

B0TB2066

卒業論文

述部機能表現に対する意味ラベル付与

上岡裕大

2014年3月6日

東北大学
工学部 情報知能システム総合学科

述部機能表現に対する意味ラベル付与*

上岡裕大

内容梗概

日本語の表現には，語彙的意味をほとんど担わず，主に統語的關係を示す機能語と，複数の機能語が組み合わさって全体として1つの機能語と同等に働く複合辞が存在する．本稿では，機能語と複合辞を総称して機能表現と呼ぶ．機能表現を正しく認識することは，拡張モダリティ解析や含意関係認識などの応用タスクに有用である．機能表現の曖昧性や規模を考慮すると，正しく解析するためには学習・評価用のコーパスが不可欠であるが，現時点で解析ツールを開発するための基礎となるコーパスが利用可能でない．そこで，本稿では，日本語コーパス中の述部機能表現に対して意味ラベルを付与した述部機能表現意味ラベル付与コーパスの作成について報告する．また，作成したコーパスを用いて機械学習による意味ラベル付与を行い，解析結果の誤り分析を行う．

キーワード

機能表現，コーパス

*東北大学 工学部 情報知能システム総合学科 卒業論文, B0TB2066, 2014年3月6日.

目次

1	序論	1
2	関連研究	3
2.1	応用タスクに関する研究	3
2.2	日本語機能表現に関する研究	4
3	意味ラベル付与コーパスの作成	6
3.1	アノテーション対象	6
3.2	意味ラベルの表現方法	6
3.3	意味ラベルの追加	8
4	意味ラベル付与コーパスの分析	10
4.1	コーパス概要	10
4.2	意味ラベルの分布	10
5	実験	13
5.1	実験条件	13
5.2	実験結果	14
5.3	分析	14
6	結語	16
	謝辞	17

1 序論

日本語の表現には、語彙的意味をほとんど担わず、主に統語的關係を示す機能語と、複数の機能語が組み合わさって全体として1つの機能語と同等に働く複合辞が存在する。本稿では、機能語と複合辞を総称して機能表現と呼ぶ。日本語の文章は、文章に記述される事象にこれら機能表現が後続し、種々の主観的情報を表す。例えば、文(1)と文(2)にはそれぞれ「たい」や「てしまったかもしれない」という表現が現れる。

- (1) 明日、本を買いに行き たい。
- (2) パソコンが壊れてしまったかもしれない。

文(1)は、助動詞「たい」が願望の意で用いられ、事象「買いに行く」の成立を望んでいることを表す。また、文(2)は、複合辞「てしまっ」と助動詞「た」が完了の意、複合辞「かもしれない」が推量の意で用いられ、事象「壊れる」の成立を推量していることを表す。

機能表現の意味を正しく特定することによって、拡張モダリティ解析や含意関係認識などの応用タスクに貢献することができる。応用研究の例として、江口ら [1] による拡張モダリティ解析の研究、および成田ら [2] による日本語の事実性解析の研究がある。いずれの研究においても、機能表現の曖昧性に起因する解析誤りがあることを明らかにしている。機能表現には、助動詞「ため」が”理由”を表すために用いられる場合と、”目的”を表すために用いられる場合があるように、同一表現で異なる意味を表す表現が存在し、曖昧性がある。拡張モダリティ解析や事実性解析などの応用タスクにおける精度向上のためには、機能表現を正しく認識する必要がある。特に、日本語では、書き手の主観的情報は述部に現れることが多いため、述部の機能表現を正しく認識することは、自然言語処理にとって重要な要素となる。

しかし、現時点では機能表現の表す意味を正しく認識する研究が十分に行われているとは言えない。Shudoら [3] は、機能表現を検出することを目的として、機能的用法と内容的用法を識別するための規則を人手で作成した。これに対して、鈴木ら [4] は、人手による検出規則ではコストが大きすぎることや網羅できる機能表現の規模に限界があることを指摘し、条件付確率場を用いて複合辞を検出する手法を提案した。しかし、いずれの研究も抽出した機能表現の意味の特定は行っていない。また、今村ら [5] は機能表現の意味を機械学習により特定する意味ラベルタガを提案したが、学習に用いたコーパスの公開やタグ付けにおける誤り分析は行われていない。

機能表現の曖昧性や規模を考慮すると、その意味を正しく認識するためには、学習・評価のためのコーパスが不可欠であるが、現時点で解析ツールを開発するための基礎となるコーパスが利用可能ではない。そこで本稿では、日本語コーパスの述部機能表現に

対して意味ラベルを付与したコーパスの作成について報告する．また，作成したコーパスを用いて機械学習による意味ラベル付与を行い，その誤り分析の結果を報告する．

本稿の構成は次の通りである．まず，第2章において，機能表現解析に関連する研究について述べる．次に，第3章で，意味ラベル付与コーパスの作成について述べ，第4章において，作成したコーパスの統計情報について述べる．さらに，第5章では，意味ラベル付与コーパスを用いた実験および実験によって明らかになった課題を述べる．最後に，第6章でまとめを述べる．

2 関連研究

この章では，機能表現解析に関する関連研究について述べる．まず，応用タスクに関する研究 [1][2] を挙げ，機能表現の重要性を明らかにする．次に，日本語機能表現を収録した網羅的な辞書を編纂した研究について述べ，それらを踏まえて機能表現の検出を試みた研究 [3][4] について述べる．さらに，機能表現の意味の特定を試みた研究 [5] について述べる．

2.1 応用タスクに関する研究

江口ら [1] は，モダリティや真偽判断，価値判断を統合して拡張モダリティとして定義し，拡張モダリティ解析における問題点を考慮した拡張モダリティ解析システムを提案した．まず，江口らは，独自の拡張モダリティタグ体系を設計し，タグ体系に基づくコーパスを作成した．江口らの設計したタグ体系は，〈態度表明者，時制，仮想，態度，真偽判断，価値判断，焦点〉の7項目から構成される．また，設計したタグ体系に基づいて，日本語文中の各事象に対して人手でタグ付与をしたコーパスを作成した．さらに，コーパス作成時に一部の内容語が拡張モダリティに影響することを明らかにし，内容語が直前の事象に与える影響を〈態度，真偽判断，価値判断〉の三つ組で記述した手がかり表現辞書を作成した．

これらに基づいて，江口らは，条件付確率場を利用して解析システムを構築し，実験を行った．実験の結果の中に多かった拡張モダリティタグ”意志”を”叙述”と誤った例についての分析を行っている．分析において，両者に共通して出現する特徴的な形態素情報を含む事象を誤る傾向があることを明らかにした．例えば，直後に機能語「ため」のような形態素を含む事例では，「ため」が”理由”を述べるために使われているのか，”目的”を述べるために使われているかを正しく把握することが重要であるとしている．

また，本タグ体系に基づくコーパスは，その規模を拡大し，拡張モダリティタグ付与コーパスとして公開されている [6] ．

成田ら [2] は，英語を対象とした事実性解析モデルをもとに，日本語に対するルールベースの事実性解析器を構築し，拡張モダリティタグ付与コーパスの一部を解析し，その誤り分析を行った．解析には江口らの作成した手がかり表現辞書や後述する機能表現辞書を用いている．特に，解析アルゴリズムにおける機能表現辞書を用いた更新のステップでは，後方からの最長一致により機能表現辞書と着目している文節とを照らし合わせ，機能表現が見つかった場合には，検出された表現の意味に応じて，ルールに基づいて各事象の事実性を更新している．さらに，実験の結果から，現在利用可能な語彙知識のカバレッジや，語彙的手がかりの持つ曖昧性はあるか，局所的な語彙知識の組み合わせだけで解決できない問題は存在するのか等の観点から次のように誤りを分類している．

1. 語彙知識のカバレッジの問題（内容語 16%，機能語 8%）
2. 語彙知識としては存在するが，表記揺れ等，辞書引きが難しい問題（2%）
3. 語彙的手がかりの多義性の問題（9%）
4. 手がかりとなる表現のスコープの問題（37%）
5. その他（28%）

誤りのうち最も多いのは手がかりとなる表現のスコープの問題であるが，同程度の割合で語彙知識に関する問題があることを指摘している．また，同手法を用いた成田らの別の文献 [7] によれば，主節における事実性解析の誤りは，そのほとんど (106/108) が機能表現の解析誤りに起因するものであるとしている．

2.2 日本語機能表現に関する研究

大規模な数の機能表現を処理する基礎として，松吉らによる日本語機能表現辞書『つつじ』 [8][9] がある．つつじでは，言語学的文献を参考にして得た見出し語 341 件について，種々の異形を考慮し，約 17000 種類の機能表現を収録している．つつじの見出し体系は 9 つの階層を持つ階層構造となっており，見出し語・意味・派生・機能語の交替・音韻的变化・とりたて詞の挿入・活用・「です/ます」の有無・表記のゆれなどを考慮している．既存の機能表現リストとの比較により，各々の見出し語に対して，ほぼすべての異形を網羅しているとされる．

機能表現を解析するにあたって，最初の問題となるのはその検出である．Shudo ら [3] は，機能表現を検出することを目的として，機能的用法と内容的用法との曖昧性がある表現に対して，用法を識別するための規則を人手により作成した．しかし，人手での規則作成はコストが大きいため，網羅できる機能表現の規模に限界がある．

鈴木ら [4] は，大規模な均衡コーパスである『現代日本語書き言葉均衡コーパス』において，機能的用法・内容的用法の曖昧性を持つ機能表現を対象として，機械学習により用法判定を行う手法を適用し，その性能を評価した結果を報告した．当文献における機能表現とは，我々の扱う機能表現のうち複合辞に相当するものである．複合辞となり得る表記を構成する形態素が，機能語となるのか複合辞の部分となるのかという曖昧性解消に取り組んだ．条件付確率場を利用したチャンキングを用いた実験の結果，97%近い F 値を達成している．

今村ら [5] は，形態素解析結果に対して，述部を同定し，つつじの意味ラベルを機能表現に付与する意味ラベルタガールを構築した．今村らの研究では，機能表現辞書つつじと識別モデルに基づく最尤選択を組み合わせることで機能表現を同定した．辞書を用いてラティスを作成することにより，異なるドメイン間において系列ラベリングより高い精度で意

味ラベルを付与することに成功した。しかし，当文献では，タグ付けに用いた学習・評価用のコーパスの公開や，タグ付けにおける課題などは明らかにされていない。

3 意味ラベル付与コーパスの作成

本研究では，機能表現の意味の同定のために，日本語コーパス中の述部機能表現に対して意味ラベルを付与し，学習・評価用の意味ラベル付与コーパスを作成した．本章では，意味ラベル付与コーパスの作成について述べる．

3.1 アノテーション対象

本コーパスの元となるテキストは，拡張モダリティタグ付与コーパス [6] である．当該コーパスは現代日本語書き言葉均衡コーパス¹のコアデータ (BCCWJ) に対して，拡張モダリティタグを付与したものである．当該コーパスには，本文およびその形態素解析結果の他に，文中の各事象に対して拡張モダリティタグが記述されている．本稿では，ルールベースで解決不可能な問題を解決することを目的として，拡張モダリティタグ付与コーパスの一部である Yahoo!知恵袋 (OC) のうち，成田らの分析 [7] において，主節に解析誤りがあると認められた 1100 文をアノテーション対象の文として選定した．今回，アノテーションの対象としたテキストの例を図 1 に示す．

アノテーション対象は，各文の主節に含まれる述部に後続する機能表現である．本稿では，これを述部機能表現と呼ぶ．もちろん主節以外にも機能表現は出現するが，書き手の意図をもっとも反映しているのは主節の述部であると仮定し，本稿では主節に含まれる述部のみに焦点を絞った．

3.2 意味ラベルの表現方法

本研究では，形態素を最小単位として意味ラベルを付与する．機能表現の範囲を示す要素の表現方法としては，以下で示す IOB2 フォーマット [10] を使用する．

- I 述部機能表現に含まれる先頭以外の形態素
- O 述部機能表現に含まれない形態素
- B 述部機能表現の先頭の形態素

また，機能表現の単位は機能表現辞書つつじ [8] に収録されるものを基準に判定し，意味ラベルには文献 [9] で定義される意味的等価クラスを用いる．文献 [9] で用いられる最も上位の階層における 45 の意味的等価クラスをもとに意味を判別し，対応するラベルを付与する．文 (2) に対する意味ラベルの付与例を以下に示す．

¹http://www.ninjal.ac.jp/corpus_center/bccwj/

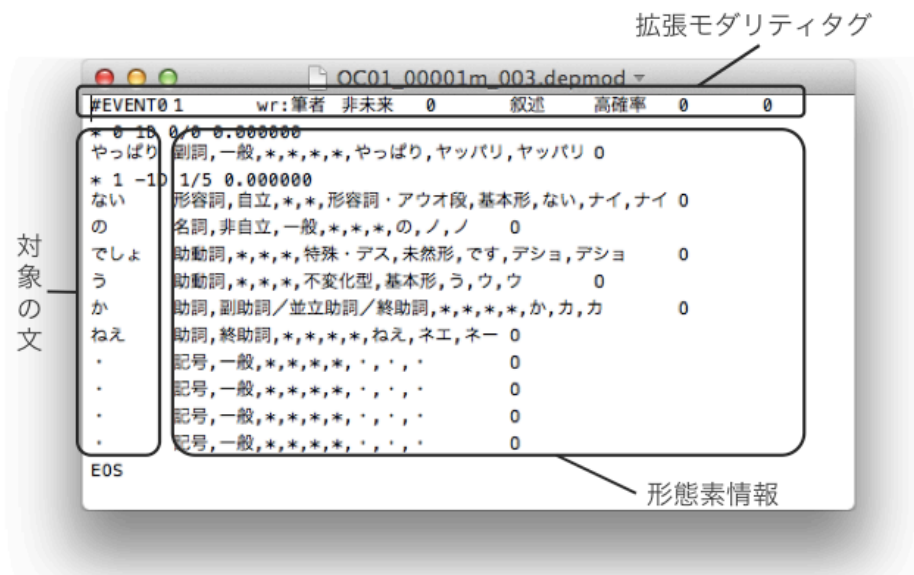


図 1: アノテーション対象のテキスト. 1 テキストに 1 文の形態素解析結果および拡張モダリティタグが記述される.

パソコン が 壊れ て しまっ た か も しれ ない .
 O O O B-過去 I-過去 B-過去 B-推量 I-推量 I-推量 I-推量 O

機能表現「かもしれない」は意味的等価クラスのうち推量クラスに分類され、形態素単位では「か/も/しれ/ない」と分割される。したがって先頭の形態素「か」にラベル”B-推量”を付与し、続く形態素「も/しれ/ない」には、それぞれにラベル”I-推量”を付与する。機能表現「てしまっ」および「た」も同様に意味的等価クラスに準拠したラベルを付与する。また、形態素「が」は格助詞であり、機能語に分類されるが、述部ではないためラベル”O”を付与する。それ以外の形態素にもラベル”O”を付与する。

以上の方法に従い、形態素単位で人手でラベル付与を行った、実際のラベル付与作業においては、1 行が 1 形態素に相当する表形式のデータを扱った。実際の作業ファイルを図 2 に示す。付与作業では、作業者が主観に従い対象となる述語を選択し、その述語に後続する形態素について、表層一致をもとに機能表現辞書引きを行った結果を参考に適

切なラベルを選択した。適切なラベルがないと判断した場合には、新たに定義し、その旨を記述した。実際にラベル付与を行ったコーパスの例を図3に示す。

3.3 意味ラベルの追加

アノテーション作業の際に、機能表現辞書つつじには収録されていないが、主観的情報を表し得る表現が存在することが明らかになった。表明者の主観的情報を読み取るという観点からは、これらの表現に対しても意味ラベルを付与することが望ましい。そこで、本研究では、これらの表現に対して新たなラベルを定義し、意味ラベル付与の対象となるすべての形態素に対して意味ラベルを付与できるようにした。具体的には、受身、自発、使役、尊敬、丁寧、困難、容易、様態、態度の9ラベルを追加した。追加したラベルと表現の例を表1に示す。これらの他に、該当する意味的等価クラスは存在するが、表記揺れなどの理由でつつじには収録されない機能表現を含め、計207エントリを新たに追加した。

表 1: 追加したラベル一覧と対応する表現の例

追加したラベル名	表現の例
受身	れる, られる
自発	れる, られる
使役	せる, させる
尊敬	せる, させる
丁寧	ます
困難	にくい, がたい, づらい
容易	やすい, よい, いい
様態	そう
態度	ね, ねえ, なあ, なあ

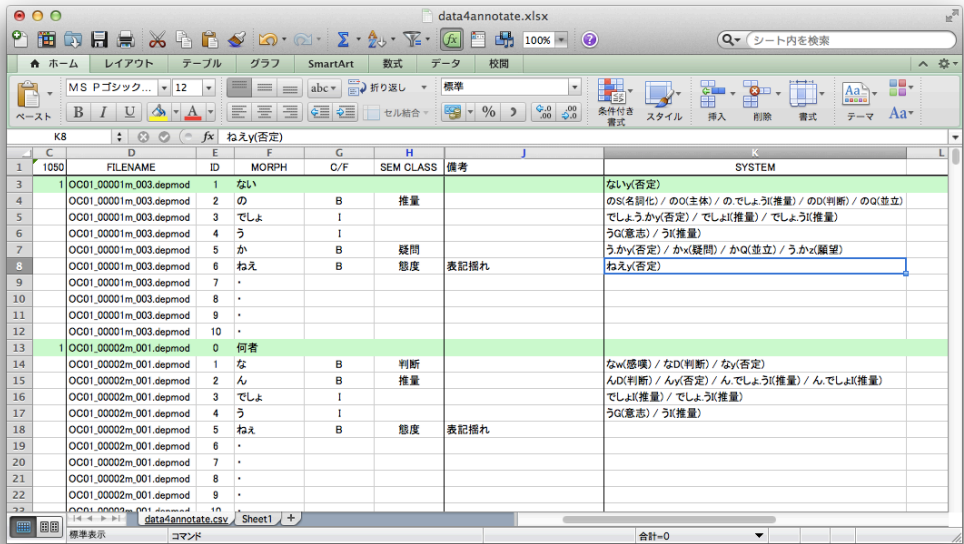


図 2: 実際の作業ファイル .



図 3: アノテーション済みのテキスト .

4 意味ラベル付与コーパスの分析

本章では，作成したコーパスについて，その内容の詳細について述べる．

4.1 コーパス概要

コーパスの全体的な統計情報を表 2 に示す．まず，対象とした 1100 文のうち方言や誤字，口語調の崩れた日本語などを主節に含む 50 文を除外した結果，収録文数は 1050 文となった．述部機能表現であるとしてラベル”B”または”I”を付与した形態素数は全体で 2853 であった．また，コーパス全体に現れる述部機能表現数は 1944 であり，その異なり総数は 280 であった．コーパス中で使用されるラベルの種類数は，文献 [9] で使用される 45 の意味的等価クラスの中の 31 に加え，本稿で新たに追加した 9 ラベルを加えた計 40 種類である．

表 2: コーパスの統計情報

収録文数	1050
意味ラベル数	2853
機能表現数	1944
機能表現異なり総数	280
ラベルの種類数	40

4.2 意味ラベルの分布

コーパス中に含まれる意味ラベルごとの表現数は表 3 のようになった．出現したラベル 40 種のうち，表現数上位の 10 種類が全体の約 9 割を占めた．最も出現頻度の高いラベルは”判断”のラベルだった．このラベルには「なければならない」のように”当為”を表す機能表現も含まれるが，コーパス中では，文 (3) や文 (4) のように「だ」や「です」のような断定の助動詞がほとんどの割合を占めた．これらの表現は一般的に広く使われる表現であり，ドメインによらず頻出すると考えられる．

- (3) 戦う戦わないにしても、自分の身になることだ。
- (4) どなたか教えていただけるとありがたいです。

また，”願望”や”丁寧”のラベルが”判断”のラベルに次いで出現頻度が高かった．”願望”のラベルを付与された表現には文 (5) の「て下さい」のように相手に働きかける表現

が多く、「丁寧」のラベルを付与された表現には文(6)の「ます」のような表現が多かった。これらの表現は、今回、ラベル付与の対象としたコーパスに特徴的であると考えられる。コーパス作成の対象としたQA 掲示板では、相手に問いかけ、何かについての答えを求める表現が多くみられた。また、相手を想定した文章であるため、「ます」のように丁寧を表す表現が多かった。

- (5) 何か良い方法があれば教えてください。
- (6) 参考になるHPや、アドバイス宜しくお願いします。

表 3: 意味ラベル別出現回数

ラベル名	表現数	ラベル名	表現数	ラベル名	表現数	ラベル名	表現数
判断	440	逆接	27	尊敬	8	比況	1
願望	200	順接	25	使役	8	自発	1
丁寧	173	理由	25	感嘆	5	名詞化	1
推量	173	受身	22	容易	5	定義	1
態度	165	伝聞	21	可能	4	継起	1
否定	148	授受	17	限定	4	起点	1
過去	125	並立	15	不必要	3	不許可	1
疑問	124	自然発生	11	困難	2	強調	1
意志	81	許可	10	範囲	2	対象	1
進行	80	例示	9	様態	2	目的	1

表 4: 1文あたりの機能表現数の分布

機能表現数	出現回数
0	63
1	402
2	335
3	168
4	53
5	19
6	9
7	1

さらに、1文あたりの機能表現数を表4に示す。例えば、「壊れてしまったかもしれない」の場合、「てしまっ/た/かもしれない」のように考え、機能表現数は3とする。1文

あたりの機能表現数が1~3の機能表現が全体の8割以上を占めていた。4つ以上の機能表現列からは出現頻度が極端に少なかった。一方で、コーパス中には対象とする述語に機能表現が後続しない文が63個含まれていた。これらは、文(7)のように動詞の活用によって”命令”を表す文や文(8)のように名詞で終わる文であり、機能表現のみでは主観的態度を認識することはできない。

(7) 興味があるならしりごみしないで当たって砕けろ！

(8) URLをコピーして、貼り付け！

5 実験

システムによる意味ラベル付与の現状を明らかにし、誤り分析を行うことを目的として、条件付確率場 (CRF) を利用した系列ラベリングによる意味ラベルの自動付与実験を行った。本章では、実験の内容およびその結果について述べる。なお、CRF 実装として CRFsuite²を用いた。

5.1 実験条件

実験は、第 2 章で作成した意味ラベル付与コーパスを正解データとして、10 分割交差検定によるクロズドメインテストを行った。機能表現を基本単位と考え、BI タグの連続をまとめて一致したラベルを正答とし、以下で定義する適合率、再現率、F 値を算出した。さらに、実用的な精度として、文全体での精度も算出した。文全体の精度では、1 文中に含まれるすべての意味ラベルが正解データと一致した場合にのみ正答とした。また、機能表現の検出が正しく行われているかを検証するため、意味を考慮せずに BI タグが一致した表現を正答とする評価も同時に行った。

$$\text{適合率} = \frac{\text{正しくラベル付与された機能表現数}}{\text{システムによっていずれかのラベルを付与された機能表現数}}$$

$$\text{再現率} = \frac{\text{正しくラベル付与された機能表現数}}{\text{評価データに存在する機能表現数}}$$

$$F \text{ 値} = \frac{2 \times \text{適合率} \times \text{再現率}}{\text{適合率} + \text{再現率}}$$

CRF を用いて系列ラベリングを行うにあたって、本稿では形態素情報 (表層・品詞・品詞細分類) を素性とした。具体的には、対象となる語の形態素情報、前後 2 単語の形態素情報、前後 1 単語の形態素情報との bigram を用いた。

ベースライン手法には、機能表現辞書を用いた人手の規則による意味ラベル付与を用いる。今回は、形態素解析結果を受け取り、前方からみた表層の最長一致によるラベル付与を用いた。なお、今回、アノテーション対象としたデータは、ルールベースによる機能表現誤りが認められた文であるため、機械学習に対してやや不利な結果になる。

²<http://www.chokkan.org/software/crfsuite/>

5.2 実験結果

機能表現検出の結果を表 5 に示す．本実験における機能表現の検出精度は 78.74%であった．

また，表 6 に実験結果を示す．結果として，機能表現単位での精度は 83.44%となり，ベースラインを約 11%程度上回った．前述の通り，ルールベースはやや不利な設定であるが，それを考慮しても意味ラベル付与の精度は大きく改善した．また，文全体の精度でもルールベースを上回る結果となった．

表 5: 機能表現検出の結果

適合率	再現率	F 値
0.9547(1558/1632)	0.7963(1548/1944)	0.7874

表 6: 実験結果

	適合率	再現率	F 値	文全体の精度
ベースライン	0.7492(1341/1790)	0.6939(1349/1944)	0.7205	0.6095(640/1050)
系列ラベリング	0.9148 (1493/1632)	0.7670 (1491/1944)	0.8344	0.7190 (755/1050)

5.3 分析

人手による規則と CRF による系列ラベリングの結果を比較するため，BI タグについて形態素単位でのラベルの正誤を表 7 にまとめた．この結果では，正解データにおいて述部機能表現でないことを表すラベル”O”が付与されるものは考慮しない．

表 7: 人手による規則と CRF の実験結果の比較

		人手による規則	
		正答	誤答
CRF	正答	1644	393
	誤答	99	71

最も多いのは人手による規則，CRF による系列ラベリングのいずれの手法によっても正解した事例である．これに該当するものは手法によらず比較的認識しやすい表現であるため，今回は分析の対象外とする．

次に事例数が多いのは，人手による規則では誤答だったが，CRF では正答となった事例である．これら 393 件は，人手による規則では検出不可能であるが，CRF を用いるこ

とによって正しくラベルを付与することができるようになった例である．例えば，機能表現「でしょうか」は，正解データにおいて「でしょう」「か」と分割され，それぞれ”推量”，”疑問”のラベルが付与されている．この表現に対して，人手による最長一致規則を的よすると「でしょうか」を一単位として”否定”のラベルを付与してしまう．一方で，CRFを用いると，これらの表現の持つ曖昧性を解消し，正しく”推量”，”疑問”のラベルを付与することができた．

表 8: CRF によって改善した例

機能表現	でしょう	か
正解ラベル	推量	疑問
人手による規則の出力	否定	
CRF の出力	推量	疑問

表 9: CRF によって悪化した例

機能表現	たい	です
正解ラベル	願望	
人手による規則の出力	願望	
CRF の出力	願望	判断

次に，人手による規則では正答だったが，CRF では誤答となった事例が 99 件あった．例えば，機能表現「たいです」は，機能表現辞書つづじに「たいです」が”願望”として収録されているため，ラベル”願望”を付与するのが正解である．しかし，コーパス中にラベル”判断”を付与された機能表現「です」の割合が多いため，CRF では「たい」にラベル”願望”を付与し，「です」にラベル”判断”を付与してしまった．これは，正解データに「たいです」よりも「です」の出現回数が多いことに起因する．したがって，より精度の高い意味ラベル付与を実現するためには，偏りの少ないコーパスを作成する必要がある．また，一方で，分割してラベルを付与した場合でも必ずしも意味を誤ったと判断することはできない．したがって，付与する機能表現の定義を改めて見直す必要がある．

6 結語

本論文では，拡張モダリティタグ付与コーパス中の述部機能表現に対して意味ラベルを付与し，機能表現解析ツールの基礎となる学習・評価のための意味ラベル付与コーパスを作成した．また，作業の過程で，機能表現辞書つつじでは網羅しきれない表現があることを明らかにし，それらの表現に対して新たなラベルを定義した．

また，作成したコーパスを用いて条件付確率場を解析モデルとした意味ラベル自動付与を行った．新たなエントリを追加した機能表現辞書を利用した人手の規則による検出をベースラインとして，10分割交差検定の結果と比較したところ，ベースラインに対して約11%の精度向上がみられた．また，文全体での一致率は71%を超え，比較的高い精度での機能表現認識が可能であることを示した．さらに，実験結果の分析から，コーパスの偏りや規模が不足していることが明らかになった．

今後の課題として，文全体に対して意味ラベルを付与したコーパスを作成することや，選択的サンプリングにより偏りの少ないコーパスを作成することが挙げられる．加えて，実用的なコーパスにするためには，規模の拡大や作業者間での一致率なども考慮する必要がある．さらに，今回の実験において，学習に用いた素性は形態素情報のみである．機能表現検出の研究ではその他の素性を組み合わせることによって精度を向上した例があるため，素性についても検討する必要がある．

謝辞

本研究を進めるにあたり，御指導頂いた乾健太郎教授，岡崎直観准教授に感謝致します．研究活動全般に渡り，直接の御指導を頂き，的確なご助言をくださいました水野淳太氏に感謝致します．最後に，本研究における議論・検討にあたって，有益な議論とご助言を頂いた研究室の皆様にご礼申し上げます．

参考文献

- [1] 江口萌, 松吉俊, 佐尾ちとせ, 乾健太郎, 松本裕治. モダリティ、真偽情報、価値情報を統合した拡張モダリティ解析. 言語処理学会第16回年次大会論文集, pp. 852–855, 2010.
- [2] 成田和弥, 水野淳太, 乾健太郎. 日本語事実性解析課題の経験的分析. 情報処理学会研究報告 第204回自然言語処理研究会, Vol. 2011-NL-204, No. 17, pp. 1–8, 2011.
- [3] Kosho Shudo, Toshifumi Tanabe, Masahito Takahashi, and Kenji Yoshimura. Mwes as non-propositional content indicators. In *In Proceedings of Association for Computational Linguistics (ACL2004) Multiword Expressions: Integrating Processing*, pp. 32–39, 2004.
- [4] 鈴木敬文, 阿部佑亮, 宇津呂武仁, 松吉俊, 土屋雅稔. 『現代日本語書き言葉均衡コーパス』における複合辞の検出と評価. 第1回『コーパス日本語学ワークショップ』予稿集, pp. 365–372, 2012.
- [5] 今村賢治, 泉朋子, 菊井玄一郎, 佐藤理史. 述部機能表現の意味ラベルタガー. 言語処理学会第17回年次大会論文集, pp. 2–5, 2011.
- [6] 松吉俊, 佐尾ちとせ, 乾健太郎, 松本裕治. 拡張モダリティタグ付与コーパスの設計と構築. 言語処理学会第17回年次大会発表論文集, pp. 147–150, 2011.
- [7] Kazuya Narita, Junta Mizuno, and Kentaro Inui. A lexicon-based investigation of research issues in japanese factuality analysis. In *In Proceedings of the 6th International Joint Conference on Natural Language Processing (IJCNLP 2013)*, pp. 587–595, 2013.
- [8] 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.
- [9] 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.
- [10] Erik F. Tjong Kim Sang and Sabine Buchholz. Introduction to the conll-2000 shared task: Chunking. In *Proceedings of CoNLL-200 and LLL-2000, Lisbon, Portugal*, pp. 127–132, 2000.