

機能表現のレキシカルプロファイリングの手法と 検索インターフェースの開発

プラシャント・パルデシ 理論・対照研究領域

国立国語研究所オープンハウス2018 平成30年12月22日(土)

機能表現とは

語は、文を構成する基本単位であり、実質的な内容を表す内容語と統語的關係を表す機能語に大別することができます。日本語では、動詞、形容詞、名詞、副詞などが内容語に相当し、助詞、助動詞が機能語に相当します。しかし、日本語にはこれ以外に、「～なければならない」、「～たことがある」など、複数の語が組み合わさって機能語としての役割を果たす複合辞が数多く存在します。機能表現とは、統語的關係を表す機能語と複合辞を合わせた総称です。

NLB とは

NLB は、国語研の現代日本語書き言葉均衡コーパス (BCCWJ) を検索するためのツールです。レキシカルプロファイリングという手法に基づいて、内容語の振る舞いを調べることができます。例えば、「お茶」という名詞を調べると、「お茶を」の後にはどのような動詞が来るか、つまり、「お茶」を目的語とする動詞にはどのようなものがあるかを調べることができます。

右図は、NLB で「お茶を」に続く動詞を調べている画面です。「お茶を飲む」が最も多く、そのほかに「お茶を入れる」、「お茶を濁す」、「お茶をする」などの表現が多いことが分かります。このように、レキシカルプロファイリングの手法を用いると、語と語の結びつき（これをコロケーションといいます）を簡単に調べることができます。

コロケーション	頻度	MI	LD
お茶を飲む	320	10.63	8.97
お茶を入れる	74	16.31	10.97
お茶を濁す	49	14.51	10.25
お茶を出す	43	22.00	10.30
お茶を出す	42	6.94	5.31
お茶をいれる	35	10.40	8.30
お茶をいただく	31	8.37	6.63
お茶をくれる	29	10.10	8.01
お茶をする	25	11.89	8.86
お茶を飲む	22	7.98	6.23
お茶をくる	20	11.94	8.67
お茶を濁す	19	8.29	6.46

NLB の機能表現プロファイリング

NLB では、これまで内容語の振る舞いが分かるプロファイリングを提供してきましたが、機能表現の振る舞いを明らかにするためのプロファイリングの手法や検索インターフェースの開発も進めています。

機能表現のプロファイリングがどのようなものなのか、実例を挙げてみましょう。「そのように感じる、振る舞う」という意味を表す機能表現に「～がる」があります。「知りたがる」のように動詞が続く場合、「欲しがる」のようにイ形容詞が続く場合、「いやがる」のようにナ形容詞が続く場合がありますが、まず最初にこれらの後続する品詞から一つを選びます。ここでは、「動詞+がる」を選んでみましょう。すると、中央のパネルに、「動詞+がる」に続く機能語も含めたパターンが表示されます。さらに、中央のパネルの「～したがる」をクリックすると、右のパネルにコーパスに現れる実例を確認することができます。

このツールでは、現在、初級レベルに相当する機能表現約 80 種類のプロファイリングを検索することができます。ただし、暫定版につき、検索できるのは、BCCWJ コーパスの図書館・書籍サブコーパス（約 2933 万語）に限定されています。

謝辞：この検索ツールは、① 科学研究費挑戦的萌芽研究「大規模コーパスに基づく日本語機能語の基礎研究と機能語検索ツールへの応用（課題番号：16K13228）」、② 科学研究基盤研究(B)「統語・意味解析情報タグ付きコーパス開発用アノテーション研究：複文を中心に（課題番号 15H03210）」、③ 国立国語研究所機能拠点型共同研究プロジェクト「統語・意味解析コーパスの開発と言語研究」の研究助成を受けて開発したものです。

詳しくは <http://nlb.ninjal.ac.jp> にアクセス