

# 助動詞「れる」「られる」の自動意味判別

福島 真也 山本 和英

長岡技術科学大学

{fukushima,yamamoto}@jnlp.org

## 1 はじめに

日本語の語義曖昧性解消のタスクの一つとして、かな漢字換言 [1] が挙げられる。このタスクではひらがなの単語を漢字に変換することで語義の曖昧性を解消している。

本研究では、換言によって語義曖昧性解消を行う考え方を日本語の助動詞「れる」「られる」に適用させる。「れる」「られる」には受け身、可能、尊敬、自発の4つの意味が存在する。これらの中で、可能と尊敬については換言が可能である。そのため、「れる」「られる」の意味を受け身と自発に限定することが可能である。換言を行うためには、「れる」「られる」の意味を識別する必要がある。そこで今回は「れる」「られる」の意味を自動で判別することを行う。

しかし、現在公開されている「れる」「られる」に対して意味を付与してあるデータセットが存在しない。そのため、機械学習を用いた判別や、作成した自動判別器を評価することが出来ない。

本稿では、「れる」「られる」に対して意味を付与したデータセットを作成し、2つの手法で意味を自動判別した結果の評価を行った。

## 2 関連研究

本研究に近い分野としては格助詞「に」の深層格推定が存在する。この研究では機能語である格助詞「に」の深層格の自動判別を行う。その推定手法として竹野らのナイーブベイズ法による手法 [2] や田辺らのルールベースによる手法 [3] が提案されている。

また、助動詞「れる」「られる」に対して意味付与を行った研究として小山田ら [4] の研究が存在する。こちらの研究では現代日本語書き言葉均衡コーパス (以下、BCCWJと称する) のコアデータから助動詞「れる」「られる」を抽出して手作業で意味のアノテーションを行っている。しかし、データは公開されていない。

本研究では、機能語である助動詞「れる」「られる」に対して意味付与を行ったデータセットを作成した。そして辞書を用いた手法と CRF を用いた手法で意味の自動判別を行った。

## 3 データセットの構築

本研究では「れる」「られる」に意味が付与されたデータが必要である。そこで、意味が付与されたデータを作成するために、ウェブコーパスから助動詞「れる」「られる」を含む文を抽出した。その後、文中に含まれる動詞と「れる」「られる」の組み合わせに対して作業員1人で意味の付与を行った。その結果、合計4,400文に対して意味の付与を行った。今回はこのデータセットを辞書作成用の400文と評価用の4,000文に分けて使用した。表1はデータセットに出現するそれぞれの意味の出現頻度を表している。

表1: データセット4,400文中に出現するそれぞれの意味の出現頻度

意味	出現頻度 (個)
受け身	4,156
可能	361
尊敬	237
自発	195
合計	4,949

## 4 判別手法

本研究では、可能または尊敬と判別された「れる」「られる」に対して換言を行うことを目的としている。そのため、「れる」「られる」を可能、尊敬、その他の3

つに分類すれば良い。そこで、自発を受け身と統合し、可能、尊敬、受け身の3分類で判別を行う。

今回は辞書を用いた手法と CRF を用いた手法で実験を行った。

#### 4.1 辞書を用いた手法

この手法では、助動詞「れる」「られる」の直前に出現する要素が意味に関係する [4] ことに注目し、動詞で判別を行う。「れる」「られる」が受け身で使用される例は多く、BCCWJ では9割が受け身で使用されているため [5]、「れる」「られる」が受け身かどうかを辞書で判別することは現実的でない。そのため、可能と尊敬で判別を行い、どちらにも当てはまらない場合に受け身とする手法を採用した。

そこで、可能の意味で使用される動詞を収録した辞書(以下、可能表現辞書と称する)と尊敬の意味で使用される動詞を収録した辞書(以下、尊敬表現辞書と称する)を作成した。また、動詞の直前に出現する格助詞、係助詞によって可能または尊敬として判別できる動詞も存在するため、直前の格助詞、係助詞と動詞で判別を行う準可能表現辞書、準尊敬表現辞書も作成した。

##### 4.1.1 辞書の構築

BCCWJ 全体から文を取り出し、辞書に unidic-mecab 2.1.2 を用いた MeCab で形態素解析を行った。そして助動詞「れる」「られる」が文中に存在する場合、その前の動詞を収集した。なお、直前に出現する動詞が「する」の場合、動詞の直前の品詞を確認し、名詞であればそれも動詞の一部として加えた。その後、収集した動詞から尊敬の意味で使用されそうな動詞、可能の意味で使用されそうな動詞をそれぞれ、尊敬表現辞書、可能表現辞書に収録した。その後、それぞれの辞書内にある動詞の用例を BCCWJ と今回作成したデータセットの一部である 400 文から検索した。そして直前の格助詞、係助詞によって意味の出現頻度に違いがあるか確認した。直前の格助詞、係助詞と動詞の組み合わせによって、可能または尊敬と判別できるならば、その組み合わせを準尊敬表現辞書または準可能表現辞書に収録した。

表 2 は辞書に収録されている動詞の数を表す。また、表 3 は各辞書に収録されている動詞の例である。

表 2: 各辞書の動詞の収録数

辞書名	収録数 (個)
可能表現辞書	137
尊敬表現辞書	269
準可能表現辞書	64
準尊敬表現辞書	82

表 3: 各辞書の収録例

辞書名	収録例
可能表現辞書	得られる 聴き分けられる
尊敬表現辞書	感動される 住まれる
準可能表現辞書	持ち上げられる, を 上げられる, を, が
準尊敬表現辞書	座られる, が 舞われる, を

##### 4.1.2 判別の仕組み

この手法では、以下の手順で判別を行う。

1. 入力された文から、動詞と「れる」「られる」の組み合わせを抽出する。
2. 尊敬表現辞書と可能表現辞書を参照し、辞書に同じ動詞が存在するなら、辞書名に沿った意味を付与する。どちらの辞書にも存在しなければ 3へ。
3. 動詞の直前の格助詞、係助詞を取得し、準尊敬表現辞書、準可能表現辞書を参照する。取得した動詞と格助詞、係助詞の組み合わせが辞書内に存在する場合、辞書名に沿った意味を付与する。それ以外については受け身と判別する。

#### 4.2 CRF を用いた手法

この手法では、作成したデータセットを学習データとしてモデルを構築し、「れる」「られる」の意味識別を行う。形態素解析器は MeCab を使用し、辞書は unidic-mecab 2.1.2 を用いた。ツールは CRF++ を使用した。判別に用いる素性は以下の 4 つである。

- 1: 「れる」「られる」とその前後の N 単語
- 2: 「れる」「られる」とその前後の N 単語の品詞
- 3: 「れる」「られる」の出現位置より以前に尊敬表現が含まれているか
- 4: 「れる」「られる」と直前の動詞の組み合わせが 4.1.1 で構築したそれぞれの辞書に存在するか

今回の実験で使用する素性の組み合わせを表 4 に挙げる。素性の欄に記述されている数字は上記の 4 つの素性と対応している。

なお,3. の尊敬表現は「様」や「氏」など,尊敬の意味を含む単語を指す。尊敬の対象となりうる単語が出現すると,「れる」「られる」が尊敬の意味で使われやすくなると考えたからである。尊敬表現の収集方法はまず,BCCWJより,「れる」「られる」が尊敬の意味で用いられる文をランダムで 100 文収集した。そして,それらの文から尊敬の意味を含むと思われる単語を収集した。

表 5 は本研究で使用した尊敬表現の一覧である。

表 4: CRF に入れる素性の組み合わせ

名称	素性
CRF(周辺文脈)	1+2
CRF+尊敬表現	1+2+3
CRF+辞書	1+2+4
CRF+尊敬表現 +辞書	1+2+3+4

表 5: 尊敬表現の一覧

氏 さん 様 家 方 殿 下 長 ご 御 患 者 お 天 皇 婦 人 教 授 先 生 将 軍 警 部 神 父 殿 神 講 師 親 族 ト ッ プ 嬢 妃 上 閣 下 大 統 領 だ す ま す
--

#### 4.3 ベースライン

本実験のベースラインとして,全て受け身と判別する手法とランダムでいずれかの意味を付与する手法を用いた。全て受け身と判別する手法は,BCCWJ で受け身が最も出現しやすい意味である [5] ため,採用した。

## 5 実験と考察

### 5.1 予備実験

CRF を用いた実験では,素性に使用する「れる」「られる」の前後の単語数を決める必要がある。そこでその単語数 N を 1 から 4 まで変化させた時の正解率を比較した。なお使用する素性は,「れる」「られる」の前後の N 単語とその品詞とした。評価については評価用のデータセット 4,000 文を無作為に 5 等分に分割し,交差検定を行った。そして,各意味での正解率と全体での正解率を算出した。

表 6: 前後の単語数 N を変化させた時の正解率

N	受け身	可能	尊敬	全体
1	<b>0.983</b>	<b>0.767</b>	0.220	<b>0.931</b>
2	0.981	0.743	0.261	0.929
3	0.980	0.754	<b>0.285</b>	0.930
4	0.980	0.757	0.252	0.928

表 6 は N を変化させた時の各意味,そして全体での正解率である。全体の正解率は N=1 の時で最も良くなることが分かる。しかし,本研究の目的は,換言によって「れる」「られる」の語義曖昧性を解消することである。そのため,可能,尊敬での正解率が重要である。これらのことを踏まえると,本研究では N=3 が最適であると言える。そのため,以降の実験では「れる」「られる」とその前後 3 単語を CRF の素性として利用した。

### 5.2 実験

辞書を用いた手法と CRF を用いた手法でそれぞれ自動判別した結果を評価する。辞書を用いた手法では,評価用のデータセット 4,000 文に含まれる「れる」「られる」について意味の自動判別を行った。そして,各意味での正解率と全体での正解率を算出した。ベースラインの手法についても辞書を用いた手法と同様に正解率を算出した。CRF を用いた手法では,予備実験と同じように,評価用のデータセット 4,000 文を無作為に 5 等分に分割し,交差検定を行った。CRF に入れる素性は表 4 で示した 4 つの組み合わせを使用する。

表 7 にそれぞれの判別手法での実験結果を示す。今回使用した手法はどれもベースラインの手法を上回っていることが確認できる。辞書による判別と CRF に

表 7: 判別手法の評価

	受け身	可能	尊敬	全体
受け身のみ	1.00	0.00	0.00	0.879
ランダム	0.335	0.312	0.347	0.334
辞書	<b>0.990</b>	0.579	0.329	0.928
CRF(周辺文脈)	0.980	0.754	0.286	0.930
+ 尊敬表現	0.983	0.740	0.345	0.934
+ 辞書	0.985	0.748	0.440	0.941
+ 尊敬表現 + 辞書	0.984	<b>0.766</b>	<b>0.471</b>	<b>0.943</b>

よる判別を比較すると、可能において CRF が辞書での結果に比べて正解率が 0.175 上昇している。辞書では「れる」「られる」より前の情報のみを使用しており、CRF では「れる」「られる」の周辺の情報を使用している。このことから可能の判別において「れる」「られる」より後の情報も有用であると考えられる。

CRF の実験結果では、尊敬表現の有無を素性として加えると、尊敬の正解率が上がったが、反対に可能の正解率が下がっている。これは学習データの偏りが影響していると考えられる。この手法では、「れる」「られる」の前に尊敬表現がある場合に、尊敬と判別されやすくなる。しかし、尊敬表現がない場合は、意味の出現頻度の偏りから、受け身と判別される可能性が高い。その結果、尊敬の正解率が下がったと言える。

一方で、CRF に辞書を素性として加えると、尊敬において正解率が 0.154 上昇している。CRF で尊敬の正解率が低い理由の一つとしてサ行変格活用の動詞の多様性がある。「来場する」や「出席する」といった、名詞に「する」が付いたサ変動詞に「れる」が付くと、「れる」が尊敬の意味を表す場合がある。サ変動詞には多くの組み合わせがあるため、学習データが少ないと CRF では正確に尊敬と判別できない。辞書には、サ変動詞も収録されている。そのため、学習データに無い動詞でも上手く判別出来たために正解率が改善したと考えられる。

また、CRF に尊敬表現と辞書の素性を同時に加えると、尊敬と可能の正解率がさらに改善した。これは尊敬表現で発生した問題点を辞書の情報で補完することができたためであると考えられる。

## 6 おわりに

本稿では、助動詞「れる」「られる」に意味を付与したデータセットを構築した。そして、辞書を用いた手法と CRF を用いた手法で「れる」「られる」の自動意味判別を行い、作成したデータセットで評価を行った。辞書と比較して、CRF では可能で正解率が 0.175 上回る結果となったことから、「れる」「られる」の直後の単語も意味の判別に有用であることが伺える。また、CRF に尊敬表現と辞書を素性として組み込むことで、可能と尊敬で改善が見られた。本研究では、CRF に尊敬表現と辞書を素性として組み込むことで尊敬と可能の正解率を改善することができたが、尊敬の正解率は受け身、可能に比べて低い。そこで、尊敬の判別に重要な情報を探して、尊敬の正解率の改善することに取り組みたい。

また、今回作成した「れる」「られる」に意味が付与されたデータセット及び自動判別に用いた辞書 4 種は公開する予定である。

## 参考文献

- [1] 山本和英, 三上侑城. 語義曖昧性解消としてのかな漢字換言システムの開発. 言語処理学会第 22 回年次大会発表論文集, pp. 180–183, 2016.
- [2] 竹野峻輔, 松田真希子, 梶原智之, 山本和英. 機械学習を用いた二格深層格の自動付与の検討. 言語処理学会第 20 回年次大会発表論文集, 2014.
- [3] 田辺利文, 吉村賢治, 首藤公昭. 格助詞「に」の深層格推定 - 格助詞の意味再考 -. 情報処理学会研究報告自然言語処理 (NL), Vol. 2008, No. 113, pp. 65–72, nov 2008.
- [4] 小山田由紀, 柏野和佳子, 前川喜久雄. 助動詞レル・ラレルへの意味アノテーション作業経過報告. 第 2 回コーパス日本語学ワークショップ予稿集, pp. 59–68, 2012.
- [5] 中俣尚己. 日本語教育のための文法コロケーションハンドブック, pp. 242–244. 株式会社 くるしお出版, 2014.