

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

Analysis of Japanese Functional Expressions in News Corpus based on a Hierarchical Lexicon

長坂泰治[†]、坂本明子[†]、宇津呂武仁[†]、
森下洋平[†]、松吉俊[‡]、土屋雅稔[§]

筑波大学大学院システム情報工学研究科[†]

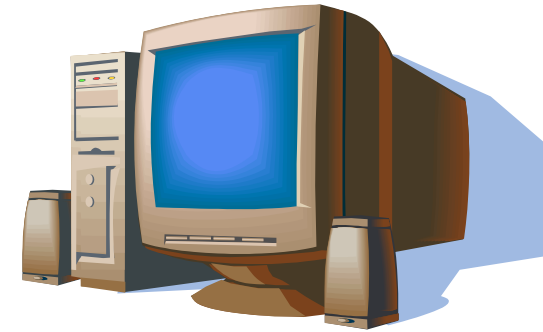
奈良先端科学技術大学院大学 情報科学研究科[‡]

豊橋技術科学大学 情報メディア基盤センター[§]

2009年9月30日(水)-10月1日(木)

NLP若手の会2009@京大

「つつじ」を用いた機能表現の検出



「つつじ」[07 松吉]

16,801表現

(1)私は自分の夢**について**話した。
機能的用法

(2)私は彼**について**走った。
内容的用法

(1) I talked **about** my dream.

(2) I ran **following** him.

機能的用法:
複数の単語から構成され、
一つの機能語のように働く

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

新聞記事における機能表現の分布

1 / 3

機能的用法と内容的用法が
適度な割合で出現する

機械学習による、用法判定

2 / 3

機能的用法に偏って出現する

一意に機能表現と決定できる

お神酒の瓶が女性にあたり, ……

内容的用法

首相は年頭にあたり首相官邸で、…

機能的用法

内容的: 60%

機能的: 40%

私は、彼の車について走った。

内容的用法

私は、自分の夢について話した。

機能的用法

内容的: 4%

機能的: 96%

検出の流れ

入力

わざわざ遠出しなくてもよい。

そんなことしなくていいです。

この辺りは騒音がなくていい環境です。

この国は治安が悪くなくていいです。



「つつじ」を利用して、
代表的表現へ言い換え

わざわざ遠出しなくてよい。

そんなことしなくてよい。

この辺りは騒音がなくてよい環境です。

この国は治安が悪くなくてよい。

検出の流れ

入力

わざわざ遠出しなくてもよい。
そんなことしなくたっていいです。
この辺りは騒音がなくていい環境です。
この国は治安が悪くなくていいです。

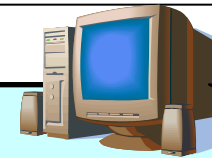


「つつじ」を利用して、
代表的表現へ言い換え

わざわざ遠出しなくてよい。
そんなことしなくてよい。
この辺りは騒音がなくてよい環境です。
この国は治安が悪くなくてよい。

出力

わざわざ遠出しなくてもよい。
そんなことしなくたっていいです。
この辺りは騒音がなくていい環境です。
この国は治安が悪くなくていいです。



機能表現検出器
によって用法判定

わざわざ遠出しなくてよい。
そんなことしなくてよい。
この辺りは騒音がなくてよい環境です。
この国は治安が悪くなくてよい。