

形態素・構文解析情報を用いた深層格分類

4R-7

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†

北海学園大学工学部††

1 はじめに

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、日本語文の解析においても動詞とそのとりうる格との意味的関係を表す深層格が重要な役割を果たす。表層構造で表しきれなかった意味のつながりを、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。このことから深層格を獲得する研究がなされている[1]。

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、意味解析として単語を深層格に分類することを考える。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[2]。

本稿では、日本語単文を形態素・構文解析して得られた情報から、最初に質問応答形式で11種類の深層格に決定し、次に得られた結果を元にして自動的に深層格に分類することを考える。

2 本手法の概要

本手法では、まず図1に示す質問応答形式の処理により、入力文に対し、形態素解析と構文解析を行ない、その結果から質問応答形式により深層格に分類する。次に、図2に示す自動分類形式の処理により、新しい入力文の形態素解析と構文解析の結果、さらに先の深層格分類の結果を用いて自動的に深層格に分類する。

2.1 形態素・構文解析

形態素解析システム[A]により日本語単文を名詞・動詞・格助詞などの品詞に分類し、その結果を用いて、構文解析システム[B]により木構造で表される日本語単文が得られる。

なお、自動分類形式では上記システムで解析できなかった文は修正した結果を使用している。

2.2 深層格への分類および自動分類

日本語単文に出てくる動詞に着目し、その動詞と動詞以外の単語との関係を、形態素解析結果と構文解析結果によって得られた情報、さらに格助詞について考察し、11種類の深層格に分類した。

まず、質問応答形式（名詞と動詞の関係を考えてどのような深層格をとるか決定する）で深層格を定義する。すべての名詞、動詞を登録することは時間的にも労力的にも大変な作業である。そこで、質問応答形式によって得られた深層格情報を用いることにより、自動的に単語を深層格に分類することを考える。まず、未登録の名詞が入力された場合（動詞は登録済み）には、動詞と格助詞の関係から深層格を決定する。このとき、複数の深層格を取り得る単語を判別するために、分類語彙表[3]を用いている。次に、未登録の動詞が入力された場合（名詞は登録済み）には、名詞と格助詞の関係から深層格が判定できるときはその深層格に決定する。

なお、現在のシステムでは名詞、動詞ともに未登録の場合は深層格自動分類を行なわない。

2.3 深層格の種類

ここで用いた11種類の深層格を表1に示す。

3 実験と結果

3.1 質問応答形式

形態素・構文解析の両解析ともに成功した日本語単文1000文について、質問応答形式による深層格分

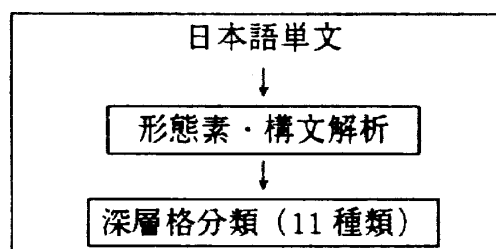


図1 質問応答形式による処理過程

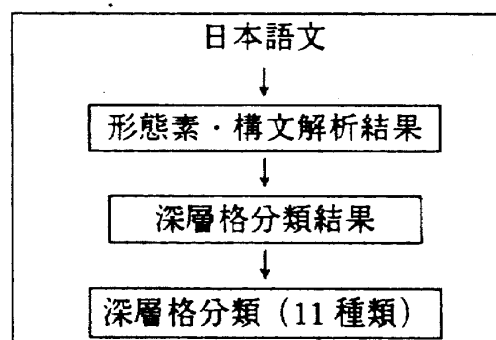


図2 自動分類形式による処理過程

Classification of Deep Case Using Information of
Morphological and Constructive Analysis

Yoshiyuki Masatomi†, Kenji Araki††
and Koji Tochinnai†

†Graduate School of Engineering, Hokkaido University

††Faculty of Engineering, Hokkai-Gakuen University

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

表 1 深層格の種類

類を行なった。日本語単文 1000 文について、文中の名詞・動詞はすべて 11 種類の深層格に分類された。ここで、文中に出てきた単語は、動詞が 432 単語、名詞が 798 単語であった。

3.2 自動分類形式

形態素・構文解析済の日本語文 200 文について、深層格自動分類実験を行なった。日本語文 200 文中で動詞と名詞との関係のべ 496 種類について、深層格に分類することに成功したものは、253 種類であった。主な失敗理由は、入力された名詞と動詞ともに未登録であった場合に分類できないためであった。

例として、入力文「私は彼にその秘密を守るように忠告を与えた。」にたいする形態素解析結果、構文解析結果そして深層格自動分類結果の例をそれぞれ図 3、4、5 に示す。この例では、「与える」という動詞に対して、「私」と「彼」は深層格に分類されたが、これは動詞・名詞とも登録済みであった。「忠告」という名詞は未登録であったが動詞と格助詞により分類された。「守る」と「秘密」は未登録であったため、深層格に分類されなかった。

4 おわりに

日本語単文 1000 文を質問応答形式により深層格分類し、その結果を用いて日本語文 200 文について自動的に深層格に分類した。限られた条件下ではあるが程度の成果が得られたと思われる。

今後は質問応答形式と自動分類形式の融合を考えていく予定である。また、深層格の種類および深層格情報の活用法についてさらなる検討を行なうつもりである。

参考文献

- [1] 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).
- [2] 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).

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 3 形態素解析結果

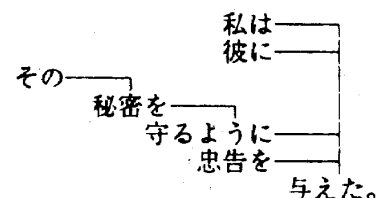


図 4 構文解析結果

動詞	名詞	深層格
与える	私	動作主
与える	彼	受益者
守る	秘密	解析不能
与える	忠告	対象

図 5 深層格分類結果

- [3] 国立国語研究所: 分類語彙表, 秀英出版.

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、以下のツールを使用させて頂きました。ここに感謝の意を表します。

- [A] 京都大学工学部 長尾研究室
奈良先端科学技術大学院大学 松本研究室
日本語形態素解析システム JUMAN
[B] 京都大学工学部 黒橋禎夫 長尾真
日本語構文解析システム KNP