

# 階層的機能表現辞書に基づく新聞記事中の機能表現の調査・分析

長坂泰治、坂本明子、宇津呂武仁、森下洋平 筑波大学大学院 システム情報工学研究科  
松吉俊 奈良先端科学技術大学院大学 情報科学研究科  
土屋雅稔 豊橋技術科学大学 情報メディア基盤センター

## ①はじめに

### 機能表現を自動的に検出したい

**機能表現**: 複数の単語から構成され、一つの単語のように振舞う表現

例:「(に)たいして」、「なければならぬ」、「からして」

**機能的用法**: 複数の単語から構成され、一つの単語のように働く

**内容的用法**: 構成する自立語が持つ本来の意味が合わさって働く

- |                    |                              |
|--------------------|------------------------------|
| (1) 私は自分の夢について話した。 | (1) I talked about my dream. |
| 機能的用法              | following                    |
| (2) 私は彼について走った。    | (2) I ran about him.         |
| 内容的用法              |                              |

## ③提案方式



## ②対象となる機能表現

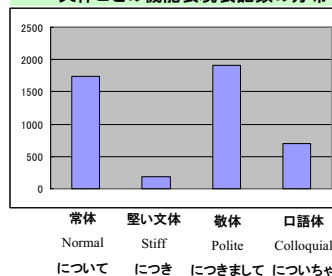
つつじ: 日本語機能表現一覧(2007 松吉)の16,801表現

表現数 階層 分類観点

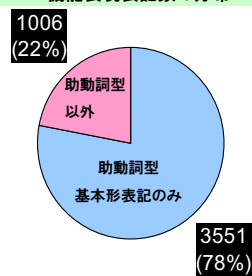
341	L <sup>1</sup>	見出し	008 (に)たいして
435	L <sup>2</sup>	意味	0081 (に)たいして <対象> 0082 (に)たいして <側面>
555	L <sup>3</sup>	文法	0081P (に)たいして <格助詞型> 0081D (に)たいしての <連体助詞型>
774	L <sup>4</sup>	交替	0081P1 (に)たいして <交替無し> 0081P2 (に)たいし <交替有り>
1,187	L <sup>5</sup>	音韻変化	0081P1x (に)たいして 0081P1x1h (に)たいしちや
1,810	L <sup>6</sup>	挿入	0081P1xx (に)たいして
6,870	L <sup>7</sup>	活用	0081P1xx01 (に)たいして
9,722	L <sup>8</sup>	です・ます	0081P1xx01n (に)たいして 0081Pxx01s (に)たいまして
16,801	L <sup>9</sup>	表記	008P1xx01n01 (に)たいして <仮名> 0081Pxx01n02 (に)たいして <漢字>

## ④日本語機能表現一覧における機能表現の分布

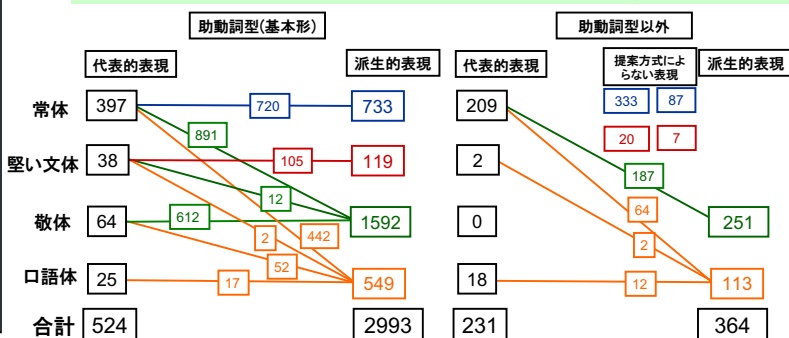
文体ごとの機能表現表記数の分布



機能表現表記数の分布



代表的表現への集約における文体ごとの機能表現表記数の分布



## ⑤新聞記事における機能表現の分布

毎日新聞1995年: 約430MB、約130万文

機能表現表記の出現数分布

回数	割合		
50回以上	12%		
10回以上 50回未満	5%		
1回以上 10回未満	13%		
0回	68%		
552	264	566	3021

50回以上出現する代表的表現と派生的表現の組数

文体の組み合わせ	組数	組の例
常体-常体	25	てよいでもよい
常体-敬体	30	なければならぬ なければならぬです
常体-口語体	20	ても仕方ない ても仕方ない
合計	75	

50回以上出現する機能表現表記数の分布

	助動詞型(基本形)		助動詞型以外			合計
	代表的表現	派生的表現	代表的表現	派生的表現	その他	
常体	164	38	87	0	178	467
堅い文体	8	3	0	0	9	20
敬体	14	42	0	1	0	57
口語体	7	37	1	13	0	58
合計	193	120	88	14	187	602

機能的用法・内容的用法の分布

両方の用法が適度に混合して出現する表記の割合が全体の約3分の1である。  
日本語複合辞用例データベース[土屋 06]では、187表記のうち、約3分の1が、混合して出現する。  
その他のみでは、約40%となる。

表中の数字は、以下の通り(X: 機能的用法の割合)

90% ≤ x ≤ 100%となる表現の割合 / 10% < x < 90%となる表現の割合 / 0% ≤ x ≤ 10%となる表現の割合

	助動詞型(基本形)		助動詞型以外			合計
	代表的表現	派生的表現	代表的表現	派生的表現	その他	
常体	66.2 / 28.1 / 5.7	0 / 100 / 0	57.1 / 32.5 / 10.4	—	53.0 / 39.1 / 7.9	58.5 / 33.9 / 7.6
堅い文体	83.3 / 16.7 / 0	—	—	—	14.3 / 57.1 / 28.6	46.2 / 38.5 / 15.3
敬体	100 / 0 / 0	85.7 / 0 / 14.3	—	—	—	94.4 / 0 / 5.6
口語体	20.0 / 60.0 / 20.0	33.4 / 33.3 / 33.3	0 / 0 / 100	0 / 100 / 0	—	58.3 / 33.1 / 8.6
合計	67.7 / 26.7 / 5.6	50.0 / 27.8 / 22.2	56.4 / 32.1 / 11.5	0 / 100 / 0	51.3 / 39.9 / 8.8	58.3 / 33.1 / 8.6

## ⑥今後の課題

・判定箇所から除外した表現について、新たに用例文を補充収集する。

・補充収集した表現、機能的用法・内容的用法の判定作業を行い、機械学習による用法判定のための訓練・評価データを整備する。

・上記をふまえて、提案方式の実装および評価を行う。