

拡張LINGOLを用いた日本語の概念分析

井佐原 均 石崎 俊

(電子技術総合研究所)

1. はじめに

自然言語を処理する場合、意味の問題は常に重要な課題である。従来は、構文解析の研究が主として行われてきたが、それだけで言語理解ができたことにはならない。さらに言えば、意味の問題を考えずして自然言語処理は行えないと言えよう。意味の問題に正面から取り組んだ研究はSchankらによるCD (Conceptual Dependency) 理論、SCRIPT及びそれらを用いたさまざまなシステムがあり⁽¹⁾、最近ではMOP (Memory Organization Packet)へと進んでいる。⁽²⁾ SCRIPTは、レストランでの食事のような日常的なテーマを、それに含まれる典型的な事象の並びで表現したものであり、その構造は基本的に一段の(階層性のない)ものである。

MOPはSCRIPTによる表現法をより柔軟に行なえるように拡張したもので、一つのゴールに到達するための一連のSCENEから成り立つ。各SCENEは一つ又は複数の事象から成り立つが、別のMOPが使われる場合もある。すなわち、MOP自身が階層性を持っている。

Schankらの考え方のもう一つの特徴は、名詞概念ばかりでなく、動詞(事象)概念も上位下位概念に分類し、階層的に取り扱っていることである。これにより、各事象の関連を明確にしている。これらの構造は、文法を用いずに、単語(動詞)の意味を中心に組み立てられてゆく。組み立てられた構造は文を越えて保存され、文脈レベルの処理にも用いられる。

一方、従来から当所で行なっている、拡張LINGOL⁽³⁾をパーザとして使い、SRL (Semantic Representation Language)による辞書記述を用いて意味を抽出する手法がある。⁽⁴⁾⁽⁵⁾ まず、拡張LINGOLで文脈自由文法を用いて構文解析を行ない、構文解析木を得る。次にその構文解析木に沿って意味解析を行なう。ここで、SRLによる意味のユニット構造がユニット間会話を行ない、スロットを埋める操作により相互に関係付けられ、全体の意味構造を構築してゆく。用いられる意味表現は各単語に固有のものである。

ここで述べる手法は、MOPの持つ世界知識(World knowledge)の表現能力と、拡張LINGOLの持つ柔軟な文脈自由文法による解析能力を組み合わせた強力な自然言語処理システムの作成を目指している。Schankらはセマンティクスオリエンテッドな自然言語解析法を採用しているが、本手法は、シンタクスとセマンティクスの融合(パラレリズム)を目標とし、その準備段階として構文解析——意味解析——MOPによる概念解析を結合したものである。このような方法は人間の行なう言語処理方法とは異なるもので、むしろ工学的に初歩のよいシステムの実現を目指したものと見えよう。

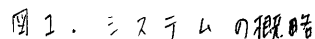
本システムの分析対象は日本語の新聞記事として実用性を考慮している。記事の内容は外電のテロリズムに関するものを扱っていく予定である。

2. システムの概略

システムは図1の様になっている。本システムで用いるMOPはSchankの

なお、オシロイで現在対象としている文は同様のものである。このように同じ内容に上つての英訳記事もある様なテロリでい関係の記事を対象としてゐる。英訳からの意味抽出について別の場合に説明する。

新聞記事はローマ字で入力された文脈自由文法で構文解析した図のような構文解析木を得る。構文解析木は原則として右下がりの木構造となっている。従って、文の意味構造と直接反映した形態とはなっていない。右下がりの木で示されてまいこの処理は、次の意味処理の部で示されている。構文解析木がありうる場合とは、それに対応する図の木の導出時には、「の」の係り先の置換



~ 2 ~

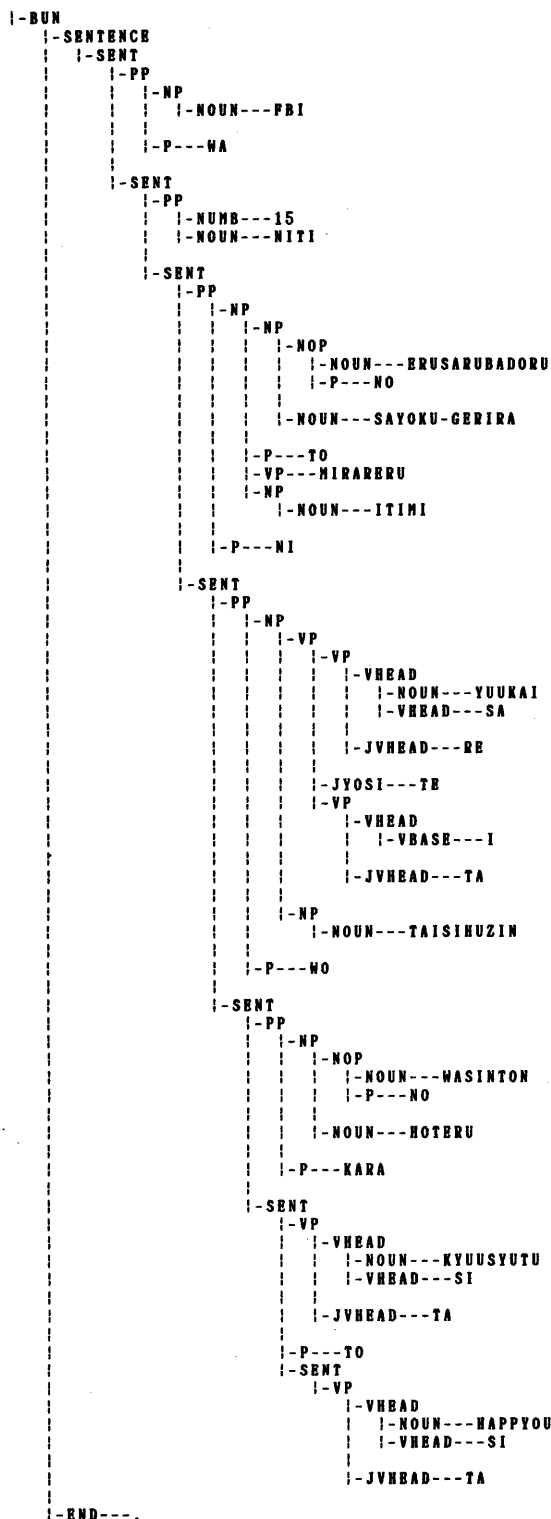


図3. 構文解析木の例

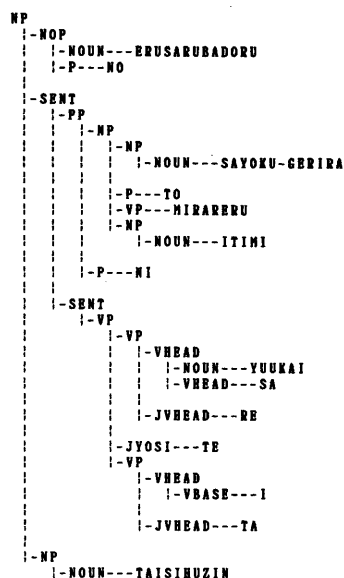


図4. 要行った部分木

文解析木も同時に作り出される。

なお、語尾変化は接続テーブル(図5)を参照しながら処理している。

意味処理は、前段で得られた構文解析木の各節点に付いている意味処理プログラムの順次起動して、ことにより行われる。処理は木の末端から、従って文の末尾から進んでいく。この過程はユニット間会話と呼ばれる。例えば、「ワシントンのホテルから救出した」という部分の処理は次の様に行われる。(図6に示した辞書記述を参照されたい。)

まず、「救出」と「する」から「救出する」という動詞を作る。これは各スロットにある動詞の拘束条件を変更することによって行なう。例えば、where スロット中の「からの」という条件を「から」に変更するなどである。これによって得られる意味記述を図7に示す。次に「ワシントンのホテル」の処理が行われる。「ホテル」の where スロットの条件が

```

(seru (((5-dan sahen) @1)))
(saseru (((1-dan kahen) @1)))
(reru (((5-dan sahen) @1)))
(rareru (((1-dan kahen seru saseru) @1)))
(nai (((5-dan 1-dan kahen sahen reru sareru seru saseru tagaru) @1)
      ((tai souda da youda mita rasi kelyousi keiyoudousi) @2))
      ((aru)))
(tai (((5-dan 1-dan sahen kahen) @2)))
(rasii (((meisi fukusi nu ta no kara made e))
        ((5-dan 1-dan kahen sahen keiyousi) @3)))

```

図5. 接続テーブル

```

=====
-----jisho-----
=====
(FBI NOUN (NIL 0)
  '(fbi unit (part-of)
    (self (a authority))
    (sf)))
(WA P ((WA) 0) '(wa))
(NITI NOUN ((TIME) 0) 0)
(ERUSARUBADORU NOUN (NIL 0)
  '(ElSalvador unit (part-of)
    (self (a country))
    (sf location)))
(NO P ((NO) 0) '(no))
(SAYOKU-GERIRA NOUN (NIL 0)
  '(leftist-guerrilla unit (part-of)
    (self (a bad-guy))
    (sf)
    (where (location /_no))))
(TO P ((TO) 0) '(to))
(ITIMI NOUN (NIL 0)
  '(gang unit (part-of)
    (self)
    (sf)
    (what (t /_eq))))
(NI P ((NI) 0) '(ni))
(YUUKAI NOUN (NIL 0)
  '(kidnap unit (part-of)
    (self)
    (sf action)
    (cact ((a bad-guy) /_no))
    (obj ((a hito) /_no))))
(TAISIHUZIN NOUN (NIL 0)
  '(wife-of-ambassador unit (part-of)
    (self (a hito))
    (sf)))
(WO P ((WO) 0) '(wo))
(WASINTON NOUN (NIL 0)
  '(washinton unit (part-of)
    (self (a capital))
    (sf location)))
(HOTERU NOUN (NIL 0)
  '(hotel unit (part-of)
    (self)
    (sf location)
    (where (location /_no))))
(KARA P ((KARA) 0) '(kara))
(KYUUSYUTU NOUN (NIL 0)
  '(rescue unit (part-of)
    (self)
    (sf action)
    (cact ((or (a hito) (a authority)) /_no))
    (obj ((a hito) /_no))
    (where (location /_karano))
    (from ((a bad-guy) /_karano))))
(HAPPYU NOUN (NIL 0)
  '(announce unit (part-of)
    (self)
    (sf action)
    (cact ((or (a hito) (a authority)) /_no))
    (obj ((a hito) /_no))))

```

図6. 辞書記述

「ワシントン」+「
の」と一致すること
から、whereスロット
への「ワシントン」
の挿入が行われる。
次に、この「ワシン
トン」+「ホテル」
の全体に「から」と
いう助詞を加えたも
のと「救出した」と
の間でユニット間会
話を試みる。今回も、
「救出した」のwhere
スロットに右致し、
そこへの挿入が行な
われる。この操作を
繰り返し行うこと
により最終的に図8
の意味処理結果を得
る。

4. MOP

MOPは世界
知識を表現するた
めの一つの記憶構
造である。すべての事象
をSCRIPTで表現し
ようとするのではなく
SCRIPT中の共通要
素を効率よくまとめ
てパッケージ化し
たものである。従って、
従来のSCRIPTはM
OPの組み合わせに
よって表現可能であ
り、また、MOP自
身が他のMOPと階
層関係を持つことが
できる。MOPは一
つのゴールに到達す
るための一連の標準
的なSCENEから成り
立つ。各SCENEは一

```
' (rescue-suru unit (part-of)
  (self)
  (sf action)
  (cact ((or (a hito) (a authority)) /_ga /_wa))
  (obj ((a hito) /_wo))
  (where (location /_kara))
  (from ((a bad-guy) /_kara)))
```

図7. 作られた動詞の意味記述

MOPには付いた若干の事象の前提と有る制約条件が親のMOP中に記述されている。このことにより、共通の事象を各MOPが様々な状況下で用いることが可能となり、又、一つのMOP中での物語の枝分かれの可能性なども示すことができる。

また、図9中の一番下に書かれた部分は、他の部分とは異なり、特定の（この場合は「悪人が人を拘束している」）条件下での枝分かれの可能性を記述している。すなわち通常は現在起動されているMOPに沿って解釈を進めてゆくが特殊な場合が考えられるような時は、このような特例処理の部分が起動され、可能な事象の一つとして探索される。これにより、E-MOP自体には標準的な事象の並びだけを書きおき、特例についてはそれらを別に記述しておくことにより、さまざまな事象のつながりの可能性を表現できる。

実際に沿って説明してみよう。「誘拐によって金を得る」という行為は通常、「誘拐」→「要求」→「解放」という過程をたどる。不幸な場合とは、「解放」のかわりに「傷つける」という事象が起る。E-MOP上では、この

つ又は複数の事象によって表現される。

「強請(エス)る」という行為に關するE-MOPは、例えば図9に示すように表現される。従来の

```
((FBI 4)
  UNIT
  (PART-OF)
  (SELF (A AUTHORITY))
  (SF))
((ELSALVADOR 6)
  UNIT
  (PART-OF)
  (SELF (A COUNTRY))
  (SF LOCATION))
((LEFTIST-GUERRILLA 7)
  UNIT
  (PART-OF)
  (SELF (A BAD-GUY))
  (SF)
  (WHERE = (ELSALVADOR 6)))
((GANG 8)
  UNIT
  (PART-OF)
  (SELF (A BAD-GUY))
  (SF)
  (WHAT = (LEFTIST-GUERRILLA 7)))
(V)
((KIDNAP-SURU 9)
  UNIT
  (PART-OF)
  (SELF)
  (SF ACTION)
  (CACT = (GANG 8))
  (OBJ = (WIFE-OF-AMBASSADOR 10)))
```

```
((WIFE-OF-AMBASSADOR 10)
  UNIT
  (PART-OF)
  (SELF (A HITO))
  (SF))
((WASHINGTON 11)
  UNIT
  (PART-OF)
  (SELF (A CAPITAL))
  (SF LOCATION))
((HOTEL 12)
  UNIT
  (PART-OF)
  (SELF)
  (SF LOCATION)
  (WHERE = (WASHINGTON 11)))
(V)
((RESCUE-SURU 13)
  UNIT
  (PART-OF)
  (SELF)
  (SF ACTION)
  (CACT = (FBI 4))
  (OBJ = (WIFE-OF-AMBASSADOR 10))
  (WHERE = (HOTEL 12))
  (FROM ((A BAD-GUY) _KARA))))
((ANNOUNCEMENT-SURU 14)
  UNIT
  (PART-OF)
  (SELF)
  (SF ACTION)
  (CACT = (FBI 4))
  (OBJ = (RESCUE-SURU 13))
  (WHEN = ((15 DAY) 5)))
```

図8. 意味解析結果

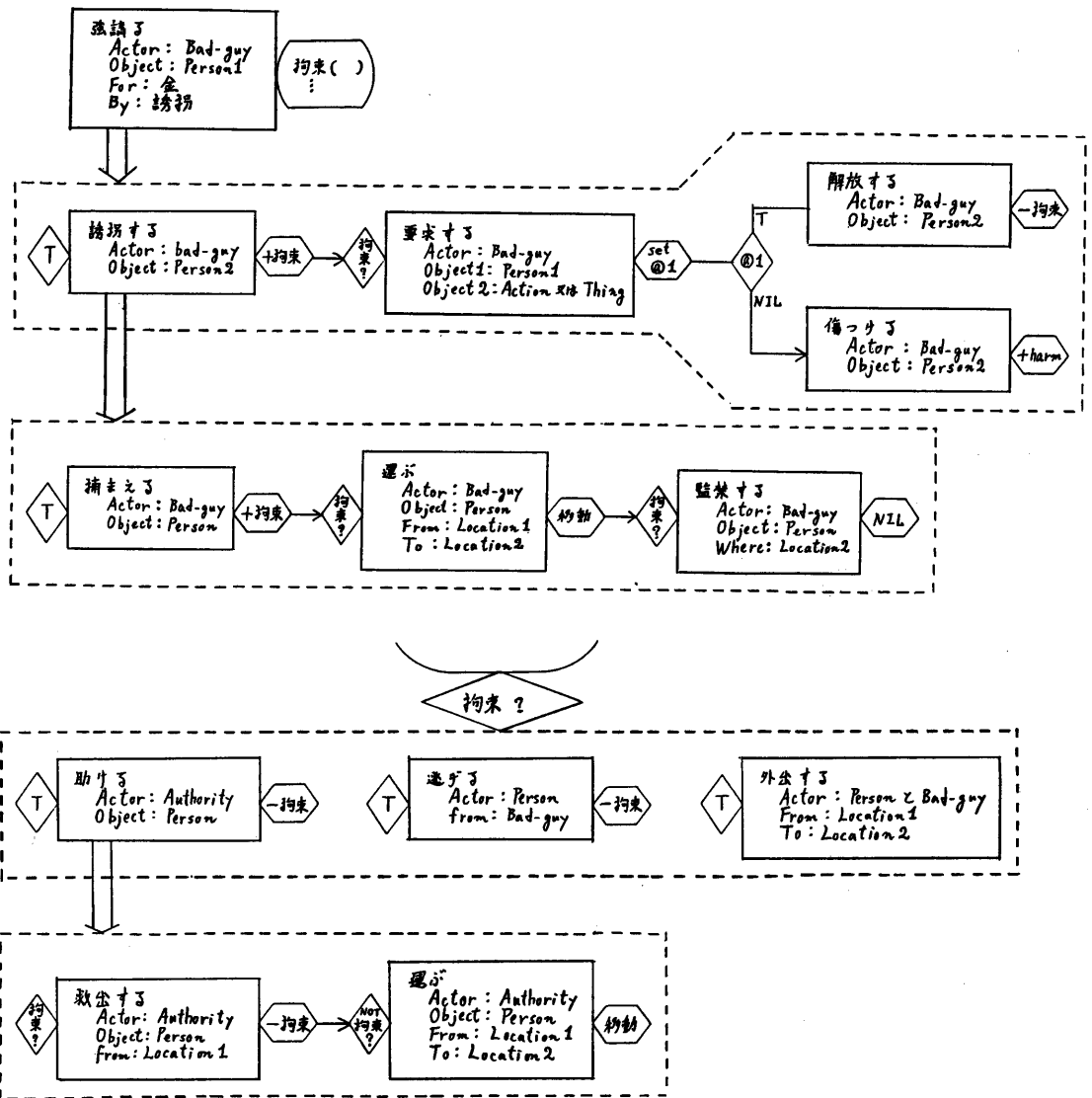


図9. E-MDP の例

「強請る」という行為に対して、「誘拐」、「要求」として「解放」又は「傷つける」という下位事象が記述されている。「誘拐」はいつでも（前提条件がT）悪人（Bad-guy）にとって起こり得て、対象（Person 2）は強請られる人（Person 1）とは異なっており、その結果として人（Person 2）が拘束されることが記述されている。

「誘拐」に続いて「要求」が行われる。この場合、「要求」はPerson 2を拘束しているという条件の下でしか起動できない。この前提条件が「強請る」のE-MOPに書かれているため、別のE-MOPからは別の条件の下で「要求」のE-MOPを起動できる。これがE-MOPの特徴の一つである。

「要求」の結果の成否が④1という変数にセットされ、その値が次の行動の前提条件となる。「要求」が成就すれば「解放」が行われ、「要求」が通らなければ、被誘拐者（Person 2）が傷つけられる。「解放」は「拘束」を解くことと示される。

「誘拐」の内容をより細かく記述するために、「誘拐する」に対応するE-MOPがある。これは「捕まえる」、「運ぶ」、「監禁する」のうちのSCENEをさらに下位に持っている。

以上により、標準的な「強請り」が表現できただけだが、現実にはこのように誘拐者の意思通りに進むわけではない。途中で警察が被誘拐者を救出したり、被誘拐者が逃げたりあったり、あるいは、何か少し別の動作を記述する必要が生じるかもしれない。これらの事象は誘拐の物語中のどこででも起こりうるもので、それらを標準のMOPの中に入れておくと煩雑になる。これらは「拘束」という状態の下で起こるわけであるから図9の下部に示すようにまとめて記述してある。これらのE-MOPは拘束が起こっている場合に常に起動されており、適当な語に対応して物語がこのE-MOPの中へ入っていく。これらのE-MOPの内、あるものは物語中の挿話として枝分かれした後再び合流し、またあるものは、上位のE-MOPからの脱び出し条件として（例えば、誘拐からの脱走など）、上位のE-MOPの中にある前提条件に反する形で合流する。

以上により標準的な物語、挿話、例外的な結果などがE-MOP上でわかりやすく表現できる。なお、E-MOPはSRLによる意味処理結果のユニット構造をも同時に取り込んで表示するため、SchankのMOPのように全ての単語を含んでいるわけではなく、事象の並びに対応するもののみが現われている。また、従来、拡張LINGOLの辞書における意味記述は、単語毎に記述されていたが、CD理論の特色である事象をまとめるという操作のため、現在は特定の辞書記述から同一の意味記述を参照できるようにしている。

5. おわりに

拡張LINGOLとSRLとを用いて行なった日本語からの意味抽出過程の延長として、より深い意味を取り扱える手法であるSchankのMOPの考え方を導入してみた。これにより、標準的な物語およびそのバリエーションが処理できるようにする。現在は兩者を分離した形で取り扱っているため、MOPの利点を十分に生かしているとはいえない。また、E-MOP関連の部分にはまだいくつか問題点が残っているが、明るい見通しを得ている。また、処理可能な文の量も増えさせてゆく予定である。自然言語処理における大玉待課題である文脈処理に

対するMOPの可能性をうまく用いて初歩のよい解析を行なう手法の開発が今後の大きな目標である。

この手法は、まだ研究を始めたばかりであり課題は数多い。「CDと文法との接点」ほどのあたりにあるか？(Top downとBottom Up)とか「意味のPrimitiveほどの程度にまで与えられるか？」あるいは「MOPの作り方、知率よい表示法および利用法はどのようなものか？」等ではまだほとんどわからなかつた。

6. 謝辞

本研究を行なう機会を与えて下さった電子技術総合研究所パソコン情報部長中島隆元氏に感謝致します。繰り返し有意義な討論をして下さった同部推論システム研究室の若氏に感謝致します。自然言語処理研究に関し、常に有益な助言と共に御指導下さった東京工業大学助教授田中穂積氏(元推論システム研究室長)に心より感謝の意を表します。

7. 参考文献

- (1) Schank, R. C. and Riesbeck, C.
"Inside Computer Understanding"
LEA (1981)
- (2) Schank, R. C.
"Dynamic Memory"
Cambridge University Press (1983)
- (3) 井佐原 均, 元吉 文男, 田中 穂積
"N進本拡張LINGOLのユーティリティ関数について"
電子技術総合研究所彙報第46巻第12号 (1982)
- (4) 田中 穂積
"計算機による自然言語の意味処理に関する研究"
電子技術総合研究所研究報告第797号 (1979)
- (5) 井佐原 均, 田中 穂積
"融合方式による日英機械翻訳"
「自然言語処理技術」シンポジウム (1983)
- (6) Ishizaki, S.
"Generation of Japanese Sentences from Conceptual Representation"
IJCAI'83 vol. 1 (1983)