

統合辞書における文型の記述

7L-1

橋本三奈子^{†1} 岡部了也^{†2} 桑畑和佳子^{†1} 齊藤初江^{†2} 佐藤幸子^{†2} 中川建司^{†2}
 山下智弥^{†2} 青山文啓^{*3}

[†]情報処理振興事業協会(IPA)
 技術センター

[†]上智大学 * 桜美林大学

1 はじめに

IPA技術センターでは、動詞、形容詞、名詞の順に辞書を作成してきた経験によって、動詞や形容詞が形成する構文を記述することが、構文に現れるあらゆる品詞についての辞書記述の基本になる、という知見を得た。現在は、「単文」を基本として、単文についての情報を蓄積していき、そこからその単文に現れるさまざまな単語についての詳細情報を参照できるような、辞書の枠組みを作成中である。このような辞書の記述作業は、見出し語数を拡張するためにも効率良く行なう必要がある。そのため、あらかじめ決められた範囲の中から文型を選択できるようにしておくことが望ましい。本稿では、品詞ごとに異なっていた文型の記載方法の統一と作成した文型一覧について述べる。

2 単文を中心とした枠組み

例えば、「雨」という名詞は、名詞辞書の見出し語「雨」として、「降る」「あがる」「激しい」などの動詞や形容詞と共に記載される一方、動詞辞書の「降る」「あがる」などの見出し語にも、その動詞と共に記載され、形容詞辞書の「激しい」などの見出し語にも、その形容詞と共に記載されている。これまでに作成した辞書は、見出し語を区分し、区分ごとに、構文や共起する単語や関連語などを記載してある。したがって、見出し語については区分番号が明確だが、記載された構文に現れる別の単語などについては区分番号が記載されていない。つまり、名詞辞書に「雨があがる」という構文が記載されるとき、「雨」の区分は一つに同定されている。けれど

も、その構文に現れる「あがる」の方は、動詞辞書で「階段をあがる」「熱があがる」などのように、複数に区分されているうちのどの「あがる」なのかという情報が欠けている。そのため、三つの辞書を統合する際には、区分まで限定してリンクを張るなどの特別な処理を加える必要があった[1]。

このことから、見出し語ごとに構文を記述するのではなく、動詞や形容詞が形成する構文を一箇所で記述・管理することが、構文に現れるあらゆる品詞についての辞書記述の基本になり、それによって、辞書の記述効率を上げることができる、という知見を得た。

IPA L統合辞書では、述語の範囲を動詞や形容詞だけでなく、名詞+スル、名詞+ダまで広げ、文の中にこれらの述語が一つだけ現れる「単文」を基本として、単文についての情報を蓄積していき、そこからその単文に現れるさまざまな単語についての詳細情報を参照できるような、辞書の枠組みを作成することにした。

3 文型一覧の作成

単文を中心とした記載で重要なのは、文型の記載方法である。ところが、記述者によって、どのような文型を記述するかということに揺れが生じる。

例えば、形容詞辞書（136語：フロピイ版）では、文型として19種類が記載されている。名詞辞書（1081語：FTP版）の述語用法欄では、11種類の文型が記載されていて、そのうち7種類は形容詞と共通している。担当者の校閲が済んでいない名詞2000語では、この11種類の他に、さらに57種類の文型が記載されていたが、これらを分析して、記載方法の誤りや語順などを整理した結果、2種類の文型を新たに認めれば良いだけで、残りはすべて先の11種類に集約できるということがわかった。

辞書の作成現場では、このような一次記述の内容を統一する校閲作業に工数がかかっている。辞書の語彙数が増え、執筆者が増えれば増えるほど、このような校閲作業が増大することになるため、記載方法の誤りや記述者による記述の揺れを防ぐことが急務であ

Description of sentence patterns in the integrated dictionary

[†]Hashimoto Minako (e-mail hasimoto@stc.ipa.go.jp)

Software Technology Center,

Information-technology Promotion Agency, Japan

3-1-38 Shiba-koen Minato-ku TOKYO 105, JAPAN

¹富士通株式会社より出向中

²情報処理振興事業協会技術センター臨時WG委員

³情報処理振興事業協会技術センターWG委員主査

番号	文型	形容詞文例 単語例	数	名詞文例 単語例	数
1	ガ	ダイヤモンドは硬い 悲しい, 辛い, 軽い…	428	風が穏やかだ 明らか, 厚め…	131
2	ガガ	彼はテニスがうまい あやしい, 大きい, 良い, 正しい…	56	彼は行動が勝手だ 主, 肝心, 最後, 最初, 花	6
3	ガニ	その先生は女生徒に甘い 良い, 等しい, 遠い…	40	彼の判断はみんなに公平だ 延期, 固有, 毒, 秘密	6
4	ガニ/カラ	私の家は彼の家に/から近い 遠い	2	その劇は主婦層に/から好評だ 不評	0
5	ニ/ガガ	日本には火山が多い 薄い, 浅い, 大きい…	43	日本語には雨に関する単語が豊富だ 豊か, 貧弱	0
6	ニガ	贈り物には手作りの品が望ましい 好ましい, 少ない, 多い…	9	ご近所づきあいには挨拶が肝心だ	1
7	(ガ)	(私は)飲み過ぎでつらい 眠い, 恥ずかしい, 暑い…	13	彼女に任せておけば(私は)安心だ	1
8	(ガ)ガ	(私は)規則が煩わしい 暑い, 痛い, 惜しい…	40	(私は)彼が哀れだ 嫌, かわいそう, 感心, 気の毒…	9

表 1: 形容詞述語文・名詞述語文における文型一覧の一部 (佐藤幸子が作成)

る。この問題を解決するため、記述者が文型欄の記述の際に、そこから選択するだけで済むような文型一覧の作成を行なうことにした。

一覧の作成にあたっては、動詞・形容詞・名詞の辞書ごとに、(1)「()」の使い方、(2)取り上げた格助詞の種類、(3)二重主格構文を取り上げる基準について違いがあるため、これらを統一する必要がある。

例えば、動詞辞書では随意格と思われるものに「()」をつけて表示していたが、形容詞・名詞辞書では、必須格と随意格との区別を廃止し、人称制限が認められる格に「()」をつけている。統合辞書でも、後者を採用することにした。

また、動詞・名詞辞書では、「ガ, ヲ, ニ, カラ, ヘ, ヨリ, デ, ト, ϕ」を格助詞として取り上げたが、形容詞辞書では、比較級・最上級を取り上げないなどの方針をたてたため、「ヨリ, デ, ϕ, ト, ヘ」を扱わなかった。形容詞においては、「彼は 彼女より 英語が うまい」のような比較級の構文は当然成り立ちうるものなので、語彙的な特徴を記載すべき辞書に収録する必要はない。しかし、「彼の英語の実力は 彼女より まさっている」「彼の英語の実力は 彼女より 上だ」に現われる、動詞の「まさる」、名詞の「上」などでは、このような構文が可能であることを語彙的な特徴として記載しておく必要がある。品詞ごとの辞書と異なり、統合辞書においては、動詞、形

容詞、名詞+スル、名詞+ダを述語とした単文を等しく扱えるようにするべきなのでこのような場合には、作業量を増やさないようにする工夫が必要である。

さらに、二重主格構文については、名詞辞書に準じて、二つのガ格名詞句を「ノ」で結ぶことが不可能なものだけを扱うことにした。

このような観点で見直すと、形容詞・名詞の述語文では、約30種類の文型に集約される。表1に示すのは、その文型一覧の一部である(名詞辞書FTP版で数が0のものは、見出し語以外の例を入れた)。

表1の文型一覧は、「ある」「まさる」「劣る」などの状態動詞にも拡張可能であると考えられる。今後は状態動詞、動作動詞などの定義づけを行なって文型一覧を拡張する計画である。

4 おわりに

自然言語処理システムを開発する際には、文型や名詞の意味素性などを記載した辞書が重要であり、このような辞書の記述作業をできるだけ効率化する必要がある。事例データベースから、係受け情報を抽出することによって、辞書の文型情報欄を効率良く記載できるようにするしくみについても、現在検討中である。

参考文献

- [1] 梁慶昇. 「IPAL辞書の自動的ハイパーテキスト化」. 北陸先端科学技術大学院大学修士論文, 1996.