

要約支援システム COGITO¹
言語解析部

5J-8

北 研 二
(沖電気工業 (株) 総合システム研究所)

安原 宏

1. はじめに

文章理解においては、従来の一文単位の構文・意味解析だけではなく、一文ごとの意味表現を連続的な文になった場合に結合して統合化する必要がある。現在、我々はそのための一つのアプローチとして、世界知識およびそれに基づく推論等を論理型表現のパラダイムで統一的に実現する方式を研究中であり、これらの技術を用いた文書要約支援システムCOGITOを開発している。

要約支援システムCOGITOは、新聞記事を入力して意味解析、要約処理をして、記事の要約として、表(テーブル)形式のリストを出力とするシステムである。

COGITOの特長の一つとして、コンピュータ上での文の意味表現形式を論理型表現の枠組で扱っており、知識表現、推論型との整合性を図りやすいということがあげられる。

以下で、COGITOでの自然言語の意味表現形式および自然言語解析方式について述べる。

2. システム概要

COGITOは、日本語の意味解析を行なう解析モジュールと、その結果を用いて文のパラグラフを抽出し、要約テーブルを作成する要約モジュールから成る。解析モジュールは、文を格成分に相当するものに分解してPrologのユニット・クローズに展開する。要約モジュールは、文と文との関係付けを行ない、パラグラフを抽出する。

また、要約処理の際に参照される知識ベースがあり、これは長期知識と短期知識とから構成される。長期知識は、世界知識、単語知識から成り、短期知識は、解析結果、要約結果等の動的知識であり、入力されたテキストが表現する知識の集合である。知識の表現形式としては、木構造(意味ネットワーク)、辞書(フレーム)、述語論理といった異種の表現を使うが、インプリメントはすべて論理型言語で行なう。

システムへの入力、当面、新聞記事の製品発表記事に限り、出力は要約テーブルと呼ぶ製品の属性と属性値から成るリスト形式のものを想定している。

システムは、現在のところ、C-Prologで記述され、VAX11/785上に開発しているが、将来的には、ICOTで開発された逐次型推論マシンPSIに移植する予定である。

3. 意味表現形式

システムは、言語解析モジュールと要約モジュールの2つに大きく分けられるが、この間での情報の受け渡しは、入力テキストの意味を表わす所定の意味表現形式が用いられる。

最近、知識をPrologのホーン節の集合で表現するという試みが小山、田中[1]で行われている。そこでは、“sem”述語と呼ばれる基本述語を用いて知識を表現しているが、以下で説明する“meaning”述語も同様の試みである。ただし、テンス、モーダリティ、接続関係等も“meaning”述語の枠組みの中で統一的に扱っているという点が異なっている。また、小山、田中[1]の場合は、概念辞書と知識表現を同一レベルで扱っているが、膨大な語彙を持つ必要のあるシステムでは、辞書と知識表現を切り離れたほうがよいと考え、辞書記述は意味表現形式と完全に分離した。

COGITOでは、文の意味は、Prologのユニット・クローズの集合で表現され、このために述語“meaning”が用意してある。述語“meaning”の一般形式は、
meaning(Concept#Id, Attr, Val, S_order).
である。ただし、“#”はオペレータとして宣言されている。また、述語“meaning”の各引数は、次のような意味を持つ。

Concept : 概念名.
Id : 識別子.
Attr : 属性名.
Val : 属性値.
S_order : 文インデックス.

例えば、述語概念を表わすものが第1引数にきた場合、第3引数の属性名には深層格名、第4引数の属性値としてその深層格となる概念が入る。また、テンス、モーダリティ等も述語に対する属性として扱い、格の場合と同様に

¹ 本研究は第5世代コンピュータプロジェクトの一環としてICOTからの委託で行なわれたものである。

meaning(XXX, tense, past, XXX).

meaning(XXX, modality, inference, XXX).

のように表わす。

接続助詞などによって単文と単文が結びつけられる場合は、概念名のところには、“causal_rel”（因果関係）等の接続の関係を示す名前が入り、属性名のところには各単文がその接続関係の中で果す役割を示す名前が入る。例えば、因果関係という関係で単文Aと単文Bが結びつけられ、Aが原因をBが結果を表わすとすると、

meaning(causal_rel#1, cause, A#XXX, XXX).

meaning(causal_rel#1, result, B#XXX, XXX).

のような意味表現が得られることになる。

また、文インデックスは、各文がテキスト中の何番目の文かとか、何番目の段落に表われたかというような文の出現順序を表わすのに用いる。

このように、COGITOでの意味表現形式は、文の意味を細かなユニットに分解するところに特徴がある。こうすることにより、要約モジュールでの概念の取捨選択が容易になると期待される。また、意味表現をPrologのユニット・クローズに設定したことにより、Prologのユニフィケーション機構を用いて、必要な項目を取り出すことが楽にできる。

4. 言語解析部

COGITOは、文法記述言語にDCG形式を採用し、解析にはBUPを用いている。

DCGによる文法記述では、各文法カテゴリを述語と対応させ、その述語にカテゴリの持つ文法的な情報を引数として持たせることができる。COGITOでは、このことを利用し、各カテゴリにそのカテゴリを認識する過程で得られた中心的な概念およびその概念の修飾子リストを持たせている。そして、最終的なカテゴリである“sentence”を認識したときに、修飾子リストを分解し“meaning”述語に変換し、Prologデータベースにassertするようになっている。

現在、DCGで記述した文法の数は約30であり、簡単な新聞記事を解析することが可能である。

図1に、入力文およびその解析結果を示す。

5. おわりに

要約支援システムCOGITOの言語解析部を中心に述べてきた。

また、基本的な文法構造を持つ文しか解析することができないが、文法規則および辞書を拡充し、より広範囲の文を解析可能とする予定である。

沖電気工業（社長橋本南海男氏）は18日32ビットスーパーパーソナルコンピュータ「if1000 UNITOPIAモデル10M」を販売した、と発表した。

入力文

```
[
  pnoun(沖電気工業, [沖電気, [org]]),
  fp(は, [は]),
  tnoun(18日, [18日, [time]]),
  cnoun(32ビットスーパーパーソナルコンピュータ,
    [コンピュータ, [apro, cprol],
     [32, ビット, スーパー, パーソナル, コンピュータ], [1]]),
  pnoun('if1000 UNITOPIAモデル10M', [if1000, [apro, cprol]]),
  kp(を, [を]),
  snoun(販売, [販売, act,
    [[[[hum], [が], agent],
      [[con, abs], [を], obj],
      [[hum], [に], ag]],
     [[[[hum], [が], は], agent],
       [[con], [を], obj],
       [[loc], [に, へ, に向けて], べ向けて], loc]]]]),
  back(し, [2]),
  aux(た, [3, [tense, past]]),
  kp(と, [と]),
  snoun(発表, [発表, act,
    [[[[hum], [が], は], agent],
      [[s], [と], obj],
      [[lang], [に], loc]],
     [[[[hum], [が], は], agent],
       [[s], [と], obj],
       [[hum], [に], ag]],
      [[[[hum], [が], は], agent],
        [[act, con, abs], [と, を], obj],
        [[hum, lang], [に], loc]]]]),
  back(し, [2]),
  aux(た, [3, [tense, past]])
].
```

形態素解析結果

```
meaning(発表#7, tense, past, '[]').
meaning(発表#7, agent, 沖電気#1, '[]').
meaning(発表#7, obj, 販売#5, '[]').
meaning(販売#5, tense, past, '[]').
meaning(販売#5, time, 18日, '[]').
meaning(販売#5, obj, if1000#3, '[]').
meaning(if1000#3, mod, [32, ビット, スーパー, パーソナル, コンピュータ]#2, '[]').
```

意味表現

図1. 入力文とその解析結果

参考文献

- [1]. 小山、田中：Definite Clause Knowledge Representation, in LPC '85.
- [2]. 安原：文書要約システムにおける世界知識の構造情報処32回全国大会6M-7.
- [3]. 小松、安原：要約支援システムCOGITO —「テキスト・パーサ」によるテキスト解析— 本大会予稿.