

## 格助詞「に」の深層格推定 ―クローズドテストによるモデル検証―

田辺利文<sup>†</sup> 吉村賢治<sup>‡</sup> 首藤公昭<sup>‡</sup><sup>†</sup>福岡大学工学部 <sup>‡</sup>福岡大学大学院工学研究科

{tanabe, yosimura, shudo}@tl.fukuoka-u.ac.jp

## 1. はじめに

論文誌「自然言語処理」巻頭言において、自然言語処理に関して、統計的な手法や機械学習手法等を用いて 8 割程度の精度・再現率は確保されるがそれ以上の精度を確保することは難しいことについて述べられている(森, 2007)。この原因の 1 つとして、インターネット上のテキストデータが爆発的に増大した(inflo-plosion, 2008)のに反して、現在の自然言語処理システムおよび方法論が、言語現象の多様性に追従できていないことがあげられる。言語現象の多様性をカバーするためには、個別の現象をきめ細かく扱える仕組みを用意しておくことが必要である。本論文では、言語現象の多様性の 1 つをカバーすることを目的とした、意味の多様性の側面をとらえることのできるモデルを提案する。

自然言語処理における意味解析においては Fillmore の格文法(Fillmore, 1975)がよく知られている。これは、文中の動詞に対して他の単語がどの役割(深層格)を付与されるかが決定される枠組みであり、深層格の種類を〈動作主格(A)〉〈経験者格(E)〉〈道具格(I)〉〈対象格(O)〉〈源泉格(S)〉〈目標格(G)〉〈場所格(L)〉〈時間格(T)〉の 8 種に分類している。日本語においては、深層格を決定するには「が」「を」「に」「で」「の」「から」「へ」「と」などの格助詞による係り受け関係を認識することが重要である。例えば『紙に書く』の意味として「紙という対象物に書く」があるが、『彼に書く』では「彼という対象物に書く」だけでなく「彼に宛てて手紙を書く」の意味も存在する。このように深層格を正確に認識するためには、係り受け関係にある述語だけでなく名詞の取り扱いが必要となる。また『大統領に選ばれる』の意味としては、「大統領として(誰かが)選ばれる」、「大統領によって(誰かが)選ばれる」、「大統領が選ぶことができる」の少なくとも 3 通りの解釈が存在するように、深層格を認識するためには名詞及び述語だけでなく「れる」「せる」などの文末表現を含めて取り扱うことが必要となる。しかし、日英翻訳ソフトに挙げられるような既存のシステムの出力結果から推測すると、既存のシステムは格助詞「に」の深層格を適切に認識しているとは考えにくい。また、(渋谷, 2006)では、頻出助詞である上位 5 位の「を」「に」「が」「は」「で」において深層格の推定を行っており、格助詞「に」の出現頻度は「を」に次いで高いこと、「に」の深層格推定は「で」に次いで困難であったことが示されている。これらのことから、格助詞「に」の深層格推定を精度良く行うことは意味処理を行う上で重要であると考え

られる。

本論文では、深層格の種類が比較的多いとされる格助詞「に」を対象にし、格助詞「に」を介した係り受け関係にある名詞と述部によって深層格を詳細に推定するモデルを提案する。

## 2. 関連研究

格助詞「に」の深層格としては、(石綿, 1999)では〈存在の場所(L)〉〈所有者(Pos/E)〉〈移動の着点(G)〉〈動作のゆきつく相手/対象(O)〉〈動作のゆきつく相手(P-A)〉〈動作が関係し、ゆきつくさきとしての相手(P-E)〉〈動作状態のゆきつく相手、対象(P-O)〉〈変化の結果(P-O)〉〈比較、異同判断の相手(P-O)〉〈動作状態のゆきつく、またはかわる受け身的な人(E)〉〈動作状態のゆきつく目的(Purp)〉〈原因(Caus)〉〈比較の基準・限界など(Comp, Lim)〉〈行為の時(Tim)〉の 14 種があるとしている。また(国立国語研究所, 1997)では、科学技術庁機械翻訳プロジェクト( $\mu$  プロジェクト)においては、〈対象〉〈受け手〉〈与え手〉〈相手 1〉〈時〉〈時・終点〉〈場所〉〈場所・終点〉〈終状態〉〈属性〉〈方式〉〈目的〉〈役割〉〈範囲規定〉〈比較〉の 15 種を定義している。(庭, 2004)では、〈目的地・到着点〉〈存在の場所〉〈対象〉〈相手〉〈恩人〉〈時点〉〈原因〉〈基準〉〈変化の結果〉〈使役の対象〉〈受身の「元の文」の動作の主体〉の 11 種があるとしており、〈使役の対象〉〈受身の「元の文」の動作の主体〉は、「せる」や「れる」のような文末表現により与えられるとしている。さらに(杉村, 2005)では、(1)存在の場所・時点の表示、(2)一方向性を持った動きの着点の表示、(3)被動的行為の動作主の表示、の 3 つの用法があるとし、それぞれを、(1)では〈存在の場所(位置)〉〈所有者〉〈能力の主体〉〈時間(時点)〉〈距離的な位置〉の 5 種、(2)では〈位置変化の着点〉〈移動の着点〉〈目的〉〈変化の結果〉〈行為の相手〉〈働きかけの対象〉〈受益者〉〈恩恵の受益者〉〈比較の基準〉〈精神行為の相手〉〈随伴の対象〉〈行為の対象〉の 12 種、また(3)では〈受身の対象〉〈授与者〉〈恩恵の授与者〉〈原因〉の 4 種としており、細分化されたものを合計すると 21 種の深層格を設定している。しかし、いずれの研究でも機械処理を行うために必要な格助詞「に」の深層格としては十分ではないと考えられる。

## 3. 格助詞「に」の深層格推定の概要

深層格推定のために作成した単語辞書、及び格助詞「に」の深層格推定の概要を説明する。

### 3.1 名詞属性と述部属性

(分類語彙表, 2004)の意味体系をベースとして、分類語彙表の見出し語に、人手で名詞属性 (N属性) または述部属性 (P属性) を付与したものを自立語辞書としてN属性自立語辞書及びP属性自立語辞書を作成した<sup>1</sup>。基本的には、分類語彙表における体の類に属する見出し語にはN属性を、用の類、相の類に属する見出し語にはP属性を与えている<sup>2</sup>。深層格を考慮し、N属性の種類は《人、人の組織性》《有意味・有機能性》《場所・位置・方向性》《サ変性》《属性》《催物性》《個数性》《一般数量性》《時間量性》《生産物・用具性》《観点・注目点・方向・意味性》《状態性》《情報性》《役割・道具・方法・用途性》《時点性》《時間的領域性》《序列性》《時・場合・際性》《心理現象》《自然現象》《社会現象》の21種、P属性の種類は《自動詞性》《他動詞性》《形容詞性》《形容動詞性》《精神・行為性》《「が」対象性》《使役性1》《使役性2》《受益性1》《受益性2》《恩恵付与性》《可能性》《変化・作用性》《変化性》《設定・仮定性》《設定存在性》《困難性》《容易性》《受動性》《「に」動作対象性》《「に」授与着点性》《「に」受容起点性》《情報伝達性》《「に」=「にとって」, 「にたいして」対象性》《「に」=「として」対象性》《「と」対象性》《「と」内容・引用対象性》《「を」場所・時間性》《「に」移動方向着点性》《「に」存在場所性》《「に」出現創出場所性》《「に」拡散場所性》《「に」所有含有場所性》《「に」接近接触離脱対象性》《「に」参加参入場所性》《「に」状態変化原因性》《「に」心理状態変化原因性》《「に」状態変化着点性》《「に」認識場所性》《「に」認識内容性》《「に」獲得受容生産選定性》《「に」動作原因性》《消費使用性》《充足不足優劣劣勢性》《「で」終了時点性》《「に」決定固定内容性》の46種と詳細に設定した。

### 3.2 属性付与

入力文に対して構文解析を行い格助詞「に」の係り受け関係にある名詞および述部を抽出し、名詞と述部それぞれに属性を付与する。

#### 3.2.1 名詞属性付与

抽出された名詞 (名詞句) がN属性自立語辞書に存在した場合にはその名詞 (名詞句) の属性を付与する。名詞句がN属性自立語辞書に存在しなかった場合には、名詞句の主辞 (head) の属性を名詞句の属性とする<sup>3</sup>。さらに、「陸上競技」の「競技」や「10個」の「個」など単語列の末尾に存在することで、N属性を付加できる単語がある。このように単語列の末尾に存在する、特定の単語の存在により属性を付加する処理を末尾処理とする。末尾処理を行ってもN属性が付与できなかった場合には、どの属性も存在する可能性が有るため、末尾処理によって付与される属

性以外の全てのN属性を付与する。

#### 3.2.2 述部属性付与

述部先頭に位置する用言の基本形をP属性自立語辞書から検索しP属性を付与する。次に、述部に存在する特定の文末表現全てに対して対応するP属性を付加する、文末表現処理を行う。ここでの文末表現とは、「せる」「ようにさせる」「てもらう」「てくれる」「れる」「られる」「ずらい」「足りない」「つく」「かかる」のような、用言に後接することで特定のP属性を付与できる表現を表す。文末表現処理を行ってもP属性が付与できなかった場合には、格関係を変更しうる文末表現によって付与される属性以外の全てのP属性を付与する。

### 3.3 深層格推定

本論文で提案する格助詞「に」の深層格は、(奥田, 1983)の分類を参考として、独自に〈起点性〉〈対象性〉〈時性〉〈場所性〉〈目標・方向性〉の5種を基本深層格として設定し、機械処理で扱いやすい体系の作成を目的にそれぞれをより精密にし43種の深層格を設定した。表1に、格助詞「に」の深層格の全リストを示す。

また(奥田, 1983)では、格助詞「に」の深層格を〈対象的なむすびつき〉〈規定的なむすびつき〉〈時間的なむすびつき〉〈原因のむすびつき〉の4種に分け、それぞれのむすびつきを成り立たせる名詞と動詞の種類を示している。本論文で提案する格助詞「に」の深層格推定は奥田氏の考え方に基づき、格助詞「に」の深層格ごとに必要なN属性とP属性の組を定義した。表2に、〈被害文の対象(動作者)〉と〈受動文の対象(動作者)〉と〈可能文の主体(可能者)〉の深層格が推定されるための、それぞれのN属性およびP属性の条件を示す。ここで〈被害文の対象(動作者)〉が推定されるには、N属性は《有意味・有機能性》であり、かつ、P属性は《他動詞性》または《自動詞性》、かつ《受動性》でなければならないことを示している<sup>4</sup>。格助詞「に」の深層格の中には、N属性、P属性のいずれかの条件を満たすだけで決定されるものもある。例えば、表2において〈受動文の対象(動作者)〉が推定されるにはP属性が《受動性》かつ《他動詞性》であればよい<sup>5</sup>。

推定された格助詞「に」の深層格が複数になることもある。文『犬に食べられる』は表2で挙げている全ての深層格が推定される例である。

<sup>1</sup> (IPADIC, 2008)に収録されている動詞の一部も抽出し属性を付与している。

<sup>2</sup> N属性、P属性は、選択制限を行う際の意味マーカー(長尾, 1996)に相当する。

<sup>3</sup> 例えば、「福岡大学」がN属性自立語辞書に存在せず「大学」が存在した場合には、「福岡大学」のN属性として「大学」のN属性を付与する。

<sup>4</sup> 被害文とは一般的には迷惑受身と呼ばれる文で、『雨に降られる』など自動詞に「れる」が後接しているのが特徴であるが、『私は魚を猫に食われる』のような文も被害文となる。

<sup>5</sup> 格助詞「に」の深層格は、名詞のN属性、述部のP属性により推定される。これは、句全体の意味は、句を構成する単語の意味を組み合わせることで求めることができる、言い換えると、構成性原理が成り立つことを前提とした深層格推定モデルである。そのため『水に流す』の解釈の1つである「なかったことにする」は、構成性原理が成り立たない解釈なので推定できない。

表1 格助詞「に」の深層格

基本 深層 格	格助詞「に」の深層格	場所 性	〈可能文の主体(可能者)〉, 〈困難文の主体(困難者)〉, 〈容易文の主体(容易者)〉, 〈充足不足優勢劣勢文の主体〉, 〈原因・理由〉, 〈存在空間的場所〉, 〈数量的場〉, 〈属性,所有物,抽象物の所有者(存在場所)〉, 〈属性の帰属領域(場所)〉, 〈出現創出・拡散場所〉, 〈所有含有場所〉, 〈認識対象のあり場所〉, 〈場合〉, 〈序列〉, 〈役割・用途〉, 〈個数による割合〉, 〈一般数量による割合〉
起 点 性	〈受動文の対象(動作者)〉, 〈受益文1の対象(動作者)〉, 〈被害文の対象(動作者)〉, 〈受容文の対象(授与者)〉, 〈情報受信文の対象(情報発信者(物))〉	目 標 ・ 方 向 性	〈授与文の対象(受容者)〉, 〈受容文の着点〉, 〈情報発信文の対象(情報受信者)〉, 〈決定・固定内容〉, 〈参加参加場所〉, 〈行為,活動の目的——一部1〉, 〈行為,活動の目的——一部2〉, 〈接近接触離脱対象(場所)〉, 〈移動方向着点〉, 〈変化の結果(着点)〉, 〈恩恵付与文2の受益者〉
対 象 性	〈使役文2の対象(動作者)〉, 〈動作の対象(対象2)〉, 〈使役文1の対象(動作者)〉, 〈恩恵付与文1の対象(受領者)〉, 〈受益文2の対象(受領者)〉, 〈対象6(一般形容詞形容動詞文の対象)〉, 〈対象7(一般形容詞形容動詞文の対象)〉, 〈認識内容〉		
時 性	〈時点〉, 〈時点(幅つき)〉		

表2 格助詞「に」の深層格を推定するためのN属性およびP属性の条件例

格助詞「に」の深層格	N属性条件	P属性条件
被害文の対象(動作者)	有意思・有機能性	受動性 & (自動詞性 + 他動詞性)
受動文の対象(動作者)		受動性 & 他動詞性
可能文の主体(可能者)	有意思・有機能性	可能性

ここで「犬」は《有意思・有機能性》のN属性を持ち、「食べる」は《他動詞性》、文末表現「られる」は受身および可能をあらわす助動詞であるから属性《受動性》《可能性》が付与され、「食べられる」のP属性は《他動詞性》《受動性》《可能性》の3種を持つことになる。よって、『犬に食べられる』の深層格は、「犬」のN属性および「食べられる」のP属性から〈被害文の対象(動作者)〉〈受動文の対象(動作者)〉〈可能文の主体(可能者)〉の3種が推定されることになる。

#### 4. 予備実験

格助詞「に」の係り受け関係の深層格推定システムを評価するために予備的実験を行った。格助詞「に」を介した係り受け関係が1文に1箇所だけ存在する453文を対象とする。入力文は(cabocha, 2008)で構文解析を行い、格助詞「に」を介した係り受け関係にある名詞(名詞句)と述部のペアを、深層格推定システムの入力とした<sup>6</sup>。これらの文は、深層格推定モデルの妥当性を検証するため人手で選定した。

##### 4.1 実験内容

実験を行った453文における、格助詞「に」を介した係り受け関係のそれぞれに対して人手で正解であると定義した深層格の総数をA、本システムが出力した深層格のうち正解であった深層格の総数をB、本システムが453文に対して出力した深層格の総数をCとする。ここで再現率 $R_r$ 、適合率 $R_p$ 、F値をそれぞれ次のように定義した。

$$R_r = B / A * 100 (\%)$$

$$R_p = B / C * 100 (\%)$$

$$F = 2 * R_r * R_p / (R_r + R_p) (\%)$$

#### 4.2 実験結果及び考察

実験結果を表3に示す。

表3 実験結果

正解と思われる深層格の総数: A	675
本システムが出力した深層格のうち正解の総数: B	639
本システムが出力した深層格の総数: C	753
入力文数: D	453
不正解を含んだ文の数: E	89
再現率: $R_r$ (%)	94.7
適合率: $R_p$ (%)	84.9
F値: F (%)	89.4

深層格が43種類と詳細であるにもかかわらず、適合率は約85%、再現率に至っては約95%となり、F値も約90%に達した。この結果から、入力文は人手で選定したものの、提案する深層格推定モデルは妥当であると考えられる。一方、深層格を過剰に出力した文は89文存在した。この原因として、ひらがな文字が入力に存在することが考えられる。例えば、用言の「かける」において、P属性自立語辞書のP属性付与の際、本来の見出しであるはずの「掛ける」以外に、「書ける」「欠ける」などの見出しにもマッチしてしまい、それらの見出しの属性も付与するため、過剰な属性付与が行われた結果、過剰に深層格を推定していることが考えられる。

<sup>6</sup> これら453文は、すべて構成性原理が成り立っている文であり、cabochaが解析を誤ったと考えられる文は含まない。

## 5. おわりに

本論文では、深層格の種類が比較的多いとされる格助詞「に」を対象にし、格助詞「に」を介した係り受け関係にある名詞と、文末表現をまとめた述部を適切に取り扱うことで、深層格を詳細に推定できるモデルを提案した。

これからの自然言語処理システムに求められることは、言語現象の多様性をできるだけカバーし得ることであると考えている。そのためには、形態素解析や構文解析などの自然言語処理の前段において推測しうる意味は可能なだけ抽出し、後段の処理で意味を絞り込むアプローチを採用することが、システム構築上メリットは大きいものと考えており、特に、前段側で高い再現率を確保することは必要不可欠であると考えている。格フレームによる意味記述は多項の係り受け関係をベースにしているのが一般的であるが、本論文では、係り受け関係にある名詞と述部の2項関係のみを使ったシンプルなモデルであり、格フレームに比べてより詳細な意味記述が可能であると考え、43種の深層格を設定し、再現率を重視したモデルを提案した。予備実験の結果、深層格の種類は43種と精細であるにも関わらず適合率は約85%に達し、さらに再現率は約95%が得られたことから、本論文で提案した深層格推定モデルはおおむね妥当であることを示している。

格助詞「に」の深層格推定の応用として、格助詞「に」を、同じ深層格をもつ他の格助詞相当表現に言い換えることがあげられる。例えば『彼に本を貰う』の意味として〈受容文の対象(授与者)〉が推定された場合には、『彼から本を貰う』と言い換えを行うことができる。特殊な用法などで稀な深層格が推定された場合には、格助詞の言い換えを行うことによって、日本語教育や、日英翻訳システムの出力の質の向上も期待できる<sup>7</sup>。他にも、英訳を試みる場合、例えば『犬に食べられる』の深層格として〈被害文の対象〉〈受動文の対象〉〈可能文の主体〉の3種が推定されることになるが、格助詞「に」に係る名詞をN、係り先の述語をP、日本語単語Xの英訳をE(X)、動詞Xの英語過去分詞形をE(X)-en とすると、〈被害文の対象〉の場合の英訳は“Subj have Obj E(P)-en by E(N)”、〈受動文の対象〉では“Subj be E(P)-en by E(N)”、〈可能文の主体〉では“E(N) can E(P) Obj”などと、トランスファー方式による日英機械翻訳を行う場合の変換規則を規定することができる<sup>8</sup>。

今後の課題をいくつか挙げる。予備実験の結果の考察でも述べたが、ひらがな文字に対する属性付与処理の検討が挙げられる。この対策としては、例えば、述語がひらがなであった場合

には名詞のN属性を使い、述部に付与するP属性の絞込みを行うことなどが考えられる。予備実験で使用した文は人手で選定したものであり、実験はクローズドテストに位置づけられる。今後はオープンテストを実施し性能を再確認する必要がある。特に、N属性自立語辞書には固有名詞を含んでいないため、オープンテストを行うには、固有名詞対策が必要であり、現在、cabochaが出力する品詞情報に依存したN属性を付与することを考えている。また格助詞「で」など、深層格の推定が困難と思われる場合についても、本モデルは有効であると考えている。さらに文中に複数の格助詞が存在し、それぞれの深層格が複数推定されるような場合には、深層格の優先順位の設定および係りの種類の非重複性の制約を組み合わせることで、深層格の曖昧さが軽減できると考えている。また、本モデルは構成性原理が成り立つ表現のみを取り扱うため、構成性原理が成り立たないような表現は別途データベースとして登録(田辺ら, 2008)し例外処理を行うことで、意味解析のさらなる精度向上ができると考えている。

## 参考文献

- 分類語彙表. 2004. 国立国語研究所. 大日本図書
- cabocha. 2008. <http://chasen.org/~taku/software/cabocha/>
- Charles J. Fillmore (田中春美, 船城道雄 訳). 1975. 格文法の原理. 三省堂
- google 翻訳. 2008. [http://translate.google.co.jp/translate\\_t](http://translate.google.co.jp/translate_t)
- info-plosion 情報爆発. 2008. 情報爆発時代に向けた新しい IT 基盤技術の研究 (文部科学省科学研究費補助金「特定領域研究」) <http://www.infoplosion.nii.ac.jp/info-plosion/>
- IPADIC. 2008. <http://sourceforge.jp/projects/ipadic/>
- 石綿敏雄. 1999. 現代言語理論と格. pp135-137. ひつじ書房
- 国立国語研究所. 1997. 日本語における表層格と深層格の対応関係. 国立国語研究所報告 113. 三省堂
- 森辰則. 2007. 末広がり (巻頭言). 自然言語処理. 第 14 巻第 1 号. pp1-2.
- 長尾真. 1996. 自然言語処理. 岩波講座ソフトウェア科学 15. pp199-230.
- 庭三郎. 2004. 庭三郎の現代日本語文法概説. 7.格助詞のまとめ. 7.3 に. <http://www.geocities.jp/niwasaburoo/07kakujosi.html#7.3>.
- 奥田靖雄. 1983. に格の名詞と動詞とのくみあわせ. 言語学会(編). 日本語文法・連語論(資料編). pp281-323. むぎ書房.
- 渋木英潔 荒木健治 桃内佳雄 柄内香次. 2006. 単語概念の深層格選考に基づく深層格推測手法. 電子情報通信学会論文誌 D. Vol. J89-D No.6. pp1413-1428
- 杉村泰. 2005. イメージで教える日本語の格助詞と構文. 名古屋大学言語文化論集. 第 27 巻第 1 号. pp49-62
- 田辺利文 吉村賢治 首藤公昭. 2008. 日本語連語候補に対する情報付与について. 福岡大学工学集報第 80 号. pp21-26

<sup>7</sup> 『私は彼に車をもらった』及び『私は彼から車をもらった』をそれぞれ (google 翻訳, 2008) の日本語入力としたところ、それぞれ “I gave him a car.” “I got a car from him.” となり、格助詞の言い換えによる訳質の向上が確認できた。

<sup>8</sup> ここで Subj は入力日本語における主語の英訳、Obj は入力日本語における目的語の英訳である。