

ユーザとの対話による深層格獲得システム

正富 欣之 荒木 健治 栃内 香次

北海道大学大学院工学研究科

〒 060-8628 札幌市北区北 13 条西 8 丁目

TEL: 011-706-6823 FAX: 011-709-6277

E-mail: {tome,araki,tochinai}@media.eng.hokudai.ac.jp

あらまし 自然言語処理において、格構造は文の意味を表示するために必要不可欠なものとして扱われ、文の解析においても動詞とそのとりうる格との意味的關係を表す深層格が重要な役割を果たす。深層格を用いて意味解析を行なう時に問題となるのが深層格の曖昧性である。深層格の種類や分類はユーザによって異なるものであると考えられる。ユーザのゆらぎに対応するシステムを構築するために、システムとユーザとの対話により深層格を獲得する。対話によって得られた結果を用いることにより、ユーザに適応した深層格による意味解析システムが生成される。本稿では上述の手法を提案する。また日本語文 1,200 文について深層格獲得処理を行ない、日本語文 400 文について本システムの評価実験を行なって、システムの有効性を示す。

キーワード 深層格、ユーザのゆらぎ、対話、意味解析

Deep Case Acquisition System by Dialogue with Users

Yoshiyuki MASATOMI, Kenji ARAKI and Koji TOCHINAI

Graduate School of Engineering, Hokkaido University

N13 W8, Kita-ku, Sapporo, 060-8628 Japan

TEL: +81-011-706-6823 FAX: +81-011-709-6277

E-mail: {tome,araki,tochinai}@media.eng.hokudai.ac.jp

Abstract In natural language processing, case frames are treated as essential in expressing the meaning of sentences. Deep cases expressing semantic relation between a verb and its case are important for the analysis of sentences. In the case of use deep cases for semantic analysis, it is an important problem that they have ambiguity. Each user has various classifications of deep cases. We acquire deep cases by dialogue with users to resolve user's ambiguity. In this report, we propose the deep case acquisition method by dialogue. Then, we carry out the experiment for evaluating the system based on our method.

Keywords Deep case, User's ambiguity, Dialogue, Semantic analysis

1 はじめに

文中において依存関係のある2つの語の間には何らかの意味的な関連がある。これらの意味関係を知るためには単語のみ、あるいは依存関係を示す表層上の表現だけを見るのでは不十分で、より深いレベルで意味関係を表す、格構造と呼ばれる意味の表現形式が用いられている。したがって、自然言語処理において、格構造は文の意味を表示するために必要不可欠なものとして扱われ、動詞とそのとりうる格との意味的關係を表す深層格が重要な役割を果たす。すなわち、表層構造で表しきれない意味のつながりを、深層構造に変換することにより表すことができるようになる。このことから自然言語処理分野では深層格を獲得する研究が行なわれている [1] [2]。

そこで、「深層格」の概念が文のあるレベルでの「意味」を表示するための基本的な考え方になっていることから、自然言語処理システムにおいて、意味解析として単語を深層格に分類することが必要である。しかし、「深層格」の概念は曖昧なものであり、どれだけの異なる格を用意し、どのような基準で認定するかについては一致した見解は得られていないのが現状である [3]。

また、実際の言語使用を見ると、すべての人の判断が一致する典型的な格関係と、文脈の存在や人によって判断が分かれるような格関係がある。特に、人によって判断が分かれるような格関係については、システムがあらかじめ用意された格関係により深層格に分類したのでは、ユーザとしての満足度は低いものとなってしまう。

この問題を解決するために、本稿で我々は、質問応答を用いた深層格決定手法を提案する。このシステムは日本語文を形態素解析、構文解析した情報を用いて、深層格に分類する。ここで、深層格に分類するのは助詞を伴う名詞と動詞の関係である。最初は、質問応答によりユーザが深層格情報を与える。質問応答を用いて深層格を決定するという処理により、ユーザ(ユーザ同士あるいはユーザ個人)のゆらぎに対応することができる。次に、質問応答により得られた深層格情報を用いて、新たな入力文における助詞を伴う名詞と動詞の関係についてシステムが深層格を決定する。このとき、新たな入力文中の名詞あるいは動詞の一方が獲得

されていない場合には、助詞の用法等を利用して深層格を決定する。以下、本手法にもとづく実験システムの概要とそれによる実験について述べる。実験では、まず最初に日本語文 1,200 文についての実験を行なって深層格情報を蓄え、次に評価実験を行なってシステムの有効性を検証する。

2 深層格について

Fillmore は語と語の間の意味関係を動詞を中心にとらえた格文法(case grammar) [4] を提案した。格という考え方は古くから存在するもので、主格や目的格、所有格などがある。これらの格は表層的に決まる格であり、表層格(surface case)とよばれるものである。格文法はこの考え方を意味的・深層的なものに拡張し、文中の語に対して他の単語がどのような役割、つまり深層格(deep case)をもつかということを考えたものである。

前節で述べたように、「格」の概念は曖昧なものであり、どれだけの異なる格を用意し、どのような基準で認定するかについての一致した見解はない。本研究では、助詞を伴う動詞と名詞の関係を深層格を用いて表すことを考える。そこで、名詞と動詞が助詞で結ばれているときの関係を表す深層格についての検討を行なった。文献 [2] の深層格を主として参考にした。

本研究では表 1 に示す 11 種類の深層格を用いている。

3 処理過程

3.1 概要

処理過程の概要を図 1 に示す。まず、入力された日本語文を形態素解析し、さらに構文解析する。次に、解析結果から得られた情報で、助詞を伴う名詞と動詞の関係が深層格決定ルールとして獲得されているかどうかの判定を行なう。深層格決定ルールは深層格獲得部での質問応答により獲得されるルールである。名詞・動詞の両方が深層格決定ルール中に獲得されていない場合には、深層格獲得部で処理を行ない、得られた結果の出力と深層格決定ルールへの追加を行なう。また、名詞あるいは動詞の少なくとも一方の関係が獲得されて

表 1: 深層格の種類

| 深層格 | 意味 |
|-----|---------------|
| 動作主 | 動作を行なう主体 |
| 対象 | 動作の対象となるもの |
| 場所 | 動作を行なう場所 |
| 道具 | 動作を実行するための道具 |
| 理由 | 動作を行なう理由 |
| 時 | 動作を行なう時 |
| 受益者 | 動作主から動作を受けるもの |
| 状態 | 動作の行なう状態 |
| 材料 | 動作を行なう材料 |
| 起点 | 動作を行なう起点となるもの |
| 終点 | 動作を行なう終点となるもの |

いる場合は、深層格決定部により深層格決定ルールに従って深層格を決定する。深層格決定部で深層格を決定できないものについては深層格獲得部で処理を行なう。以下では各処理の詳細について述べる。

3.2 形態素解析

本システムの形態素解析の処理については、日本語形態素解析システム JUMAN [6] を使用している。表 2 に形態素解析結果例を示す。形態素解析の結果から、それぞれの語の「読み」「原型」「品詞」の情報が得られる。なお、表 2 には示されていないが助詞でも「格助詞」「副助詞」というように「品詞」をさらに細かく分類した「品詞細分類」の情報も得られている。

3.3 構文解析

本システムにおける構文解析処理は、日本語構文解析システム KNP [7] を使用している。図 2 に構文解析結果例を示す。本構造により、「与える」という動詞に「私は」「彼に」「守るように」「忠告を」がかかっていることが分かる。また、「守るように」には「秘密を」が、「秘密を」には指示詞の「その」がそれぞれかかっている。これにより、助詞を伴う名詞と動詞の情報が得られる。

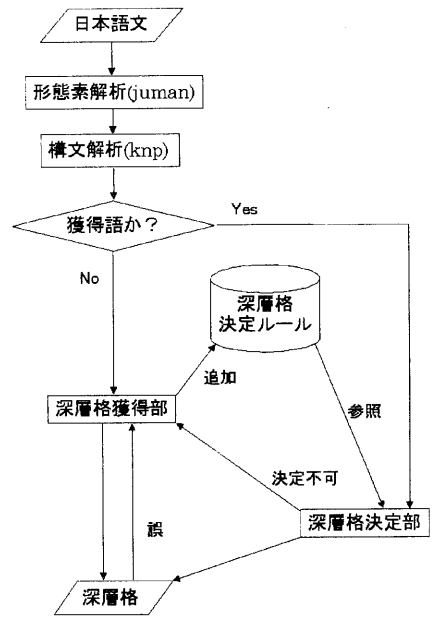


図 1: 処理過程

表 2: 形態素解析結果

| 原文 | 読み | 原型 | 品詞 |
|-----|-------|-----|-----|
| 私 | わたし | 私 | 名詞 |
| は | は | は | 助詞 |
| 彼 | かれ | 彼 | 名詞 |
| に | に | に | 助詞 |
| その | その | その | 指示詞 |
| 秘密 | ひみつ | 秘密 | 名詞 |
| を | を | を | 助詞 |
| 守る | まもる | 守る | 動詞 |
| ように | ように | ようだ | 助動詞 |
| 忠告 | ちゅうこく | 忠告 | 名詞 |
| を | を | を | 助詞 |
| 与えた | あたえた | 与える | 動詞 |
| 。 | 。 | 。 | 特殊 |

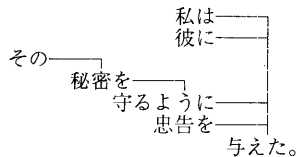


図 2: 構文解析結果

3.4 深層格決定ルール

「深層格決定ルール」とは、本研究で取り上げている助詞を伴う名詞と動詞の深層格を決定するためのルールである。本研究では、「名詞-助詞-動詞」という関係から深層格を決定できるものを対象としている。獲得されたルールは、名詞については [名詞-助詞-深層格]、動詞も同様に [動詞-助詞-深層格] という形式で蓄積される。深層格決定ルールの例を名詞について表 3 に、動詞については表 4 に示す。

表 3: 深層格決定ルール (名詞)

| 名詞 | 助詞:深層格 |
|-----|--|
| 私 | が:動作主 に:受益者 は:動作主 を:対象 |
| 彼 | が:動作主 に:受益者 は:動作主 を:対象 から:起点 |
| 計算機 | で:道具 を:対象 |
| 銀行 | に:対象 に:場所 |
| 昆虫 | は:動作主 を:対象 |
| 電話 | が:対象 は:対象 で:道具 を:対象 を:対象 |
| 道路 | を:対象 を:場所 で:場所 |
| 富士山 | が:場所 で:場所 に:場所 へ:終点 |
| 戦争 | で:原因 を:対象 |
| 駅 | で:場所 に:終点 を:場所 |

3.5 深層格獲得部

深層格獲得部では質問応答を用いることにより深層格を獲得する。ここでは、システムがユーザ

表 4: 深層格決定ルール (動詞)

| 動詞 | 助詞:深層格 |
|------|---|
| 与える | が:動作主 に:受益者 は:動作主 を:対象 |
| 渡る | で:道具 は:動作主 を:場所 |
| 焼く | で:道具 は:動作主 を:対象 |
| 食べる | が:動作主 に:場所 で:場所 で:道具 に:時 は:動作主 を:対象 |
| 掃除する | で:道具 は:動作主 を:場所 |
| 入る | が:動作主 に:場所 に:時 は:動作主 で:道具 |
| 折る | で:材料 は:動作主 を:対象 |
| 輸出する | に:終点 は:動作主 を:対象 |
| 動く | が:動作主 で:原因 で:道具 に:場所 に:時 は:動作主 |
| 植える | に:場所 は:動作主 を:対象 |

に対して、どの深層格であるか質問する。ユーザが深層格を応答することにより、その結果を出力するとともに得られた深層格を深層格決定ルール中に追加する。この処理を行なうのは、以下の3つの場合である。

1. 入力文中の助詞を伴う名詞と動詞の関係が名詞・動詞とも深層格決定ルール中に含まれない場合
2. 深層格決定ルール中にはあるが深層格決定部で決定できない場合
3. 結果として出力された深層格が誤っているとユーザが判断した場合

3.6 深層格決定部

深層格を決定するときに、すべての名詞と動詞の関係について質問応答により深層格を決定することは現実問題として非常に困難である。そこで、新しい入力文における名詞あるいは動詞の少なくとも一方の関係が深層格決定ルール中に獲得されている場合についても、システムが深層格決定ルー

ルおよびシソーラスを用いて深層格を決定する。深層格決定の手順を以下に示す。

1. 助詞と名詞あるいは助詞と動詞の関係から深層格決定ルールからとり得る深層格を選ぶ

- (1) とり得る深層格が1種類するとき → 3へ
- (2) とり得る深層格が2種類以上あるとき → 2へ

2. 分類語彙表を用いて未獲得の名詞あるいは動詞に近い獲得語を探す

- (1) 近い語があるときは近い語の深層格決定ルールを用いる → 3へ
- (2) 近い語がない場合には深層格決定不可能 → 4へ

3. 深層格を決定する

4. 深層格獲得部で処理を行なう

手順4でいう「近い」とは分類語彙表での意味的距離である。なお、分類語彙表ではそれぞれの単語に10桁の数を与えている。本研究では、この数字の上位5桁が一致しているものについて意味的距離が「近い」と判断している。

4 評価実験

初期状態では深層格決定ルールは空の状態である。入力文を与えていき、それに対してシステムが深層格についての質問を行ない、それに対してユーザーが答えることにより深層格決定ルールを獲得していく。ここでは、与えた入力文に対して、ユーザーが決定した深層格に基づく深層格決定ルールによる深層格決定の評価をするために、まず日本語文1,200文について実験を行ない、次に本システムの有効性を評価する実験を行なった。

4.1 実験方法

まず、日本語文1,200文を入力し、形態素解析および構文解析を行なった。得られた形態素解析結果と構文解析結果を用いて深層格獲得部による処理を行なった。なお、使用した日本語文は英和辞書 [8] 中の英文の訳文等である。また、形態素

解析、構文解析に誤りが含まれるものについては人手により修正を加えた。処理終了時には名詞・動詞各々について、表5に示されるような深層格決定ルールが得られた。

表5: 評価実験前の名詞・動詞の数

| | 深層格決定ルール中の数 |
|----|-------------|
| 動詞 | 548 |
| 名詞 | 1,041 |

上述の処理によって得られた深層格決定ルールを用いて、日本語文400文についての評価実験を行なった。使用した日本語文は英和辞書 [8] 中の英文の訳文等である。これらの文についても、まず形態素解析・構文解析の両解析を行ない、解析結果に誤りが含まれる場合には人手による修正を加えている。得られた形態素解析結果と構文解析結果を用いて本システムによる評価実験を行なった。なお、深層格決定の正誤については、ユーザーの判断によるものである。

4.2 実験結果

実験結果を表6~9に示す。表6におけるケースの「質問応答のみ」というのは、名詞、動詞ともに深層格決定ルール中にない場合である。この場合には、深層格決定部での処理を行わず、直接深層格獲得部での処理が行なわれたものである。「質問応答なし」というのは、名詞あるいは動詞の少なくとも一方が深層格決定ルール中に存在し、深層格決定部のみにより処理が行なわれた場合である。さらに、「質問応答あり」というのは、名詞あるいは動詞の少なくとも一方が深層格決定ルール中に存在し、深層格決定部により処理を行なったが、深層格の決定が不可能であったか、あるいは誤った結果を出力し、深層格獲得部での処理を行なった場合である。日本語文400文中に助詞を伴う名詞と動詞の組合せがのべ941種類あり、深層格決定ルール中に獲得されていない名詞と動詞の組合せは253種類(26.9%)であった。

次に表7であるが、これは深層格決定ルール中に獲得されている名詞あるいは動詞があるときの

正解率を示した表である。全体として得られた結果は正解率が66.2%となった。名詞または動詞の少なくとも一方が深層格決定ルール中に獲得されていたときの正解率を表8に示す。この場合の正解率は90.6%という結果が得られた。

最後に、日本語文400文についての評価実験終了後の深層格決定ルール中の名詞と動詞の数を表9に示す。評価実験前と比較すると、動詞が252種類、名詞が358種類増加した。

表 6: 評価実験結果 1

| ケース | 名詞と動詞の組合せ |
|--------|-----------|
| 質問応答のみ | 253 |
| 質問応答なし | 623 |
| 質問応答あり | 65 |
| 合計 | 941 |

表 7: 評価実験結果 2

| 深層格決定ルール | 正 | 誤 | 合計 | 正解率 (%) |
|----------|-----|-----|-----|---------|
| 動詞無名詞無 | 0 | 253 | 253 | 0 |
| 動詞有名詞無 | 142 | 38 | 180 | 78.9 |
| 動詞無名詞有 | 239 | 21 | 260 | 91.9 |
| 動詞有名詞有 | 242 | 6 | 248 | 97.6 |
| 合計 | 623 | 318 | 941 | 66.2 |

表 8: 評価実験結果 3

| 深層格決定ルール | 正 | 誤 | 合計 | 正解率 (%) |
|----------|-----|----|-----|---------|
| 一方有 | 623 | 65 | 688 | 90.6 |

4.3 考察

評価実験によって得られた結果により表6中の「質問応答あり」の場合についての考察を行なう。

表 9: 評価実験結果 4

| | 深層格決定ルール中の数 |
|----|-------------|
| 動詞 | 800 |
| 名詞 | 1,399 |

この結果は深層格決定ルール中に名詞または動詞が獲得されているが、最終的に深層格決定を行なうときに質問応答により深層格を決定したものである。深層格決定ができなかった理由としては、出力結果に誤りが含まれる場合と取り得る深層格が複数種類あって一つに決定できなかったことが挙げられる。以上の理由を1~5に細かく分類する。

1. ユーザのゆらぎによる深層格の曖昧性
2. 深層格獲得段階での入力ミス
3. 動詞または名詞の一方のみ深層格決定ルール中に含まれるときの判定誤り
4. 獲得された名詞あるいは動詞の中に入力された単語と意味的に近い語がシソーラス中に入らない場合
5. シソーラス中に判定したい単語が含まれない場合

1の例を図3に示す。この例では、「寝る」という動詞に対して、名詞「ベッド」に助詞「で」が付属している場合の深層格は「道具」であると図3.aの文を処理したときに決定されていた。したがって、新たな文図3.bという入力に対しても、出力された深層格結果は「道具」である。しかし、ユーザーはこのときの深層格は「道具」ではなく「場所」とあると判断し、訂正した。1、2については、出力された深層格の結果を見て、ユーザーが正解かそうでないかについての判断を下すことにより修正が加えられる。次に、3についての例を図4に示す。この例では、「出かける」という動詞は助詞「に」について深層格決定ルール中で深層格「終点」と図4.aの文の処理により獲得されていた。一方、図4.bの名詞「サイクリング」については未獲得であった。このとき、システムは動詞「出かける」についての深層格決定ルールから深層格を「終点」

a. 彼女はいつもベッドで寝る。
b. 彼は朝までベッドで寝た。

| | | |
|---------|----|----|
| 名詞 | 助詞 | 動詞 |
| ベッド | で | 寝る |
| 道具 → 場所 | | |

図 3: ユーザのゆらぎによる深層格の曖昧性

と決定した。しかし、この場合「サイクリング」は「終点」ではなく「対象」であるとユーザが誤りを訂正した。同様に、名詞が深層格決定ルール中に獲得されているが動詞が未獲得の場合についても、深層格決定が誤ることがあった。3 については誤りを正すことにより深層格決定ルールに新たに情報が追加される。4 の例を図 5 に示す。こ

a. 彼は飛行機で外国に出かけた。
b. 彼は自転車でサイクリングに出かけた。

| | | |
|---------|----|------|
| 名詞 | 助詞 | 動詞 |
| サイクリング | に | 出かける |
| 終点 → 対象 | | |

図 4: 動詞のみ深層格決定ルール中に含まれる時の誤り

の例では動詞「打つ」は深層格決定ルール中に獲得済みである。名詞「ハンマー」は未獲得であり、シソーラス中で「ハンマー」に「近い」語も未獲得である。したがって、システムは「道具」と「場所」の判定ができない。ただし、4 については入力文の増加により解消されていくと考えられる。5 の例を図 6 に示す。この例では名詞「ニューヨーク」は深層格決定ルール中に獲得されている。また、「ニューヨーク」の助詞「に」が付属する場合の深層格は「終点」または「場所」としてルール中に獲得されている。したがって、シソーラスを用いて動詞「出発する」に「近い」語を探して深層格を決定したいのだが、シソーラス中に「出発する」はない。よって、深層格決定ができない。このように、シソーラス中に含まれない語の主なもの、

● 彼はくぎをハンマーで打つ。

| | | |
|------------------|-----------|--------|
| 動詞 | 助詞「で」の深層格 | シソーラス中 |
| 打つ | 道具 or 場所 | |
| で「ハンマー」に「近い」語未獲得 | | |

図 5: 獲得された名詞あるいは動詞の中に入力された単語と意味的に近い語がシソーラス中にない場合

● 彼らはニューヨークに出発する。

| | | |
|-----------------|-----------|---|
| 名詞 | 助詞「に」の深層格 | シ |
| ニューヨーク | 終点 or 場所 | |
| ソーラス中に「出発する」はない | | |

図 6: シソーラス中に判定したい単語が含まれない場合

地名などの固有名詞やサ変動詞などである。したがって、5 の解決策としては、深層格決定ルールの単語の獲得データの変更や、シソーラスへの単語の追加などが考えられる。1~5 の結果について表 10 に詳細を示す。表 10 からシソーラスによる判定ができないものの数が多いことがわかる。このことから、質問応答をしないで正解率を上げるためには、シソーラスの使い方や質の向上などの対策が有効であると考えられる。

表 10: 誤りの分類

| case | 数 | 比率 (%) |
|------|----|--------|
| 1 | 5 | 7.7 |
| 2 | 5 | 7.7 |
| 3 | 10 | 15.4 |
| 4 | 21 | 32.3 |
| 5 | 24 | 36.9 |
| 合計 | 65 | 100.0 |

5 おわりに

本稿では質問応答を用いた深層格分類システムを提案した。このシステムを用いて深層格獲得処理を行ない、得られた結果を用いてシステムの評価実験を行なった。

評価実験の結果から、助詞を伴う名詞と動詞の組合せで一方の単語が獲得されていない場合についての深層格決定におけるシステムの有効性が示された。誤りや判定が不可能となる主な原因は、シソーラス中に判定したい単語が含まれないことである。

今後の研究課題としては、本システムでユーザのゆらぎを吸収できるかについての検証を行なうことを予定している。また、深層格決定ルール中に獲得した名詞も動詞もないときの表6「質問応答のみ」の場合での深層格決定についても、シソーラスや語の周辺情報から決定できないか検討中である。さらに、誤りの原因となるシソーラス中にない語については、得られた深層格情報からシソーラスへの語の追加によるシソーラスの拡張を検討している。

謝辞

なお、この研究の一部は文部省科学研究費補助金（課題番号 10680367）により行なわれた。

また、本研究では、ツールとして日本語形態素解析システム JUMAN ならびに日本語構文解析システム KNP を使用させて頂きました。開発者の方々に感謝の意を表します。

参考文献

- [1] 浪岡保男, 浦谷則好, 相沢輝明: “ニュース文における深層格抽出手法”, 情報処理学会 自然言語処理研究会, Vol.77, No.3 (1990).
- [2] 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 (1995-11).
- [3] 辻井潤一, 山梨正明: “格とその認定基準”, 情報処理学会 自然言語処理研究会, Vol.52, No.3 (1985).
- [4] C.J.Fillmore, 田中春美 他訳: “格文法の原理”, 三省堂, 1975.
- [5] 国立国語研究所: 分類語彙表, 秀英出版.
- [6] 黒橋禎夫, 長尾真: “日本語形態素解析システム JUMAN version 3.5”, 京都大学大学院工学研究科, 1998.
- [7] 黒橋禎夫: “日本語構文解析システム KNP version 2.0 b6 使用説明書”, 京都大学大学院情報学研究科, 1998.
- [8] 小川芳男, 桃沢力, 稲見芳勝, 太田浩, 齊藤次郎, 樽田真, 堀内克明: “サンライズ英和辞典”, 旺文社, 1986.
- [9] R.L.Liu and V.W.Soo: “An empirical study on thematic knowledge acquisition based on syntactic clues and heuristics”, Proceedings of the 31th Annual Meeting of ACL, pp.243-250, 1993.
- [10] 長尾真 編: “自然言語処理”, 岩波講座ソフトウェア科学 15, 岩波書店, 東京, 1996.