お腹, 雨粒	鳴らす, 鳴る, する, 聞こえる, いう, ひびく
語義②	かなり重量のある物体や肉体が連続してこがるようす。	台車, 蒲団, 地面, 上, 身, 回転, 斜面, 石	転がる, 転がす, 寝返る, 横たえる
語義③	働かないで時を過ごすようす。何もしないで時を過ごすようす。	家, 部屋, 毎日, 元旦, リビング, 和室, ベッド, 日曜日, こたつ, お家, 休み	する, 過ごす, できる, やる, 寝る, 休む, くつろぐ
語義④	たくさんありすぎて珍しくも貴重でもないようす。	男, 石, 連中, 岩, 人, 肉, 砂利, かたまり, じゃがいも, にんじん	する, いる, 転がる, ある, 出る, あふれる
語義⑤	かたまりや異物がはいり込んでいて違和感を感じるようす。	目, ヤニ, 糸, 右眼, 眼, 口, コンタクト	する, 痛む

人手で抽出された語（寄与コロケーション）の分析

- 辞書情報からだけでは読み取れない特徴がある語
→ 寄与コロケーションを収集することで、より語義や使い方を理解しやすくと考えられる
- 語義の定義文と密接に関係した語
→ 寄与コロケーション抽出の自動化が期待できる

抽出された寄与コロケーションは各語義に対応したクラスタを形成する？

各語義ごとの特徴が解釈できるように寄与コロケーションをまとめあげたい！
k-means++で自動的にまとめあげることができるか？

Word2Vec¹を用いて寄与コロケーションの単語分散表現を取得し、k-means++によるクラスタリングを行う。

k = 語義の数として、各語義の寄与コロケーショングループと形成されたクラスタを比較する。

「あっさり」の名詞クラスタ（一部抜粋） 平均Dice係数：0.42

語義①	語義②	語義③	語義④
色、味などの濃度が薄いようす。	形態、様式が複雑でなく、簡素であるようす。	態度、性格がしつこくなく、思いきりのよいようす。	事態が簡単に変化するようす。
語義①グループ 味, 醤油, スープ, 薄味, ハンバーグ, 麺, 甘味, プリン, タン, 味付け, 下味, レモン, ドレッシング, 冷奴, オリーブ, 調理, シソ, 杏仁, 香り, 山菜, 白菜, 風味	語義②グループ 顔, メール, 供述, 切れ, 表現, 作品, 形状, 風景, 説明, 名前	語義③グループ 別れ, 承諾, 同意, 放棄, スルー, 態度, 返事, 断念, 会話, OK, 告白, 受諾, 結論, 辞任, 退陣, 他人事, 拒否, 妥協, 一蹴, 説明, 承知, 撤回, 注射, 了解, 納得, 関係	語義④グループ 解決, 敗退, 確定, 実現, 接続, 紐, キープ, 倒産, 突破, 姿, 復活, 問題, 決着, 衰退, 連勝, 死, 一般, 成功, 失敗, 終了
麺, ハンバーグ, 鶏, コンソメ, 薄味, 味付け, 下味, 醤油, スープ, 納豆, 鰻, 調理, ラーメン, 豚骨, 肉, 鮭, 葱, 山菜, 油揚げ, 舞茸, 白菜	味, 甘味, プリン, 味わい, レモン, ドレッシング, 杏仁, 香り, 風味	敗退, 態度, 別れ, 供述, 返事, 承諾, 同意, 放棄, 告白, 受諾, 党, 辞任, 退陣, 拒否, 一蹴, 撤回, 決着, 連勝, 了解, 死	顔, 確定, 実現, 系, 断念, 表現, スルー, 解決, 結論, 倒産, タン, 突破, 他人事, 妥協, 復活, 形状, 説明, 承知, オリーブ, 白, 成功, 納得, 関係, 失敗, 終了
クラスタ① Dice係数：0.65	クラスタ② Dice係数：0.00	クラスタ③ Dice係数：0.52	クラスタ④ Dice係数：0.51

クラスタリング結果

- 食べ物に関係する語義は比較的きれいにまとめあがった
→ コロケーションを手掛かりとした自動語義分類が期待できる
- 複数の語義グループの語が混在したクラスタが形成された
→ 静的な単語分散表現だけでは、各語義のニュアンスを上手くまとめあげることができない

まとめ

- 語義決定に寄与する語（寄与コロケーション）を人手で抽出し、語義との関係を分析した
→ より語義を理解しやすくなる寄与コロケーションが抽出された
→ 定義文から容易に推測できる寄与コロケーションは、抽出の自動化に役立てられそう

抽出した寄与コロケーションがクラスタリングで語義ごとに分かれるか確認した
→ 特定のドメインに関係する語はクラスタを形成し、他は混在したクラスタを形成した

今後の課題

- 寄与コロケーションの分析
 - 表層格情報と語義の関係の分析
 - 寄与コロケーションの自動抽出方法の検討
- 寄与コロケーションのクラスタリング
 - 自動語義分類システムへの適用方法の検討
 - 文脈を考慮した単語分散表現の活用

¹ 「日本語 Wikipedia エンティティベクトル」を使用。（鈴木正敏, 松田耕史, 関根聡, 岡崎直観, 乾健太郎. Wikipedia 記事に対する拡張固有表現ラベルの多量付与. 言語処理学会第 22 回年次大会発表論文集, pp.797-800, 2016.）