HOKKAIDO UNIVERSITY

Title	汎用文章生成システムによる日本語主題表現の生成方略の実現
Author(s)	高橋, 晃; 桃内, 佳雄; 宮本, 衛市
Citation	北海道大學工學部研究報告, 140, 133-142
Issue Date	1988-05-30
Doc URL	https://hdl.handle.net/2115/42097
Type	departmental bulletin paper
File Information	140_133-142.pdf

Hokkaido University Collection of Scholarly and Academic Papers : HUSCAP

北海道大学工学部研究報告
第140号（昭和63年）

Bulletin of the Faculty of Engineering,
Hokkaido University. No. 140 (1988)

汎用文章生成システムによる日本語主題表現の生成方略の実現

高橋　晃　　桃内　佳雄　　宮本　衛市

（昭和62年12月26日受理）

An Implementation of Strategies for Generating Topic Expressions in Japanese Text by Generalized Text Generator

Akira TAKAHASHI, Yoshio MOMOUCHI and Eiichi MIYAMOTO

(Recieved December 26, 1987)

Abstrast

The need for good text generation is rapidly increasing in man-machine interface through natural language in various systems such as database systems, expert systems and machine translation systems. A sequence of sentences must exhibit cohesion to be considered a well-formed text. Ellipsis of topic expressions is one of the most important devices to create cohesive Japanese texts. To decide how to eliminate a topic expression, key factors such as focus, coherence, perspective and text style must be considered effectively through the process of generation. In this paper, we argue about strategies for generating topic expressions using these factors. We also describe the implementation of a computer text generation system Hi-GTG (Hokkaido University Information Engineering-Generalized Text Generator), designed to incorporate our strategies for controlling the generation of the topic expressions.

1．はじめに

文章生成は，人間の発話行為のモデルという見地からも興味深い研究対象であるが，また実用的なマン・マシン・インタフェースの構成における重要な構成要素でもあり，その研究の有用性は広く認められている。

文章生成における重要な課題のひとつとして，生成文章の質の向上の問題がある。生成文章の質の向上を考慮して文章を生成するシステムの開発における重要な問題は，

(1) 生成文章の質の評価をどのように行うか。

(2) 生成文章の質を向上させる処理方略をどのように実現するか。

の2点である。

文章の質を評価する指標として，文章の曖昧性，冗長性，明瞭性，統一性，調和性，目的適合性などがある。われわれは，これらの指標の全ての評価に関連し，基礎となる性質として，文章の結束性に着目した。従って，質の高い文章を生成するためには，生成される文章の結束性を考慮することがきわめて重要であると考える。また，文章の結束性を保持するための処理は，文章

情報工学科　言語情報工学講座

生成の様々なレベルで考慮しなければならず，結束性を実現する具体的な言語的手段も多岐に亙るため，その処理は包括的に見通し良く構成しなければならない。われわれは，そのために文章生成を生成規則により形式的に記述し，そのなかに結束性の処理方略をモジュール化して効率よく埋め込むことのできる方法を採用し，その実現のためのシステムとして汎用文章生成システムを作成した。

1.1 結束性を考慮した文章生成

文章は単なる文の並びではなく，各文の間には有機的なつながり，すなわち結束性(cohesion)が必ず存在する。われわれは，結束性を適切に考慮することが生成文章の質を向上させると考える。この結束性には以下の2つのレベルが存在する。

(1) 文章中の表現に陽に現れる表層的なつながり(連結性：connectivity)

(2) 文章中の表現には陽に現れない意味的なつながり(首尾一貫性：coherence)

(1)のレベルの結束性は，省略(ellipsis)，指示(reference)，接続(conjunction)，語彙結合(lexical cohesion)等の言語的手段を用いて達成される。

(2)のレベルの結束性は，文相互の意味的関係，論理的関係，あるいは，先行文と後続文の名詞表現の指示対象の間の連合関係，または前後の文脈の間接的な関連性や推論によって達成される[8]。

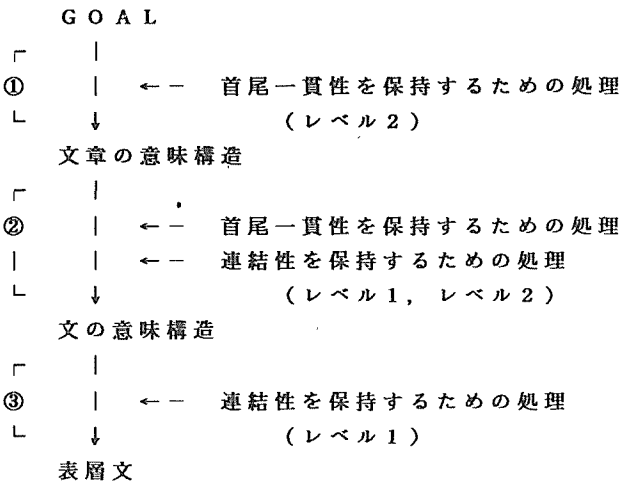


図1 文章生成モデル

生成文章の結束性を考慮するためには，(1)の表層レベルでの結束性と，(2)の意味レベルでの結束性の両方を考慮しなければならない。われわれが想定している文章生成モデルの全体構成を図1に示す。これまでに開発された，質の向上を考慮して文章を生成するシステムは，Appelt[1]のようにその出発点を発話のゴールに求め，意味レベルのつながりを重視するものと，Granville[2]やDerr & Mckeown[3]等のように，あらかじめ生成する文章の意味構造を与え，意味レベルでのつながりを保証したうえで，表層レベルのつながりの処理を重視するものに分けられる。

Derr と Mckeown は，複文や単文などの文構造の選択を適切に行うことにより生成文章の質の向上を図り，PROLOG の DOG 形式を用いてシステムを記述した。しかし，このシステムはデータベースに対する質問応答システムの応答生成部のためのシステムであり，生成文章は2，3文からなる比較的小さなまとまりの文章に限定されている。また，Appelt のシステムでは表層レベルの結束性を考慮する処理は行われていない。一方，Granville により開発されたシステム PAUL は，拡張された文脈自由生成規則である APSG(Augmented Phrase Structure Grammar)[4][5] により文章生成アルゴリズムを記述するもので，その記述の中に，焦点情報の予測を行う処理や，焦点情報と名詞句の距離を測る処理を埋め込み，文章中の適切な語彙間の結合を考慮して文章生成を行うシステムである。PAUL は文章の質を結束性という観点から択えたシステムであるが，結束性の考慮の処理は，図1における③の連結性のレベルでの，語彙間の結合の処理のみで，文

章全体におよぶ結束性の考慮の処理は行われていない。

1.2 本研究の目標

本研究は，PAUL においては実現されなかった，図1における②のレベルからの結束性の考慮の処理を行おうとするものである。また，それらの文章の結束性を考慮する処理としては，主題表現の省略に焦点をしぼって考察した。主題表現の省略は，日本語文章において非常に多く生じる省略のタイプであり，文章の結束性を保持するために基本的な役割を果たしているものである。結束性を考慮する処理を実現する方法としては，PAUL で用いられた方法を踏襲して，APSG により生成規則を形式的に記述して文章生成を行う汎用文章生成システム Hi-GTG(Hokkaido University Information Engineering-Generalized Text Generator)を作成した[16)17)]。

2．主題表現生成の制御

日本語文章における主題表現の省略は，文章の結束性を保持するための重要な手段である。日本語主題表現の省略の一般的に原則は，久野[9)]，畠[10)~12)]，永野[13)]，寺島[14)]等で論じられているが，それらを参考にしながら，本章では，主題表現の生成を制御する要因について検討し，さらに，1.1で示した文章生成モデルに沿って主題表現生成の制御方略を構成する。ここで検討する主題表現の生成を制御する要因は，一般に文章生成を行うときに必要な要因であり，特に，主題表現生成のために必要なものとは仮定していない。

2.1 日本語主題表現の出現パターン

日本語の主題表現は，新しい主題については「…が」という表現で表され，既知の主題については「…は」という表現で表されることが多い。さらに，同一の主題が繰り返される場合には，繰り返される主題表現を省略することにより文章の結束性を高めている。以下では，文の主題となる要素はその意味構造においてその文の焦点情報として示されると考え，さらに主題表現の出現パターンを，主題導入，主題維持，主題変化，主題回復の4つに分類して考える。それらの4つのパターンの各々について，主題表現をどのように表現するか，あるいは省略するかを決定する要因については次節で考察することとして，以下では教育出版しょうがくこくご1上の“むしのはなし”を例として，主題表現の省略によって文章の結束性が強化され，文章の質の向上が図られる様子を示すことにする(図2)。

図2(a)について説明する。ϕ の部分が主題表現が省略されている所である。この例では，文1で主題“あげはちょう”が導入され，文2から文4までその主題が維持される。文5で主題は“たまご”に変化し，文6では主題は“あげはちょう”に回復している。ただし，文1から文4までの主題は，インスタンスとしてのあげはちょう，文6の主題はクラスとしてのあげはちょうであるとも考えられる。

1．あげはちょうが，みかんのきにとんできました．
2．φ　ときどきおなかのさきをまげて，はに，なにかつけています．
3．φ　いったい，なにをしているのでしょう．
4．φ　たまごをうんでいるのです．
5．φ　うすいきいろのちいさなたまごです．
6．あげはちょうは，いちまいのはに，たまごをひとつずつうみます．

（a）むしのはなし（一部）［主題表現省略あり］

1．あげはちょうが，みかんのきにとんできました．
2．あげはちょうはときどきおなかのさきをまげて，はに，なにかつけています．
3．あげはちょうはいったい，なにをしているのでしょう．
4．あげはちょうはたまごをうんでいるのです．
5．たまごはうすいきいろのちいさなたまごです．
6．あげはちょうは，いちまいのはに，たまごをひとつずつうみます．

（b）むしのはなし（一部）［主題表現省略なし］

図2　むしのはなし

このような短い文章においても，もし主題表現の省略が行われなければ，図2(b)のように文章はひどく間延びしたものになり，結束性と言う観点から文章の質が低下することは明らかである。

2.2 主題表現生成を制御する要因

文章の主題表現を生成する要因となる情報を次の4種類に分類して考えた。

（1） 文章の意味構造レベルで記述される情報

文章のスタイル(STYLE)：文章のスタイルとしては，説明文，物語文，談記文，日話文等があると考えられる。

（2） 段落の意味構造レベルで記述される情報

大域的焦点(GFOCUS)：段落全体に及ぶ発話者の話題の中心であり，1段落に一つ存在すると考えられる。

（3） 文の意味構造レベルで記述される情報

直接的焦点(FOCUS)：文における話題の中心であり，1文に一つ存在すると考えられる。文の主題となる要素はこの焦点として示されると考える。

視点(PERSPECTIVE)：文の発話者の視点である。文中で言及される各要素と発話者の距離関係が問題となる。

強調(STRESS)：文中で特に強調する要素である。

場面(TIME, SPACE)：文の背景となる場面情報。

文間の連接関係(COHERENCE)：段落中での文と文の意味的なつながりである。順接，逆接，添加，対比，転換，同列，補足，連鎖等があり，さらに細かく分類できる[15]。なお，“むしのはなし”の文章の意味構造に現れる連接関係を表1に示す。

（4） 表層文が生成される時に生じる文脈的な情報（1），（2），（3）のそれぞれの履歴と表層レベルにおける名詞表現の履歴を主題表現生成のための文脈的な情報と考える。

表1 文間の連接関係(一部)
Table 1 A part of coherence relations.

TOP	文頭
ATT	添加・単純
SEQ	添加・継起
REL	連鎖・連係
SPP	補足
TUR	転換
ANS	（疑問に対する答えとしての） 連鎖・補足説明
TRA	（新しい話題への）転換・推移

2.3 主題表現生成の制御方略

主題表現生成の制御は，2.2のそれぞれの情報を文脈情報としてリストに保持し，それらのリストを参照したり，操作したりすることによって行う。現在考えている文脈情報リストを表2に示す。このうちSTYLEには文章全体レベルの情報が格納され，GFOCUSには段落レベルの情報が格納される。FLIST, PLIST, CLIST, TLIST, SLIST, NLISTには，文レベルの文脈情報の履歴が格納される。また，リストに対する操作としては表3のようなものを考えている。

表2 文脈情報リスト
Table 2 Context information lists.

STYLE	文章のスタイル
GFOCUS	大域的焦点（global focus）
CLIST	文間の連接関係リスト （coherence list）
FLIST	直接的焦点リスト（focus list）
SLIST	空間的場面リスト（space list）
TLIST	時間的場面リスト（time list）
PLIST	視点リスト（perspective list）
NLIST	名詞リスト（noun list）
STRESS	強調される要素

表3 リストに対する操作

CLEAR(LIST)	LISTの内容をクリアする。
PUSH(LIST, RECORD)	LISTにRECORDをプッシュする。
POP(LIST)	LISTからRECORDをポップする。
TOP(LIST)	LISTの先頭要素を返す。
SECOND(LIST)	LISTの第2要素を返す。
THIRD(LIST)	LISTの第3要素を返す。
MEMBER(LIST, RECORD)	RECORDはLISTの要素であるか。
MEMBER$(LIST, RECORD)	RECORDまたはRECORDの上位概念はLISTの要素であるか。
$MEMBER(LIST, RECORD)	RECORDはLISTまたはLIST要素の上位概念の集合の要素であるか。

これらのリストとリストに対する操作を用いて，文章中の主題表現生成の制御方略を以下に述べるように構成した。ここで〔……〕は，それぞれの場合の条件をリストに対する操作などを用いて形式的に記述したものである。

（1） 文章全体レベルでの処理　文章のスタイルが，日記文のような特別のものであれば（〔STYLE＝'DIARY'〕），以下のレベルでの主題表現は原則として省略する。

（2） 段落レベルでの処理　主題参照の新しい環境を作るため，文脈情報のクリアを行う（CLEAR(FLIST)，CLEAR(NLIST)……）。

（3） 文レベルでの処理　まず，主題表現の出現パターンの場合分けを FLIST を参照することで行う。これは畠[10~12]，永野[13]のいう主題の展開，主語の連鎖に対応している。

主題提示―FLIST も NLIST も空である。

〔FLIST＝ϕ，NLIST＝ϕ〕

主題維持―現在の文の主題が直前の文の主題と等しい。

〔TOP(FLIST)＝SECOND(FLIST)〕

主題変化―現在の文の主題が，同一段落のそれまでに出現したどの主題とも異なる。

〔TOP(FLIST≠SECOND(FLIST)，MEMBER$(FLIST，TOP(FLIST))＝$\phi$〕

主題回復―現在の文の主題が直前の文の主題とは異なるが，同一段落のそれまでに出現したどれかの主題と等しい。

〔TOP(FLIST)≠SECOND(FLIST)，MEMBER$(FLIST，TOP(FLIST))≠$\phi$〕

以上のような場合分けの後，一般に，主題維持中の主題表現は省略し，主題変化，主題回復時の主題表現は省略しないという基本的な方略を適用することになる。しかし，例外が存在して，主題維持中の主題表現で省略しない場合や，主題変化，主題回復時の主題表現で省略する場合がある。

そのような場合の主題表現の省略の制御は，視点の考え方を取り入れた久野[9]の方略だけでなく，寺倉[14]のいう継続文なら省略，非継続文なら非省略という原則を文間の連接関係によって調べるという方略をも含めた，以下に示すような方略によって行う。

（ア） 主題維持中の主題表現で省略しない場合

① 主題表現が強調要素である場合
〔STRESS＝TOP(FLIST)〕

② 場所が大きく変化した場合
〔TOP(SLIST)≠SECOND(SLIST)〕

③ 時間が大きく変化した場合
〔TOP(TLIST)≠SECOND(TLIST)〕

④ 前文との連接関係が特別の関係である場合
〔MEMBER((補充1，制御1，……)，TOP(CLIST))〕

（「補充1」は1文前の文に対する連接関係が「補充」であることを表す。以下同様）

（イ） 主題変化時の主題表現で省略する場合

① 前文も現在の文も現在の文の主題の視点からみた文である場合（新主題条件[9]）
〔TOP(FLIST)＝TOP(PLIST)，TOP(PLIST)＝SECOND(PLIST)〕

② 前文も現在の文も前文の主題の視点からみた文である場合（異主題条件[9]）
〔SECOND(FLIST)＝SECOND(PLIST)，TOP(PLIST)＝SECOND(PLIST)〕

③ 前文との連接関係が特別の関係である場合
〔MEMBER(連鎖・連係1，……)，TOP(CLIST))〕

（ウ） 主題回復時の主題表現で省略する場合

① 3文間の連接関係が特別の関係である場合

〔TOP(CLIST)＝'補足２', SECOND(CLIST)＝'連鎖・連係１'〕

または

〔TOP(CLIST)＝'添加・序列２', SECOND(CLIST)＝'連鎖・連係１'〕

３．文章生成システム　Hi-GTG

文章では汎用文章生成システム Hi-GTG と，そのデータの表記法である NLP 記法について簡単に述べる。

3.1　文章生成システム(Hi-GTG)の概要

文章生成を形式的に記述するためのシステムとして，北大大型計算機センタの UTILISP および apollo-Domain の CommonLisp 上に文章生成システム Hi-GTG を作成した[16,17]。Hi-GTG は，あらかじめ与えられている生成文法および辞書を用いて，文章の意味構造から文章を生成する。

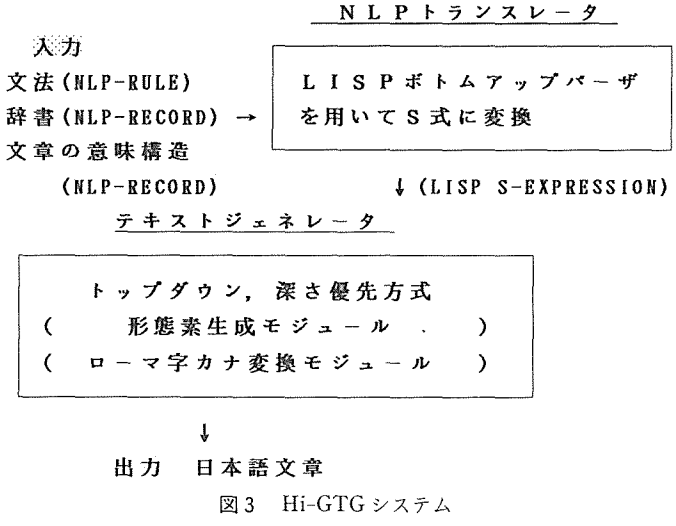


図３　Hi-GTG システム

文章の意味構造や文法，辞書を記述するための言語の仕様は MIT で開発された NLP システム[2]における NLP 記法の仕様をほぼ満たすものとなっている。Hi-GTG は，NLP 記法で書かれた辞書，文法記述，文章の意味記述を LISP の S 式に変換する NLP トランスレータと，トップダウン，深さ優先方式で文章を生成するテキストジェネレータ，および形態素生成モジュール，ローマ字やカナ変換モジュール等[6),7)]のユーティリティから構成されている。図３に Hi-GTG の概略図を示す。

3.2　NLP 記法

NLP 記法は NLP レコードと NLP ルールの２種類からなる。

3.2.1　NLP レコード　本システムの基本的なデータ構造が NLP レコードである。NLP レコードにより，辞書，文および文章の意味構造を記述する。この NLP レコードは格フレーム形式と同等の意味記述能力があり，一般的な形式は

レコード名(代入操作部)：

で表すことができる。

レコード名は，そのレコードの名前であり，フレームシステムにおけるフレーム名に相当する。また，代入操作部でレコードの属性に値の代入が行われる。

代入操作部の形式は基本的には

属性名：＝属性値

属性名：＝{LIST 関数呼び出し}

の２通りである。属性名は任意のアトムであり，属性値はアトムを"'"で囲んだもの，もしくは関数呼び出しが返す値である。属性の記法は，以下の２つの簡略形式がある。

属性値(属性名を省略した形式)

これはSUPという特別な属性への代入を表す。すなわち，「属性値」は，SUP：＝属性値と等価である。属性名SUP上位概念を示し，階層的な概念構造を表すのに用いられる。

属性名(属性値を省略した形式)

これは属性の存在が重要であって，属性値が意味を持たない，あるいは興味がない場合の記法である。システムはこの属性値に定数「T」を与える。すなわち，「属性名」は，属性名：='T'と等価である。この記法は本システムでは辞書の品詞やアスペクトを表すのに用いられている。なお，関数呼び出しについては，次のNLPルールのところで説明する。

3.2.2　NLPルール　NLPルールは文章の生成規則を記述する。一般的な形式は，

〈テスト〉——＞〈クリエーション〉

である。〈テスト〉は「レコード名(テスト部)」，〈クリエーション〉は「レコード名(アクション部)」からなる。

NLPルール中のテスト部およびアクション部の記述には，レコードの属性にアクセスするためのさまざまな機能を用いることができ，NLPルールは拡張文脈自由生成規則となっている。以下にそれらの機能の一部についての簡単な説明を行う。

(1)　参照　テスト部およびアクション部では，自分のレコードの属性を直接探索していく直接参照，他のレコードの属性を探索していく間接参照，指定した属性が見つかるまで，ある属性に沿って探索していく連続参照の3通りの記述ができる。

(2)　代入およびコピー　アクション部では，属性への代入，および他の属性値をコピーする記述ができる。

(3)　その他の機能　主なものに関数呼び出しがある。NLPレコードおよびNLPルールのテスト部とアクション部には関数呼び出し

LISP関数〈引数，……〉

を記述でき，引数には上記の(1)で示される属性値を用いることができる。

3.3　文章生成アルゴリズム

文章生成のアルゴリズムは文章生成規則を記述したNLPルールと，文章の意味構造を記述したNLPレコードから，トップダウン，深さ優先方式の処理で文章を生成するという簡潔なものである。以下で，文「はなこがあるく。」の生成例により文章生成アルゴリズムについてその概略を説明する。

図4にNLPルール，NLPレコード，生成の出力およびスタックの変化の様子を示す。生成の開始は，データ(SENT SENT1)をスタックに積むことにより行う。これは「レコードSENT1を用いてSENT(文)のレベルから生成を始めよ」ということを意味する。〈テスト〉がSENTにマッチするルールの適用が試みられる。この場合，SENTから始まるルールは1つで，かつテスト部がないので無条件にこのルールが適用され，データ(NP ※REC1※)，(VP ※REC2※)，(※MAMU※ T)が作られ，スタック

NLPルール	SENT --> NP(%SUBJECT(SENT)) VP(%SENT) *MARU* ; NP --> OUTPUT(%NP) P(SUP := 'GA') ; VP --> OUTPUT(%VP);
NLPレコード	SENT1('ARUKU'. SUBJECT := 'A1'); A1 ('HANAKO') ;
出力	生成のためのスタックの内容のダンプ
HANAKO	((SENT SENT1)) ((NP *REC1*) (VP *REC2*) (*MARU* T)) ((OUTPUT *REC3*) (P *REC4*) (VP *REC2*) (*MARU* T))
GA	((P *REC4*) (VP *REC2*) (*MARU* T))
ARUKU	((VP *REC2*) (*MARU* T)) ((OUTPUT *REC5*) (*MARU* T))
.	((*MARU* T))
	()

図4　「はなこがあるく」の生成例

に積まれる。ここで，※REC2※，※REC3※はアクション部を評価した結果作られる新しいレコードである。例えば，※REC1※には属性 SUP：＝'HANAKO'が与えられるが，これは，アクション部%SUBJECT(SENT)の評価の結果，(SENT SENT1)に含まれるレコードに SENT1 の，属性 SUBJECT の属性値 A1 で参照されるレコードの属性および属性値('HANAKO')がコピーされることによる。次のルールの適用はスタックの先頭のデータ(NP ※REC1※)に対して行われる。そして，表層表現の出力は対応するルールの存在しないレコード名(OUTPUT)のデータ，および句点を出力する特別なレコード名(※MARU※)のデータがスタックの先頭にあるときに行われる。

4．Hi-GTG による文章生成例

2.1 の例で用いた“むしのはなし”に対応する文章を Hi-GTG によって生成した例について述べる。なお，焦点情報の決定を含めて，文章の意味構造の作成は人手によって行った。

4.1 文章の意味構造の記述

文章の意味構造は，文章レコード，段落レコード，および文レコードの 3 つのレベルで構成される。文章レコードは，段落レコードのリストと，文章のスタイルを表す情報からなる。段落レコードは，文レコードのリストと，大域的焦点情報からなる。そして，文レコードには文の格フレームによる意味記述と，直接的焦点情報，視点情報，文間の連接関係の情報，強調情報などの情報が書き込まれる。

```
TEX1 ( PARAS := LIST < 'P1' , 'P2' > ,
       STYLE := 'NORMAL' );
```

図5(a) 文章の意味構造

```
P1 ( SENTS:=LIST< 'SENT1','SENT2',
          'SENT3','SENT4','SENT5','SENT6' > ,
     GFOCUS := 'AGEHATYOU' );
P2 ( SENTS:=LIST< 'SENT7','SENT8','SENT9',
       'SENT10','SENT11','SENT12','SENT13' > ,
     GFOCUS := LIST<'ARI','TENNTOUMUSHI'> ) ;
```

図5(b) 段落の意味構造

```
SENT1 ( 'KURU' , PST , DECLARATIVE , POL ,
        AAG := 'N1' ,
        SPT := 'N2' ,
        S*MET := 'S*SENT1' ,
        FOCUS := 'N1' ,
        P*P := 'NORMAL' ,
        S*R := 'TOP' ) ;
N1 ('AGEHATYOU');
N2 ( 'KI' , SORT := 'MIKANN' );
S*SENT1 ( 'TOBU' , PRS , DECLARATIVE ,
          AAG := 'N1' ,
          SPT := 'N2'   );
. . .
SENT6 ( 'UMU' , PRS , DECLARATIVE , POL ,
        AAG := 'N1' ,
        OOB := 'N8' ,
        OAC := 'TAMAGO' ,
        DEG := 'M3' ,
        FOCUS := 'AGEHATYOU' ,
        P*P := 'NORMAL' ,
        S*R := 'SPP2' );
```

図5(c) 文の意味構造

4.1.1 文章の意味構造 図5(a)に“むしのはなし”の文章の意味構造の例を示す。この文章 TEX1 は 2 つの段落 P1，P2 からなり文章のスタイルは，「普通(STYLE：＝'NORMAL')」である。

4.1.2 段落の意味構造 図5(b)に“むしのはなし”の段落の意味構造の例を示す。段落 P1 は 6 つの文からなり，大域的焦点は，「あげはちょう(GFOCUS：＝'AGEHATYOU')」である。

4.1.3 文の意味構造 図5(c)に“むしのはなし”の文の意味構造(の一部)の例を示す。文は属性 SUP として述語をとり，アスペクト等の情報や格情報，焦点情報，文の連接関係の情報等を含んでいる。SENT1 が表している内容は，述語が(来る('KURU')」，時制が「過去(PST)」，動作主が「N1(AAG：＝'N1')」，直接的焦点が「N1(FOCUS：＝'N1')」等々ということである。

4.2 主題表現生成の制御方略の実現

2.3 で述べた主題表現生成の制御方略は自然な形で本システムで実現することができる。

(1) 3 文間の連接関係を考慮する方略の実現例 図6(a)に焦点情報により主題表現の出現パターンの類別を行い(「2，3(3)文レベルでの処理」を参照)，さらにその表現を省略するか否かを

決定する処理を埋め込んだ(NP → NP)という名詞句の生成規則を示す。

(2)　前文と現在の文の焦点を考慮する方略の実現例　図6(b)に3文間の連接関係を考慮する処理(「2，3(3)(ウ)主題回復時の主題表現で省略する場合」を参照)を埋め込んだ(SENT → NP VP)という文レベルの生成規則を示す。

```
NP(SUBJECT,NULL<FLIST('PREVIOUS')))
    --> NP( -SUBJECT , P := 'GA' );
NP(SUBJECT,REF = CAR*<FLIST('PREVIOUS')))
    --> NP( -SUBJECT , P := 'HA' , OMIT ):
NP(SUBJECT,
    MEMBERFLIST<REF,FLIST('PREVIOUS')  > )
    --> NP( -SUBJECT , P := 'HA' );
NP(SUBJECT)
    --> NP( -SUBJECT , P := 'GA' );
```

図6(a)　主題出現パターンの分類

```
SENT(SUBJECT ,
      CAR*<FLIST('CURRENT')) =
      CADR*<FLIST('PREVIOUS')) ,
   +( CAR*<FLIST('CURRENT')) =
      CAR*<FLIST('PREVIOUS')) ) ,
      CAR*<CLIST('PREVIOUS')) = 'REL1' ,
      CAR*<CLIST('CURRENT')) = 'SPP2' )
 --> NP (%SUBJECT(SENT) ,
      REF := SUBJECT(SENT) , SUBJECT , OMIT
   VP (%SENT) ;
```

図6(b)　3文間の連接関係の処理

4.3　文章生成例

“むしのはなし”の生成例を図7に示す。文章の意味構造に対応して主題表現の生成の制御が適切に行われ，教科書中の“むしのはなし”の文章とほぼ同じものが生成されている。教科書中の文章と異なる生成がなされているのは，主題以外の省略，数詞表現の選択等で生じており，これらを適切に制御することは今後の課題である。

```
>(GENLP '(TEXT TEX1))
  ｱｹﾞﾊﾁｮｳ ｶﾞ ﾐｶﾝ ﾉ ｷ ﾆ ﾄﾝﾃﾞ ｷﾏｼﾀ.ﾄｷﾄﾞｷ ｵﾅｶ ﾉ ｻｷ
 ｵ ﾏｹﾞﾃ, ﾊ ﾆ ﾅﾆｶ ｵ ﾜｹﾃ ｲﾏｽ.ｲｯﾀｲ ﾅﾆ ｵ ｼﾃ ｲﾙﾉﾃﾞｼｮｳ.
ﾀﾏｺﾞ ｵ ｳﾝﾃﾞ ｲﾙﾉﾃﾞｽ.ｿﾉ ﾀﾏｺﾞ ﾊ ｳｽｲ ｷｲﾛ ﾉ ﾁｲｻﾅ ﾀﾏｺﾞﾃﾞ
ｽ.ｱｹﾞﾊﾁｮｳﾊ 1ﾏｲﾉ ﾊ ﾆ ﾀﾏｺﾞｵ ﾋﾄﾂｽﾞﾂ ｳﾐﾏｽ.
  ｱﾘ ﾄ ﾃﾝﾄｳﾑｼ ｶﾞ ｲｿｶﾞｼｿｳﾆ ﾊﾞﾗ ﾉ ﾀｷ ｵ ｱﾙｲﾃｲﾏｽ.ｲｯﾀｲ ﾅﾆ
 ｵ ｼﾃ ｲﾙﾉﾃﾞｼｮｳ.ﾄﾞﾁﾗﾓ ｱﾌﾞﾗﾑｼ ｵ ｻｶﾞｼﾃ ｲﾙﾉﾃﾞｽ.ｱﾘ ﾊ ｱﾌﾞﾗ
ﾑｼ ｵ ﾐﾂｹﾙﾄ, ﾋｹﾞ ﾃﾞ ｱﾌﾞﾗﾑｼ ｵ ｶﾙｸ ﾀﾀｷﾏｽ.ｿｼﾃ, ｱﾌﾞﾗﾑｼ ｶﾗ
ﾃﾞﾙ ｱﾏｲ ｼﾙ ｵ ｽｲﾏｽ.ﾃﾝﾄｳﾑｼ ﾊ ｱﾌﾞﾗﾑｼ ｵ ﾐﾂｹﾙﾄ, ｱﾌﾞﾗﾑｼ ｵ
ﾀﾍﾞﾃ ｼﾏｲﾏｽ.1ﾆﾁﾉ ﾋ ﾆ 10ﾋｷﾞﾗｲ ｱﾌﾞﾗﾑｼ ｵ ﾀﾍﾞﾏｽ.
NIL
>
```

図7　文章生成例

以下の(1)から(3)では，意味構造の変化が生成される文章に及ぼす効果を示す例として，名詞句をクラスと見なすかインスタンスと見なすかにより生成文章に違いが生じる例を示す。

(1)　第6文の焦点をクラスとしての“あげはちょう”と考えて文章を生成。SENT6においてAAG：='AGEHATYOU'とすると，生成文章の第6文は次のようになる。

アゲハチョウ　ハ　1マイノ　ハ　ニ　タマゴ　オ　ヒトツ　ズツ　ウミマス。

(2)　第6文の焦点をインスタンスとしての“あげはちょう”と考えて文章を生成。SENT6においてAAG：='N1'とする。

1マイノ　ハ　ニ　タマゴ　オ　ヒトツ　ズツ　ウミマス。

(3)　第6文の焦点をインスタンスとしての“あげはちょう”と考え，さらに，その焦点を強調して文章を生成。SENT5においてAAG：='N1'，STRESS：='N1'とする。

ソノ　アゲハチョウ　ハ　1マイノ　ハニ　タマゴ　オ　ヒトツ　ズツ　ウミマス。

(強調のため，指示代名詞'その'が生成されている。この処理は，本論文で考察した主題表現の生成制御とは別の処理によって行っている。)

5．むすび

日本語主題表現の生成方略について考察し，それを汎用文章生成システム Hi-GTG によって実現した。小学校低学年の国語教科書の説明文章程度の主題表現の生成の制御を実現することができた。

さらに，多くの文章の生成実験を行い，生成方略の整備を行いたいが，次のステップの結束性の制御としては

（1） 主題表現の代名詞化の制御

（2） 接続詞の選択の制御

等を考えている。

謝　辞

本研究は文部省科学研究費補助金（特定研究番号　62210002）の援助による。

文　献

1) Appelt, D. E.: "Planning English Sentences", Cambridge University Press (1985).
2) Granville, R. A.: "Cohesion in Computer Text Generation: Lexical Substitution", MIT/LCS/TR-310 (1983).
3) Derr, M. and McKeown, K.: "Using Focus to Generate Complex and Simple Sentences", Proc. of COLING-84, pp. 8319-326 (1984).
4) Heidorn, G. E.: "Natural Language Inputs to a Simulation Programming System", Tech. Rep. SPS-55HD72010A, Naval Postgraduate School (1972).
5) Heidorn, G. E.: "Augrented Phrase Structure Grammars", IBM Research RC 5787 (# 25076) (1975).
6) 草薙　裕：マイコンによる自然言語処理入門，工学図書，pp. 219-224 (1984).
7) 佐伯　胖(監修)，田中穂積，元吉文男，山梨正明：LISP で学ぶ認知心理学 3，東京大学出版会(1983).
8) 山梨正明：言語理解と情報処理，Computer Today, 1985/1, No. 5, サイエンス社，pp. 44-51 (1985).
9) 久野　暲：談話の文法，大修館書店(1978).
10) 畠　弘巳：文とは何か―主題の省略とその働き―，日本語教育 41 号，pp. 198-208 (1980).
11) 畠　弘巳：主題の展開と談話分析，国際商科大学論叢(商学部編)第 31 号，pp. 104-117 (1985).
12) 畠　弘巳：接続詞と文章の展開，日本語教育 56 号，pp. 13-27 (1985).
13) 永野　賢：文章における主語の連鎖，日本語教育 56 号，pp. 1-12 (1985).
14) 寺倉弘子：談話における主題の省略について，言語，Lol. 15, No. 2, 大修館書店，pp. 98-105 (1986).
15) 市川　孝：国語教育のための文章論概説，教育出版(1978).
16) 高橋　晃，桃内佳雄，宮本衛市：結束性を考慮した文章の生成について，情報第 31 回全国大会，pp. 1363-1364 (1985).
17) 高橋　晃，桃内佳雄，宮本衛市：汎用文章生成システムによる日本語主題表現生成方略の実現，情処研究報 NL-56-2 (1986).