

TM-0992

意味表現の階層に基づく生成システム

池田 光生

December, 1990

© 1990, ICOT

**ICOT**

Mita Kokusai Bldg. 21F  
4-28 Mita 1-Chome  
Minato-ku Tokyo 108 Japan

(03)3456-3191~5  
Telex ICOT J32964

---

**Institute for New Generation Computer Technology**

## 意味表現の階層に基づく生成システム

池田光生

(財) 新世代コンピュータ技術開発機構

〒108 港区三田 1-4-28 三田国際ビル 21 階

*Tel:* 03-456-3194

*E-mail:* [ikeda@icot.or.jp](mailto:ikeda@icot.or.jp)

## 1 はじめに

これまで、自然言語処理の研究は解析を中心として行われてきた。伝統的に解析の手法は、形態素解析、構文解析によって一文の統語構造を抽出し、意味解析によってその意味を求め、さらに文脈解析によってその解釈を行うというものである。これらの処理はほぼ独立して行われ、形態素解析の出力が構文解析の入力となるというように、ある処理の出力が別の処理の入力となるという構成が前提とされている。

生成においてもこのような処理の手順を前提として、解析に類似したモジュール構成が取られることが多い。どのような生成システムでも、生成処理全体は、主に生成すべき文の内容を決定するプランニング部と、主に文の内容を適切に表現する表層文生成部とに分けて行なわれることが多く、両者の間で一つの文の表現に対応する中間構造を介して情報が伝達される。そこで用いられる中間構造は、述語を中心とした格構造のような比較的単純な構造が想定され、単純な文の場合には効率良く意味を記述することができた。

しかしながら、人間らしい自然な表現を生成するためには、多様な統語構造や表層表現を生成しなければならないならず、生成処理が多様化、複雑化するにしたがって、中間構造に含めなければならない情報が増大し、中間構造が非常に複雑なものになってしまう。

本稿では、多様な表現を生成するための中間構造として、五つの階層からなる意味表現を導入した。各階層は、助詞「は」の役割や従属節の統語的な分類、あるいはテンス、アスペクト、話し手の判断、態度など、文の統語的な階層構造を反映したものであり、統語的な制約を自然な形で意味の階層の中に埋め込むことにより統語的な制約と意味的、文脈的な制約を統合している。本稿で設定した、統語的な制約を含めて階層化した意味表現を、意味表現の階層と呼ぶことにする。意味表現の階層を用いることにより、多様な従属節の接続表現の違いや提題の「は」の役割を、記述する階層の違いとして、より自然に明示的に表現することができる。

次章では従来の生成システムの中間構造の問題点を明らかにし、三章では問題点を解決するための基盤となる意味表現の階層構造について述べる。最後に四章でまとめと今後の課題について述べる。

## 2 従来の中間構造の問題点

多くの生成システムのモジュール構成ではプランニング部で文脈的、意味的な処理を行い、中間構造を用いて表層文生成部に情報を伝達する。したがって、主題化や取り立ての対象オブジェクトの記述や、統語構造や従属節の接続表現を決定するための情報などを、表層文生成部への入力中に記述することになる。

生成の多くの処理は文脈情報を考慮しなければならないことや、統語的な制約と意味的、文脈的な制約は密接に関係しあっていて明確に分離できないことから、多様で複雑な生成処理を行う場合には、中間構造の記述が非常に複雑なものになってしまう。このことを示すために、特に助詞の「は」と統語構造や従属節の接続表現との関係に着目して、例を示す。

日本語で助詞「は」は、主題や対照を表す場合に用いられるとされている。従来の生成では、述語を中心とした格構造中の“主題”や“取り立て”のマークの付けられた格要素を、統語構造の生成過程で助詞「は」を付加し、文頭に置くという処理をすることが多い。しかしながら、“主題”や“取り立て”といった概念は話し手が何について述べているか、何に着目しているかを示すものであり、それらは話し手の信念状態から生まれてくるものである。

話し手の着目対象を示す働きのある助詞「は」は、統語構造を生成する上で大きな役割を果たしている。これを示すために南[10]による従属節の構成要素に基づく分類をもとに、統語構造と話し手の着目対象との関係について述べる。南は従属節の述語的部分と述語的部分以外の要素との関係を見ることによって、従属節を A 類、B 類、C 類の三つに分類している。A 類従属節は、テンス、アスペクトを持たない述語的部分などを含む従属節で、継続の「～ながら」「～つつ」などの接続表現が用いられる。B 類従属節は条件的な意味を持ち、テンス、アスペクトを持つ述語的部分と時間、場所の修飾語などだけを含む従属節で、「～ので」「～たら」などの接続表現が用いられる。C 類従属節は、意志や推量や提題の「は」などを含む従属節で、「～が」「～から」などの接続表現が用いられる。

以下の例文は、イラクのクウェート侵攻に関するものであるが、事実的内容は同じであるにもかかわらず、三通りの接続表現が用いられている。この三つの例文の生成処理について考える。

- (1) イラクはクウェートに侵攻したので、西側に非難された
- (2) イラクはクウェートに侵攻したが、西側は直ちに非難するべきだ
- (3) イラクがクウェートに侵攻し、西側がイラクを非難したのだ

(1) は話し手はイラクに着目し、イラクについての記述を行っている。「クウェートに侵攻したこと」「西側に非難されたこと」はイラクについての記述であり、「イラクは」は文全体にかかっている。このような従属節を生成するためには、従属節が主節の理由であることを示すために、従属節に「理由」などと表現しておくのが普通である。しかし、「ために」「から」など他の理由を表す接続表現との区別を行なうためには、別の情報が必要である。(1) では、話し手の着目対象が従属節を越えて主節にまでおよんでいる。このような従属節は、南の分類における B 類従属節に相当し、そのために理由を表す接続表現として、B 類従属節に用いる「ので」が使われている。

これに対し、(2) では前半はイラクについての記述で、次に西側に注意を移して西側についての記述を行っている。このように話し手の着目対象が従属節内にとどまるような従属節は、南の分類の C 類従属節に相当し、C 類従属節に用いる接続表現「が」が使われている。

(1) と (2) の接続表現を生成するためには、従属節と主節のそれぞれについて、話し手が何に着目して記述を行なっているかを示す必要があり、さらに、従属節と主節とで別々の着目対象について記述しているか、あるいは一つの対象についての記述であるかといった情報が必要である。

(3) はイラクや西側についての記述というよりは、二つの事柄を述べることにより何か別な主題(例えば、「これまでの経過は」、「1990 年には」など)についての記述であり、このことは「イラク」および「西側」に助詞「が」が用いられることに現れる。(3) を生成するためには、話し手は「クウェート侵攻」や「西側の非難」以外の対象に着目していて、(3) の文全体がその着目対象についての記述であるという情報が必要である。

次の例文(4)と(5)では事実的内容は同じであるが、従属節の持つ時相が異なる。

- (4) イラクは経済制裁に耐えながら、西側の動きをうかがっている
- (5) イラクは経済制裁に耐えていて、西側の動きをうかがっている

(4) では、「イラクが経済制裁に耐えること」と「西側の動きをうかがうこと」とは同一の時相で成立する事態であり、結果状態を表す「ている」というアスペクト辞は両方の事態のアスペクトを表している。従属節がテンス、アスペクトを含まず、従属節と主節とが同一の時相で成立するような従属節は、南の従属節の分類における A 類従属節に相当する。そのため、(4) では A 類従属節に用いる接続表現である「ながら」が使われている。

「ながら」という接続表現を生成するためには、従属節の表す事態と主節の表す事態とが同時に起こることを表すために、従属節に「同時」などと表現しておくのが普通である。しかし、(4)と(5)の接続表現を生成するためにはそれ以外にも、従属節と主節のそれぞれの時相についての情報と、(4)ではそれらが同一であり、(5)では別々であるという記述が必要である。

さらに、助詞「は」は、以下の例のように一文内に複数現れる場合がある。

(6) セロリは根元は横にスライスする

(7) 根元はセロリは横にスライスする

(6)と(7)は、主題化の対象が複数の場合で、話し手の信念内ではそれらの間には構造があり、その構造は統語的には語順に反映される。(6)はまず「セロリ」という話し手の着目対象を提示し、次に「根元」という着目対象を提示していて、(6)の二つの対象の語順に反映されている。語順を入れ替えた(7)の「根元はセロリは」の場合は「(一般的に)根元について言うときセロリの場合には」という意味を持つ。したがって、(6)と(7)を生成するためには、話し手はまずセロリに着目し次にその根元について述べているというような、話し手の着目対象の推移やその構造を情報として与えなければならない。

以上のように、何を主題化や取り立ての対象とするかを決定することは、話し手の着目対象の移り変わりなどの文脈的情報によって行うものであり、同時に統語構造や従属節の接続表現を決定することでもある。また、主題化や取り立てと、省略、代名詞化、態変化、接続表現の決定などは互いに関連があり、各々の処理モジュール間で情報を伝達しなければならない。

比較的単純な文に対しては、これまで中間構造を容易に記述することができていたが、複雑な文を生成しようとするとき上に示したように、中間構造に様々な情報を含めなければならない。その場合には中間構造が複雑になり、表層文生成部の各処理モジュール間の多様な情報の流れの制御も困難になる。

このことを解決するためには、統語的な制約の記述と意味的、文脈的な制約の記述を統合し、同じ環境で文脈処理、意味処理、統語処理が行えることが望ましい。そのためには、文脈情報を形式化すること、意味的な制約と統語的な制約の両方を統合した意味表現を設定することが必要である。

### 3 意味表現の階層

前章で述べた問題点を解決するために、統語的な制約を反映した五つの層からなる意味表現を設計した。意味表現の階層によって、例文(1)～(3)で示した着目対象の違いや、(4)と(5)で示した時相の違いなどを、階層の違いとして自然に明示的に表現することができる。意味表現は以下のように階層化されている。

第一層 基本概念の記述

第二層 基本概念間の関係の記述

第三層 基本概念間の関係が成立する時相状況と時相状況に対する話し手の判断の記述

第四層 時相状況間の関係の記述

第五層 話し手が着目する対象と着目対象についての記述内容の組と、それらと談話構造との関係の記述

一方で、日本語の構造の分析として、文の構成要素間の関係を階層構造として見る見方が多く見られる。渡辺 [15] は、日本語文の構成要素を叙述と陳述に分け、それらの中間的存在も含めた述語の階層構造を示した。南 [10] は従属節の構造の違いに基づいて文全体の階層構造が考えられることを示している。上で示した意味表現の階層は、これらの統語的な日本語文の階層構造とほぼ対応していて、助詞「は」の役割や統語的な分析による従属節の分類などが自然に階層構造に埋め込まれている。

第一層は、“イラク”、“侵攻する”など基本的な概念の言語的な知識の記述であり、いわゆる辞書に相当する。第二層は、基本概念間の関係を表す項の記述で、“侵攻する”の agent が“イラク”であることなどが記述される。第二層で記述した項はテンス、アスペクトなどの時相の値を持たない。

第三層の時相状況は、あることがらを状況として見ることにより時間的な幅を考えるために取り入れた概念である。時相状況の考え方は東条 [14] による。第三層で、第二層の項が成立する時相状況を考えることによりテンス、アスペクトなどの時相が生じる。

第四層では、時相状況間の関係を記述する。時相状況間の関係は南の分類の B 類従属節を用いた文を表す。B 類従属節とは、統語的には提題の「は」や話し手の態度などを含まない従属である。提題の「は」や提題の名詞句などは、第五層で着目対象として記述する。着目対象とは話し手の注意が置かれている対象であり、その着目対象についての記述の表現を記述内容と呼ぶ。第四層の時相状況間の関係の記述は一つの着目対象の記述の範囲内にあると考える。このことは提題の「は」や提題の名詞句などが B 類従属節を越えて、文全体にかかることに対応する。

このように、統語的な分析による提題の「は」や提題の名詞句の役割を、着目対象とその記述内容という考え方で、意味表現の階層の違いとして表現することにより、従来の提題化の処理に比べて、より自然に明示的に表現することができる。同様に、多様な従属節の接続表現の違いを意味表現の階層の違いとして表現することで、様々な接続表現を生成することができる。

以下で、意味表現の各階層とその統語的な階層構造との対応の詳細を述べ、二章であげた例文の、意味表現の階層における記述を示す。

### 3.1 基本概念間の関係

第一層は“イラク”、“侵攻する”など最も基本的な概念の言語的知識の記述であり。第二層は“イラクがクウェートに侵攻する”のような基本概念間の関係を表す項の記述である。第二層の記述例を示す。

侵攻する [agent= イラク, goal= クウェート]

非難する [agent= 西側, object= イラク]

第二層の記述のうち用言に対応する項は、態の値を持っている。しかし、時間的な概念は含んでいないため、テンスやアスペクトについての値は持たない。したがって、第二層では様態、程度などの副詞は記述するが、テンス、アスペクトの副詞は上位の階層で記述する。第二層の項は A 類従属節の記述に相当する。事実的内容の等しい例文 (1) ~ (3) の記述および (4) と (5) の記述は第二層では全く同じである。<sup>1</sup>

### 3.2 時相状況

第二層で記述された項はテンス、アスペクトを含んでいない。これに対し第三層では、項が成立する時相状況を考えることにより、アスペクトが生じる。ある時間的視点の集合  $S$  があって、その中の

<sup>1</sup>ここでは、事実的内容は時相を含まないことにする

任意の時点から見たときにある項 $\sigma$ が成立するとき、その最大の $S$ を $\sigma$ の時相状況 $S^\sigma$ とする [14]。時相状況はある時間的視点からある視野をもって見ることにより、アスペクトが決まる。本システムでは、時相状況に対して“継続”、“非継続”の二値を視野の値として設定している。

例えば、「イラクがクウェートに侵攻する」という第二層の記述 $\sigma$ を第三層で状況として見ることによりアスペクトが発生する。この状況は「イラクがクウェートに侵攻する」ことが成立する時相状況 $S^\sigma$ を表す。この時相状況は「イラクがクウェートに侵攻する」という事態に時間的な幅を与えたもので、アスペクトの値を持っている。

次に、例えばこの状況内に視点を置いてこの状況を非継続的に見ることにより、新たな時相状況を定義することができる。新たに定義した状況は、ここでは、「イラクがクウェートに侵攻している」ことが成立する時相状況 $S_{iciru}^\sigma$ を表す。

次に、時相状況に対してテンスを考えるために、時相状況間の関係を定義する。時相状況 $S^{\sigma_1}$ に含まれるすべての時点が、時相状況 $S^{\sigma_2}$ に含まれるどの時点よりも先行するとき、

$$S^{\sigma_1} < S^{\sigma_2}$$

と書く。時相状況を、その状況に後続するある時間的視点から見た場合を過去(完了)とする。現在を表す状況を $S_{now}$ としたとき、

$$S^\sigma < S_{now}$$

ならば、過去(完了)として生成する。過去(完了)、非過去(不完了)の判断を、ある時相状況のある時間的視点から見て行うという点において、テンスとアスペクトは時相状況の考え方の中で同じように処理される。テンス、アスペクトについての詳細な分析は本稿の範囲を越えるので、ここではこれ以上詳しくは述べてない。

時相状況を用いて、

(4) イラクは経済制裁に耐えながら、西側の動きをうかがっている

のようなテンスの入らない従属節を用いた文を記述することができる。この従属節は南の分類のA類従属節に相当する。A類従属節とは従属節を表す項が主節と同一の時相状況中で成立するような従属節であるといえる。以下のように同一の時相状況内に複数の項を記述することによって表現する。

$S_1$ :: 耐える [agent= イラク, object= 経済制裁]

$S_2$ :: うかがう [agent= イラク, object= 西側の動き]

$S_1$  は「イラクが経済制裁に耐えながら西側の動きをうかがう」という状況を表している。これに対して、

(5) イラクが経済制裁に耐えていて、西側の動きをうかがっている

の場合には従属節が主節とは別に時相を持っているため、以下のように別々の時相状況を用いて記述する。

$S_1$ :: 耐える [agent= イラク, object= 経済制裁]

$S_2$ :: うかがう [agent= イラク, object= 西側の動き]

(4) の記述は第三層の同一の時相状況の記述に相当し、(5) の記述は第四層の時相状況間の関係の記述に相当する。このように、絶対的な時間の同時性以外に、(4) では従属節の時相が主節の時相と等しく(5) では異なるということを意味表現の階層の違いとして表現できる。

### 3.3 着目対象と記述内容

第五層では、話し手の注意が置かれている対象とそれについての記述が表現される。話し手の注意がおかれている対象を着目対象と呼ぶ。本稿では、発話において、話し手はある対象に注意を置き、その対象についての記述を行うものとしている。したがって、着目対象は発話の基本単位を形成する中心となる。

着目対象についての記述の表現を記述内容と呼び、意味表現の階層では第四層以下で記述されたこととがらに相当する。着目対象は統語的には記述内容全体にかかり、提題の「は」、取り立て詞、提題の名詞句などとして生成される。

着目対象とその記述内容という考え方をを用いて、それらを意味表現の階層の違いとして表現することにより、提題の「は」の役割を明示的に表現する。意味表現の階層による記述は、提題の「は」の統語的な制約も含み、従来の提題化の処理に比べて、より自然に直接的に表現することが可能である。

(1)～(3)で示したように着目対象の違いは、従属節の接続表現の違いとして現れる。

- (1) イラクはクウェートに侵攻したので西側に非難された
- (2) イラクはクウェートに侵攻したが、西側は直ちに非難すべきだ
- (3) イラクがクウェートに侵攻し、西側がイラクを非難したのだ

(1) は一つの着目対象“イラク”についての記述である。二つの時相状況間の関係は一つの着目対象についての記述であり、B類従属節を用いた文に対応するため、B類従属節の接続表現である「ので」が使われている。B類従属節とは、従属節と主節の時相状況において着目対象が変わることがなく着目対象が従属節を越えて文全体にかかるような従属節であるといえる。

(2) は二つの着目対象“イラク”と“西側”についての記述である。着目対象が文全体を支配するB類従属節に対して、C類従属節とは、着目対象が従属節の外へ出て、文全体にかかることができない従属節であるといえる。ひとつの記述内容の中では着目対象は変化できないから、C類従属節を用いた文は二つの着目対象とその記述内容を持つと考える。第四層の時相状況間の関係と第五層の発話単位同士の関係は、B類従属節を用いた文の記述であるか、C類従属節または接続詞を用いた文の記述であるかの違いとして区別される。したがって、(1)の記述と(2)の記述の違いは、意味表現の階層の第四層の記述と第五層の記述の違いとして表現することができる。

また、(3)では、話し手の注意は「イラクがクウェートに侵攻したこと」や「西側がイラクを非難したこと」に現れる概念以外の対象に置かれていて、(3)の文全体はその着目対象(例えば、「90年には」)についての記述である。このように、(1)(2)と(3)の違いは、話し手の着目対象の違いとして意味表現の階層の上で表現することができる。

次に、(6)(7)のように一文内に複数の着目対象が含まれる場合について述べる。

- (6) セロリは根元は横にスライスする
- (7) 根元はセロリは横にスライスする

(6)では、話し手はまず“セロリ”を着目対象としていて、「根元は横にスライスする」はその記述内容である。さらにその記述内容自身は“セロリの根元”を着目対象として、「横にスライスすること」を記述内容としている。この構造を以下のように表現する。

[着目対象 = セロリ,  
記述内容 = [着目対象 = セロリ. 根元,  
記述内容 = スライスする [向き = 横, object = セロリ. 根元]]]

これに対して、(7)は逆に、まず一般の根元について着目し、次にセロリの根元に着目している。これを以下のように表現する。

[着目対象 = 根元,  
記述内容 = [着目対象 = セロリ, 根元,  
記述内容 = スライスする [向き = 横, object = セロリ, 根元]]]

このように着目対象と記述内容の組の構造を表現することによって、(6)と(7)の着目対象の違いを表現することができる。

以上で述べてきたように、五つの階層を設定することによって、(1)～(3)の従属節の分類による接続表現の違いや、(4)、(5)の従属節の時相の違い、(6)、(7)の着目対象の構造を表現することができる。(1)と(2)の違いは第四層と第五層の階層の違いで表現され、(1)(2)と(3)との違いは第五層の着目対象の違いによって表現される。(4)と(5)の時相の違いは、第三層と第四層の階層の違いで表現される。(6)と(7)では第五層における着目対象と記述内容の組の構造が異なる。

伝統的に生成処理過程は、主に生成すべき文の内容を決定するプランニング部と、その内容を適切に表現する表層文生成部に分かれる。本稿では表層文生成部について述べたが、プランニング部については触れていない。階層化された意味記述は表層文生成部に対する入力であるかのように見えるが、以下の点において、意味表現の階層は通常の入力とは異なる。

第一に、意味表現の階層における記述は部分的に与えてもかまわない。第五層では、着目対象と記述内容の他に話し手の視点、現在の話題などを記述するが、これらの着目対象、話題、視点などは必ずしも値がわからなくても、直前の発話と同じ値を用いるなど、文脈情報から推測して決定することができる。第二に、意味表現の階層が文の内容の決定に影響を与える場合がある。例えば、話題、視点、着目対象はできるだけ保持する、あるいは、できるだけ関連のあることを続けて発話する、などの制約が考えられる。以上のように、文の内容を決定する過程と内容を表現する過程を統合することにより、両者の処理過程を柔軟に行うことができる。

#### 4 おわりに

生成処理過程におけるプランニング部と表層文生成部との間で、複雑な情報を伝達するために、統一的な制約と意味的、文脈的な制約を統合した五つの階層からなる意味表現を設定した。このことにより、複雑な文を生成するために必要な多様な情報を、複雑な中間構造を用いることなく、自然に簡潔に記述することができた。

現在、この枠組に基づいて生成システムを開発し、実験を行っている。意味表現の階層の記述には ICOT で開発中の知識表現言語 *QUILT* [16] を用いている。また、文の内容の決定には、「イラクのクウェート侵攻」に関する事実から、いくつかのルールを用いて文章の構造を生成するモジュールを用いている。

本稿では、否定表現については述べていない。これをどの階層でどのように記述するかについてを検討する必要がある。また、話し手の聞き手に対する態度は第五層で記述するが、どのように記述しどのように処理するかについては考えていない。話し手の態度の意味表現は、ある文が疑問に使われるか、命令に使われるかなどの違いを表すものであり、いわゆる疑問文、命令文などの生成に必要である。生成処理過程についても、例題は比較的単純な文でできており、例題に現れない、多様な従属節、使役、可能などの態、複雑なアスペクト、多様な話し手の判断を表すムードなどを用いた文の生

成が、この枠組の上で行えることを確かめる必要がある。

## 参考文献

- [1] Douglas E. Appelt. Planning natural-language referring expressions. In David D. McDonald and Leonard Bolc, editors, *Natural Language Generation Systems*. Springer-Verlag, 1988.
- [2] Laurence Danlos. Conceptual and linguistic decisions in generation. In *the Proceedings of 10th COLING Conference*, 1984.
- [3] Barbara J. Grosz and Candace L. Sidner. Attention, intentions, and the structure of discourse. *Computational Linguistics*, Vol. 12, No. 3, 1986.
- [4] 橋田浩一. 制約と言語. ディスコースと形式意味論ワークショップ, 1989.
- [5] Eduard H. Hovy. *Generating Natural Language under Pragmatic Constraints*. Lawrence Erlbaum Associates, Publishers, 1988.
- [6] 池田尚志. 助詞「が」の働きについて - 認知的なレベルからの考察 -. 電子情報通信学会論文誌, Vol. J72-D-II, No. 11, pp. 1904-1909, 1989.
- [7] 金田一春彦. 日本語動詞のテンスとアスペクト. 日本語動詞のアスペクト. むぎ書房, 1976.
- [8] 益岡隆志, 田窪行則. 基礎日本語文法. くろしお出版, 1989.
- [9] K. R. McKeown and W. R. Swartout. Language generation and explanation. In Michael Zock and Gérard Sabah, editors, *Advances in Natural Language Generation*, volume 1. Ablex Publishing Corporation, 1988.
- [10] 南不二男. 現代日本語の構造. 大修館書店, 1974.
- [11] 野口直彦, 鈴木浩之. 「が」と「は」の語用論的機能について. 言語理解とコミュニケーション研究会 *NLC90-16*, 1990.
- [12] 佐野洋. 述部の階層分析と文脈情報. 談話理解モデルとその応用シンポジウム論文集. 情報処理学会, 1989.
- [13] 田野村忠温. 現代日本語の文法 I 「のだ」の意味と用法. 和泉選書, 1990.
- [14] 東条敏. Temporal situations and an analysis of verb aspects. Draft, 1990.
- [15] 渡辺実. 国語構文論. 塙書房, 1971.
- [16] H. Yasukawa and K. Yokota. An overview of a knowledge representation language *QUIXOTE*. Draft, 1990.