

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

高橋晃 桃内佳雄 宮本衛市
北海道大学工学部

文章生成において生成される文章の結束性を考慮する事は、質問応答システムの応答文や機械翻訳システムの訳文などの生成文章の質の向上にとって本質的な課題である。

文章の結束性を保持するための基本的な役割を果しているものの一つに、文章の主題表現の連鎖がある。この主題表現の連鎖において、連続した同じ主題を省略することにより、さらに文間に強いつながりを持たせる事もでき、文章生成において主題表現の生成の制御は重要な問題である。

本報告では、日本語文章の主題表現の生成を制御する要因と制御方略について考察し、主題表現生成の制御方略を埋め込んだ文章生成システムの実現について報告する。

CONTROLLING SUBJECT GENERATION BY GENERALIZED TEXT GENERATOR

Akira TAKAHASHI Yoshio MOMOUCHI Eiichi MIYAMOTO

Division of Information Engineering, Graduate school of engineering, Hokkaido University,
Kita 13, Nishi 8, Kita-ku, Sapporo 060, Japan

The need for good text generation is rapidly increasing. One requirement for generated output to be considered text is to exhibit cohesion. Subject ellipsis is one of the most important device to create Japanese cohesive text.

This paper describes how to control subject generation with information such as focus, perspective and so on, and shows an implementation of the subject generation by Hi-GTG (Hokkaido university Information engineering - Generalized Text Generator) system. The system uses tests on information such as focus of attention within the APSG (Augmented Phrase Structure Grammar) formalism to determine how to omit the subject.

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。

英語文章生成に関しては、文章の質を考慮して文章を生成するシステムがいくつか開発されている。例えば、DerrやMcKeownらは、焦点情報を用いて、その内容にふさわしい構造をもつように複文／単文の生成を制御して文章を生成するシステムをPROLOGのDCGの形式を用いて記述した。また、GranvilleによるPAULは、文章生成をAPSG(Augmented Phrase Structure Grammar)^{1), 2), 3)}により記述し、その記述の中に、焦点情報の予測を行うアルゴリズムや、焦点情報と名詞句の距離を考慮する処理をうめこみ、語彙結合を考慮して文章生成を行うシステムである。

一方、日本語文章生成においては、生成文章の質を考慮して文章を生成するシステムの開発は、従来あまりなされていなかったと思われる。その原因としては、まず生成文章の質を評価する判定基準が定めにくいということがあげられる。また、生成文章の質を考慮する処理は様々な要因が関係していると考えられ、それを文章生成のどの段階で、どのような方略で行うかを決定することも困難な問題となっている。

われわれは、文章の質を文章の結束性(つながり)という観点から捉えることにした。そして、文章の結束性を保持するための具体的な言語的手段に対する処理をひとつずつ構成することによって、質の良い文章が生成されると考えている。

また、文章の質に関する考慮の処理を見通し良く構成する一つの方法として、文章生成を形式的に記述し、そのなかに質の考慮の方略をうめこんでいく方法が考えられ、さらにそのような形式的な記述は、種々の質の考慮の処理方略をそれぞれモジュール化して記述することを可能とする。前述のPAULは、この方法に基づいて構成されたシステムである。われわれも、生成文章の質を考慮した文章生成システムの構成にあたり、この方法を採用して、汎用文章生成システム Hi-RTGシステム(北大情報-Generalized Text Generator)を作成した。本報告では、質の考慮の処理を効率よくうめこんだ日本語文章生成をAPSGにより形式的に記述し、それをHi-RTGシステムにより実現したことについて述べる。

生成文章としては、小学校一年生の国語の教科書中の説明および物語文章を対象とした。それらの文章中で用いられている、結束性を保持するための言語的手段は省略、接続、指示など多岐に渡るが、本報告では、主題表現の省略に焦点をしばり、その制御をスタックを用いて処理する方法について論じる。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ク

イプであり、日本語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。

2. 結束性を考慮した文章生成

2.1 文章における結束性

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(cohesion:結束性)が必ず存在する。このつながりには、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のレベルの結束性を保持するための具体的な言語的手段としては、

省略(ellipsis)-文脈から復元可能な言語表現を後続文で省略するという表層的な手がかりによって文章のつながりを保証する。

代用(substitution)-先行文中の言語表現を後続文において別の言語表現で置き換えるという表層的な手がかりによって文章のつながりを保証する。

(「大きいのを頂戴」の「の」などがある)

指示(reference)-先行文中の言語表現が指示する対象を後続文において別の言語表現で指示するという表層的な手がかりによって文章のつながりを保証する。(代表的なものに、代名詞がある)

接続(conjunction)-文と文を接続する表現を用いることによって文章のつながりを保証する。

(接続詞、接続助詞が用いられる)

語彙結合(lexical coherence)-先行文中の語彙に対して、その同一語、類義語、上位語等を後続文で用いる事で文章のつながりを保証する。

等がある。

2のレベルの結束性は、文相互の意味関係、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を介して保証される。

2.2 文章生成モデル

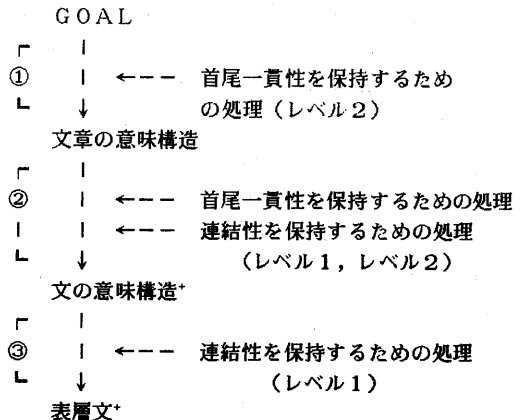


図1. 文章生成モデル

生成文章の結束性を考慮するという事は、意味レベルでの結束性(すなわち首尾一貫性)の考慮と、表層レベルでの結束性(すなわち連結性)の考慮の両方を考慮することであるが、文章生成のモデルとしては、Appeltのようにその出発点を発話者のゴールに求め、意味レベルのつながりを重視したものと、GranvilleやMckeown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したものに分けられる。文章生成モデルの全体構成を図1に示す。

本報告においては、主題表現の省略という表層レベルの処理を扱うために、後者の立場を取っており、現在の所、文章生成の出発点としては文章の意味構造を設定し、また結束性を保持するための処理は、主に②、③の段階の処理を中心に考えている。

3. 主題表現生成の制御

3.1 日本語主題表現の出現パターン

日本語の主題表現は、新たな主題提示の場合は、「・・・が」という表現、また既知の主題について述べる場合には「・・・は」という表現で表されることが多いが、同一の主題表現が繰り返される場合は、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ることが多い。以下では、文の主題となる要素は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おいて、主題表現を、「・・・は」で表現するか、「・・・が」で表現するか、或は省略するかを決定する要因について考察しなければならず、次節でそれを行う。

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ている例を調べてみると次のようになる。

(φの部分为主题表現が省略されている所である。)

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か。
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、たまごをひとつだけうみます。
7. ありとてんとうむしが、ばらのくきをいそがしそうにあるいています。
8. φ いったい、なにをしているのでしょうか。
9. どちらも、あぶらむしをさがしているのです。
10. ありは、あぶらむしをみつけると、ひげであぶらむしをかくたたきます。
11. そして、φ あぶらむしからでるあまいしるをすいます。
12. てんとうむしは、あぶらむしをみつけると、た

べてしまいます。

13. φ [ありを] いちにちにじっぴきぐらいたべます。

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。

同様に、文7で主題“ありとてんとうむし”が導入され、それは文8、文9まで維持される。文10で主題は“あり”に変化し、それは文11まで維持される。文12で主題は“てんとうむし”に変化し文13まで維持される。

(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。また文7のありとてんとうむしはインスタンスとしてのありとてんとうむしで、文10のありは、クラスとしてのありであり、文12のてんとうむしもクラスとしてのてんとうむしであると考えられる。)

この様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、文章はひどく間延びしたものになることは明らかである。

3.2 主題表現生成を制御する要因

文章の主題表現の生成を制御する要因としては、

- ア) 文章の意味構造の記述のレベルで定まる情報
- イ) 段落の意味構造を記述するレベルに必要な情報
- ウ) 文の意味構造を記述するレベルに必要な情報
- エ) 表層文が生成される時に記録される文脈的な情報の4種類を考えた。

ア. 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)

文章生成のトップレベルでの情報をたくわえる。スタイルとしては、説明文、物語文、談話文、日記文等があげられる。

イ. 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焦点で、一段落に一つ存在する。

*段落間の接続構造(S_COHERENCE)

起・承・転・結、序・破・急などの段落の役割を記述する。

ウ. 文の意味構造レベルで記述される情報

*直接的焦点(FOCUS)

一文に一つ存在すると考えられる。文における話題の中心である。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の距離関係が問題となる。全ての要素と等距離にある視点はいわゆる“神の視点”といわれる。

*強調(STRESS)

文中で特に強調する要素。

*場面 (時間的, 空間的)

文中の背景となる場面情報。

*文間の接続構造(COHERENCE)

段落中での各文の機能。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。この分類は主に文献14によった。

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

エ. 文脈的な情報

*ア, イ, ウのそれぞれの履歴

*表層レベルにおける主題表現の有無の履歴

*表層レベルにおける名詞表現の履歴

これらの情報は文章の階層的な意味構造に対応している。

3.3 スタックを用いた主題表現生成の制御方略

主題表現生成の制御方略は, 3.2のそれぞれの情報をスタックに保持して, それらの複数のスタックを参照したり, 操作したりする事によって行う。現在考えているスタックは次の様なものである。

STYLE : 文章のスタイル

GFOCUS : 大域的焦点 (global focus)

CLIST : 文間の接続関係リスト (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)

ELIST : 主題表現リスト (expression list)

NLIST : 名詞リスト (noun list)

STRESS : 強調される要素。

このうち文章全体レベルの情報が格納されるものはSTYLEで, 段落レベルの情報が格納されるのはGFOCUS, 文レベルの文脈情報の履歴が格納されるのがFLIST, PLIST, CLIST, TLIST, SLIST, ELIST, NLIST である。

また, スタック (リスト) に対する操作として次のようなものを考えている。

• CLEAR(STACK)

STACKの内容をクリアする。

• PUSH(STACK, RECORD)

STACKにRECORDをプッシュする。

• POP(STACK)

STACKからRECORDをポップする。

• TOP(STACK)

STACKの先頭要素を返す。

• SECOND(STACK)

STACKの第2要素を返す。

• THIRD(STACK)

STACKの第3要素を返す。

• MEMBER(STACK, RECORD)

RECORDはSTACKの要素であるか。

• MEMBER\$(STACK, RECORD)

RECORDもしくはRECORDの上位概念はSTACKの要素であるか。

• \$MEMBER(STACK, RECORD)

RECORDはSTACKまたはSTACK要素の上位概念の集合の要素であるか。

これらのスタックの操作に基づいて文章中の主題表現の生成の制御は次のように構成される。

*文章全体レベルでの処理

STYLE = 'DIARY'

TOP(FLIST) = '1人称'

STRESS ≠ TOP(FLIST)

なら主題表現を省略。

*段落レベルでの処理

文脈的な情報のクリアを行う

(CLEAR(FLIST), ...)

*文レベルでの処理

主題表現のパターン分類はFLISTを参照することで行う。

主題提示

FLIST=φ, NLIST=φ

主題維持

TOP(FLIST)=SECOND(FLIST)

主題変化

TOP(FLIST)≠SECOND(FLIST)

MEMBER\$(FLIST, TOP(FLIST)) = φ

主題回復

TOP(FLIST)≠SECOND(FLIST)

MEMBER\$(FLIST, TOP(FLIST)) ≠ φ

一般に, 主題維持中の主題表現は省略され, 主題変化, 主題回復時の主題表現は省略されない。しかし, 主題維持中の主題表現で省略されない場合や, 主題変化, 主題回復時の主題表現で省略される場合がある。

• 主題維持中の主題表現で省略されない場合

1 TOP(SLIST)≠SECOND(SLIST)

2 TOP(TLIST)≠SECOND(TLIST)

3 MEMBER(補充1, 制約1, ...),

TOP(CLIST))

(**1は1文前に対する接続関係を表す)

・主題変化時の主題表現で省略される場合

- 1 TOP(FLIST)=TOP(PLIST)
TOP(PLIST)=SECOND(PLIST)
- 2 SECOND(FLIST)=SECOND(PLIST)
TOP(PLIST)=SECOND(PLIST)
- 3 MEMBER(連鎖・連係1, . . .), TOP(CLIST))

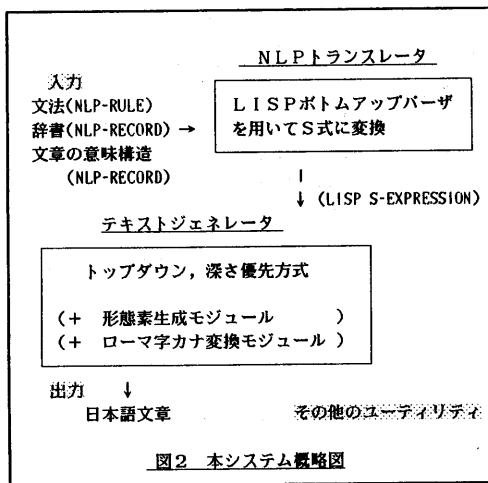
・主題回復時の主題表現で省略される場合

- 1 TOP(CLIST)='補足2'
SECOND(CLIST)='連鎖・連係1'
 - 2 TOP(CLIST)='添加・序列2'
SECOND(CLIST)='連鎖・連係1'
- (**2は2文前に対する関係を表す)

4. 文章生成システム Hi-GTG

4.1 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システムとして、北海道大学大型計算機センターのUTILISP上に、文章生成システムHi-GTGをインプリメントした^[1,2,3]。Hi-GTGシステムの言語仕様はMITのNLPシステム(PAULバージョン)^[3]の仕様をほぼ満たしている(プリプロセッサ部を除くすべての機能を含んでいる)。システムは、NLPの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で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、ローマ字カナ変換モジュール等のユーティリティから構成される。図2にシステムの概略図を示す。



4.2. NLPトランスレータ

NLPトランスレータはNLPの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。変換の方法はLISPのボトムアップパーザを用いて宣言的に記述されているため、仕様の変更、機能の拡張が比較的容易である。

NLPトランスレータの構成は

- ・字句解析のための辞書
- ・構文解析のための文法(意味解析部も含んでいる)
- ・パーザ本体

の3つからなっており、NLPの記法をLISPのS式に変換するのは、パーザの意味解析部で行われる。したがって、新たな機能を付け加えるには、辞書と文法および補助的な関数の変更を行う事になる。

字句解析のための辞書は

(辞書項目 カテゴリ メッセージ 意味解釈部)の形式をしている。ここで、メッセージは構文解析の過程を制御するためのものであるが、普通はNILであり、意味解釈部でその辞書項目の意味を与える。

また構文解析のための文法は

- (a ((b . 拡張部)
(c . 拡張部)) (意味解釈部))

で、a → b c の文脈自由規則に対応し、拡張部で文法規則の適用を制御する。構文解析が終わった後にその解析木の末端から意味解析部が働きS式を生成する。

なお、このボトムアップパーザの詳細については、文献6に詳しい。

4.3 NLP記法

- ・NLPレコード

以下では、シンボルとは英数字および_ (アンダーバー)と* (アスタリスク)からなるLISPアトムとする。

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、任意のシンボルからなる。これは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 := {LISP関数呼び出し}

の2とおりである。属性名は任意のシンボルであり、属性値はシンボルを"'"で囲んだもの、もしくはLISP関数呼び出しが返す値である。LISP関数呼び出しは、関数名<引数1, 引数2, . . . >の形式をしており、各引数は属性値である。なお、属性の記述は記法の簡便のために次の二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という特別の属性への代入を表す。すなわち、'属性値'は陽に書けば、SUP := '属性値'である。属性名SUPは階層的な概念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、その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。

・ NLPルール

NLPルールはAPSGを記述する。一般的な形式は
 テスト部 --> アクション部

という一種のプロダクションルールとなっている。

NLPルールの表記法にはレコードの属性にアクセスするためのさまざまなファシリティが備わっている^[3]。以下でそれらについての説明を行う。本システムにおいては、これらのファシリティの追加、変更は容易である。

属性を参照するとき、その上位概念の属性や、別のレコードの属性を参照出来る。

イ) 直接参照

自分のレコードの属性を直接手繰って参照する。

ロ) 間接参照

他のレコードの属性を手繰って参照する。

ハ) 連続参照

指定した属性が見つかるまで、ある属性に沿って手繰っていき参照する。

ニ) その他の機能

主なものにファンクションコールがある。テスト部およびアクション部にNLPレコードと同様にファンクションコール LISP関数< 引数, ... >を記述出来る。引数には、イからハを用いることができる。

(例) SENT(SUBJECT) --> NP(%SUBJECT(SENT),
 REF := SUBJECT(SENT),
 SUBJECT)
 VP(%SENT);

現在のデータのレコード名がSENTでその属性SUBJECTの値がNILでなければこのルールが適用され、NPというレコード名のデータとVPというレコード名のデータが作られる。データNPにはデータSENTの属性SUBJECTが示すデータの属性値がすべてコピーされ、また、属性REFにデータSENTの属性SUBJECTの内容がコピーされ、属性SUBJECTに「T」がセットされる。

4.4 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APSR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シンプルなものとなっている。本システムの特徴は、活用語の形態素生成を次節に述べる形態素生成モジュールを用いて行っている所にある。文章生成アルゴリズムは文献1による。図3に文章生成の概念図を示す。

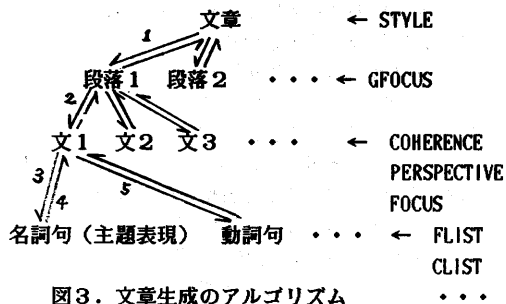


図3. 文章生成のアルゴリズム

4.4.1 形態素生成モジュール

日本語文生成においては、動詞や助動詞等の活用語は接続に共起関係があり、その接続関係によって形態素が変化する。このような活用処理を見とおし良く記述するために、本システムでは動詞や助動詞などのアスペクトや態を表す付属語には一つの形態素を割り当て、活用の処理は用言の形態素に付属語の形態素を付加していくという方法をとっている。形態素付加の部分はNLPルールを用い、音便処理などの音韻の細かな調節はLISP関数によって行っている。この形態素生成の基本的なアイデアおよび助動詞の形態素や接続関係などの処理は主に文献5によった。

本システムの中でこのモジュールはテスト部のセグメント名がYOUGENのルールに対応している。動詞の形態素生成は、ラベル名YOUGENのデータを作り、動詞名は属性SUPに、その他の助動詞に対してはその機能名をコピーすればよい。

例) 現在のデータのラベル名がYOUGENで、属性名SUPの値が「あるく」、属性PST(過去)とPOL(丁寧)の値がTだったら、このデータがカレントスタックのトップにある時、「あるきました」という形態素を生成する。助動詞の接続関係によりPSTの形態素を「*JTA」、POLの形態素を「*IMAS」と決定できる所がこの方式のポイントである。

表2 「あるきました」の形態素生成

形態素付加	語幹	POL	PST
	ARUK	*IMAS	*JTA
音韻処理	ARUK	IMAS	ITA
ローマ字カナ変換	あるきました		

4.4.2 その他のユーティリティ

*ローマ字カナ変換

北大大型計算機センターのUTILISPは平仮名および漢字が使えないため、ローマ字を片仮名に変換する関数を作成した。

関数名 (R-K)

例) (R-K 'AGEHATYOU) --> アゲハチヨウ

本システムでは大域変数#RK#をTにセットすると文章生成の出力がローマ字カナ変換される。

*デバッグユーティリティ

デバッグユーティリティとしてNLPルールの適用ごとに現在の属性を出力する機能、NLPルールの適用ごとに現在のカレントスタックの内容を出力する機能や、現在のNLPルール、NLPレコードをファイルにセーブする関数(PRINT-RULE)(PRINT-RECORD)などNLPルールやNLPレコードを編集するユーティリティが幾つかあるが、開発環境の整備は今後の課題である。

5. Hi-GTGによる文章生成例

3. 1の例で用いた“むしのはなし”をHi-GTGによって生成した例について述べる。なお文章の意味構造の作成および焦点情報の決定などは、現在の所、人手によって行っている。

5. 1 文章の意味構造の表現

文章の意味構造は、文章レコードと段落レコード、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と同様の意味記述と、直接的焦点情報、視点情報、文間の接続関係の情報、強調情報などの情報が書き込まれている。

*文章の意味構造

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

```
TEX2 ( PARAS := LIST < 'P1' > ,  
      STYLE := 'NORMAL' );
```

*段落の意味構造

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

*文の意味構造(一部)

```
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' );
```

...

```
N6 ( 'TAMAGO' );
```

```
SENT5 ( 'DA' , PRS , DECLARATIVE , POL ,  
       TAG := 'N6' ,  
       TOB := 'N7' ,  
       FOCUS := 'N6' ,  
       P#P := 'N1' ,  
       S#R := 'REL1' );
```

```
N7 ( 'N6' , ADT := 'TIISANA' ,  
     ATR := 'C1' );
```

```
C1 ( 'KIIRO' , ADJ := 'USUI' );
```

...

5. 2 主題表現の生成方略の実現

3. 2で述べた主題表現の生成方略は自然な形で本システムで実現することができる。

・3文間の接続関係を考慮する方略の実現例。

```
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 ) ;
```

・前文と現在の文の焦点を考慮する方略の実現例。

```
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 , REF = CAR#<PPLIST('CURRENT')> ,  
     CAR#<PPLIST('CURRENT')> =  
     CAR#<PPLIST('PREVIOUS')> )  
--> NP( -SUBJECT , P := 'HA' , OMIT );  
NP ( SUBJECT ,  
     CAR#<FLIST('PREVIOUS')> =  
     CAR#<PPLIST('PREVIOUS')> ,  
     CAR#<PPLIST('CURRENT')> =  
     CAR#<PPLIST('PREVIOUS')> )  
--> NP( -SUBJECT , P := 'HA' , OMIT );  
NP ( SUBJECT ,  
     MEMBERNLIST<REF,NLIST('PREVIOUS') > )  
--> NP( -SUBJECT , P := 'HA' , AMB );  
NP ( SUBJECT , SUP = GFOCUS('CURRENT') )  
--> NP( -SUBJECT , P := 'HA' );  
NP ( SUBJECT )  
--> NP( -SUBJECT , P := 'GA' );
```

5.3 文章生成例

・生成例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。

文6の意味構造

SENT6 ('UMU', PRS, DECLARATIVE, POL,

AAG := 'AGEHATYOU',

OOB := 'N8',

OAC := 'TAMAGO',

DEG := 'M3',

FOCUS := 'AGEHATYOU',

P*P := 'NORMAL',

S*R := 'SPP2');

生成結果1

>(GENLP '(TEXT TEX2))

アケハチョウカ ミカノキニトテ、キマシ。トキトキオカノサ
キオマケテ、ハニナカオツケイマス。イタイナニオシテイルデシ
ヨウ。タマコオウンテイルデス。ウスイキロノチイナタマコデス。
アケハチョウハ 1マイノハニタマコオヒツタケウミス。

・生成例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。

文6の意味構造

SENT6 ('UMU', PRS, DECLARATIVE, POL,

AAG := 'N1',

OOB := 'N8',

OAC := 'TAMAGO',

DEG := 'M3',

FOCUS := 'N1',

P*P := 'NORMAL',

S*R := 'SPP2');

生成結果2

>(GENLP '(TEXT TEX2))

アケハチョウカ ミカノキニトテ、キマシ。トキトキオカノサ
キオマケテ、ハニナカオツケイマス。イタイナニオシテイルデシ
ヨウ。タマコオウンテイルデス。ウスイキロノチイナタマコデス。
1マイノハニタマコオヒツタケウミス。

・生成例3

第6文の焦点をインスタンスとしての“あげはちょう”
として考え、さらに、その焦点を強調して文章を
生成。

文6の意味構造

SENT6 ('UMU', PRS, DECLARATIVE, POL,

AAG := 'N1',

OOB := 'N8',

OAC := 'TAMAGO',

DEG := 'M3',

FOCUS := 'N1',

P*P := 'NORMAL',

S*R := 'SPP2',

STRESS := 'N1');

生成結果3

>(GENLP '(TEXT TEX2))

アケハチョウカ ミカノキニトテ、キマシ。トキトキオカノサ
キオマケテ、ハニナカオツケイマス。イタイナニオシテイルデシ
ヨウ。タマコオウンテイルデス。ウスイキロノチイナタマコデス。
1マイノハニタマコオヒツタケウミス。

6. おわりに

日本語主題表現の生成方略を考案し、Hi-GTG
システムによって実現した。今後の課題としては、

- (1) 主題表現の代名詞化の制御を行う。
- (2) 主題表現生成の制御要因をさらに整理し、首尾一貫性を保持するための処理の制御を行う。

等があげられる。

参考文献

- 1) Heidorn, G.E.: "Natural Language Inputs to a Simulation Programming System", Tech. Rep. NPS-55 HD72010A, Naval Postgraduate School, 1972.
- 2) Heidorn, G.E.: "Augmented Phrase Structure Grammars", IBM Research RC 5787 (#25076), 1975.
- 3) Granville, R.A.: "Cohesion in Computer Text Generation: Lexical Substitution", Massachusetts Institute of Technology, 1983.
- 4) Derr M., & McKeown K.: "Using Focus to Generate Complex and Simple Sentences", Proceedings of COLING-84, pp. 319-326, 1984.
- 5) 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.
- 6) 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.
- 7) 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.
- 8) 久野晴: 談話の文法, 大修館, 1978.
- 9) 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.
- 10) 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.
- 11) 畠弘巳: 接続詞と文章の展開, 日本語教育56号, 日本語教育学会, pp. 13-27, 1985.
- 12) 永野賢: 文章における主語の連鎖, 日本語教育56号, 日本語教育学会, pp. 1-12, 1985.
- 13) 寺倉弘子: 談話における主題の省略について, 言語, VOL. 15, NO. 2, 大修館書店, pp. 98-105, 1986.
- 14) 市川孝: 国語教育のための文章論概説, 教育出版, 1978.
- 15) 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, 情報処理学会第31回 (昭和60年後期) 全国大会, 1985.
- 16) 高橋晃, 桃内佳雄, 宮本衛市: 文章合成システムの自動生成について, 電気関係学会北海道支部連合大会, 1985.