

格構造モデルに基づく日本語意味構造の表現について

内藤昭三 島津明 野村浩郷

(日本電信電話公社 武蔵野電気通信研究所)

1. はじめに

日本語の格構造の設定や相、時などの法情報の分類を行い(1)(2)、それを日英機械翻訳システムLUTE-JEの日本語解析部に応用してきた(3)(4)。本報告では、日本語の意味構造モデルとその表現形態(5)について述べる。ここで述べる意味構造モデルはLUTE-JEの日本語辞書の表現や文解析結果の日本語内部表現として応用されているが、それは現在研究中のもの簡略版である。本稿ではその簡略版について述べる。

2. 意味理解過程

2.1 意味理解モデル

言語理解の過程を、概念的に捉えると、図1のように表わすことができる(6)。すなわち、理解システムがあらかじめ持っている種々の知識(ことばや概念に関する知識、言語構造的知識、一般常識、専門知識など)が、そのときの「場」に照らされて解釈され、テキスト中の情報を引き金として起動され、テキストに作用し、テキストが表わす意味構造に対応する内部表現を作りあげていく。ここで、場とは、話題や話の焦点の内容を示すもので、具体的事象などを解釈する場合の共通の知識を与える。場は一般に階層構造をとる。テキスト全体の話題やテキストの提示されている状況に関するものや、テキストの流れの個々の部分に対応する局所場などがある。内部表現が得られるとそれを使って場の内容は更新され、引き続き文の解析に使われる。内部表現とは、テキストの意味内容を計算機内部に表現したものである。この内部表現は、質問に回答したり、翻訳したりするために使われる。

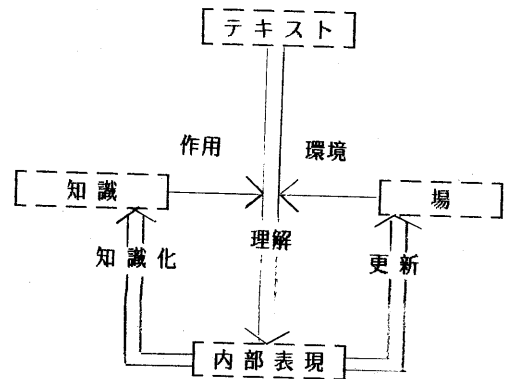


図1. 意味理解モデル

内部表現は、新しい外部条件下での知識のインスタンスとも見なすことができる。新しい知識の取り込みもこの内部表現から行われる。

2.2 内部表現がつくられる過程

下に示す文1から文5までの文の系列を使って説明する。

(文1) 急行電車の窓から通過駅の駅名表示板を見る時、簡単な駅名ならばひと目で読みとることができるが、字数が多いとなかなかむずかしい。

(文2) しかし、文字数がいくつの駅名であったかは、5、6文字までならば、比較的たやすくわかる。

(文3) プラットホームにいる人の人数も5、6名なら一瞬に把握できよう。

(文4) 8、9名だと不確かになる。

(文5) 10名以上の人であると、急行電車の窓からは無理であり、同じ車内にいる人でも、数えてみないとわからない。

まず、このテキストの出典やタイトルなどから図2のような場の情報が与えられているとしよう。このテキストのタイトルは「ひと目でいくつのものが見えるか」であるが、この中のキーワード「ひと目」、「見える」からもある程度の情報を自動的に抽出することができる。

このような場の設定が行なわれたうえで、この文に対する解析が進められ、意味構造が以下のように作られていく。

文1の第一文「急行電車の窓から通過駅の駅名表示板を見る」の解析によって得られる意味構造から図3のような局所場が得られる。「見る」の動作主は提示されていないが、この場合科学技術解説文であるという場の情報を利用して、動作主はデフォルトとして“不特定の人”であると設定する。この場の情報の下で後続の部分が解析される。急行電車が動いていることは陽に示されていないが、「通過駅」の「通過」などを鍵にし、推論によって抽出される。

文2では、「見る」対象が駅名から駅名の文字数に変わる。また、文字数が5、6文字ということに焦点があてられる。照応処理により文2中の「駅名」は文1中の「駅名」と同一対象を指していることが抽出される。

文3では、見る対象が駅名からプラットホームにいる人に変わる。その他の局所場の情報は保存される。

このように新しい情報がくるごとに局所場は少しずつ変更され後続のテキストの解析に使われる。

3. 意味構造の分類

理解の過程では、種々のタイプの意味構造が関与する。意味構造は例えば表1のように分類される。まず、理解システ

分野：心理学

タイトル：ひと目でいくつのものが見えるか

著者：大山正

出典：***

テーマ：視覚、実験

書かれた時：***

読み手：一般読者

図2. 場の情報

事象：見ること

動作主：不特定の人

対象：通過駅の駅名表示板

場所：急行電車の中

状況：急行電車は動いている

図3. 局所場の情報

表1. 意味構造の分類

- (1) システムのもっている知識
 - a. ことばや概念に関する知識
 - b. 言語構造的知識
 - c. 一般常識
 - d. 話題に関する専門知識
- (2) テキストの解析結果として得られるepisodicな意味構造
 - a. 各々の文の意味構造
 - b. 文脈情報
 - c. 場の情報

ムがあらかじめもっている知識と、テキストと知識との相互作用の結果として得られるepisodicな意味構造に分類される。知識は、ことばや概念に関する知識、言語構造的知識、一般常識、専門知識などからなる。episodicな意味構造は各々の文の意味構造、文脈情報、および場の情報などからなる。以下、ことばや概念に関する知識、文の意味構造、文脈情報、および場の情報について節を分けて述べる。

4. ことばや概念に関する知識

これらの知識は、普通、辞書に記されているものに対応する。以下のような情報を含んでいる。

- | | |
|--|---|
| <p>1) ことばと概念をつなぐ情報</p> <p>a. 文字コード</p> <p>2) 文法情報</p> <p>b. 品詞</p> <p>c. 活用型</p> <p>d. 接続情報</p> <p>3) 概念記述</p> <p>e. 意味素性あるいは上位概念 (意味カテゴリ)</p> <p>f. 格フレーム</p> <p>g 他のことば・概念との関係</p> | <pre>(ENTRY (*CODE "8+") 図4. 5. 6. 「見る」、「窓」、 (*CODEa "mi") 「よう」の辞書記述 (上より) (*ID MIRU) (*HINSHI DOUSHI) (*RENSETSU DOUSHI-1-GOKAN) (*KATSUYOU MA KAMI1) (*gogi (1 (*CASE-FRAME (DOUSASYU (*CATEGORY DOUBUTSU) (*DEFAULT WAREWARE) (*CASE-MARKER GA-K)) (TAISYOU (*CATEGORY NOT-SPECIFIED) (*CASE-MARKER WO-K)) (GENSEN (*CATEGORY BASYO) (*CASE-MARKER KARA-K))) (*CATEGORY ACTION CONTINUAL PERCEPTION)) (2 (*CASE-FRAME (DOUSASYU (*CATEGORY DOUBUTSU) (*DEFAULT WAREWARE) (*CASE-MARKER GA-K)) (TAISYOU (*CATEGORY NOT-SPECIFIED) (*CASE-MARKER WO-K)) (TAISYOU2 (*CATEGORY NOT-SPECIFIED) (*CASE-MARKER TO-K) (*OBLIGATORY T)) (*CATEGORY MENTAL ACTION MOMENTAL))))</pre> |
|--|---|

ことばは品詞分類によって大まかに分類される。以下、格構造モデルに基づき以下のものの概念記述について述べる。

- (1) 用言：事象記述の中心となる
- (2) 名詞、副詞：格要素となる
- (3) 付属語：関係情報や法情報を表わす

(1) 用言：

一般に、一つの用言は複数の語義をもつ。図4には、「見る」に対する二つの語義が表現されている。語義1は、「目で物を見る」の意味に対応し、語義2は、「AをBとみなす」の意味に対応する。

各々の語義は次のような情報を含んでいる。

- (A) 格フレーム
- (B) 意味カテゴリ
- (C) 上位用言、類義用言およびそれとの意味の違い

```
(ENTRY (*CODE "Ak")
(*CODEa "mado")
(*ID MADO)
(*HINSHI MEISHI)
(*RENSETSU MEISHI)
(*CATEGORY BUTTAI)
(*MATERIAL GARASU)
(*KANKEI
(ZENTAI-BUBUN
NORIMONO TATEMONO)))
```

```
(ENTRY (*CODE "$h$&")
(*CODEa "you")
(*ID YOU-ZD)
(*HINSHI ZYODOUSHI)
(*RENSETSU YOU-SYUUSHI)
(*CATEGORY SUIRYOU
ITO-MENTAL)
(*MODALITY
MODAL
(*PROCEDURE j:a-m-u)))
```

(D) コメント情報

以下各々の項目に関して説明する。

(A) 格フレーム

格フレーム中には語義に対する固有の格フレームが記される。例えば、「見る」の語義1は固有の格フレームとして動作主、対象、源泉の3つの格をとることが記されている。格関係の種類の設定に関しては既に報告した(2)。各々の格中には以下のような情報を含む。

a) 格要素の意味カテゴリに対する制限

例：「見る」の語義1の動作主は動物

b) この語義をとるために格要素の出現が必須であることの情報

例：「見る」の語義2中の*obligatory情報は、この語義が選ばれるためには、taisyou2の格要素の出現が必須であることを意味する

c) デフォルト

d) 格要素の例

例：「手相を見る」は「手相」という格要素に対する固有の意味をもつ

e) 格マーカ

(B) 意味カテゴリ

動作、状態、瞬間、継続、知覚などの意味素性の情報が記される。

(C) 上位用言、類義用言

上位の用言、類義の用言、およびそれらの意味の違いなどの情報を記す。

例：「読みとる」に対して「読む」は上位の用言であり、「読みとる」の格フレームとして「読む」の格フレームを使うことができる。

(D) コメント情報

コメント情報は格関係を固有格フレームとは別の観点からみた情報を記述するために使う。例えば、「説得する」という用言は、「AがBにCすることを説得する。」という使い方をする。この場合、普通は、Cの動作主はBである。この情報を表わすために「説得する」の格フレーム中に、次のような情報を与えておく。

(*comment (taisyou1 dousasyu) = taisyo2)

この情報は、「説得する」のtaisyou1の動作主はtaisyou2に等しいことを表わす。

(2) 名詞、副詞：

辞書項目としての名詞は、固有名詞以外はプロトタイプとみなす。図5に「窓」に対する辞書記述例を示す。一般には名詞や副詞も多義であり、各語義に対する意味構造を記述する必要がある。例えば、「窓」には、図5中に記述された「物体としての窓」の他に「内と外とをつなぐもの、例えば心の窓」という、より抽象的な第2の意味がある。

(3) 付属語：

これらのことばに対する意味記述はおのおのの品詞に応じて個別的である。格助詞は格関係を表わすマーカの働きをし、その機能は格フレーム中に記されている。「ている」、「れる」などの補助用言、あるいは助動詞は種々の法情報を意味する。あいまいでない場合はその意味が、あいまいな場合には、あいまい性を

解消するための手段が辞書中に記されている。助動詞「よう」に対する辞書記述の例を図6に示す。解析の過程で「よう」が出現するごとに手続き「a-m-u」が起動され、「意志」か「推量」のいずれの意味かを決定する。

5. 文の意味構造表現

文レベルの意味構造は事象や事象間の関係を記述する。図7に文1の意味構造表現例を示す。以下この表現について説明する。

(1) 単位文

単位文は一個の事象を記述する。単位文は一個の述部といくつかの格要素からなる。unfilled-slotは対応する格要素が出現していないことを示しており、補完の処理で埋められる。

(1-a) 述部

述部の情報は以下のものよりなる。

- ・ 自立用言（どの語義かの情報を含む）
- ・ 法情報（相、時、判断・態度）

(1-b) 格要素

格要素情報は以下のものよりなる。

- ・ 格名（動作主、場所など）
- ・ 格インスタンス
- ・ 格マーカ

格インスタンスは、格要素がプリミティブかコンプレックスかいずれかによって異なった構造をとる。プリミティブはほぼ単語に対応する。例えば、「窓」はプリミティブである。これに対してコンプレックスは格助詞「の」や埋込み構造によって修飾された表現に対応する複合概念を表わす。「急行電車の窓」、「簡単な駅名」はコンプレックスの例である。

格インスタンスはプリミティブの場合、以下の情報よりなる。

- ・ タイプ（値はプリミティブ）
- ・ カテゴリ
- ・ インスタンス

```
(SETQ
 SSENT214
 '(SSENT
  (*PREDICATE PREDICATE112)
  (*CASE CASE109 CASE108)
  (*UNFILLED-SLOT
   (DOUSASYU (*CATEGORY DOUBUTSU)
              (*DEFAULT WAREWARE)
              (*CASE-MARKER GA-K)))
  (SCORE 1)))
```

(a) 単位文

```
(SETQ
 PREDICATE160
 '(PREDICATE (*VERB YOMITORU 1)
              (*ASPECT)
              (*TENSE)
              (*MODAL KANOU)
              (SCORE 0)
              (WST WST33)
              (CO-WST WST35 WST36
                WST37 WST39 WST41)))
```

(b) 述部

```
(SETQ
 CASE261
 '(CASE (*CASE-NAME TOKI)
        (*CASE-INSTANCE CASE-INSTANCE216)
        (*CASE-MARKER NIL)
        (SCORE 2)
        (WST WST17)
        (CO-WST WST18)))
```

(c) 格要素

```
(SETQ
 CASE-INSTANCE98
 '(CASE-INSTANCE (*TYPE PRIMITIVE)
                  (*CATEGORY BUTTAI)
                  (*INSTANCE MADO)
                  (*MANNER-MODIFIER NIL)))
```

(d) 格インスタンス（プリミティブ）

```
(SETQ
 CASE-INSTANCE216
 '(CASE-INSTANCE
  (*TYPE COMPLEX)
  (*CATEGORY TOKI)
  (*INSTANCE
   (COMPLEX
    (*MODIFIERS
     (MODIFIER
      (*RELATION-CATEGORY UMEKOMI-TYPE3)
      (*MODIFIER-INSTANCE SSENT214)
      (*RELATION-INSTANCE ZERO)))
     (*MODIFICANT CASE-INSTANCE217)))
  (*MANNER-MODIFIER)))
```

(e) 格インスタンス（コンプレックス）

図7. 文1の意味構造表現例（単位文）

・ manner-modifier

manner-modifier のスロットには副助詞の表わす法情報が書かれる。副助詞は形態的には小さいが意味的には文全体の命題に影響を与える。例えば、

(文6) シロは肉だけ食べる。

という文は

(文7) シロは肉を食べる。

という情報に加えて

(文8) シロは肉以外は食べない。

という情報も表わしている。これは、

(文9) シロは肉しか食べない。と同義である。

格インスタンスがコンプレックスの場合にはインスタンス・スロットはさらに構造をもつことになる。この値は、modifierとmodificantのスロットをもつ。このコンプレックスは再帰的構造をとる。relation-categoryには、修飾関係が記述される。埋込み構造となっている場合には、埋込みタイプの1, 2, 3が、「の」による修飾の場合には、全体・部分、所有、場所などの意味的關係が記される。埋込みタイプの1, 2, 3は修飾される名詞句が(1)埋込み文中の格要素となっているもの、(2)埋込み文中の格要素を修飾しているもの、(3)同格的なものにそれぞれ対応する。埋込みの場合には、modifier-instanceの値は文になる。

(2) 複合文

複合文はmodifierとmodificantより構成される。modifierの部分にはrelation-categoryとともに、time-relationの情報も記す。時間情報は、修飾と被修飾文中の主文間の時間順序関係を記す。図7の複合文フレームのtime-relation中には、修飾文csent256中の主文ssent238と被修飾文csent257中の主文ssent253に対して、どちらも「現在の事象であり、2つの事象間にとくに時間的前後関係はない」ことが記されている。

以上に述べた表現は以下のような情報を加えることによってより深い理解の状態を表現した意味構造になる。

- (1) 文の意味構造を単位文間の修飾関係として捉えるのではなく、事象間の関係として捉えるもう1つの視点からの表現を記す。
- (2) 照応処理が行えるように、格要素や事象をインスタンス化する。
- (3) 格要素や事象に対するタイプ(総称か特定かなど)の記述を行なう。

6. 文脈情報

文脈情報はテキストの連続によって得られる情報の総体である。代名詞の指示対象の決定や、別々の文にでてくる同一表現や言い換え表現が同じものを指して

```
(SETQ
 CSENT258
 '(CSENT
  (*MODIFIERS
   (MODIFIER
    (*RELATION-CATEGORY GYAKUSETSU)
    (*MODIFIER-INSTANCE CSENT256)
    (*RELATION-INSTANCE GA-S)
    (*TIME-RELATION
     (CSENT256 SSENT238 ((? GENZAI) ACTION))
     (CSENT257 SSENT253 ((? GENZAI) STATE))))))
 (*MODIFICANT CSENT257)
 (SCORE 3)))
```

図7. 文1の意味構造表現例(複合文)

いるか否かの表現は理解の観点から不可欠である。例えば、

(文10) 昨日、急行電車に乗った。

という文では、「電車」は特定の電車を指す。この文に続いて、

(文11) 電車は10両編成であった。

という文が続けば、この「電車」は文10の電車と同一対象を指している。しかし、

(文12) 電車はかつて最もポピュラーな交通手段だった。

という文が続けば、この「電車」は総称としての電車を指しており、文10中の特定の電車を指してはいない。このような照応の情報を表現するために5節で述べた格要素のインスタンス化が必要である。

照応関係の処理に時間、場所、様態などの格情報を使うことができる。例えば、文11を変更して

(文13) 電車は10両編成が普通だった。昨日の電車は16両編成だった。

という文脈になると、文13中の第一文の「電車」は総称的になり、第二文中の「電車」は文10中の電車と同一対象を指すことになる。この可能性は、「普通」や「昨日」などの格要素に着目することによって得られる。

複合名詞を構成する部分名詞のレベルでの照応処理が必要な場合もある。文1において、「通過駅」中の「駅」と「駅名表示板」中の「駅」は同一対象を指している。また、「駅名表示板」中の「駅名」と「簡単な駅名」の「駅名」は同一対象を指している。複合名詞を構造体として捉えておけばこのように複合名詞中の構成要素レベルでの照応関係の表現を行うことができる。

7. 場の情報

場の情報は、テキストの話題に関する情報や、テキストの提示されている状況に関する情報の総称である。

場の情報は、(1) 語義の選択、(2) 省略された格要素の補完や判断・態度情報の当事者あるいは対象者の判定において有効である。以下この二つに対して述べる。

(1) 語義の選択

以下に例を示す。

(文14) 昨日、駅へ定期券を買いに行った。売場の前には長い行列ができていた。

(文15) 計算機は種々の数学上の問題を解くために使われてきた。大規模な行列の計算はその1つである。

文14は日常の話題を扱っており、この場合の「行列」は普通の「人の連なり」を意味している。文15は数学という特殊な話題に関しての話であり、「行列」は数学のmatrixを意味している。このような区別は場の情報を用いると比較的に判定できる。

(2) 省略された格要素の補完や判断・態度情報の当事者あるいは対象者の判定

2節の例テキストは、科学技術解説文であり、著者が読者に、ある視覚心理の実験に関する解説を行うことを目的として提示されている。この場合テキスト中の陳述の主体は著者である。例えば、「字数が多いとなかなかむずかしい」と判

断しているのも「一瞬に把握できよう」と推量しているのも著者である。

また、科学技術解説文では、一般事実を述べる文が多く、しばしば動作主は省略される(2)。この場合、この省略された動作主は、“不特定の人”であることが多い。例えば、「通過駅の駅名表示板を見る」の動作主も、「ひと目で読みとる」の動作主もこの“不特定の人”と考えられる。

このような性質を解析のためのheuristicsとして利用することができる。例えば、文6の「同じ車内にいる人でも、数えてみないとわからない」において、「同じ車内にいる人」は「数える」の動作主あるいは対象の2つの可能性がある。この場合このheuristicsを使って対象の可能性が高いと推測することができる。

8. むすび

言語理解の過程を知識とテキストとの相互作用であるという観点から捉え、この過程に関与する種々の意味構造の表現について述べた。日本語文に対する意味モデルとして格構造に基づく意味モデルを設定した。本報告では、この格構造モデルに基づく日本語文の意味構造表現を中心に述べた。また、ことばや概念に関する知識、文脈情報、場の情報の表現や利用に関しても述べた。今後は、これらの情報および知識の構造をさらに追及していく。

本稿で示した知識は、全体部分、上位下位、意味カテゴリなどの断片的な知識である。事象間の関係に関する知識や、例えば、「認識」といった多くの側面からなる複雑で抽象的な事象に関する知識のようなより大きい単位の知識の表現や利用については別の機会に述べる。

謝辞 日頃、研究の進め方に関してご指導いただく基礎研究部畔柳部長、第一研究室山下室長に感謝します。

文献

- (1) 島津、内藤、野村：“格構造モデルに基づいた日本語文の分析と解析”自然言語処理29-1、1982
- (2) 内藤、島津、野村：“日本語文における法情報の解析”自然言語処理26-2、1981
- (3) 島津、内藤、野村：“機械翻訳システムLUTEの日本語意味解析”自然言語処理33-6、1982
- (4) 片桐、野村：“機械翻訳システムLUTEにおける意味構造変換処理”自然言語処理35-5、1983
- (5) 内藤、島津、野村：“日英機械翻訳システムにおける日本語文の意味表現について”情処24回大会、1G-6、1982
- (6) 野村、島津、飯田：“機械翻訳への知能処理アプローチ”自然言語処理30-5、1982