

## スロット表現による複合機能語の処理

5F-10

兵藤安昭 池田尚志  
岐阜大学工学部

## 1 はじめに

日本語の機能語には、学校文法に出てくるような単純な助詞、助動詞だけでなく、「ている」「なければならない」のような複合的な機能語が多数存在する。そしてこれらの複合的な機能語には、「てはいる」「てもいる」「ねばならない」のように意味的には同様なものであるが、派生的なものが多数ある。形態素解析の段階でこれらを「て+は+いる」のように単に単語に分割するのではなく、「【ている】に【は】が付加されたもの」というように、意味的なまとまりとして解析してしまおうことが望ましい。

我々は、スロット表現による辞書記述を用いて処理することによってこのことを実現した。さらに、このような表現法を用いたことにより、派生的な複合機能語を取り込み機能語辞書を系統的に拡張することができるようになった。これを用いて、先に開発した形態素解析用辞書構築支援システムによって実際のテキストから抽出された複合的な機能語を、整理、分類することを行なった。

## 2 形態素解析用辞書構築支援システム

我々は先に、形態素解析を行ないながら、自動的に辞書を構築するシステムを作成した[1]。これは、実際のテキストを用いて形態素解析を行ない、その結果に後編集を加えることにより、辞書にない語や接続規則が不備なものについては自動的に辞書登録する辞書構築支援システムである。解析結果や語が出現するテキストについての情報等は自動的に登録され、機能語を整理、分類する際にこれらの情報は参照される。インターフェースには、日本語 Emacs (Nemacs) を用いており、後編集や辞書登録を非常に容易に行なうことができる。

```
(((@ @ い @) (て で)(は も φ)(! る !ます))-(1)
(SRT VMR FR) _____(2)
((class ている*1) _____(3)
((vpar1 @2)(vpar2 @3))) _____(4)
```

図1 スロットモデルによる辞書表現

## 3 スロットモデルによる機能語辞書の表現

複合的な機能語の中には、意味的には同様なものであっても、表記の違いや、活用、助詞の付加による意味の添加などにより、派生的なものが多数存在する。そこで我々は、複合的な機能語のこのような変動・派生を吸収するために図1に表すようなスロットモデルを用いて機能語辞書の表現を行なった。

図1は、機能語「ている」とその派生語に対する辞書表現である。(1)式の第1項、(@ @ い @)の[@]をここではスロットと呼んでおり、第2項、第3項、第4項がそれぞれのスロットに入り得る要素を示している。すなわち、1番目のスロット(@1)に入り得る要素は「て」または「で」であり、2番目のスロット(@2)には「は」または「も」または「φ」、3番目のスロット(@3)には「る」または「ます」の活用形が入る。ここで、「φ」は空を、「!」は活用する語であることを示す(活用表を図2に示す)。また図1の(2)式は「ている」の接続規則である。

```
((! る) (φ fmr る fst れ fk))
((! ます) (ませ fm まし fr ます fs ましよ fm))
((! ない) (なく fr なかつ fr ない fst なければfk))
```

図2 活用表

これにより、「ている」のグループとして8語（活用語を含めると24語）が表現されたことになる。

本システムでは形態素解析の結果はCLOS (Common Lisp Object System)のオブジェクトとして表現することになっているが、図1の(3)式は、これにより解析されるものが、クラス「ている\*1」のインスタンスとなることを示している。

また、図1の(4)式は、そのインスタンスの2つのスロットに、@2の(はも中)と@3の(いる!ます)の中の実際に現われたものが入られることを示している。この例では、@1の(てで)は単に表音上の違いであって、意味的な内容を持っていないので、解析結果のオブジェクトの中には表現しないようにしてある。このようにして、例えば「てはいます」は、「「ている」+「は」+「ます」」のように解析できる。

辞書表現については以上のようなものであるが、形態素解析システムでは、この表現から作られる機能語を全てメモリ上に展開して登録している。

```

(((@@@) (てで)(いゆ)(くき!ます))
(SRT VMR FR)
((class ていく*1)
 (vpar1 @3)))

(((@@お@) (てで)(はも中)(いる!ます))
(SRT VMR FR)
((class ておる*1)
 (vpar1 @2)
 (vpar2 @3)))

(((を おいて @ ほかに @@) (はも中)(はも中)(!ない))
(N)
((class をおいてほかにない*1)
 (vpar1 @1)
 (vpar2 @2)))

(((だけで @ @) (はも中)(!ない))
(VST AST N FS PK)
((class だけでない*1)
 (vpar1 @1)))

```

図3 整理分類された機能語の例

```

(class 事象自体の助述 sup-class 助述表現)
(class アスペクト sup-class 事象自体の助述)
(class ている*1 sup-class アスペクト)
(class ていく*1 sup-class アスペクト)
(class ておる*1 sup-class アスペクト)
(class をおいてほかにない*1 sup-class 限量)
(class だけでない*1 sup-class 限量)

```

図4 機能語の意味分類

#### 4 朝日新聞記事に基づく機能語の抽出と分類

我々は先に形態素解析用辞書構築支援システムを用いて、朝日新聞の社説1カ月分を解析した。この結果、当初から登録していた機能語約100語に加えて新規に約400の機能語が登録された。これらをスロット表現を用いて整理、分類、拡張を行った結果、約200の機能語のグループに整理、抽出された(図3)。これらは現在のところ、さらに約60の機能語クラスに分類してある(図4)。

また、この約200の機能語グループは解析システム上では約1500語の機能語に展開されている。

このようにして作成された機能語辞書を用いて、朝日新聞の天声人語1カ月分を解析してみたところ、機能語の異なり語数168語のうち、新規に登録された機能語は10語のみであった。

#### 5 おわりに

スロット表現を用いた、機能語、特に複合的機能語の処理方法について述べた。新聞記事を対象とした解析実験においてこの手法の有効性を確認できた。今後はさらに他のテキストに対して実験し、機能語辞書を充実させたい。さらにこの解析結果を用いた支援型の構文解析システムの構築を行なっていく予定である。

#### 参考文献

- [1] 兵藤、池田：形態素解析の後編集による解析用辞書知識の獲得、情報処理学会第44回全国大会、1992