

## 文脈情報を利用した格文法による日本語文解析\*

1E-6

児玉 尚文<sup>†</sup> 上原 徹三<sup>†</sup> 石川 知雄<sup>†</sup>武蔵工業大学工学部<sup>†</sup>

## 1 はじめに

構文解析において格文法解析は、係り受け文法では解決されない構文の曖昧性を低減するために用いられている。格文法解析とは文を動詞(述語)を中心に捉え、動詞が取り得る格の種類と、これらの格に成り得る名詞の意味(概念)を制約する範囲とから成る、文型パターン(格フレーム)を基に解析を行なう手法である。

しかし、格文法において【多義語】、【未登録語】、【代名詞】など個々の単語の概念が特定できない語があると、係り先の曖昧性を低減する効果が薄れてしまうという問題がある[1]。このような時人間は、文脈上から適当な情報を獲得することで問題の解決を行っている。

論文[2]は構文解析に文脈から得られる情報を利用することで、多義語を含む文の構文の曖昧性を低減させることができ、文脈情報の利用が構文解析に有効であることを示している。しかし、単語の概念の情報は取り扱っておらず、単語の概念の不確定さから生じる格文法解析での同様な問題には対応していない。

そこで本研究では、当該文の格文法解析で不足する概念情報を文章頭側の文の格解析情報から獲得し、これを文脈情報として利用することで、概念の特定できない【未登録語】などがあっても格文法解析をより高精度に行なえる可能性があるとの観点から、格文法解析に文脈情報を利用する方法を提案する。

## 2 文脈情報の利用

格文法解析を行う時、ある単語が【多義語】であった場合、可能性のある多義語の全ての概念について解析するので、ある概念の時は係り先が一意にきまるが別の概念の時は別の係り先があり、結果的に係り先の候補を絞り切れず、格文法の効果が薄れることになる。【未登録語】、【代名詞】なども概念の特定できない語であり、同様なことが言える。

これは人間が文の理解(解析)を進める上でも発生しうる問題である。しかし人間の場合、文の理解をそ

の文一つのみにより行っている訳ではなく、文章全体から得られる知識全てを用いて文の理解を進めている。つまり、【知らない単語(未登録語)】が現れた場合、その語の現れた文の周りからこの語の意味(概念)を推定できる適当な何かを見つけることで、この問題に対処している。例えば次のような【知らない単語「NO<sub>x</sub>】を含む例文がある。

「NO<sub>x</sub>を削減する方法について提案する。」...(例1)  
この時、この文一つのみでは【知らない単語「NO<sub>x</sub>(を)】の係り先は「削減する」、「提案した」いずれに係るのか判断できない。これは格フレームで言えば「提案する」が【思考内容など】を【object(ヲ)格】として取り得るので「NO<sub>x</sub>」の意味(概念)を知らない限り、係り先の候補として残さざるを得ないためである。

「NO<sub>x</sub>を排出する量は年々増えている。」...(例2)  
ここで例2のような文が例1と同じ文章内で現れたとすれば、例2の理解(解析)から「NO<sub>x</sub>」が「排出される」何かであることが推定できる。この推定から人間は例1「NO<sub>x</sub>」の係り先を「削減する」とし、「【排出される何か】を「削減する」と理解するであろう。

## 2.1 格文法における文脈情報の利用

計算機による文解析(理解)において単語の概念の不確定さから生じる同様な問題に対しても人間と同様に文脈情報を利用するプロセスを設けることを考える。

- 例1の未登録語「NO<sub>x</sub>」の係り先としては、係り受け解析で「削減する」、「提案した」の2つの候補が挙がる。次に格文法解析を行いたい【NO<sub>x</sub>】の概念が解らないので格文法解析を用いても係り先の候補が減らない。この時、解析に不足する情報は「NO<sub>x</sub>」の概念情報であり、この情報を例2の文から推定することを考える。
- 例2において「NO<sub>x</sub>」は「排出する」に係ると判断される。よって「排出する」の格フレーム情報から「NO<sub>x</sub>」の概念を推定すると、「NO<sub>x</sub>」が【金属、鉱物、化学物質など】の意味を持つ語であると推定される。
- 一方、例1において「削減する」は【化学物質など】の概念を持つ語を格として取り得るが、「提案

\*Japanese Sentence Parsing by Case Grammar Using Context Information

<sup>†</sup>Takafumi KODAMA, Tetuzou UEHARA and Tomoo ISHIKAWA

<sup>†</sup>Faculty of Engineering, Musashi Institute of Technology

する」は「化学物質など」の概念を持つ語を格として取り得ないことが、各々の格フレームからわかる。よって、例1「NOx」の係り先を「削減する」のみに限定できる。

## 2.2 文脈テーブル

文脈情報は以降の解析で利用可能なように、以下に示すような情報を当該文の格文法解析の結果を用いて作成する、以下これを文脈テーブルと呼ぶ。文脈テーブルは動詞エリアと名詞エリアに分けられる。ここでは、例2を用いて文脈テーブルに保持する情報について説明する。作成されたテーブル(一部)を表1に示す。

### ● 文脈テーブルの動詞の格フレーム表：

動詞「排出する」の格フレームより得られる情報「深層格 [agent][object]」と、その格が取り得る「名詞の概念範囲 [人間など][化学物質など]」の情報、及び「解析情報 [「NOx」が深層格 (object) であり得る]」が記載される(表1：上段)。

### ● 文脈テーブルの名詞の概念表：

語の「表記 [NOx]」と、概念辞書とシソーラスから名詞の「概念 [未登録]」とこの概念の最上位概念までの「ルート (枝) [未登録]」、及び「解析情報 [「排泄する」の深層格 (object) であり得る]」が記載される(表1：下段)。

表1:例2の文脈テーブル(一部)の情報

NOxを排出する量が...(例2)			
動詞の格フレーム表			
	排出する(概念)		
深層格	agent(が)	object(を)	source(から)
概念範囲	人間など	化学物質など	建物など
解析情報		NOx(を)	
名詞の概念表			
表記	NOx(を)		
概念(枝)	*****		
解析情報	排出する(object)		

「NOx」の係り先が動詞「排出する」のみに限定されるのは、表1においては省略されているが、格フレームにより「増える」が「ヲ格」を取り得ないことが判明するからである。

## 3 解析方法

### 3.1 処理手順

ここでは『文脈情報を利用した格文法解析』で行う一文の処理手順を述べる。

### 1. 係り受けの可能性：

品詞わけ(形態素解析)された文情報を入力し、文節間の係り受け可能性(係り先候補)を求める。

### 2. 一文内格文法解析：

手順1の結果を入力として格文法解析を行う。この時、当該文の格解析情報、動詞の格フレーム情報、名詞の概念情報を文脈テーブルとして作成する。

### 3. 文脈情報を利用した格文法解析：

手順2を受けて対象語(多義語、未登録語、代名詞)があれば文章頭側の文脈テーブル中で必要とする情報を探索し、探索が成功すれば獲得した情報を元に当該文の格文法解析を文脈情報を利用してさらに進める。

## 3.2 文脈情報の内容と探索方法

ここでは、「多義語」、「未登録語」、「代名詞」それぞれについて獲得すべき文脈情報の内容とその探索方法について述べる。探索は当該文から文章頭側にさかのぼりながら行なわれる。

### ● 多義語 (辞書に複数の概念が記述されている語)

(探索箇所)：多義である名詞表記を文章頭側の文脈テーブルの名詞の概念表の名詞表記欄から探索する。

(獲得内容)：探索が成功すればこの名詞(表記)の概念を獲得する。

### ● 未登録語 (辞書に概念の登録がされていない語)

(探索箇所)：未登録である名詞の表記を文章頭側の名詞の概念表の名詞表記欄から探索する。

(獲得内容)：探索が成功すれば、多義語の時とは違い名詞の概念ではなく、文脈テーブルの動詞格フレーム表からこの名詞に対応する格の概念範囲を獲得する。

### ● 代名詞 (指示する語により概念の異なる語)

(探索箇所)：当該文の動詞の格フレーム表より、この代名詞を格に取り得る動詞の格の概念範囲を調べ、この概念範囲に入る概念を文章頭側の名詞の概念表の概念欄から探索する。

(獲得内容)：探索が成功すればこの名詞の概念を獲得する。

## 参考文献

- [1] 長尾 真 編.「自然言語処理」岩波書店(1996)
- [2] 那須川 哲哉.「文脈情報を利用した自然言語文における構造的曖昧性の解消」情報処理学会論文誌 Vol.36 No.10 (October.1995)