

代表・派生関係を利用した日本語機能表現の解析

鈴木 敬文^{†1} 宇津呂 武仁^{†1}
松吉 俊^{†2} 土屋 雅稔^{†3}

本稿では、「機能表現一覧」の階層性を利用し、階層において下位に位置する派生的表現について、用法が類似するより上位の代表的表現の用例を参照して、用法判定を行う手法について述べる。特に、前後の形態素の品詞が代表・派生間において不変の場合には、代表的表現と派生的表現の間で用法の傾向に相関がある、という特徴を利用する方式を提案する。提案方式に基づいて、派生的表現の用法の分析を行った結果、代表的表現の表記の用法判定済み用例集合(約 38,000 例)を参照して、派生的表現の表記の用法判定を行うことにより、80%以上の用例の用法を正しく判定できることを示す。

Analysis of Japanese Functional Expressions using Canonical/Derivational Relation

TAKAFUMI SUZUKI^{†1} TAKEHITO UTSURO^{†1}
SUGURU MATSUYOSHI^{†2}
and MASATOSHI TSUCHIYA^{†3}

The Japanese language has various types of functional expressions. In order to organize Japanese functional expressions with various surface forms, a lexicon of Japanese functional expressions with hierarchical organization was compiled. This paper proposes how to design the framework of identifying more than 16,000 functional expressions in Japanese texts by utilizing hierarchical organization of the lexicon. In our framework, more than 16,000 functional expressions are roughly divided into canonical / derived functional expressions. Each derived functional expression is intended to be identified by referring to the most similar occurrence of its canonical expression. In our framework, contextual occurrence information of much smaller number of manually identified canonical expressions are expanded into the whole forms of derived expressions, to be utilized when identifying those derived expressions. We also empirically show that the proposed method can correctly identify more than 80% of the functional / content usages only with less than 38,000 training instances of manually identified canonical expressions.

1. はじめに

機能表現^{*1}とは、「にあたって」や「をめぐって」のように、2つ以上の語から構成され、全体として1つの機能的な意味をもつ表現である。一方、この機能表現に対して、それと同一表記をとり、内容的な意味をもつ表現が存在することがある。例えば、文(1)と文(2)には「にあたって」という表記の表現が共通して現れている。

(1) 出発する にあたって、荷物をチェックした。

(2) ボールは壁 にあたって、跳ね返った。

文(1)では、下線部はひとかたまりとなって、「機会が来たのに当面して」という機能的な意味で用いられている。それに対して、文(2)では、下線部に含まれている動詞「あたる」は、動詞「あたる」本来の内容的な意味で用いられている。このような表現においては、機能的な意味で用いられている場合と、内容的な意味で用いられている場合とを識別する必要がある。

我々はこれまでに、現代語複合辞用例集²⁾(以下、用例集)中の代表的複合辞一覧に基づいて、それらの派生形である337種類の機能表現を規定し、その用例データベース(日本語複合辞用例データベース¹⁵⁾、以下、用例データベース)を作成した。また、それらの用例データベースを訓練事例として、機械学習により機能表現の検出・係り受け解析を行う方式を提案した^{13),14)}。また、機能表現の異形の語構成パターンを網羅することにより、日本語機能表現一覧⁵⁾(以下、「機能表現一覧」^{*2)}を作成した。

ここで、文献13)、14)の機械学習による機能表現検出においては、一つの表現あたり50例程度の訓練用例に対して、人手で機能的・内容的等の用法判定を行う必要がある。しかし、「機能表現一覧」の全機能表現16,801種類に対して、それだけの規模の作業を行うことは容易ではない。そこで、文献10)では、「機能表現一覧」の階層性を利用し、階層におい

†1 筑波大学大学院 システム情報工学研究科

Graduate School of Systems and Information Engineering, University of Tsukuba

†2 奈良先端科学技術大学院大学 情報科学研究科

Graduate School of Information Science, Nara Institute of Science and Technology

†3 豊橋技術科学大学 情報メディア基盤センター

Information and Media Center, Toyohashi University of Technology

*1 機能表現は、複数形態素からなる複合辞と一つの形態素からなる機能語から構成されるが、本稿では、複合辞と同等の意味で機能表現という用語を用いる。

*2 <http://kotoba.nuee.nagoya-u.ac.jp/tsutsuji/>

て下位に位置する機能表現 (以下, 派生的表現) について, 用法が類似するより上位の表現 (以下, 代表的表現) の用例を参照して, 用法判定を行う方式を提案した. そして, 文献 10) の提案の実現に向けて, 代表的表現・派生的表現の間の用法の派生関係について分析を進めてきた^{7)-9), 16)}.

文献 8) では, 文献 10) の提案をふまえて, 「機能表現一覧」中の情報のうち, 特に文体の情報に注目し, 代表的表現および派生的表現の区別を整理した. さらに, 毎日新聞 1995 年分のテキストデータ中において, 「機能表現一覧」の機能表現の出現頻度調査を行い, 文献 13), 14) の機械学習による機能表現検出において必要となる訓練事例 (出現頻度 50 以上) が存在する機能表現の規模を推定した. また, 文献 7) では, 毎日新聞 1995 年分のテキストデータ中に 50 回以上出現する代表的表現の表記を対象として, 人手で機能的用法・内容的用法の判定作業を行った. さらに, 各機能表現表記に対して, 機能的用法・内容的用法の両方が適度な割合で混合して出現し, 機械学習によって機能表現検出を行う必要のある機能表現表記の割合を求めた結果について報告した.

文献 9) では, 機能的用法として偏って出現する代表的表現に対して, 毎日新聞 1995 年分のテキストデータ中に 50 回以上出現しない派生的表現の出現個所を対象として, 機能的用法・内容的用法の判定作業を行った. さらに, それらの派生的表現の各出現個所の前後の形態素の品詞の組み合わせ, および, 代表的表現の出現個所の前後の形態素の品詞の組み合わせについて, 代表・派生間の傾向の差異を分析した. 分析の結果, 前後の形態素の品詞が代表・派生間において不変の場合には, 派生的表現が機能的用法である割合が高く保たれることが分かった.

文献 16) では, 文献 9) において明らかになった特性として, 前後の形態素の品詞が代表・派生間において不変の場合には, 代表的表現と派生的表現の間の用法の傾向に相関がある, という点に注目して, 機能的用法・内容的用法の両方が適度な割合で混合して出現する代表的表現, およびその派生的表現を対象として, この特徴をより詳細に分析した. 具体的には, 機能的用法・内容的用法の両方が適度な割合で混合して出現する代表的表現に対する派生的表現の表記の出現個所を対象として, まず, 機能的用法・内容的用法の判定を人手で行った. 次に,

派生的表現の表記の各出現個所と同様の用法となる用例が, 対応する代表的表現の

表記の用例中に存在するか否か, あるいは, そのような用例が作例可能か否か, を人手で調査した. 実際に, 毎日新聞 1995 年分において, 用法判定済みの用例数が 10 例以上となる派生的表現 56 表現, 約 2,000 用例を対象とした調査を行ったところ, 代表的表現

の表記の用法, および, その前後の形態素の品詞が, 派生的表現の表記の前後の形態素の品詞と同一となる用例が, すでに用法判定済みの 50,000 用例中に存在する場合は, 約 69%強となった. また, 同様に, 代表的表現の表記の用法, および, その前後の形態素の品詞が, 派生的表現の表記の前後の形態素の品詞と同一となる用例が容易に作例可能である場合は, 約 19%となった. このように, 調査対象となった派生的表現の表記の用例のうちの約 88%強については, 文献 9) で示された特性を一般化した, 「前後の形態素の品詞が代表・派生間において不変の場合には, 代表的表現と派生的表現の間の用法の傾向に相関がある」という仮説に従うことが分かった. 一方, 残りの 12%弱については,

代表的表現から派生的表現へと表記上の派生が生じた結果, 偶然, 別の機能表現や内容的表現, あるいは, それらが複合した表現 (の一部) と同一の表記となってしまった. 一方, 代表的表現の表記においては, 派生的表現の表記と類似の現象は起こらない.

という用例であった.

以上の調査結果に基づき, 本稿では, 「機能表現一覧」の階層性を利用し, 階層において下位に位置する派生的表現について, 用法が類似するより上位の代表的表現の用例を参照して, 用法判定を行う具体的な手法について述べる. 特に, 文献 9), 16) において明らかになった特性として, 前後の形態素の品詞が代表・派生間において不変の場合には, 代表的表現と派生的表現の間の用法の傾向に相関がある, という特徴を利用する方式を提案する. さらに, 前後の形態素品詞に加え, 代表的表現と派生的表現の間に, 機能表現の表記を構成する形態素列の品詞パターンの中に派生関係があるという特性を利用する方式を提案する. 提案方式に基づいて, 派生的表現の用法の分析を行った結果, 代表的表現の表記の用法判定済み用例集合 (約 38,000 例) を参照して, 派生的表現の表記の用法判定を行うことにより, 80%以上の用例の用法を正しく判定できることが分かった.

2. 階層的機能表現辞書

「機能表現一覧」⁵⁾ は, 9 つの階層構造をなしており, 各階層は, 表 1 に示されるような観点によって分類されている. 同表に, 各階層における機能表現数が示されており, 図 1 に階層構造の一部をそれぞれ示す.

また, 機能表現の文体に着目し, 文体ごとの機能表現の振る舞いについて述べる. 文体とは, 「機能表現一覧」中の表現に付与されている情報であり, 常体, 堅い文体, 口語体, 敬体の 4 種類がある. 表 2 にそれぞれの文体における表現例を示す.

表 1 機能表現辞書の 9 つの階層

階層		分類数	表現数		
			合計 (L^9 表現数)	助動詞 型以外	助動詞型
L^1	見出し語	—	341 (488)	281	207
L^2	意味	45/128/199	435 (488)	281	207
L^3	派生 (格助詞型, 接続助詞型, 連体助詞型, 接続詞型, 助動詞型, 形式名詞型, とりたて詞型, 提題助詞型)	8	555	348	207
L^4	機能語の交替	—	774	492	282
L^5	音韻的变化	38	1,187	633	554
L^6	とりたて詞の挿入	18	1,810	659	1151
L^7	活用	—	6,870	659	6211
L^8	「です/ます」の有無	2	9,722	895	8827
L^9	表記のゆれ	—	16,801	1360	15411

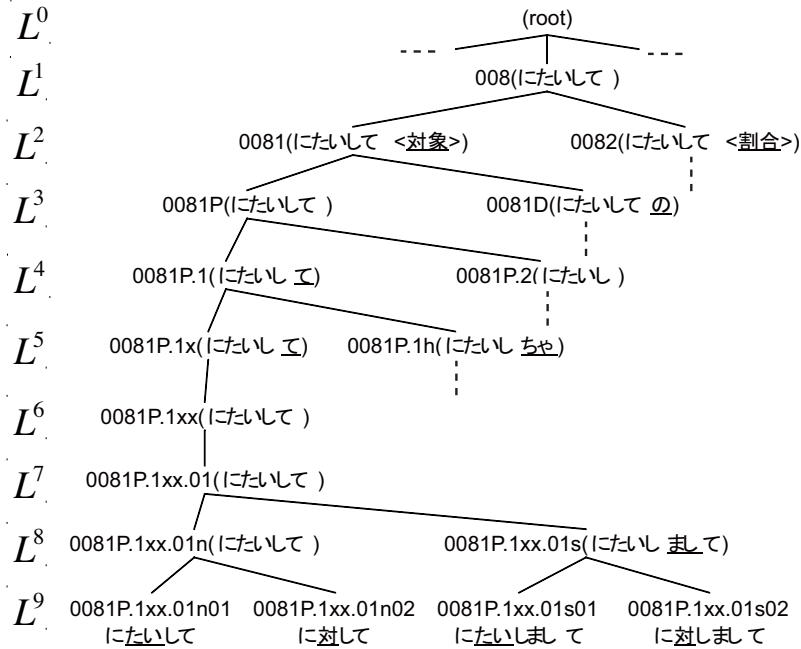


図 1 機能表現辞書階層構造の一部

表 2 文体の種類

文体	表現例
常体	について
堅い文体	につき
口語体	についちゃ
敬語体	につきまして

3. 派生関係を利用した日本語機能表現の解析

3.1 代表的な表現の選定

文献 10) で提案した代表的表現への集約方式においては、階層の上位に位置する代表的表現は、 L^4 階層相当の 1,000 表現程度の規模とする。そして、「機能表現一覧」において、代表的表現を除く表現をを派生的表現と定義する。ただし、代表的表現を選定する際には、以下の制約を課す。

- 機能表現の語頭の無声・有聲の制約により前接する活用語の活用型が制限される場合は、この制限を保持する。
- 機能表現の仮名表記・漢字表記の違いを保持する。
- 助動詞型の機能表現の場合には、活用形を保持する。

3.2 派生的な表現の解析方式

本節では、代表的表現の表記の用法判定済み用例集合 S_e^{tr} を参照して、派生的表現の表記の用法判定を行う方式について述べる。

用法判定の手順の概要を以下に示す。ただし、以下では、派生的表現の表記を F_d 、対応する代表的表現の表記を F_c とする。

- (1) F_d の用例に対して、 F_d の前後の形態素の情報 (主として、品詞情報と活用形)、及び、 F_d を構成する形態素列の情報 (主として、品詞情報と活用形) を抽出する。
- (2) F_c について、用法判定済みの用例集合 S_c^{tr} を対象として、(1) で得た情報との間で一定の基準以上の類似性を持つ用例を検索する。
 - (2a) (i) 一定の基準以上の類似性を持つ用例が存在する場合は、その用例を参照し、(ii) その用例に対して付与されている用法判定結果 l^{tr} (機能的用法もしくは内容的用法) を、(1) の F_d の用例に付与する。
 - (2b) 一定の基準以上の類似性を持つ用例が存在しない場合は、以下のいずれかの場合に該当する。したがって、(1) の F_d の用例には、この段階では、用法に関する情報は付与しない。
 - 他の機能表現の表記の用例として取り扱い、代表的表現の表記の用例を検索することにより、上記の (2a) によって判定が行える場合。
 - その他の場合は、いずれの機能表現の機能的用法にもならないと考えられるので、用法に関する情報は付与しないことにより、結果的に内容的用法と判定したことと同等となる。

上述の用法判定手順にしたがって、派生的表現の表記の用法が機能的用法であると判定した例を図 2 上半分に、内容的用法であると判定した例を図 2 下半分に、それぞれ示す。

上半分の例においては、派生的表現の表記の前後の形態素の品詞は、いずれも、「動詞」および「記号」となる。また、派生的表現の表記を構成する形態素列の品詞情報は、派生のタイプによって差異があるが、それらを包含する品詞列パターンは「助詞-(助詞)-動詞-(助動詞)」と表現することができる。そして、これらの情報と十分に類似する前後文脈、および、機能表現表記を構成する形態素列を持つ代表的表現の表記の用例を検索し、その用法判定結果を参照することにより、これらの派生的表現の表記の用法は、「状態が継続している」意味をもつ機能的用法であると判定できる。同様に、下半分の例においては、派生的表現の表記の用法は、派生的表現の表記を構成する動詞の内容的用法であると判定できる。

4. 機能表現の表記の用法の分析

4.1 分析対象

まず、毎日新聞 1995 年の 1 年分を対象として、代表的表現 497 表記の用例を収集し、人

手で用法判定を行った用例、37,769 用例中から、用法判定済みの個所が 50 個所以上となり、機能的用法の割合が 1~9 割となる代表的表現の表記を選定したところ、94 表記となった。次に、毎日新聞 1995 年の 1 年分のうち、上で述べた用法判定済みの 37,769 用例以外から、代表的表現 94 表現の派生的表現の用例を収集し、人手で用法判定を行った後、用法判定個所数が一定頻度以上となる派生的表現の表記およびその用例 2,058 用例を収集した。なお、これらの用例の中には、代表的表現の機能的用法の用例と判定されたものも含まれ、機能表現の表記の総数は 116 表記 (代表的表現 22 表記、派生的表現 94 表記) となった。また、これらの表記の中には、内容的用法でしか出現しないものも存在し、実際に、機能的用法として出現した表記は、99 表記、1,539 用例であった。次節では、これらの機能的用法、内容的用法の全用例、2,058 用例を対象として、3.2 節で提案した解析方式を適用した結果を分析する。

4.2 分析手順

代表的表現の表記の用法判定済みの 37,769 用例の集合 S_c^{tr} を参照して、前節で述べた 2,058 用例を対象として 3.2 節で提案した解析方式を適用し、用法判定および分析を行う手順の詳細を以下に示す。

- (1) 「機能表現一覧」⁵⁾ の全 16,801 表現のうち、表記が一字のものを除外し、入力文中の任意の位置 (ただし、形態素解析^{*1}の形態素境界と機能表現表記の境界が一致しない場合を除く) と照合する機能表現表記を列挙する。
ここで列挙された機能表現表記のうち、ある派生的表現 F_d に対応する代表的表現を F_c とし、 F_d の前接形態素、後接形態素をそれぞれ、 m_{-1} 、 m_{+1} とする。また、入力文に対して、代表的表現の表記が直接照合した場合も、その代表的表現の表記を F_c とし、 F_c の前接形態素、後接形態素をそれぞれ、 m_{-1} 、 m_{+1} とする。
- (2) (1) で求めた代表的表現の表記 F_c に対して、前後の形態素 m_{-1} および m_{+1} の品詞大分類が一致する用例を S_c^{tr} から収集する。
 - (i) (2) を満たす代表的表現の表記 F_c のうち、入力文中での派生的表現の表記 F_d (入力文に対して、代表的表現の表記が直接照合した場合は F_c) が最長となるものが一意に決まり、 S_c^{tr} における F_c の用法判定結果 l^{tr} も一意に決まる場合は、 l^{tr} を採用する。
ここで、 l^{tr} が正解となる場合が、表 3 中の (i) に該当する。

*1 MeCab および IPAdic を用いた。

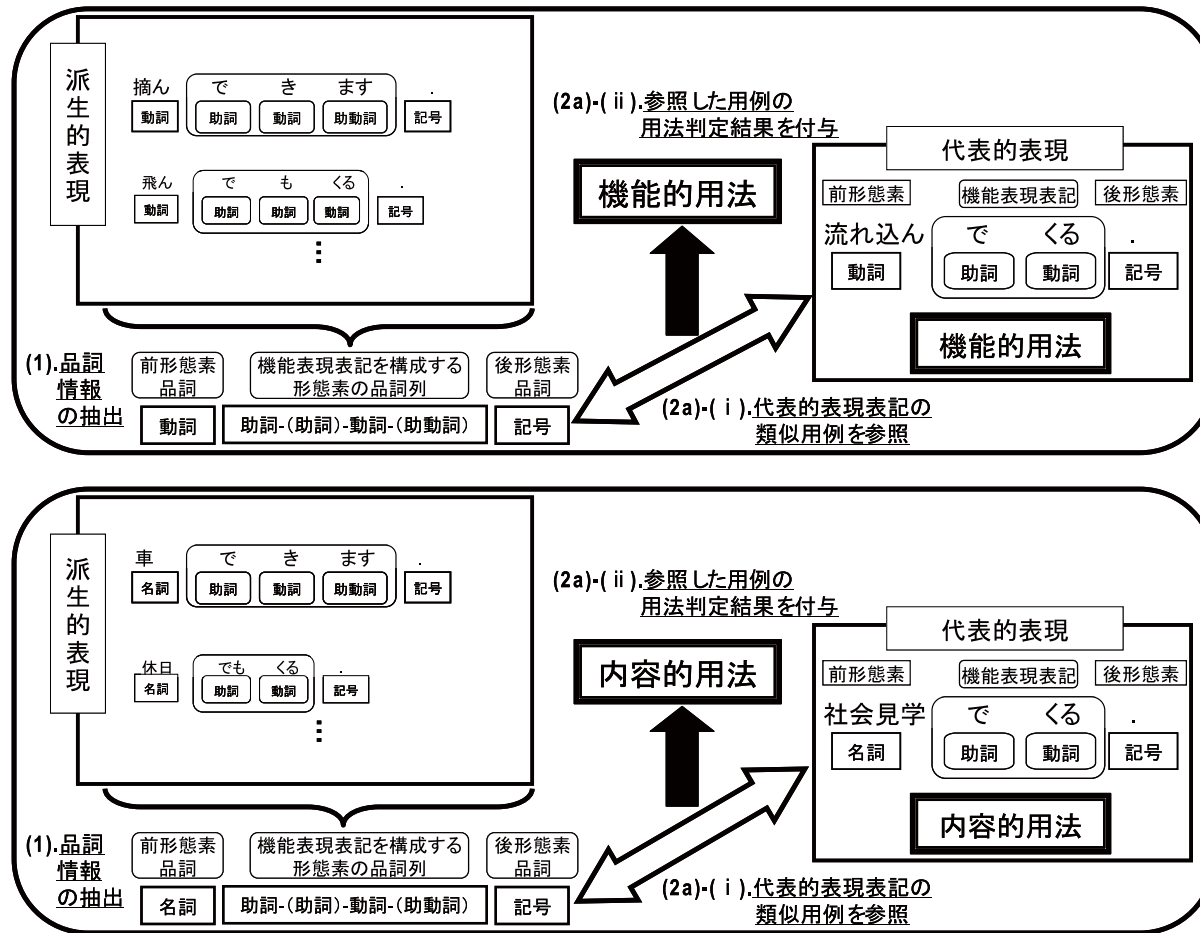


図 2 模式図: 「代表的表現の表記の用例」を参照して「派生的表現の表記の用例」の用法を判定

- (ii) (2) を満たす代表的表現の表記 F_c のうち、入力文中での派生的表現の表記 F_d (入力文に対して、代表的表現の表記が直接照合した場合は F_c) が最長となるものが複数存在する場合、あるいは、最長となる派生的表現の表記 F_d に対して、 S_c^{tr} における F_c の用法判定結果 l^{tr} が複数存在する場合。

これらの場合、前後の形態素 m_{-1} および m_{+1} の品詞細分類、または、機能表現表記 F_d および F_c を構成する形態素品詞列に対する条件を課す。

S_c^{tr} において、代表的表現 F_c の用法判定結果 l^{tr} が一意に決まる場合は、 l^{tr} を採用する。

ここで、 l^{tr} が正解となる場合が、表 3 中の (ii) に該当する。

- (iii) (2) を満たす代表的表現の用例が S_c^{tr} に存在しない場合、入力文の形態素解析結果に対して機能表現の検出を行わず、内容的用法と判定する。
ここで、この判定が正解となる場合が、表 3 中の (iii) に該当する。
- (3) (2) の (i)~(iii) によって正解とならない場合には、以下を行う。
- (iv) (1) で求めた代表的表現の表記 F_c に対して、前後の形態素 m_{-1} および m_{+1} の品詞大分類が一致する用例を作例し S_c^{tr} に追加することにより、(2) の (i), (ii) の手順を経れば正解可能かどうかを検証する。
ここで、正解可能となる場合が、表 3 中の (iv) に該当する。
- (4) (2) および (3) によって正解もしくは正解可能とならない場合には、以下を行う。
- (v) (1) で求めた派生的表現の表記 F_d に対して、前後の形態素 m_{-1} および m_{+1} の品詞大分類が一致する用例を作例し、(2) の (i), (ii) の手順と同様の手順を経れば正解可能かどうかを検証する。
ここで、正解可能となる場合が、表 3 中の (v) に該当する。
- (5) (4) の (v) において派生的表現の表記 F_d の用例を作例しても、入力文中の表記最長の他の機能表現、もしくは、その代表的表現に対応する用法判定結果 l^{tr} と競合するかどうかを検証する。
- (vi) 両者が競合し、用法判定結果を一意に絞り込めない場合が、表 3 中の (vi) に該当する。この場合は、用法判定結果を絞り込むために何らかの方式を導入することにより、正解する可能性がある。
- (vii) 一方、入力文中の表記最長の機能表現を優先するというヒューリスティクスを用いることによって、 l^{tr} が優先され不正解となる場合が、表 3 中の (vii) に該当する。

4.3 分析結果

前節の手順にしたがって、2,058 用例のすべて、および、そのうちの派生的表現の用例 1,880 用例を分類した結果を表 3 に示す。

この結果から分かるように、「代表的表現の用例の作例なしで正解」の割合は、約 82% となった。この場合に該当する用例のうち、前節の (i) および (ii) に該当し、代表的表現の表記の用法判定済みの 37,769 用例の集合 S_c^{tr} を参照することにより用法判定が正解する個所の例を表 4 (i) および (ii) に示す。一方、前節の (iii) に該当し、「 S_c^{tr} 中に対応する代表的表現の用例が存在せず、入力文の形態素解析結果に対して機能表現の検出を行わずに内容的用

法と判定し正解」の個所の例を表 4 (iii) に示す。

また、前節の (iv) に該当し、「代表的表現の適切な用例を作例することにより、正解可能」の割合は約 12% 前後となった。この個所の例を表 4 (iv) に示す。これらの用例は、代表的表現の表記の用法判定済み用例集合として、十分な規模を収集することにより正解できる可能性のある用例に相当し、これらを加算すると、正解率の推定値は約 94% となる。

最後に、前節の (v) および (vi) に該当し、「派生的表現の作例、または、候補補の絞り込み方式の導入により正解する可能性あり」の割合は 5.7% となった。このうち、(v) の「派生的表現の適切な用例を作例することにより、正解可能」の個所の例を表 5 (v) に示す。これらの用例は、派生的表現の表記に対して用法判定済み用例集合を収集しないと正解することができない。一方、(vi) の「派生的表現の用例を作例しても、入力文中で表記最長の機能表現、もしくは、その代表的表現に対応する用法判定結果と競合し、候補を一意に絞り込めない」の個所の例を表 5 (vi) に示す。これらの用例は、入力文中の派生的表現の用法とほぼ同等の用法の作例を参照したとしても、その用法とは異なる用法の代表的表現の用例が存在する場合に相当する。これらの多くは、文意まで汲み取らないと用法の判定が困難な場合に相当するが、候補補の絞り込み方式を導入することにより正解する可能性がある例も存在する。

5. 関連研究

文献 4) においては、「機能表現一覧」⁵⁾ 中の機能表現を対象として、意味を保存する言い換えが可能な機能表現の分類を規定している。また、代表的表現への言い換えを介した機械翻訳の研究としては、内容語と口語的な機能表現を扱った文献 17)、 「機能表現一覧」⁵⁾ の機能表現を対象とした文献 3)、 6)、 11)、 12) が、機能表現の検出・係り受け解析等の解析を対象とした研究としては、文献 1)、 13)、 14) がある。

6. おわりに

本稿では、「機能表現一覧」の階層性を利用し、階層において下位に位置する派生的表現について、用法が類似するより上位の代表的表現の用例を参照して、用法判定を行う手法について述べた。特に、文献 9)、 16) において明らかになった特性として、前後の形態素の品詞が代表・派生間において不変の場合には、代表的表現と派生的表現の間で用法の傾向に相関がある、という特徴を利用する方式を提案した。提案方式に基づいて、派生的表現の用法の分析を行った結果、代表的表現の表記の用法判定済み用例集合 (約 38,000 例) を参照して、

表 3 機能表現の表記の用法の分析 (代表的表現+派生的表現 / 派生的表現のみ)

分類		個所数 (割合 (%))			
(i)~(iv) 作例を含めて、 代表的表現の 用例を参照する ことにより正解。	(i)~(iii) 代表的表現の 用例の作例 なしで正解。	(i) 前後の形態素の品詞大分類が一致する代表的表現の用例が存在。そのうち、入力文中で表記最長の機能表現が一意に決まり対応する代表的表現の用法判定結果を採用し正解。	1,356 (69.5) / 1,203 (64.0)	1,704 (82.8) / 1,543 (82.1)	1,939 (94.2) / 1,773 (94.3)
		(ii) 前後の形態素の品詞大分類が一致する代表的表現の用例が存在。そのうち、入力文中で表記最長の機能表現が複数存在する場合、または、対応する代表的表現に対する用法判定結果が複数存在する場合。これらの場合で、前後の形態素の品詞細分類、または、機能表現表記を構成する形態素の品詞列に対する条件を課し、一意に絞られた派生表現に対応する用法判定結果を採用し正解。	90 (4.4) / 85 (4.5)		
		(iii) 前後の形態素の品詞大分類が一致する代表的表現の用例が存在しない。入力文の形態素解析結果に対して機能表現の検出を行わず内容的用法と判定し正解。	258 (12.5) / 255 (13.6)		
	(iv) 代表的表現の適切な用例を作例することにより、正解可能。	235 (11.4) / 230 (12.2)			
(v), (vi) 派生的表現の作例、または、 候補の絞り込み方式の導入に より正解する可能性あり。	(v) 派生的表現の適切な用例を作例することにより、正解可能。	96 (4.7) / 96 (5.1)	21 (1.0) / 11 (0.6)	117 (5.7) / 107 (5.7)	
	(vi) 派生的表現の用例を作例しても、入力文中で表記最長の機能表現、もしくは、その代表的表現に対応する用法判定結果と競合し、候補を一意に絞り込めない。				
(vii) 派生的表現の用例を作例しても、入力文中で表記最長の機能表現、もしくは、その代表的表現に対応する用法判定結果が優先され、不正解。		2 (0.1) / 0 (0)			
合計		2,058 (100) / 1,880 (100)			

派生的表現の表記の用法判定を行うことにより、80%以上の用例の用法を正しく判定できることが分かった。今後の課題としては、以下が挙げられる。

- 提案方式では、代表的表現の用例、派生的表現の用例のいずれについても、できるだけ多くの用法の用例を収集し、用法判定結果を付与した用例集合を蓄積することが性能改善の鍵を握る。そこでは、大規模な未解析テキストコーパスを情報源として、機能表現表記の前後の形態素の品詞のバリエーションをできるだけ多く収集し、サンプリングして用法判定結果を付与することが最も効果的である。
- 代表的表現と派生的表現の間で、機能表現表記を構成する形態素列の代表・派生関係を網羅的に収集しておき、これを代表的表現の用法判定済み用例との照合の際に参照する

ことによって、代表的表現と派生的表現の間の照合の性能を改善できると考えられる。

- 新聞記事を対象として構築した代表的表現の表記の用法判定済み用例集合を参照して、新聞記事以外の多様なジャンルのテキスト中の機能表現表記の用法判定を行うタスクにおいて、提案方式の有効性を評価する。

参 考 文 献

- 1) 小早川健, 関場治朗, 木下明徳, 熊野 正, 加藤直人, 田中英輝: 単語格子とマルコフモデルによる日本語機能表現の解析 — 日本語機能表現辞書「つつじ」を用いて—, 電子情報通信学会技術研究報告, NLC2009-1, pp.15-20 (2009).

表 4 機能表現の表記の用法の分析: 「(i)~(iv) 作例を含めて, 代表的表現の用例を参照することにより正解」の例

(i) 前後の形態素の品詞大分類が一致する表記最長の派生的表現あり. 対応する代表的表現の用法判定結果を採用し正解

前形態素の品詞 (表記)- 代表的表現表記- 後形態素の品詞 (表記) (新聞記事から収集)	前形態素の品詞 (表記)- 派生的表現表記- 後形態素の品詞 (表記) (新聞記事から収集)	意味的等価クラス	用法
動詞 ((覚悟で) 臨んで)- でいく- 記号 (「」)	動詞 (膨らん)- でいきます- 記号 (「」)	進行-継続-テイク	機能的用法
名詞 (マンツーマン)- でいく- 記号 (「」)	名詞 ((ゲリラ) 作戦)- でいきます- 記号 (「」)	(「でいきます」が機能的用法の場合は「進行-継続-テイク」)	内容的用法

(ii) 前後の形態素の品詞大分類が一致する表記最長の派生的表現に対して, 代表的表現の用法判定結果が複数存在.
前後の形態素の品詞細分類, または, 機能表現表記を構成する形態素の品詞列に対する条件を課し, 一意に絞られた候補が正解

前形態素の品詞-活用形 (表記)- 代表的表現の品詞-活用形 (表記)- 後形態素の品詞 (表記) (新聞記事から収集)	前形態素の品詞-活用形 (表記)- 派生的表現の品詞-活用形 (表記)- 後形態素の品詞 (表記) (新聞記事から収集)	意味的等価クラス	用法
動詞-自立-連用形 (受け止め)- 助詞-接続助詞・形容詞-非自立-基本形 (て・よい)- 名詞-非自立-一般 (の (か))	動詞-自立-連用形 ((縮小) し)- 助詞-接続助詞・助詞-係助詞・形容詞-非自立-基本形 (て・も・よい)- 名詞-非自立-一般 (の (では))	許可-許可-テヨイ	機能的用法
動詞-自立-連用形 (かかり)- 名詞-非自立-副詞可能 (きり)- 助動詞-基本形 (です)- (「内容的用法」となるが, 機能表現表記を構成する 形態素の品詞列が異なる代表的表現の用例 「歌い-きり (動詞-非自立)-たい」)	動詞-自立-連用形 (こもり)- 名詞-非自立-副詞可能 (つきり)- 助動詞-基本形 (です)-	限定-そのもの-ノミ	機能的用法

(iii) 前後の形態素の品詞大分類が一致する代表的表現の用例が存在しない.
入力文の形態素解析結果に対して機能表現の検出を行わず内容的用法と判定し正解.

前形態素の品詞 (表記)- 派生的表現表記- 後形態素の品詞 (表記) (新聞記事から収集)	意味的等価クラス	用法
助詞 ((月) に)- でも行っ- 助動詞 (た)	(「でも行っ」が機能的用法の場合は「進行-継続-テイク」)	内容的用法
助詞 ((会談) など)- でもよく- 動詞 (使わ (れる))	(「でもよく」が機能的用法の場合は「許可-許可-テヨイ」)	内容的用法

(iv) 代表的表現の適切な用例を作例することにより, 正解可能.

前形態素の品詞 (表記)- 代表的表現表記- 後形態素の品詞 (表記) (作例)	前形態素の品詞 (表記)- 派生的表現表記- 後形態素の品詞 (表記) (新聞記事から収集)	意味的等価クラス	用法
動詞 (進ん)- でよい- 名詞 (こと (になった))	動詞 (休ん)- でもよい- 名詞 (こと (にしている))	許可-許可-テヨイ	機能的用法

2) 国立国語研究所: 現代語複合辞用例集 (2001).

3) 劉 颯, 長坂泰治, 宇津呂武仁, 松吉 俊: 意味的等価クラスを用いた日本語機能表現の集約的日中翻訳規則の作成と分析, 言語処理学会第 16 回年次大会論文集, pp.194-197

(2010).

4) 松吉 俊, 佐藤理史: 文体と難易度を制御可能な日本語機能表現の言い換え, 自然言語処理, Vol.15, No.2, pp.75-99 (2008).

表 5 機能表現の表記の用法の分析: 「(v), (vi) 派生的表現の作例, もしくは, 解候補の絞り込み方式の導入により正解する可能性あり」の例

(v) 派生的表現の適切な用例を作例することにより, 正解可能.

前形態素の品詞 (表記)- 派生的表現表記- 後形態素の品詞 (表記) (作例)	前形態素の品詞 (表記)- 派生的表現表記- 後形態素の品詞 (表記) (新聞記事から収集)	意味的等価クラス	用法
動詞-自立-連用形 ((考慮) し)- てもよかつ- 助動詞-基本形 (た) (「内容的用法」となるが, 前後の形態素の品詞大分類等の条件が 同一となる代表的表現の用例 (新聞記事から収集) 「親孝行-でき (動詞-自立-連用形)-て-よかつ-た (助動詞-基本形)-ね」)	動詞-自立-連用形 (教え)- てもよかつ- 助動詞-基本形 (た)	許可-許可-テヨイ	機能的用法

(vi) 派生的表現の用例を作例しても, 入力文中で表記最長の機能表現, もしくは, その代表的表現に対応する用法判定結果と競合し, 候補を一意に絞り込めない.

前形態素の品詞-活用形 (表記)- 代表的表現表記- 後形態素の品詞 (表記) (作例)	前形態素の品詞-活用形 (表記)- 派生的表現表記- 後形態素の品詞 (表記) (新聞記事から収集)	意味的等価クラス	用法
動詞 ((たくさん) あっ)- て・いい- 助詞 (ね) (「機能的用法」となるが, 前後の形態素の品詞大分類等の条件が 同一となる代表的表現の用例 (新聞記事から収集) 「使っ-て-いい-よ-と、開放して」)	動詞-((しやすく) なっ)- ていいです- 助詞 (ね)	(「ていいです」が機能的用法の場合は 「許可-許可-テヨイ」)	内容的用法

- 5) 松吉 俊, 佐藤理史, 宇津呂武仁: 日本語機能表現辞書の編纂, 自然言語処理, Vol.14, No.5, pp.123-146 (2007).
- 6) Nagasaka, T., Shimanouchi, R., Sakamoto, A., Suzuki, T., Morishita, Y., Utsuro, T. and Matsuyoshi, S.: Utilizing Semantic Equivalence Classes of Japanese Functional Expressions in Translation Rule Acquisition from Parallel Patent Sentences, *Proc. 7th LREC*, pp.1778-1785 (2010).
- 7) 長坂泰治, 坂本明子, 宇津呂武仁, 森下洋平, 松吉 俊, 土屋雅稔: 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, NLP 若手の会 第4回シンポジウム (2009).
- 8) 長坂泰治, 宇津呂武仁, 松吉 俊, 土屋雅稔: 大規模階層辞書を利用した日本語機能表現の集約と解析, 言語処理学会第15回年次大会論文集, pp.328-331 (2009).
- 9) 長坂泰治, 宇津呂武仁, 松吉 俊, 土屋雅稔: 階層的機能表現辞書に基づく日本語機能表現の分析と検出, 言語処理学会第16回年次大会論文集, pp.970-973 (2010).
- 10) 長坂泰治, 宇津呂武仁, 土屋雅稔: 大規模日本語機能表現辞書の階層性を利用した機能表現検出, 言語処理学会第14回年次大会論文集, pp.837-840 (2008).
- 11) 坂本明子, 宇津呂武仁, 松吉 俊: 日本語機能表現の集約的英訳, 言語処理学会第15回年次大会論文集, pp.654-657 (2009).
- 12) 島内 蘭, 長坂泰治, 坂本明子, 宇津呂武仁, 松吉 俊: 日英特許翻訳における日本語機能表現の集約的英訳可能性の調査, 言語処理学会第16回年次大会論文集, pp.611-614 (2010).
- 13) 注連隆夫, 土屋雅稔, 松吉 俊, 宇津呂武仁, 佐藤理史: 日本語機能表現の自動検出

- と統計的係り受け解析への応用, 自然言語処理, Vol.14, No.5, pp.167-197 (2007).
- 14) 土屋雅稔, 注連隆夫, 高木俊宏, 内元清貴, 松吉 俊, 宇津呂武仁, 佐藤理史, 中川聖一: 機械学習を用いた日本語機能表現のチャンキング, 自然言語処理, Vol.14, No.1, pp.111-138 (2007).
 - 15) 土屋雅稔, 宇津呂武仁, 松吉 俊, 佐藤理史, 中川聖一: 日本語複合辞用例データベースの作成と分析, 情報処理学会論文誌, Vol.47, No.6, pp.1728-1741 (2006).
 - 16) 宇津呂武仁, 鈴木敬文, 島内 蘭, 松吉 俊, 土屋雅稔: 平成22年度研究進捗状況報告: 日本語機能表現班大規模階層辞書を用いた日本語機能表現解析体系の研究, 特定領域研究「日本語コーパス」平成22年度全体会議予稿集, pp.37-44 (2010).
 - 17) 山本和英, 白井 諭, 坂本 仁, 張 玉潔: SANDGLASS: 両言語換言機構を基軸とする音声翻訳, 言語処理学会第7回年次大会発表論文集, pp.221-224 (2001).