

機能・意味・形態にもとづく助詞型機能表現の分類

松 吉 俊 佐 藤 理 史 宇 津 呂 武 仁

京都大学大学院 情報学研究所

matuyosi@pine.kuee.kyoto-u.ac.jp, {sato, utsuro}@i.kyoto-u.ac.jp

1. はじめに

日本語には、例えば、「に関して」、「からには」のように、2形態素以上から構成され、全体として1つの機能語のように働く機能表現が存在する。かなりの数の機能表現に対して、それと同一表記をとる内容表現が存在する。例えば、「をめぐり」という表現は、「貿易をめぐり対立する」という文では「について」に相当する機能表現であるが、「四国をめぐり旅をする」という文では内容表現である。このため、このような機能表現は、内容表現と区別して認識する必要がある。

しかしながら、現在の自然言語処理においては、機能表現の認識は不十分であり、その誤認識が致命的な誤りをもたらす場合も多い。以下に、機能表現「からして」を含む文を市販の2つの異なる機械翻訳システムで翻訳した結果を示す。

- (1) 彼の態度 からして 彼は来るつもりだ。
(1A) Judging from his attitude, he is going to come.
(1B) It will have it from his attitude, and he will come.

この文の「からして」は、「から判断して」の意味で使われている。翻訳出力(1A)から推察すると、機械翻訳システムAはこれを正しく認識していると考えられるのに対し、機械翻訳システムBは、「からして」の認識に失敗し、これが明らかな誤訳(1B)の出力の原因となっていると考えられる。

機能表現を正しく認識するためには、次の2つのことが必要である。

- (1) 認識すべき機能表現の網羅的リストを作成すること。
(2) それぞれの機能表現の表記に対して、機能表現として認識すべき場合と、そうでない場合を区別する能力をもった、機能表現の検出システムを実現すること。

本論文では、(1)の網羅的リスト作成の基礎となる、機能表現の分類体系を提案し、この分類体系に基づいて作成した助詞型機能表現の辞書について報告する。まず、第2章で、機能表現を定義し、機能・意味・形態にもとづく日本語機能表現の分類を示す。次に、第3章において、前章の分類に従って作成した、助詞型機能表現辞書について述べる。最後に、第4章でまとめを述べる。

2. 日本語機能表現の分類

2.1 機能にもとづく分類

機能表現とは、おおよそ、文中においてなんらかの機能を持つ表現ということができよう。しかしながら、機能表現全体に対して、「なんらかの機能」をより精密に書き下すことはほとんど不可能であろう。

そこで、我々は、機能表現の定義として、次のような定義を採用する。

表1のいずれかの機能を持つ語・表現を、それぞれの機能型に属する機能表現と呼び、その総称として機能表現という用語を用いる。

この表に示す機能は、森田・松木¹⁾による機能分類に、接続詞相当の機能を加えたものである。なお、助動詞型機能表現と接続詞型機能表現を除く残り6つの総称として、助詞型機能表現という用語を用いる。

なお、上記のように機能表現を定義することは、同時に、機能表現の「機能にもとづいた分類」を与えていることにもなる。

本研究では、森田・松木¹⁾でとり上げられている、271個の助詞相当の複合辞を独自に整理したものに、いくつかの助詞を追加した、326個の助詞型機能表現(以下、対象の326表現)を研究対象とする。なお、この数は、次節で述べる同表記異義語を区別して数えている。それぞれの機能型に属する機能表現の数は、表1に示すとおりである。

2.2 意味にもとづく分類

機能表現の中には、同じような意味を持つものがある。これらをグループ化しておくことは、言い換えなどの応用を想定した場合に有用である。そこで、機能にもとづく分類の下位分類として、意味にもとづく分類(意味分類)を導入する。

意味分類表は、以下の手順に従って作成した。

まず、森田・松木¹⁾の意味・機能分類にならって、各々の機能の下に意味カテゴリを設定し、対象の326表現を、いずれかの意味カテゴリに分類した。このとき、表記は同じであるが、意味が異なるものを、数字を添えることにより区別した。例えば、「をもって」は、<状態>、<仲介>、<基準>という異なる意味を持っているので、それぞれ「をもって₁」、「をもって₂」、「をもって₃」として区別し、異なる機能表現として扱う。

表 1 機能にもとづく機能表現の分類

機能	名称	例	表現数
前件の体言を後件の用言に係属付ける	格助詞型機能表現	に対して、によって、に至るまで	57
前件の体言を後件の体言に係属付ける	連体助詞型機能表現	に関しての、といった、における	42
前件を話題化する	係助詞型機能表現	といったら、としては、としても	34
前件に付加的なニュアンスを与える	副助詞型機能表現	に限らず、ならでは、をよそに	27
用言で終わる節を後件の用言に係属付ける	接続助詞型機能表現	からには、おかげで、かと思ったら	123
文末に付加的なニュアンスを与える	終助詞型機能表現	ものだ、ことだ、ないかい	43
前件の述語に付加的なニュアンスを与える	助動詞型機能表現	なければならない、てもいい、かもしれない	-
前の文を後ろの文に係属付ける	接続詞型機能表現	と思いきや、にしても、とはいいいながら	-

表 2 意味にもとづく分類の概要

	意味カテゴリ数と例	A1	A2	B	C	F	計
格助詞型	12 「対象」 ₁ 、「根拠」 ₁ 、「状況」 ₁	3(5%)	8(14%)	27(47%)	14(25%)	5(9%)	57(100%)
連体助詞型	11 「対象」 ₁ 、「根拠」 ₁ 、「状況」 ₁	4(9%)	10(24%)	18(43%)	8(19%)	2(5%)	42(100%)
係助詞型	7 「定義」 ₁ 、「話題」 ₁ 、「は観点」 ₁	0(0%)	4(12%)	16(47%)	10(29%)	4(12%)	34(100%)
副助詞型	7 「強調」 ₁ 、「非限定」 ₁ 、「添加」 ₁	0(0%)	1(4%)	14(52%)	11(40%)	1(4%)	27(100%)
接続助詞型	18 「時-継起」 ₁ 、「逆-確定」 ₁	17(14%)	14(11%)	60(49%)	22(18%)	10(8%)	123(100%)
終助詞型	11 「感嘆」 ₁ 、「願望」 ₁ 、「伝聞」 ₁	2(5%)	6(14%)	16(37%)	13(30%)	6(14%)	43(100%)
計	66	26(8%)	43(13%)	151(46%)	78(24%)	28(9%)	326(100%)

表 3 接続助詞型機能表現の意味分類 (一部)

意味	A1	A2	B	C	F
時-継起	てから あと(で/に) ところ		うえ(で) っきり きり すえ(に) あげく(に)		ところが
逆-確定	が けれど(も) けど	と(は)いって(も) ₁	とはいえ とはいうもの のもの としても ₂ にしても ₂ にしたって ₂	にしろ ₂ からといって ところで ₂ にせよ ₂	とはいいいながら ど(も)

表 4 それぞれの機能型における、表現の中心になっている品詞の割合

	名詞	用言	辞	小計
格助詞型	3(5%)	51(90%)	3(5%)	57(100%)
連体助詞型	1(2%)	37(88%)	4(10%)	42(100%)
係助詞型	0(0%)	30(88%)	4(12%)	34(100%)
副助詞型	7(26%)	10(37%)	10(37%)	27(100%)
接続助詞型	37(30%)	37(30%)	49(40%)	123(100%)
終助詞型	11(26%)	3(7%)	29(67%)	43(100%)
計	59(18%)	168(52%)	99(30%)	326(100%)

次に、各々の機能表現に難易度を設定し、意味カテゴリ内で難易度順に並べた。難易度は、JCore²⁾の難易度に準拠し、やさしい方からA1、A2、B、C、Fの5段階を設定した。この難易度の設定においては、「日本語能力試験出題基準 [改訂版]」³⁾の文法の項を参考にした。

こうして作成した意味分類表の概要を表2に示す。また、接続助詞型機能表現の意味分類表の一部を表3に示す。

2.3 形態にもとづく分類

機能表現がどのような形態素列から構成されているかという情報は、機能表現を検出するときに役立つ情報である。例えば、<状況>という意味を表す格助詞型機能表現「に当たって」は、動詞「当たる」を中心にして構成されたものである。動詞「当たる」は、「天井に当たって」のように内容語として用いられることがあるので、機能表現「に当たって」の検出は難しい。一方、接続助詞型機能表現「からには」は、助詞「から」を中心にして

構成されており、構成要素が内容語である可能性を持たないため、その検出は容易である。このような背景から、機能にもとづく分類のもう一つの低位分類として、形態にもとづく分類(形態分類)を導入する。

形態分類表は、以下の手順に従って作成した。

まず、対象の326表現を、その構成要素に基づいて、次のいずれかに分類した。

名詞中心 名詞が表現の中心になっているもの
用言中心 動詞・形容詞が表現の中心になっているもの
辞中心 助詞・助動詞・判定詞が表現の中心になっているもの

次に、各々の機能型に対して、上記の分類が同じものを集めて、3つの形態分類表を作成した。それぞれの機能型における、表現の中心になっている品詞の割合を表4に示す。また、形態分類表の一例として、格助詞型機能表現の形態分類表の一部を表5に示す。

表 5 格助詞型機能表現の形態分類 -用言中心- (一部)

構成要素			機能表現
と	する	テ	として ₁
に	する	テ	にして ₁
	とる	テ	にとつて
	至る	まで	に至るまで
から	いう	と	からいうと
		たら	からいったら
		ば	からいえば
		テ	からいって

形態分類表は、左側に機能表現を構成する要素を、右側にその構成要素からなる機能表現を示している。構成要素の欄において、表現の中心になっている形態素を太字で示している。また、比較しやすくするために、適宜、用言を基本形で表示し、その活用語尾を一つの要素として示している。

3. 助詞型機能表現辞書

前章で述べた分類表に基づき、助詞型機能表現の辞書を作成した。この辞書には、(1) 機能表現を自動的に検出する処理と、(2) 機能表現を自動的に言い換える処理、の2つの機械処理の実現を考慮し、それぞれの機能表現に対して次の3種類の情報を記述した。

- 見出し情報
- 検出のための情報
- 言い換えのための情報

以下では、これらの情報について説明する。

3.1 見出し情報

機能表現には異形が多いため、機能表現辞書において何を見出しとするかということ自身が問題となる。異形に関わる主な問題として、以下のものがある。

- 「に対し₁」、「対して₁」など、動詞の連用形とテ形
- 「あげく₁」、「あげく_に」など、末尾の助詞の有無
- 「となれば₁」、「と_もなれば₁」など、表現内への語の挿入
- 「に_当たり₁」、「に_あたり₁」など、漢字表記

今回作成した辞書では、これらの差異を表記のゆれと見なし、一つの見出しにまとめることにした。具体的には、佐藤²⁾によって提案されている、表記と読みの対からなる、「表示」を取り入れ、どの範囲の表示が、一つの見出し(機能表現)の異表記であるかということを示した。

辞書の見出し情報は、ID、見出し形、機能、表示、例文の5項目から構成される。なお、見出し形は、単なる便宜上のものであり、表記に関する情報を与えるものではない。一例として、「に対し(て)₁」の見出し情報を表6に示す。

3.2 検出のための情報

文中に存在する機能表現を検出するためには、その機能表現を構成する形態素列、およびその表現の周囲に出現しうる語の情報が必要である。この情報は、次の5つの項目から構成される。

表 6 「に対し(て)₁」の見出し情報

ID	JC1:jA210017
見出し形	に対し(て)
機能	格助詞型
表示	〈 に対し, にたいし 〉, 〈 にたいし, にたいし 〉, 〈 に対して, にたいして 〉, 〈 にたいして, にたいして 〉
例文	私の質問に対して何も答えてくれなかった。

表 7 「によると₁」の検出のための情報

形態素列	(に 格助詞)(よる 動詞 基本形)(と 接続助詞)
左接続	(に 格助詞)
右接続	(と 接続助詞)
判定指標	c2
判定条件	係り先可 (伝聞表現 推測表現)

- 形態素列
- 左接続
- 右接続
- 判定指標
- 判定条件

一例として、「によると₁」の検出のための情報を表7に示す。

以下では、これらのうち、「形態素列」以外の項目について説明する。

3.2.1 左接続・右接続

これらの項目には、機能表現の左(もしくは右)にどのような語が接続可能かを記述する。

左接続の記述法は、2種類ある。一つ目の方法は、形態素を記述する方法である。この場合は、機能表現の左接続は、指定された形態素の左接続と同一であることを表す。例えば、「によると₁」の左接続の欄には、表7に示すように「に 格助詞」と書かれており、これは、「によると₁」の左接続は、格助詞「に」の左接続と同じであることを示している。

もう一つの方法は、左接続に、佐藤^{4),5)}で提案されている境界を指定するというものである。例えば、接続助詞型機能表現「にしても₁」の左接続の欄には、9055と書かれており、これは「にしても₁」の左に境界 9055(動詞の基本形もしくは動詞のタ形の右に来ることができる境界)が来ることを示している。

右接続の記述法も、上記の左接続の場合と同様である。

3.2.2 判定指標と判定条件

機能表現の検出は、それぞれの機能表現において、検出に必要な情報と難易度が異なると考えられる。そこで、本辞書では、それぞれの機能表現に対して、その機能表現の検出を文解析処理のどの時期に行なうべきかを「判定指標」として記述することとした。設定した5種類の判定指標を表8に示す。

判定指標 0, c0 は、形態素解析中に検出することが適切であると考えられるものに付与した。「からには」のように、「その中に内容的に使用される要素を含まない」ものには 0 を、「に関して」のように、「その中に内容的に

表 8 判定指標

判定指標	適切な判定時期
0, c0	形態素解析中に判定
c1	形態素解析後、局所的な文脈を見て判定
c2, c2-	係り受け解析後、広い文脈を見て判定

使用され得る要素を含んでいるが、この形態素列においては全体として機能的な意味でのみ使用される」ものには c0 を付与した。

判定指標 c1 は、局所的な文脈からその候補となる表現が機能表現であるかどうかを判定可能であると考えられるものに付与した。さらに、このとき使用すべき判定条件を辞書に記述した。判定条件としては、以下の 4 種類の集合を記述している。

- 前集合: 直前に来ることができる語の集合
- 非前集合: 直前に来ることができない語の集合
- 後集合: 直後に来ることができる語の集合
- 非後集合: 直後に来ることができない語の集合

なお、この判定条件を用いた機能表現の検出は、別稿⁶⁾で述べる。

判定指標 c2, c2- は、機能表現であるかどうかの判定により広い文脈を考慮する必要があるものに付与した。判定ができずに曖昧性が残った場合に機能表現であるとみなすべきものには c2 を、機能表現ではないとみなすものに c2- を付与した。さらに、このとき利用できると思われる判定条件として、以下の 2 種類の集合を記述した。

- 係り先可: 機能表現が係りうる表現
- 係り先不可: 機能表現が係りえない表現

これらの情報は暫定的なものであり、今後、モグリティ副詞など、機能表現の判定に利用できそうな情報を検討していく予定である。

なお、これらの判定条件の記述にあたっては、「日本語文型辞典」⁷⁾を参考にした。

3.3 言い換えのための情報

言い換えのための情報として、次の情報を記述した。

- 意味分類表の位置と難易度
- 文体 – 常体、敬体、口語体

その見出しの機能表現の文体の項目に“1”を、他の文体の項目に、対応する表現を記述した。

- 派生
「に関して」が「に関する」、「に関しての」として用いられるように、派生して連体助詞型機能表現として用いることができる場合に、それらの派生形を記述した。
- 兼終助詞型
「くせに」のように、終助詞型として用いることができるものに“1”を記述した。
- 兼接続詞型
「とすると₂」のように、接続詞型として用いることができるものに“1”を記述した。

例として、「とすると₂」の言い換えのための情報を表 9

表 9 「とすると₂」の言い換えのための情報

分類表の位置	順-確定-3/9
難易度	A2
常体	1
敬体	としますと、といたしますと
口語体	としたら
派生	0
兼終助詞型	0
兼接続詞型	1

に示す。

4. おわりに

本研究では、助詞型機能表現を機能・意味・形態という観点から体系的に分類し、それにもとづいて機能表現辞書を作成した。この辞書には、機能表現の検出・言い換えを行なうに当たって必要となる情報を記述した。

今回対象とした機能表現は、助詞型機能表現 326 個である。今後は、文献やコーパスなどを用いて、さらなる機能表現を収集し、辞書を豊かにしていく必要がある。

本研究の一部は、次の研究費による；基盤研究(A)「円滑な情報伝達を支援する言語規格と言語変換技術」(課題番号 16200009)、21 世紀 COE プログラム「知識社会基盤構築のための情報学拠点形成」、京都大学-NTT コミュニケーション科学基礎研究所共同研究「グローバルコミュニケーションを支える言語処理技術」。

参考文献

- 1) 森田良行, 松木正恵: 日本語表現文型 用例中心・複合辞の意味と用法, アルク (1989).
- 2) 佐藤理史: 異表記同語認定のための辞書編纂, 情報処理学会研究報告 2004-NL-161, pp. 97-104 (2004).
- 3) 国際交流基金, 財団法人日本国際教育協会: 日本語能力試験出題基準【改訂版】, 凡人社 (2002).
- 4) 佐藤理史: 境界認定の提案: (1) コンセプトと実現法, 情報処理学会研究報告 2004-NL-164, pp. 25-32 (2004).
- 5) 佐藤理史: 境界認定の提案: (2) 背景と思想, 情報処理学会研究報告 2004-NL-164, pp. 33-40 (2004).
- 6) 松吉俊, 佐藤理史, 宇津呂武仁: 接続情報にもとづく助詞型機能表現の自動検出, 言語処理学会第 11 回年次大会 C5-3 (2005).
- 7) グループ・ジャマシイ編: 教師と学習者のための日本語文型辞典, くろしお出版 (1998).