

Fuzzy System Symposium 2015

現代オノマトペ・コロケーション・ データベース構築のための 係り先動詞の分析

高丸圭一, 内田ゆず, 乙武北斗, 木村泰知

宇都宮共和大学

北海学園大学

福岡大学

小樽商科大学

オノマトペの語義

- 擬音語, および, 擬態語

【辞典に記述された語義】

- 雷がごろごろ鳴る
- いつも家でごろごろしている
- 石がごろごろ転がっている

雷の鳴る音

何もせずに無駄に暮らすさま

雑然と転がっているさま

- 辞典の記述から外れて(派生して)使われる

- そういう人がごろごろいる
- 状況がごろごろ変わっていく

数が多いさま

変わりゆくさま

- 人手の語義分類でも判断に迷うことがある

- 既存の語義か? 新しい項目を立てるか??

オノマトペの派生語義・新しい語義

- 日本語母語話者は、直感的に理解可能
- 語義を厳密に記述するのは困難

- 現代の日本語にけるオノマトペの用法

現代オノマトペ・コロケーション・データベース



実際の用例(日本語文)において
オノマトペと結びつく語を抽出したデータベース

オノマトペのコロケーション

- オノマトペとよく結びつく語
 - 係り先動詞のコロケーション

例

「ごろごろ」

「転がる」

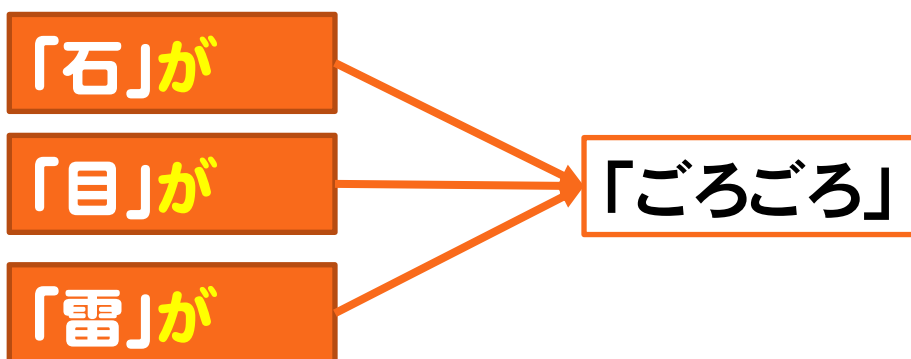
「する」

「鳴る」

オノマトペのコロケーション

- オノマトペとよく結びつく語
 - 係り先動詞のコロケーション
 - 表層格のコロケーション

例



オノマトペのコロケーション

- オノマトペとよく結びつく語
 - 係り先動詞のコロケーション
 - 表層格のコロケーション
 - 表層格・係り先動詞の組み合わせ (=文脈)

例



オノマトペの派生語義・新しい語義

- 日本語母語話者は, 直感的に理解可能
- 語義を厳密に記述するのは困難

- 現代の日本語にけるオノマトペの用法

現代オノマトペ・コロケーション・データベース

- **日本語非母語話者**
 - オノマトペの習得に利用
- **オノマトペの工学的利活用**
 - 文生成における前方・後方文脈のデータ

現代オノマトペ・コロケーション・ データベースの構築の流れ

現代の日本語文の収集

コーパス
Web文書

オノマトペを含む用例の抽出

FSS2014
JSAI2015

用例からコロケーションの抽出

抽象化／語義に基づく整理

データベース化

本発表では

- コロケーションのうち“係り先動詞”に着目
 - ① 係り先動詞の抽出
 - ② 係り先動詞とオノマトペの語義の関係を分析
- 分析対象データ
 - 地方議会会議録コーパス (2010年度)
 - 先行研究でオノマトペの抽出・語義分類済
 - 語義分類済みのオノマトペ 153語 (約1万文)

高丸圭一, 内田ゆず, 乙武北斗, 木村泰知(2015)
「地方議会会議録コーパスにおけるオノマトペ
—出現傾向と語義の分析—」
人工知能学会論文誌, Vol.30, No. 1, SP2-K, pp.306-318

本発表の分析対象単語

- 14語 (総文数 1,734)

- 条件
 - 複数の語義
 - 各語義が10以上
 - 4文字

#	オノマトペ	文数	語義数
1	がたがた	132	4
2	がちり	120	3
3	ぐずぐず	47	3
4	ごろごろ	66	6
5	とんとん	112	3
6	のんびり	244	2
7	ばたばた	230	3
8	ばっさり	132	3
9	ぱらぱら	78	4
10	ふらふら	94	3
11	ぼこぼこ	40	3
12	ぼつぼつ	98	3
13	ぽつぽつ	72	4
14	ゆったり	269	2
	合計	1,734	—

係り先動詞・表層格の抽出

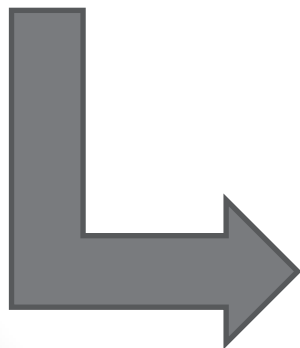
- 日本語係り受け解析器 CaboCha

【入力文】

私が本会議でもって指摘してから初めて農地法違反という問題でがたがたしだしてきて... (後略)

【入力文】

盆前からこんなにがたがたすることはなかったかもしれません。



オノマトペ	係り先動詞	カラ格	デ格
がたがた	しだす	指摘して	問題
	する	盆前	Φ
⋮	⋮	⋮	⋮

係り先動詞の抽出

- 係り先動詞が抽出された
 - 1,734文中1,428文
- 「オノマトペ—係り先動詞」
 - **397組**
 - 出現回数1回：238組 (60%)
 - 出現頻度上位の組を見

オノマトペ	係り先	頻度
ゆったり	する	119
のんびり	する	107
ばたばた	する	95
がたがた	する	63
ふらふら	する	58
ばっさり	切る	33
がちり	する	23
ごろごろ	する	23
がたがた	言う	19
ぐずぐず	する	18
ばっさり	削る	18
がちり	やる	17
ぽつぽつ	出る	17
とんとん	なる	15
がちり	組む	14
ばたばた	やる	14
ばっさり	切り捨てる	13
のんびり	やる	12
のんびり	構える	12
ゆったり	過ごす	12

係り先動詞と語義数(出現回数2以上)

- **単一語義**の「オノマトペー係り先動詞」

– 24.4% (97組 / 397組)

– 例

「がたがた一言う」

19例

「騒々しく不平を言ったり,
騒ぎ回ったりするさま」

- **複数語義**の「オノマトペー係り先動詞」

– 15.7% (62組 / 397組)

– 例

「ごろごろーある」

2例

「たくさんあるさま」

1例

「変化するさま」

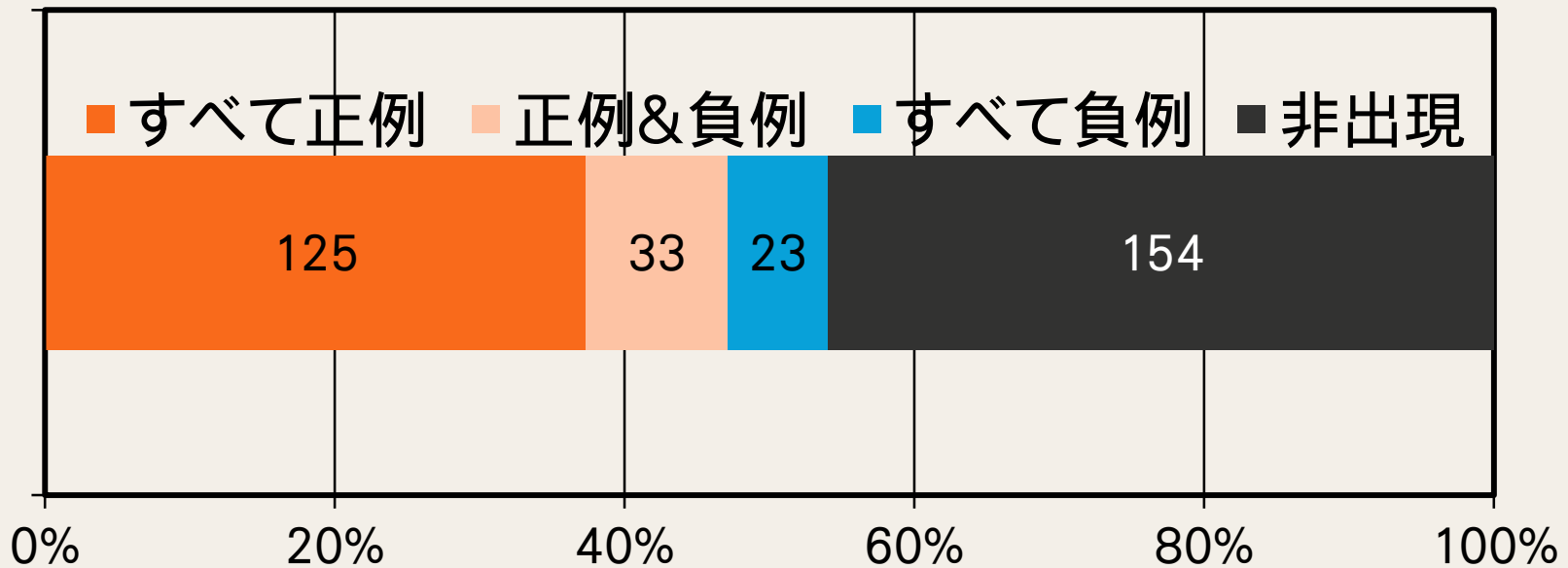
- 複数語義の係り先動詞:「する」「やる」「ある」が多い

単一語義で出現する組

- 出現頻度の高かった「オノマトペー係り先動詞」

オノマトペー係り先動詞	頻度
がたがたー言う	19
ばっさりー削る	18
がちりーやる	17
とんとんーなる	15
がちりー組む	14
ばっさりー切り捨てる	13
のんびりーやる	12
ぐずぐずー言う	11
のんびりー暮らす	11
ばたばたー倒れる	11
ぱらぱらー見る	11

データ数を
増やして
評価してみたい



- 分析データにおいて
 - 単一語義で出現した組
 - 出現頻度が1回の組
- ↓
- 評価データから抽出
 - 語義を手手で確認

計335組

181組(1,216例)

評価の結果(組数)

- すべてが正例の組(使用頻度上位10組)

オノマトペー係り先動詞	頻度
とんとんーなる	162
ばっさりー削る	59
ぐずぐずー言う	38
ばっさりー切り捨てる	38
のんびりーやる	29
ばたばたー倒れる	26
がっちりーやる	24
ふらふらーなる	24
ぱらぱらー見る	23
のんびりー過ごす	19

負例の要因分析

① 解析誤り → 技術的課題

② 語義の不足

③ 語義分類が困難な文

オノマトペの語義の
難しさの問題

④ 固有表現中のオノマトペ → 対象外の用例

「のんびり村」「とんとん会館」など

⑤ オノマトペ自体の格によって決定される語義

語義の不足, 語義分類が困難な文

- 辞典にない語義 【例】**ばたばた**—出る

ばたばた	ものが続けざまに倒れたり衰えたりするさま
	慌ただしくことを行うさま

- 「日が落ちるころに、**ばたばた**とシカが出てくるということです。」
- 「やめる保育士さんが**ばたばた**出たりとか」

- 語義判断に迷う用例 【例】**ぼつぼつ**—入る

ぼつぼつ	ものごとを少しずつ進めたり, しはじめるさま
	それをする時刻, また, 時期になりつつあるさま

- 「将来的な見通しというのも**ボツボツ**視野に入っているのかどうか」
- 「そしたらもう**ぼつぼつ**その年限に入ってくるんじゃないかな」

⑤オノマトペ自体の格によって決定される語義

• 「とんとん—いく」 (正例率: 38%)

- 正例 (12例): 「ものごとが順調にはかどるさま」
- 負例 (19例): 「ちょうど同じくらいであるさま」

• 「とんとんと—いく」

- 「こっちの思うとおりにとんとんといけばいいですけども」

• 「とんとん {で / まで / に} いく」

- 「当初はとんとんで収支がいくと言われるにもかかわらず」
- 「収支とんとんまでいかないというのも現状ですし」
- 「到底とんとんにいく、黒字になるということは、私はないと」

まとめ

データベースの有用性
オノマトペコロケーション

頻度が高いコロケーション
による語義の記述

新しい用例・用法の
コロケーションの提示

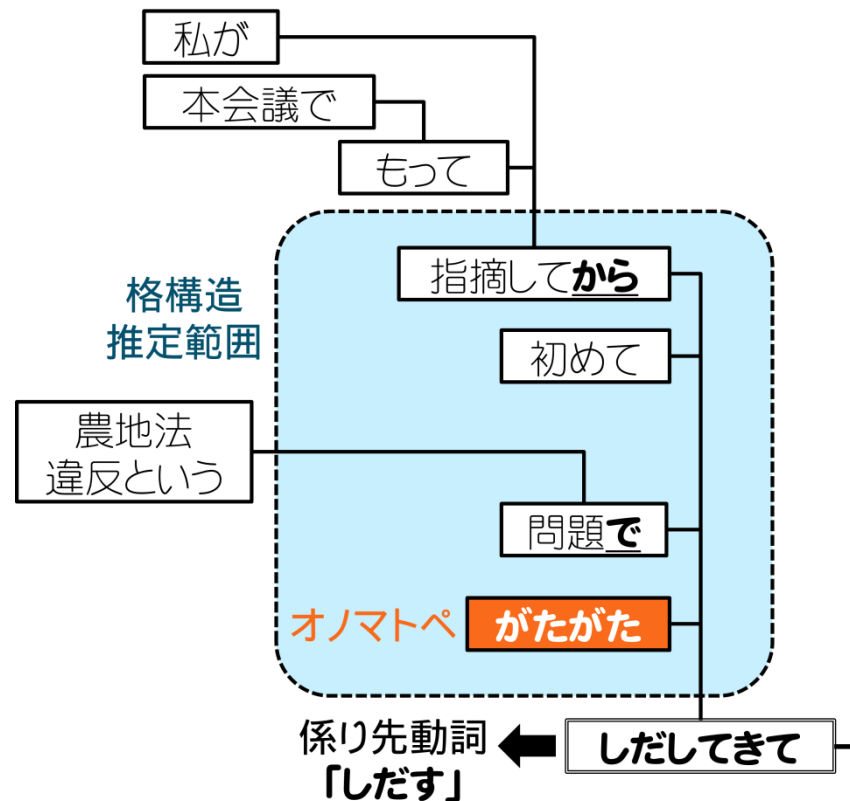
今後の課題

- 本発表
 - 語義分類済み地方議会会議録コーパス
 - 文書のドメイン特有の語義の偏り
 - 係り先動詞を分析
- 今後の予定
 - オノマトペの用例をさらに収集
 - **文書ドメインの拡充** (ウェブコーパスの利用)
 - **表層格のコロケーション**を分析

格構造抽出の例 (基本的な処理)

【入力文】

私が本会議でもって指摘してから初めて農地法違反という問題でがたがたしだしてきて... (後略)



格構造抽出の例 (基本的な処理)

【入力文】

私が本会議でもって指摘してから初めて農地法違反という問題でがたがたしだしてきて... (後略)



【結果】

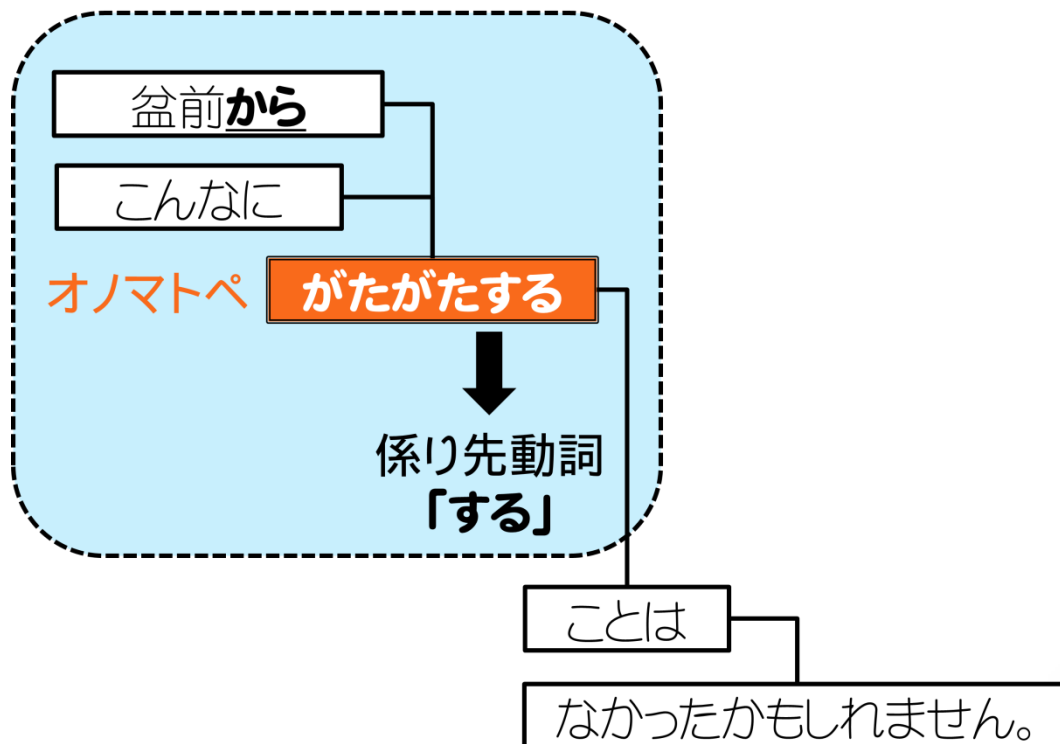
- オノマトペ: 「がたがた」
- 係り先動詞: 「しだす」
- カラ格: 「指摘して」
- デ格: 「問題」

格構造抽出の例(オノマトペ+する)

【入力文】

盆前からこんなにがたがたすることはなかったかもしれません。

格構造推定範囲



格構造抽出の例(オノマトペ+する)

【入力文】

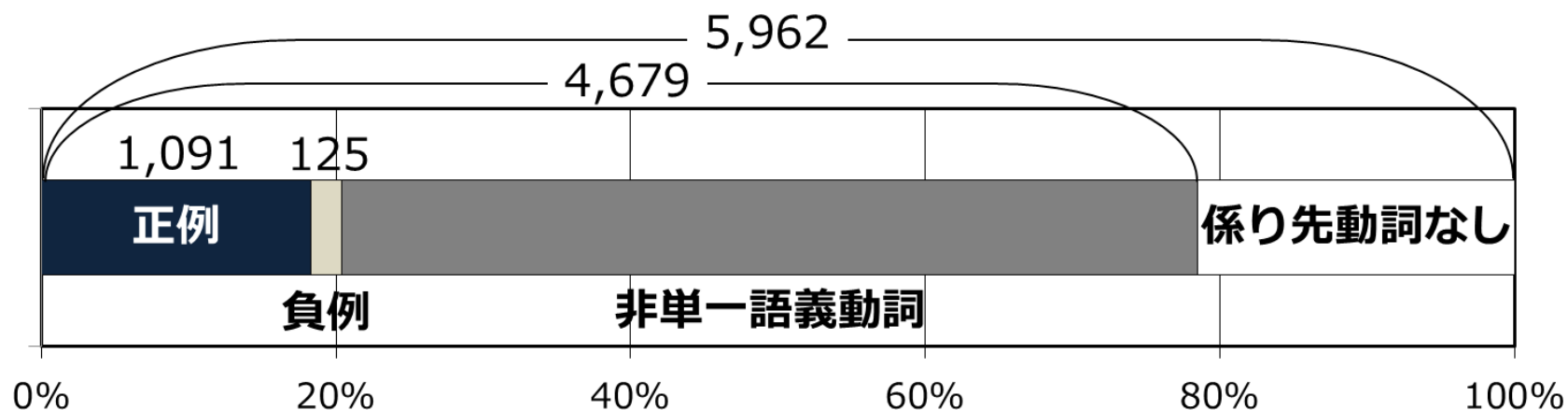
盆前からこんなにがたがたすることはなかったかもしれません。



【結果】

- オノマトペ: 「がたがた」
- 係り先動詞: 「する」
- カラ格: 「盆前」

評価の結果(文数)



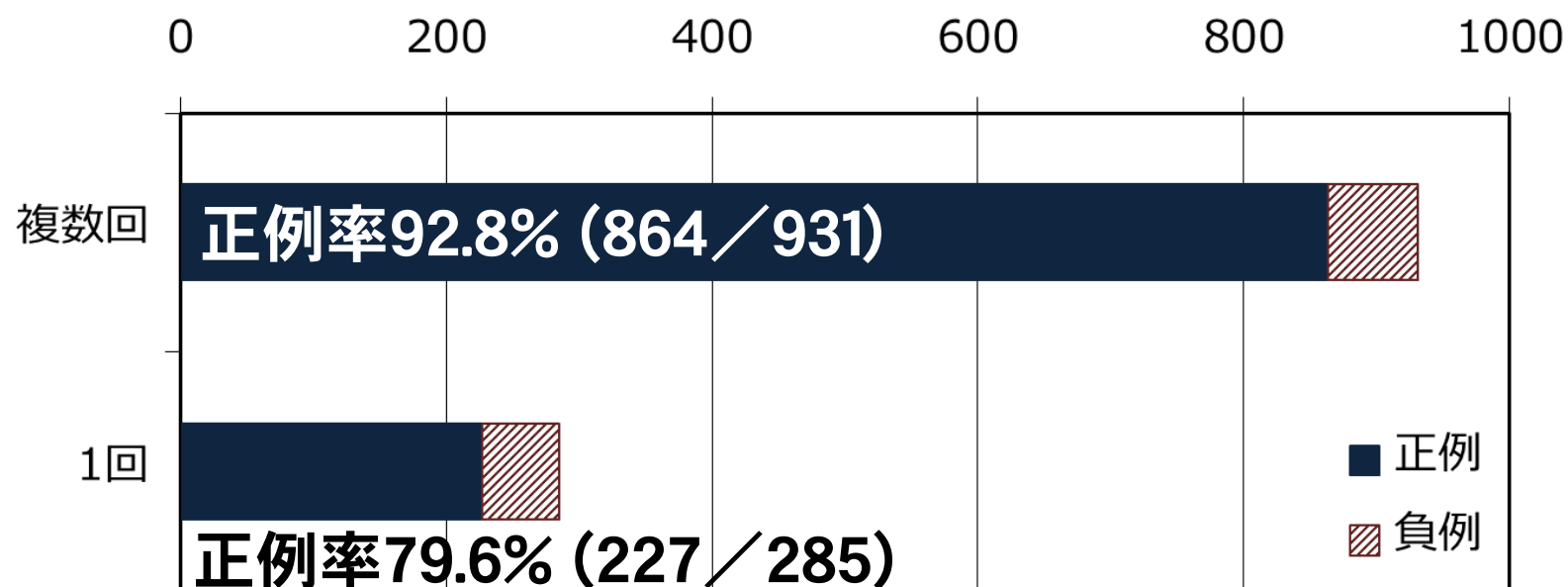
- 分析データの単一語義の組と一致する用例

- 1216文

- 正例(分析時と同じ語義の用例) :1091文
 - 負例(分析時とは異なる語義の用例) :125文

評価の結果(文数)

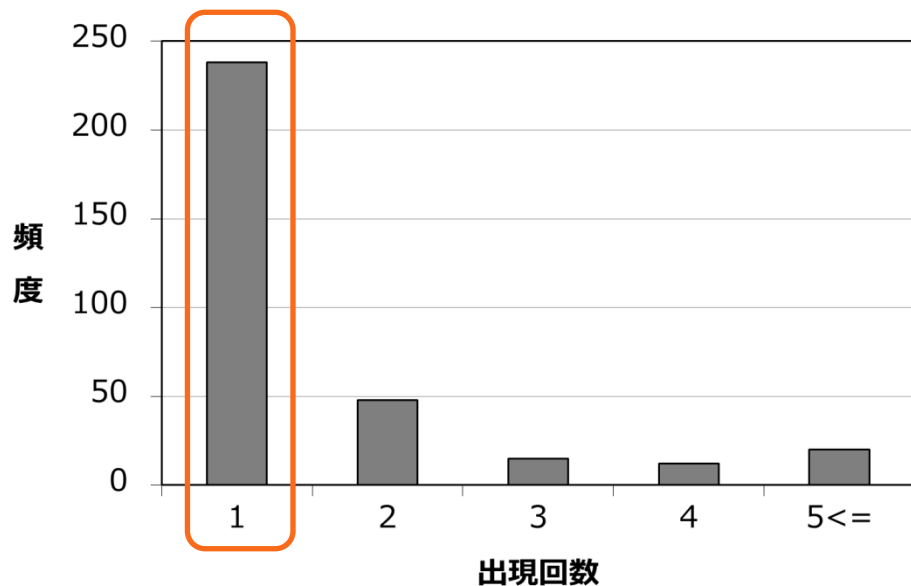
- 分析データの出現頻度別の評価



— データ数が多くすることで、語義と関係の強い係り先動詞を得ることができる

単一語義で出現する組 (335組)

- 70%が出現頻度 1
 - 用例数 1 であれば, 当然単一語義となる



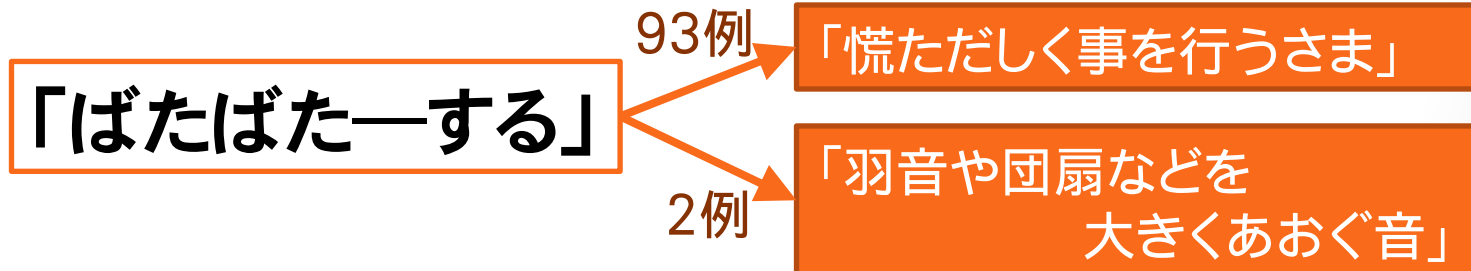
- より多くの用例によって検証する必要がある

複数語義で出現する組(62組)

- 係り先動詞

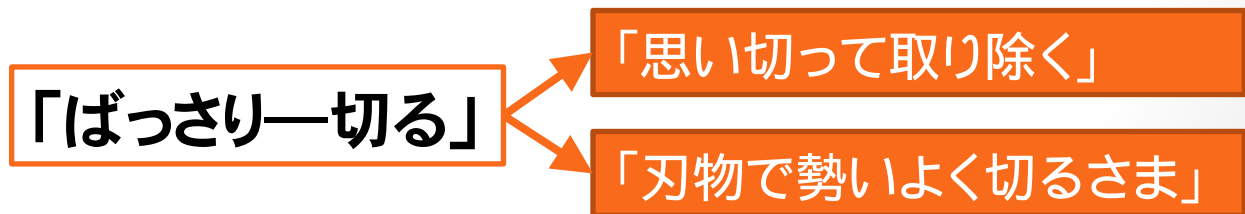
- 「する」 13組 (32語義)
- 「やる」 5組 (10語義)
- 「ある」 4組 (8語義)
- ⋮

- 出現頻度に偏りがある例もある



複数語義で出現する組(62組)

- 「する」「やる」「ある」が高い頻度で共起する
 - さまざまな文脈で使用可能
 - 語義の決定には寄与しづらい
- 表層格によって語義が決まる可能性



複数語義で出現する組(62組)

- 「する」「やる」「ある」が高い頻度で共起する
 - さまざまな文脈で使用可能
 - 語義の決定には寄与しづらい
- 表層格によって語義が決まる可能性



係り先動詞・表層格の抽出

- 日本語係り受け解析器 CaboCha
- **オノマトペを含む節の係り先動詞**を抽出
(※「オノマトペ+する」の場合は「する」を抽出)
- 係り先動詞の**係り元文節の助詞**を手がかりとして**表層格**を抽出

ガ格, ヲ格, ニ格, デ格, カラ格,
へ格, ト格, ヨリ格, マデ格

評価の結果(組数)

- 分析データにおいて単一語義の組
 - 335組中181組が評価データに出現
 - このうち125組が分析データと同じ単一語義

