ICOT Technical Report: TR-690

TR-690

意味表現の階層に基づく
自然言語生成システム

池田　光生

September, 1991

© 1991, 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自然言語生成システム

Natural Language Generation System based on a Semantic Representation Hierarchy

池田光生

(財) 新世代コンピュータ技術開発機構

IKEDA, Teruo

Institute for New Generation Computer Technology

E-mail: ikeda@icot.or.jp

要旨

多くの生成システムでは，プランニング部と表層文生成部との間で中間構造を介して情報の伝達を行う．生成処理が多様化するにつれて複雑になる中間構造を，簡潔に記述するために，五つの階層からなる意味表現を導入した．各階層は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．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自然に明示することができる．

1 はじめに

これまで自然言語処理の研究は解析を中心に行われてきたが，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取られている．

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を取ることが多い．多くの場合，自然言語生成システムでは，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，また，両者の間では一つの文の表現に対応する中間構造を介して情報を伝達することが多い．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を持ち，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，人間の発話のような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する．その場合には単純な構造を持つ中間構造では，生成に必要な情報を十分に記述することができない．

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る．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自然に明示的に表現することができる．

次章では，よりきめ細かい生成処理をしようとする場合に，中間構造が非常に複雑になってしまうことを示す．そのために，助詞の「は」と従属節の接続表現との関係を例としてあげる．これにより中間構造に必要な条件を明らかにする．三章ではその条件を満たすための基盤となる意味表現の階層構造について述べて，二章であげた例文が意味表現の階層を用いて簡潔に表現できることを示す．最後に四章でまとめ，本稿の意味表現の階層とプランニング部との関係について述べる．また，話し手の態度の表現などの課題や今後の研究の方向について触れる．

2 中間構造に必要な条件

多くの自然言語生成システム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というモジュール構成を取っている．したがって，あるオブジェクトが主題化や取り立ての対象となっていることや，従属節の接続表現や文の統語構造を決め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．

生成の多くの処理では文脈情報を考慮しなければならない．また，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．したがって，文脈に応じて多様な表現を生成する場合には，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．従来の生成で主題を扱う場合には，述語を中心とした格構造中の格要素に対して主題や取り立てなどのマークを付け，統語構造の生成過程で助詞「は」を付加して，その格要素を文頭に置くという処理を行うことが多い．しかしながら，“主題”や“取り立て”などの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，以下で示すよ

うに従属節の接続表現や統語構造を生成する上で大きな役割を果たしている．そのため，話し手の着目対象を表す助詞「は」の複雑な役割を表すためには，格要素に対する主題などのマークだけでは不十分である．

助詞「は」の役割を示すために，南 [14] 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 (図 1).

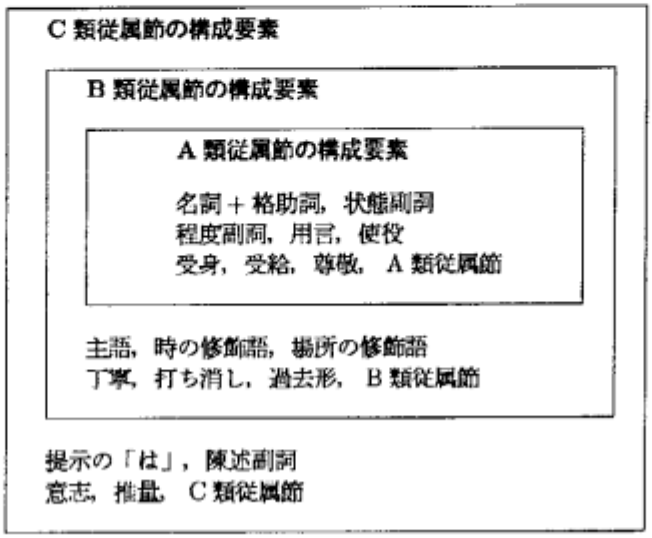

図 1: 南の従属節の分類

A 類従属節はテンス・アスペクトを持たない述語的部分や，状態，程度の副詞などだけを含む従属節で，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，意志や推量を含む述語的部分と陳述副詞や提題の「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，三通りの従属節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) で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” はいずれもイラクについての記述であり「イラクは」という語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となっていることを示すために，中間構造中の従属節に対応する部分に理由などとマークしておかなければならない．しかし，「ために」「から」など理由を表す他の接続表現との区別を行なうためには，さらに別の情報が必要である．例えば (1) では，話し手の着目対象であるイラクについての記述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，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．(1) の従属節が B 類従属節であることを知るためには，従属節がテンス・アスペクトを持っていることや，話し手の着目対象についての記述が主節を含めた文全体であることなどを中間構造中に記述しなければならない．

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についての記述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．(2) がC類従属節であることを知るためには，話し手の着目対象が従属節と主節とで独立に存在することを表現しなければならない．

このように，(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ことや，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を示す必要がある．

(3) はイラクや西側についての記述ではなく，二つの事柄を述べることにより何か別な着目対象 (例えば「これまでの経過は」，「1990 年には」など) についての記述を行っている．このことは「イラク」および「西側」に助詞「が」が用いられていることによって示される．(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．

中間構造の記述およびその処理の複雑さを示す例として，筆者らが以前に開発した文生成ツール [9] における入力形式を用いた例文 (1)，(2) の記述を図 2に示す．

例文 (1) の記述

```
{関係 /{語彙 / 非難},
 テンス / 過去,
 ヴォイス / 受動,
 主題 / イラク,
 ロール /{
   行為者 /{語彙 / 西側},
   対象 /{語彙 / イラク},
   理由 /{関係 /{語彙 / 侵攻},
          テンス / 過去,
          ヴォイス / 能動,
          主題 / イラク,
          ロール /{
            行為者 /{語彙 / イラク},
            対象 /{語彙 / クウェート}}}}}
```

例文 (2) の記述

```
{関係 /{語彙 / 非難},
 モーダル / 必然,
 ヴォイス / 能動,
 主題 / 西側,
 ロール /{
   行為者 /{語彙 / 西側},
   対象 /{語彙 / イラク},
   順接 /{関係 /{語彙 / 侵攻},
          テンス / 過去,
          ヴォイス / 能動,
          主題 / イラク,
          ロール /{
            行為者 /{語彙 / イラク},
            対象 /{語彙 / クウェート}}}}}
```

図 2: 従来の文生成ツールの入力形式による例文 (1) と (2) の意味表現

この形式では，入力となる文の意味を“属性名 / 属性値”という対の集合を用いて表現する．(1) では，理由という属性が与えられた従属節内の主題属性の値が，主節の主題属性の値と同一であることから，この従属節をB類従属節であると判断し，「ので」という接続表現を生成することができる．同様に (2) では，主節と従属節の主題属性の値が異なることから，従属節がC類従属節であると判定できる．

しかし，全体の構造が一様であり，従属節の統語的な構造の違いを区別して書くことができない．そのため，生成処理過程で意味表現中のさまざまな情報を参照しながら，統語的な構造の違いを判定していかなければならない．

次に従属節の持つ時相について考える．次の例文 (4) と (5) は事実的内容は同じだが，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，進行を表す「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．テンス・アスペクトを含まず従属節の内容と主節の内容とが同一の時相で成立する従属節は，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っていることを表すために，従属節に相当する中間構造中に同時などとマークしておかなければならない．しかし，(4) と (5) の接続表現を生成するためには，同時であるという情報の他に，従属節と主節のそれぞれの時相についての情報と，(4) ではそれぞれの時相が同一であり，(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) では話し手の着目対象が一文内に複数現れている．話し手の心的状態においてはそれらの着目対象間には構造があり，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，これらの着目対象間の構造が (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」という表現は「(一般的に) 根元について言うとセロリの場合には」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て次にその根元について述べているなどの，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，同時に統語構造や従属節の接続表現をも決定することである．また，主題化や取り立ての対象は，話し手の着目対象の移り変わりなど文脈的情報によって生じる構造を反映しなければならない．さらに，主題化や取り立ての処理と，省略，代名詞化，態変化，接続表現の決定などの処理と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．

従来，文生成用中間構造として用いられてきた属性集合などの形式では，比較的単純な文の意味を記述するには十分であるが，複雑な文を生成しようとすると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，中間構造の記述や，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が困難になる．

以上で述べたような中間構造に必要な条件を満たすためには，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定した．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による従属節の構造の違いを，システムが規定する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その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

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する考え方が多く見られる．渡辺 [20] 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 [14] 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．

意味表現の階層を用いた文の記述を図示したものを図 3 に示す．記述例は「クウェートはイラクに攻撃されていた」という文の意味表現である．

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”という概念の agent ロールが“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・アスペクトの値を持たない．

第三層で記述する時相状況とは，あることがらを時間的視野・視点によって時間的に相対化した状況として見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他 [10][18] による．この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間的な情報を記述する．

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によって表される．B 類従属節とは，提題の「は」や話し手の態度の表現などを含まない従属節である．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について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と比べて，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により，様々な接続表現を生成することができる．

以下では，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，意味表現の階層を用いた記述を示す．

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 [**agent**= **イラク** , **goal**= **クウェート**].
非難 [**agent**= **西側** , **object**= **イラク**].

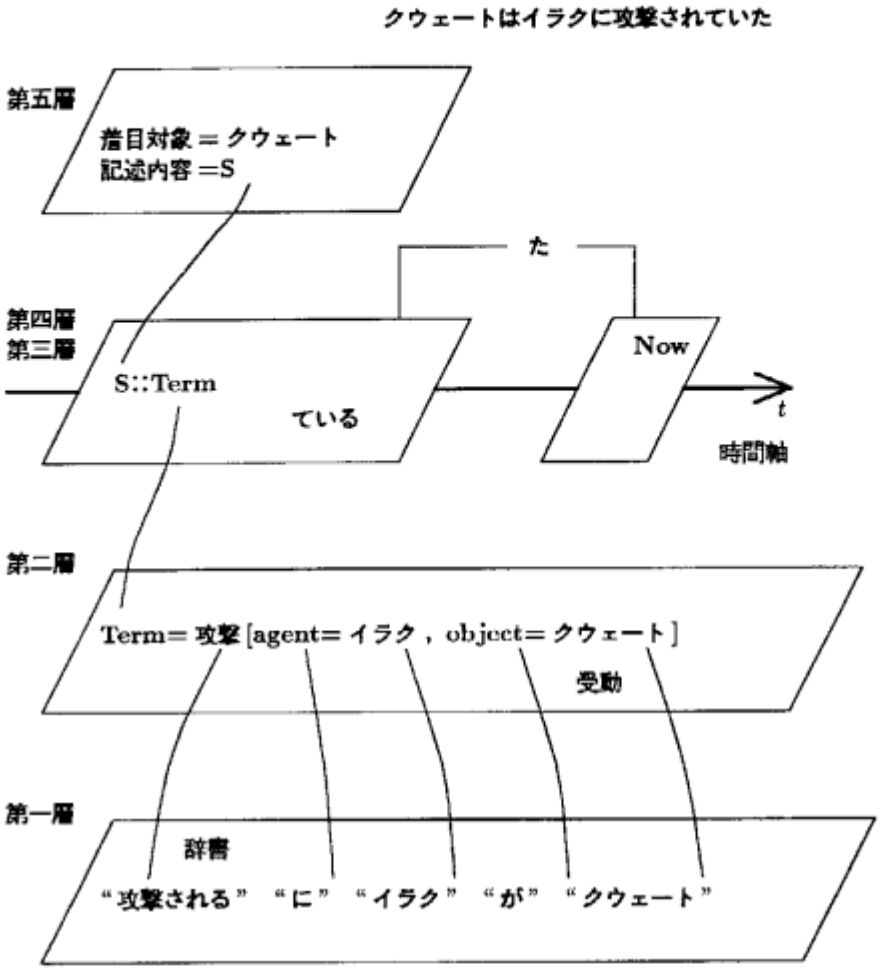

図 3: 意味表現の階層を用いた文の記述

上記の例は，“侵攻する”，“非難する”という動詞概念を表したものである．第二層の項の記述では，動作や状態を表す概念と事物を表す概念との区別はしていない．ある項が用言に対応するか体言に対応するかは，意味表現の階層による記述から表層文を生成する過程で，統語的に決定する．第二層の記述のうち，用言に対応する項はヴォイスの値を持っている．しかし，時間の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[1]．

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．第三層では，項が成立する時相状況を考えることにより，これにテンス・アスペクト情報を付加する．時相状況とは，ある事実的内容に対して時間的視野・視点を与えることにより時間の相対化を図ったものである．言語化しようとする内容のどの部分を，どこからどのように見ているかといった情報を時間的視野・視点とみなし，これらの情報を時相状況に持たせることによりテンス・アスペクトを表現する [10][18]．

[1] ここでは，事実的内容はテンス・アスペクトを含まないことにする．

例えば，「イラクがクウェートを攻撃する」という第二層の記述が，「ている」というテンス・アスペクト表現に対応する時相状況で成立することを以下のように記述する．

S[fov= 継続 , pov= 非過去]:: 攻撃 [agent= イラク , object= クウェート].

"::" の左側が，「イラクがクウェートを攻撃する」という事態に時間的な幅を与えるための時相状況の記述である．"::" の右側にはその時相状況で成立する第二層の項を記述する．時相状況は S[...] の形式で表現し，"..." の部分に時相状況が持つ属性を書く． fov (field of view) という属性は時間的視野を表し， pov (point of view) という属性は時間的視点を表す． fov の値が継続ならば，「ている」が生成され， pov の値が過去ならば，「た」が生成される． Field of view は言語化しようとする事象や行為を継続的な視野で見るか非継続的な視野で見るかを設定し， point of view は言語化しようとする事象や行為を心的に完了したと捉えるか，未完了と捉えるかを設定する．現在，第三層で用いる時相状況には四種類あり，「る」「た」「ている」「ていた」という表現に対応している．時相状況によって表された時間的視野・視点によって言語表現を決定するという点において，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ない．

この時相状況を用いて，

(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・アスペクトを含ま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従属節である．同一の時相状況内に複数の項を記述するために，時相状況の属性として id を与え， id の値が同一の時相状況は複数の項を同時に成立させる単一の状況であると考える．以下のように "::" の左側が同一の時相状況であるような複数の時相状況記述によって，右側の項がすべて同じ時相であることを表現できる．ここでは記述を簡単にするため， S[id=1 , fov= 継続 , pov= 非過去] という時相状況を S_1 で表している．同一の時相状況内に複数の項を記述することによって，それらの項が同じ時相を持つことを表現する．これらの項は A 類従属節を用いて表現する．

S_1=S[id=1 , fov= 継続 , pov= 非過去].

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

S_1:: うかがう [agent= イラク , object= 西側の動き].

上記の時相状況は「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[2]．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，時相状況間の関係を表す第四層の項として記述する．

S_1=S[id=1 , fov= 継続 , pov= 非過去].

S_2=S[id=2 , fov= 継続 , pov= 非過去].

include[1=S_1 , 2=S_2].

[2] ここで用いる意味表現では "イラク" という名前は同一のものを指している．仮に "イラク" という表現で指されるものが複数ある場合には別々の意味表現を用いる．

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].

include という名前の項は二つの時相状況の関係を表し，第四層の記述であることを示している．第四層の記述であることを示す項は二つの時相状況間の関係を表す[3]．第四層の項として，include の他に，因果関係を表す cause，時間的前後関係を表す t_prec などを用意している．

(4) は第三層で同一の時相状況中に記述され，(5) は第四層で時相状況間の関係として記述される．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の対象についての記述の組を表現する．話し手の注意が置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いて，その対象について何らかの記述を行うと考える．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．

着目対象と記述内容の組が発話の基本単位となり，複数の組を一列に並べることによって文章を表現する．第五層では，文章を構成する個々の発話について，着目対象と記述内容の組の表現と共に，直前の組に対する接続表現を記述する．第五層で記述する接続表現には，「が」「けれど」など C 類従属節に用いる接続表現や「したがって」「しかし」など文頭に置かれる接続表現がある．他の発話との関係を表す語句の中には「ところで」「第一に」「さきほど」など，意味的には直前の発話との関係ではなく談話構造を反映する関係を表すものが存在する．しかし，意味表現の階層では文章は構造ではなく発話の一次元の列として表されるから，文章を形成する発話列はその順序がすでに決められ，他の発話との関係を表す語句はすでにその表現が定まっているものと考える．

着目対象とその記述内容という考え方を用いて，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ことができる．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．

二章で例文 (1) ～ (3) を用いて示したように，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の記述は一つの着目対象についての記述であり，これは B 類従属節を用いた文に対応するため，「ので」という B 類従属節の接続表現が使われている．B 類従属節を用いた文では，従属節および主節を表す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．

(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．B 類従属節を用いた文は着目対象が文全体を支配するのに対して，C 類従属節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．ひとつの記述内容の中では着目対象は変化しないから，C 類従属節を用いた文は従属節と主節のそれぞれが着目対

[3] 第四層の項は時相状況を引数として持つが，cause[1=cause[1=... , 2=...] , 2=...] のように第四層の項をネストしてもよい．

象とその記述内容を持つと考える．ある接続関係が第四層の接続関係であるか第五層の接続関係であるかは，B類従属節を用いた文であるか，C類従属節または接続詞を用いた文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．(1)，(2) の中間構造の違いを図4に示す．(2) では従属節および主節を表す着目対象と記述内容の組の間に「が」という接続関係がある．

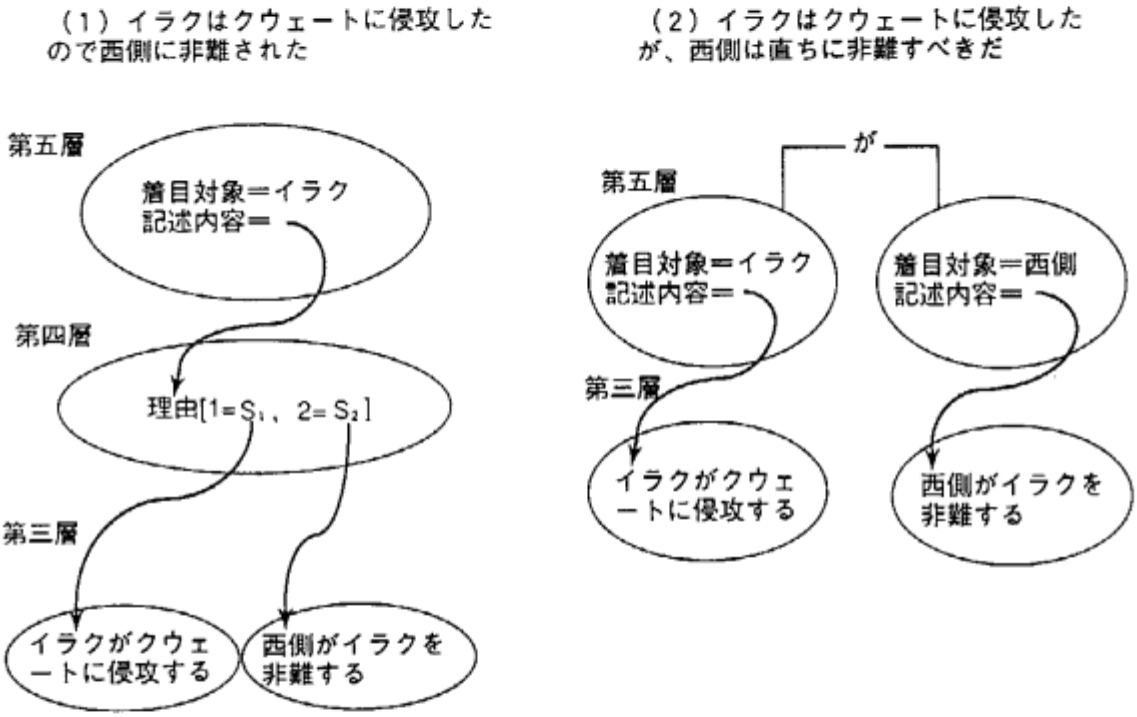

図 4: 例文 (1) と (2) の中間構造の違い

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”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[4]．

[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 [向き = 横，
object= セロリ.根元]]].

第五層で記述する項は着目対象，記述内容というラベルを持っている．記述内容というラベルの値として，二重に第五層の項を書くことによって，「は」を二個含む文を表現する．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 [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．

4 おわりに

生成処理過程において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，統語的な制約と意味的，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，多様な接続表現を生成するために必要な情報を自然に簡潔に記述することができた．

意味表現の階層は，日本語文を生成する上で必要な，日本語の持つ統語的制約を中間構造中に明示的に表現するために設定したものである．日本語には文法的な制約が少なく，日本語文の生成は，異なるレベルの情報を担う短い形態素を次々に連結していくことによって行われる．意味表現の階層は，このような特徴を持つ日本語文生成に適している．

また，プランニング部と表層文生成部との結合という点で，本稿で示した意味表現の階層は重要な役割を果たす．第五層では，話し手の着目対象と記述内容という概念を記述するが，これは生成プランニングにおける視点や話題などの文脈的制約と深い関係がある．着目対象と記述内容を考えることによって，談話構造生成を中心としたプランニングと，統語構造生成を中心とした表層文生成とを自然に結び付けることができる．

現在，この枠組に基づいて日本語文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 *QUIXOTE*[21][22] を用いている．*QUIXOTE* が持つ知識表現を階層化する機

[4] 実際の記述はもっと複雑であるが，(6)，(7) の違いの説明のために簡略に記述している．

能を用いることにより，意味表現の階層の記述をシステムとして規定している．表層文の表出も *QUIXOTE* を用いて規定された階層に基づいて行われる．また，システムの信念を聞き手に主張するための文章の構造を生成するプランニングモジュールを用いて，プランニング部との結合を行っている．プランニング部で文章の構造を生成し，次にその文章構造から一文ごとの区切りとそれらの順序を決め，一文ごとに意味表現の階層を用いて表現する．実際に文章構造生成モジュールから得られた中間構造と生成結果を図5に示す．

```
m_content(id=1,sid=1):: 奪う (object= クウェート,source= オスマントルコ,inst= 武力).
m_sb::t_prec(1=m_content(id=1,sid=1),2=m_now).
m_content(id=1,sid=2):: 正当 (cont=
                              m_judge(cont= 属す (object= クウェート,goal= イラク))).
m_sb::include(s1=m_content(id=1,sid=2),s2=m_now).
m_gen(id=1)::top(first=top(attention= クウェート,
                           content=m_content(id=1,sid=1)),
                 second=top(attention= クウェート,
                            content=m_content(id=1,sid=2), connective= から)).
```

クウェートは武力でオスマントルコから奪われたから，オスマントルコに属すべきだ．しかし，オスマントルコは現存しない．しかし，イラクはオスマントルコを受け継いでいる．したがって，クウェートはイラクに属すべきだ．したがって，イラクのクウェート攻撃は正当だ．

図 5: 文章構造生成モジュールから得られた中間表現の一部と生成結果

文章構造から一文ごとの区切りを決める処理は，話し手の着目対象とその記述内容をどのように推移させていくかということや，接続表現をどのように決定するかという処理と密接な関係がある．したがって，文章構造と本稿の意味表現の階層との関係をさらに検討する必要がある．

また，話し手の聞き手に対する態度についての情報を，どのように記述し，どのように処理するかについて検討する必要がある．意味表現の階層では，第五層，あるいはより上位の層で記述する．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．また，着目対象と同じようにプランニング部との関係が深いため，意味表現の階層中に話し手の態度についての情報を表現することにより，プランニング部と表層文生成部とをより自然に結合することができる．

謝辞

本研究に有益なコメントをくださいましたICOTの安川秀樹，佐野洋の両氏，三菱総合研究所の東条敏氏，KDDの福島秀穎氏に感謝いたします．本稿作成にあたっては松下電器産業の鈴木浩之氏，ICOTの田中裕一第六研究室長に助言をいただきました．

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 the International Conference on Computational Linguistics*, 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] E. H. Hovy. Pragmatics and natural language generation. *Artificial Intelligence*, Vol. 43, pp. 153–197, 1990.

[6] Eduard H. Hovy. Integrating text planning and production in generation. In *the Proceedings of the International Joint Conference on Artificial Intelligence*, 1985.

[7] Eduard H. Hovy. *Generating Natural Language under Pragmatic Constraints*. Lawrence Erlbaum Associates, Publishers, 1988.

[8] 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.

[9] 池田光生, 幡野浩司, 福島秀顕, 重永信一. 汎用日本語処理系 (LTB) 文生成部の生成方式. 日本ソフトウエア科学会第 5 回大会論文集, 1988.

[10] 池田光生, 安川秀樹, 東条敏. 心的時間の分割とテンス・アスペクト表現. 情報処理学会 自然言語処理研究会 *NL-83-9*, 1991.

[11] 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.

[12] 益岡隆志, 田窪行則. 基礎日本語文法. くろしお出版, 1989.

[13] 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.

[14] 南不二男. 現代日本語の構造. 大修館書店, 1974.

[15] 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.

[16] 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.

[17] 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.

[18] 東条敏, 池田光生, 安川秀樹. 時相状況の構造と表現. 90 年度 ICOT NLU-PSG-STS-ETR 合同ワークショップ, 1991.

[19] 徳永健伸, 乾健太郎. 1980 年代の自然言語生成. 人工知能学会誌, Vol. 6, No. 3–5, 1991.

[20] 渡辺実. 国語構文論. 塙書房, 1971.

[21] H. Yasukawa and K. Yokota. Labeled graphs as semantisc of objects. 情報処理学会, データベースシステム研究会, 1990.

[22] 横田一正. Qu'xote: an adventure on the way to dood. 90 年度 ICOT NLU-PSG-STS-ETR 合同ワークショップ, 1991.