日本語深層格の自動抽出のためのコーパス開発

松田　真希子*　　森　篤嗣**　　川村　よし子***
庵 功雄****　　山口昌也*****　　山本和英******

*金沢大学, **帝塚山大学, ***東京国際大学,
****一橋大学, *****国立国語研究所, ******長岡技術科学大学

mts@staff.kanazawa-u.ac.jp,moria24@gmail.com, kawamura@tiu.ac.jp,
isaoiori@courante.plala.or.jp, masaya@ninjal.ac.jp, yamamoto@jnlp.org

1　はじめに

文の解析において，深層格は重要な役割を果たしている．意味情報を正確に獲得するには深層格の理解が不可欠である．また，テキスト間の含意関係の判定[1]や機械翻訳の精度向上等，深層格の自動判定技術の応用範囲は広い．

名詞と述語の関係のあり方は表層格である格助詞が規定しているが，その一方で，当該格助詞の有する深層格は名詞とその述語によって決定されているという側面もある．そのため，過去の研究も表層格の前後の情報から深層格の自動判定を試みている[2][3][4]．しかし，その精度判定はEDR の概念関係子との照合で行われることが多い．ところが EDR が提供する深層格タグは網羅性に問題があり，[5]のような修正を施して用いられているというのが現状である．本研究ではより網羅的で妥当性の高い深層格タグの提案を目指し，深層格タグと日本語語彙体系の意味属性体系を人手で付与したコーパスを開発している．さらに，それらの情報がどの程度深層格判定に有効かを，格助詞のニを例に検討する．

2　深層格の自動判定

深層格の自動判定の研究には[2][3][4]等がある．[3]では，文の表層に現れる格助詞とそれに置換されうる語句のパターンに基づいて動詞を細かく分類する格パターン分析という手法が提案されている．しかし，助詞の深層格に関してはEDRの「概念関係子」の分類に準拠しているため，任意に付加される格に対する分類が十分とは言えず，日本語学で分類されている深層格が網羅されていない．一方日本語学分野でも深層格についての研究としては[7]をはじめいくつか存在するが，最終的な共通見解は得られていない．

3　研究手法

本研究ではタグリストの作成にあたり，[7]を採用した．[7]は日本語学分野で提案された助詞「ニ」の深層格リストで，EDR の深層格よりも詳細である．このリストに基づいて実際にタグ付けを行う過程でタグリストの問題点を検討できると考えた．

コーパスについては BCCWJ[9]を使用し，受け側単語の品詞が動詞，係り側単語の品詞が名詞となる例をすべて抽出した．今回抽出したニ格を伴う文は 57,807 例であった．そのうち 500 例を無作為抽出し，[7]の深層格リストをもとに「ニ」の深層格を人出で判定し入力した．助詞の前に位置する名詞については日本語語彙体系[8]に基づき，意味属性の番号を付与した．[8] は 30 万語の収録語を 3000 種の意味分類を用いて定義している．[7]の深層格リストと[8]の意味属性体系を表 1, 2 に示す．

表1　助詞「ニ」の深層格リスト[7]

目的	Purpose	見送りに来る
相手1	Person to Object	かぼちゃは栄養に富む
相手2	Person to Agent	コンサルタントに相談する
		こどもに菓子を与える
		メアリは親に泣きついた
原因	Cause	酒に酔う
	Cause	結果に失望する
限度・基準	Limit (Basis)	基準に足りない
対象	Object	信頼の回復につとめる
		私はこの結果に満足だ
		メアリは数学に強い
着点	Goal	下田に着く
動作主	Agent	先生にしかられる
時	Time	三時に知らせが入った
場所	Location	研究室に助手がいる
		ここに幸あり
範囲	Range	勝負に負ける
		わくにあてはまる
		矢がよろいにあたった
方向	Direction	東にむかう道をたどる

表2　日本語語彙体系の意味属性体系[8]

1 名詞
　2 具体
　　3 主体
　　4 人
　　362 組織
　　388 場所
　　389 施設
　　458 地域
　　468 自然
　　533 具体物
　　534 生物
　　706 無生物
　1000 抽象
　　1001 抽象物
　　1002 抽象物(精神)
　　1154 抽象物(行為)
　　1235 事
　　1236 人間活動
　　2054 事象
　　2304 自然現象
　　2422 抽象的関係
　　2423 存在
　　2432 類・系
　　2443 関連

表3にタグ付けコーパスの例を示す．文例（1）と（4）は同じ「人」という単語だが，文脈から（1）は一般的な人間としての「人」（5），（4）は相手となるため他人としての「人」(33)となる．

深層格付与が困難な例については unknown とした．

表3　タグ付きコーパス例[i]

文例	語彙体系	深層格
（1）ん|恋愛|は|すべて|の|パワー|に|なっ|て|くれ|ます|好き|な|ひと|が|い|て|，|その|(ひと)|**に**|(愛さ)|れ|てる|って|ベース|が|あれ|ば|，|たいてい|の|こと|は|楽しく|なる|もの|です|	5 人間	Agent
（2）者|→|ビキナー|に|アドバイス|ファースト|ピアス|　|ココ|が|ポイント|　|ピアス|って|軽く|見|すぎる|と|痛い|(目)|**に**|(あう)|けれど|，|必要|以上|に|こわ|がる|こと|は|ナシ|．|正しい|知識|を|身|に|つけよう|ね|	2608 程度	Object
（3）ん|．|5|，|六十|番|と|いう|ところ|です|．|ところ|が|，|実力|テスト|と|なる|と|(急)|**に**|(上がる)|の|です|．|しかも|，|一|ケタ|台|に|なっ|た|こと|も|あり|ます|．|夏休み|や|	2710 新旧 遅速	(Condition)
（4）．|林|　|ほんと|です|か|．|周防|　|そう|．|僕|が|書い|た|シナリオ|を|ポン|と|(人)|**に**|(預ける)|と|，|あらゆる|スタッフ|が|自分|で|考え|て|具体|的|な|もの|を|返し|て|くれる|．|	33 他人	Person to Agent

*Condition は新たに追加したタグ

4 結果と考察

4.1 深層格付与について

BCCWJ 上にある「名詞＋ニ＋動詞」の 500 例について日本語教育の専門家 3 名でタグを付与した結果，問題があることが明らかになった.

まず，[7]では分類できない深層格が見られた. 図 1 に深層格毎の出現比率を示す. 最も多かったのは，「本当に死ぬ」「実際にやっている」のような動詞の様態を「名詞（形動）＋ニ」で接続する副詞的用法や，「赤字になる」のように動詞の状態を表す用法(27%)であった. ところが，これについては深層格が設定されていなかった.

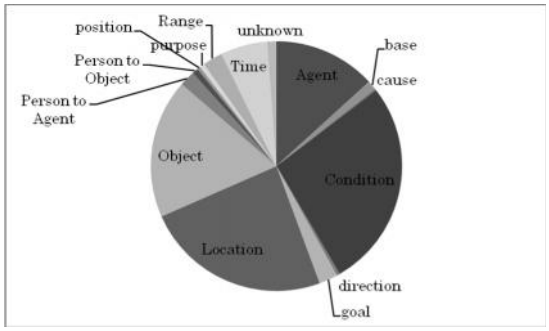

図 1　深層格の出現割合

また，Location では次のような例があった.

1) マンション(447:居住施設) に（Location）いる
2) 立場(2537:立場) に(Location?→Position)いる
3) 横(2645:側）に(Location)立つ
4) 横(2650:向き）に(Location?→Condition)寝かせる

このうち 2)のような例を Location と判定するには問題がある. 状態を表す名詞が前接する場合においては Position のような深層格を立てる必要性がありそうだ. さらに 3), 4)においても, 4)の「横」は場所を意味しないため，Location と判定するのにも疑問がもたれ，新たに加えた Condition という深層格に分類した.

また[7]では Object と Person to Object の二つの深層格が設定されているが，この二者の区別は明確ではなく, 二つにわける必然性が不明である.

Agent については主として動詞に助動詞「れる」「られる」が後続しているときに付与される. そのため後続の動詞だけではなく助動詞も含めて判定する必要があることが分かった.

また，前接の名詞については形態素解析で複合語が分割されてしまうこともあり，5)のように直前のの形態素だけでは深層格の判定ができない場合もいくつか見られた.

5) ララさん|が|不|(用意)|に|(言っ)|た|こと|が|ラン|さん|を|立腹|さ|せ|て|しまい|,

4.2 意味属性付与について

4.2.1 同一の深層格の場合の意味属性

Goal については, 深層格の意味範囲が不明瞭であったため，471（陸>具体物）から 2604（限り>抽象的関係）まで前接の名詞の範囲が非常に広く限定できないという結果となった. また Goal を目的地のみとするか，変化の結果も含むのかについて作業者間で判断の相違があった.

Location については，基本は「上」「下」等の抽象的な位置を表す 2600 番台の前半の名詞多く出現した. はずれているものには特殊な用法と固有名詞が多いが，動詞の種類によっては，かなり広範囲の名詞（ただし無生物）が用いられていた.

Time については「時」「日」等，2600 番台の後半しか出現しなかったため，前接の名詞の属性で分類を行うことが可能である.

4.2.1 同一の動詞の場合の深層格のバリエーションと意味属性

同一の動詞と表層格に複数の深層格が認められる場合の名詞の意味属性を調べたところ，いくつかの動詞については，名詞の意味属性と深層格の異なりに一定の傾向が認められた. 一例として「来る」の例を示す.

6) 家(447:居住施設>具体物）に(Location)来る
7) 品目（1120:目録>抽象物）に(Location)来る
8) 見送り（1708:見送り>人間活動）に(Purpose)来る
9) 相談(1514:会談>人間活動)に(Purpose)来る
10) 応援(1816:援助>人間活動)に(Purpose)来る
11）絶対(2465:対応>抽象的関係)に(Condition)

来る

このように，Location の場合は具体・抽象の差はあれ「物」に，Purpose の場合は「人間活動」に，Condition の場合は「抽象的関係」となっている．

4.2.2 意味が拡張した用法の深層格判定

前接の名詞が基本義から拡張しているため，深層格の判定が難しい場合が見られた．たとえば「上に立つ」の「上」については，「上」だけを見れば Location であるが，役目として「上に立つ」という意味の例文であれば，Condition となる．

その他，「口にする」「手にする」「気になる」「力になる」「耳に入る」「手に入る」等もの慣用句は深層格の判断が難しく，今回は unknown で処理した．こうした意味が拡張した場合の深層格付与については今後検討する必要がある．

5　まとめと今後の課題

今回は 500 の例文についてタグ付けを行い，深層格判定を試みた．その結果，同じ深層格の場合は，深層格の自動抽出の際に前の名詞のカテゴリーの情報が与えられれば深層格が分類できそうなもの（Location, Time）もあるが，そうでないものも多くみられ，より詳細な検討が必要であることが分かった．

また，同一の動詞で深層格が異なる場合については，名詞の意味属性である程度自動的に判定可能である可能性を指摘した．

今回は500例という限られた例文のみで検討を行ったが，生のデータをもとに深層格付与を行うことで，既存の深層格リストの問題点も明らかになった．深層格リストを整備したうえで，残りのデータに関してもタグ付けを行い，深層格リストを確定する予定である．

深層格は前接の名詞（体言）と後続の動詞・形容詞（用言）の組み合わせで意味が決定されるため広範囲，かつ，詳細に検討するほど深層格の種類も増えていく可能性がある．そのため，単に細かく分類するのではなく，その異なりが言語処理をはじめとする他分野に応用する際に有意な差と認識される異なりを見つける作業も必要である．

今後は，さらに他の助詞の深層格についても同様の作業を行い，より言語学的，言語処理的に妥当性の高い深層格タグリストを設定し，タグ付け方法の提案を行いたい．

謝辞

本研究は科学研究費補助金基盤研究（B）[課題番号 23320105]の助成を受けて行われた．

参考文献

[1] 梅基宏,杉原大悟,大熊智子,増市博. LFG 解析と語彙資源を利用した日本語含意関係判定, 情報処理学会研究報告. 自然言語処理研究会報告 2008(113), 57-64, 2008.

[2] 渋木英潔, 荒木健治, 桃内佳雄, 栃内香次 . 単語概念の深層格選好に基づく深層格推測手法，電子情報通信学会論文誌．J89-D(6), 1413-1428, 2006.

[3]大石亨, 松本裕治: 格パターン分析に基づく動詞の語彙知識獲得，情報処理学会論文誌, Vol.36, No.11,pp.2597–2610 , 1995.

[4] 小山正太, 乾伸雄, 小谷善行:「名詞と表層格」パターンに対する深層格対応の推測, 情報処理学会研究報告, NL-154-22, 2003.

[5]意味解析システム Aya
http://www.jsa.co.jp/contents/GG/group/Aya/aya.htm#siryo

[6]国立国語研究所.分類語彙表.秀英出版,1993

[7]城田俊.日本語形態論.ひつじ書房, 2002

[8] NTT コミュニケーション科学基礎研究所　監修，日本語語彙大系　CD－ROM 版 1999

[9] 国立国語研究所『現代日本語書き言葉均衡コーパス』（BCCWJ）

---

[i] *｜｜は単語境界をあらわす．(　)は対象となる品詞（名詞と動詞）を現す．紙面の都合により，実際のコーパスとは異なったレイアウトで提示している．