

## 共起関係に着目した動詞分析について

3B-8

荻野孝野、前葉玉緒、山本由紀雄、清野正樹、綿田路子  
 (株)日本電子化辞書研究所

## 1. はじめに

本研究の主たる目的は、文内における動詞と他の単語との共起関係に着目して、動詞の概念分類のためのパターンを抽出しようとするものである。著者らは、具体的に単語を三つの要素(関係子、意味情報、格表示)で記述した後、記述された種々の組み合わせでソートし、様々な角度から検討している。今回は、この研究の記述状況、ソート内容について報告する。

この研究は、EDR概念辞書開発における概念分類項目の設定や概念体系構築のために観念的な方法ではなくデータ依存型で検討及び検証を行っているものである。

## 2. 研究概要

2.1 分析対象データ : 新聞データ(文献1)

分析対象語数 : 約4000語

## 2.2 分析手順

## 2.2.1 異なり語別記述

これは、分析対象語彙を、(関係子、意味情報、格表示)の三つの要素で記述するものである。同じ表記で1語として扱われていても、語義が異なると、(関係子、意味情報、格表示)の組み合わせ情報が異なる語は別語として扱っている。関係子は、EDR設定のもの(文献3)、意味情報は、荻野の分類(文献4)、格表示は、とりうる格助詞を記載した。記述にあたっては、既存辞書および文例データ(文献1,2)などを参考にした。

[記述形式] : 識別番号/関係子/意味情報/格表示/関係子/意味情報/格表示/.../語例/\*

[記述例] : 00110/agent/1.1/ガ/孤立化する/\*

00120/agent/1.1/ガ/agent/1.1/ト/けんかする/\*

## 2.2.2 ソート

2.2.1の作業によって作成したデータを、要素の組み合わせでソートする。

## ① 意味情報のソート

[目的] : 共起する単語の意味情報の使用頻度を把握し、共起関係の記述に役立つ概念分類項目設定のための資料とする。

[ソートキー] : 意味情報でソートし、意味情報別に、記述の際その意味情報を使用した単語の数を出す。

[出力] :

意味情報	語数
1.421(集団)	844
1.11(人間)	570
.	.

## ② (格表示、意味情報)によるソート

[目的] : 結合価パターンを抽出する。これは、文生成、仮名漢字変換(同音異表記の判定)などへの利用を想定する。

[ソートキー] : 3 1 4 2 (ソート順)  
 識別番号/関係子/意味情報/格表示/関係子/意味情報/格表示/...

[出力] : \*\*\*\*1\*\*\*\*

ガ 1.1(有意志体)

孤立化する、横行する、無理する、油断する

語数=4

\*\*\*\*2\*\*\*\*

ガ 1.1(有意志体), 1.2(具象物)

倒れる

語数=1

Analisis of Japanese Verbs from the point of Cooccurrence

Takano OGINO, Tamao MAEBA, Yukio YAMAMOTO, Masaki KIYONO, Michiko NAWATA

Japan Electronic Dictionary Research Institute, Ltd.

## ③ (関係子、意味情報)によるソート

[目的]: 概念パターンの抽出を行う。これは、概念分類項目設定の参考とするもので、本研究の主たる目的にかかわる部分である。

[ソートキー]: 1 3 2 4 (ソート順)

識別番号/関係子/意味情報/格表示/関係子/意味情報/格表示/……

[出力]: \*\*\*\*1\*\*\*\*

agent 1.1(有意志体)

孤立化する、横行する、無理する、油断する

語数=4

\*\*\*\*2\*\*\*\*

agent 1.1(有意志体), agent 1.1(有意志体)

けんかする、ぶつかる

語数=2

## ④ 関係子によるソート

[目的]: 深層格パターンの抽出を行う。

[ソートキー]: 1 2 (ソート順)

識別番号/関係子/意味情報/格表示/関係子/意味情報/格表示/……

[出力]: \*\*\*\*1\*\*\*\*

agent

孤立化する、無理する、油断する、横行する、大騒ぎする、騒ぐ、反乱する、働く、作業する、待機する、沈黙する、合宿する、かけひきする、振る舞う、奮起する、活躍する、……

語数=163

## ⑤ (関係子、意味情報、格表示)によるソート

[目的]: 単語に依存しないで「意味情報+格助詞(表層情報)→関係子」を規則化できる部分があるかどうかの検討を行う。

[ソートキー]: 異なり語別の記述を(関係子、意味情報、格表示)の単位に分割したものをソート対象とする。

3 2 1 (ソート順)

識別番号/関係子/意味情報/格表示/語例/

[出力]: \*\*\*\*1\*\*\*\*

ガ 1.1(有意志体) agent

孤立化する、無理する、けんかする、追い上げる、追い払う、動かす、支配する……

語数=15

\*\*\*\*2\*\*\*\*

ガ 1.11(人間) agent

大騒ぎする、騒ぐ、作業する、待機する、振る舞う、活躍する……

語数=457

## 3. 最後に

異なり語別の記述の際、記述者によって、同表記単語の語義の境界のとりかた、記述の際の意味情報のとりかたなど、ばらつきが出た部分がある。今後はこれらの調整を行った上で、ソート結果を検討し、動詞的概念のパターン化や、格表示・意味情報・関係子の関係の形式化を行う予定である。

## [参考文献]

- (1) 田中穂積、荻野孝野、荻野綱男:「新編 日本語品詞列集成」電総研(1979)
- (2) 石崎俊、荻野孝野:「用言に関する結合価の自動抽出資料」電総研推論システム研究室(1985)
- (3) 内田裕士:「電子化辞書の開発」自然言語処理技術シンポジウム、情報処理学会(1988)
- (4) 荻野孝野:「日本語の意味分類体系」、計量国語学16巻3号(1987)