

7F-2 物語知識ベースFAIRY — 構文解析系における辞書の設計 —

上田世志、岡田和久、高田正之、小谷善行
(東京農工大学工学部電子情報工学科)

1. はじめに

われわれは、物語世界の知識を表現する知識ベースFAIRY^[1]の設計と開発を行っている。この知識ベースは、フレーム表現を基礎とし、Prologの節として保存される。構文・意味解析系は、知識ベースの自然言語処理インタフェースとして機能するものである。本稿では、知識ベース(図1.1)の構文・意味解析系における辞書の拡張、辞書の扱いについて述べる。

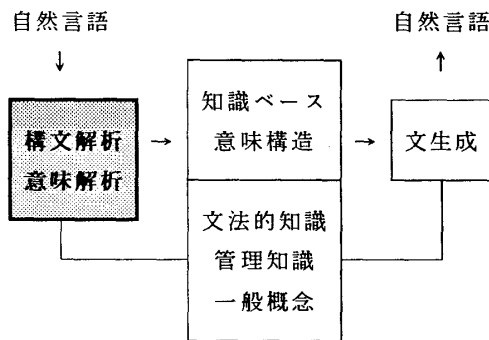


図1.1 知識ベースFAIRYの全体像^[1]

2. 概要

構文解析系は、青島^[2]のシステムを基礎として、改良されている。このシステムは、フレーム形式で記述された表記領域(単語の表記)と言語領域(その文法的な知識)を辞書として用いる。

これまで、あらゆる文に対応できるように、辞書の情報量を増やすことを検討してきた。一つの方法として、既存の情報源であるIPAL^[3]をFAIRYのデータ形式に変換するという方法をとる。IPALの情報をどのように役立たせているかについて述べる。

そして、大容量の辞書内容を扱うための、実際の辞書の参照方式について述べる。

3. IPAL情報の取り込み

構文解析システムは、使用機器SONY-NEWS NWS820 (UNIX 4.2 BSD)という環境のもとで、K-Prologにより動作する。IPAL情報の磁気テープを、UNIX上のテキストファイルに変換した。このファイルから、必要とされる辞書情報を抽出し、利用した。そのために、IPALの記述内容の部分とFAIRYの辞書項目との対応を取り、それに基づいてテキストデータを変換するプログラムを作成、実行した。

IPALの辞書項目とFAIRYの辞書項目における対応を、大まかに表3.1に示す。

表3.1 IPAL辞書項目とFAIRY辞書項目との対応

IPAL辞書記載項目	FAIRYの辞書での扱い
1. 見出し情報	異なる述語名で表現
2. 意味情報	不要(クラス概念)
3. 形態情報	活用型、活用形の情報
4. 統語情報	動詞格支配 名詞意味属性
5. ヴォイス	変形格支配情報
6. テンス・ アスペクト	様態語句の情報
7. ムード	様態語句の情報
8. その他の情報	不要

辞書情報として次のものを取り入れた。

- ・活用型 … 動詞の活用型
- ・語幹 … 動詞の活用形

- ・文型 … 動詞の格支配
- ・意味素性… 名詞の意味属性
- ・述語素 … 動詞の格支配
- ・ヴォイス… 動詞の変形情報
- ・テンス・アスペクト、ムード… 様態情報^[4]

例えば、動詞「あげる」では、次の図3.2 に示すように変換される。ただし不要なものは省いてある。

IPALの情報

《見出し》あげる

《表記》上(あ)げる, 揚(あ)げる

《活用》下一

《語幹》a g e

《文型》N1ガN2ヲ(N3カラ)
(N4ニ/へ)

《述語素》LS2, LG2/
DR2, O2c, A

《格1》ガ《素1》HUM《名1》彼

:

《ヴ1》ニ使役

:

↓ 変換

FAIRYの辞書

上げる(正体,value,上げる_1,_).

上げる_1_(表記,value,上げる,_).

上げる_1_(終止,value,上げる,_).

上げる_1_(活用型,value,下一,_).

上げる_1_(語幹,value,age,_).

上げる_1_(格支配,value,[[n1,ガ],
[n2,ヲ],[n3,カ], [n4,ニ],[n4,へ],_]).

上げる_1_(述語素,value,[ls2,lg2],_).

上げる_1_(ヴォイス,value,ニ使役,_).

:

図3.2 実際の変換例「あげる」

4. 辞書ファイルの扱い

辞書の扱いは次の方法による。辞書はPrologの節で構成されているファイルである。構文解析系を実行する前に、Prologプログラムによるツールを実行しておくことにより、ファイルがディスク上の特定のディレクトリに展開される。

ファイル(meisi.frm)の内容

/*読み領域*/

場合(syoutai,value,場合_1,_).

/*単語領域*/

場合_1_(hinsi,value,名詞,_).

:

↓ 変換

あるディレクトリの内容

40763571.frm

40763571111813171118.frm

:

40763571.frmの内容

場合(syoutai,value,場合_1,_).

図4.1 構文解析システムでの辞書の扱い

このことにより、ファイルは、その単語の名前(述語名)をキーとする名前のファイル群に変換される。その例を図4.1に示す。

ファイル名として、述語名の内部コードを順番に連結したものを使用する。例えば、「場合」という述語名において、「場」と「合」の内部コードはそれぞれ4076、3571である。よって、ファイル名は、“40763571”となる。この際、拡張子“.frm”をつけた。

5. まとめ

FAIRYにおける意味・構文解析系の辞書を、IPALの情報をもとに拡張した。また、実際のプログラムでの辞書ファイルの扱いを示した。

6. 参考文献

- [1] 佐々木泰：物語知識ベース — 概要と適用 —、情報処理学会第36回全国大会、3P-7(1988)
- [2] 青島正明：自然言語文から意味構造を生成する方式、情報処理学会第34回全国大会、4X-1、(1987)
- [3] 情報処理振興事業協会技術センター：計算機用日本語基本動詞辞書 I P A L (Basic Verbs)、情報処理振興事業協会、(1987)
- [4] 小林勉：様態概念を知識ベース上に表現する方式、情報処学会第36回全国大会、(1989)