自然言語処理　62－3
（１９８７・７・２４）

文章生成における接続詞の生成方略について

高橋晃　桃内佳雄　宮本衛市
北海道大学工学部

文章生成の過程は，生成文章の内容を決定する，いわゆる「What to say」の過程と，その内容をいかに適切に表層の表現とするかを決定する「How to say」の過程に分けて考えることができる．われわれは，日本語文章の生成における「How to say」の過程について，「文章の結束性をいかに高めるか」という見地から研究を進めてきた．

本報告では，文章中の結束性を実現するための言語的手段としての接続表現に着目し，文章生成における接続表現の生成の枠組みについて考察する．さらに，小学校低学年の国語の教科書の接続表現の解析にに基づいて，接続表現生成の制御方略を構成し，その制御方略を汎用文章生成システム（Ｈｉ－ＧＴＧ）上に実現して生成実験を行ったことについて述べる．

HOW TO CONTROL THE GENERATION OF CONJUNCTIVE EXPRESSION

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 process of text generation is divided into two phases. One is determining "what to say." The other is the process of "how to say" that transforms the content of "what to say" into an appropriate surface text. We have been studying the process of "how to say" in the Japanese text generation from the point of view of "how to generate the cohesive texts."

In this paper, we focus on conjunctive expressions as linguistic devices to generate the cohesive texts.

We examine the usage of conjunctive expressions in Japanese language textbooks and construct the strategies for controlling the generation of conjunctive expressions in the text generation.

Finally, we present the implementation of some of strategies on the general-purpose language generation system (Hi-GTG).

1．はじめに

文章生成の過程は，生成文章の内容を決定する，いわゆる「What to say」の過程と，その内容をいかに適切に表層の表現とするかを決定する「How to say」の過程に分けて考えることができる[1]．われわれは，日本語文章の生成における「How to say」の過程について，「文章の結束性[7]をいかに高めるか」という見地から研究を進めてきた[15,16]．

本報告では，文章中の結束性を実現するための言語的手段としての接続表現に着目し，文章生成における接続表現の生成の枠組みについて考察する．さらに，小学校低学年の国語の教科書の接続表現の解析に基づいて，接続表現生成の制御方略を構成し，その制御方略を汎用文章生成システム上に実現して生成実験を行ったことについて述べる．

2．文章の結束性と接続表現

文章は単なる単文の集合ではなく，文章の構成要素である段落，文，単語等の間には有機的なつながり，即ち結束性が存在する．文章生成においてこの結束性を考慮して文章を生成することは，生成文章の質を向上させるために本質的な課題である．

接続表現は，段落と段落，文と文，語と語の間の結束性を実現する具体的な言語手段である．接続表現を用いることにより，文章の論理的構成や意味的なつながりが明確にされる．しかし，接続詞の多用はかえって読者に冗長な印象を与えることにもなりかねない（例１）．

（例１）接続表現が多用された文章

<1a>空が暗くなった．そして，風がふいてきた．そして，雨が降ってきた．

<1b>空が暗くなって，風が吹いてきた．そして，雨が降ってきた．

（1aに比べ1bは「そして」の繰り返しが冗長である．）

また，接続表現の生成は省略等の他の結束性を考慮するための処理とも関わっている（例２）

（例２）主題の省略と接続表現

<2a>少年は傘をさした．少年は歩きだした．

<2b>少年は傘をさした．彼は歩きだした．

<2c>少年は傘をさした．そして，歩きだした．

<2d>少年は傘をさした．そして，彼は歩きだした．

（2aは結束性の考慮がなされていない文章で，２文の関係は表層表現に現れず，２文間の関係は読み手に委ねられている．2bは代名詞の使用によって２文の関係が表現されているが，その関係がどのような種類であるかは表層表現には現れていない．2cおよび2dは接続詞の挿入により２文の関係が表層表現に現れた文章で，2b，2c,2dの文章の生成には結束性の考慮がなされている．）

これらの例からも，接続表現の生成は文章の結束性を総体的に考慮する処理の枠組の中で考えなければならない．

3．文章における接続表現

文章における接続表現は，段落間の接続，文間の接続および名詞句間の接続をするものといった３つのタイプに分類される．

（１）段落間の接続：段落の切れ目は改行，字下げにより表層文章に陽に示される．また接続詞や接続助詞を伴った副詞句等によりその構造が示される．

（例３）第一に，・・・・・．
次に，・・・・・．
最後に，・・・・・．

（２）名詞句間の接続：接続詞による結合

（例４）関東，東北および北海道の上空に雨雲がある．

（３）文間の接続：文間の接続の手段としては以下の３通りの手段がある．

（ａ）接続詞による結合

（例５）
雨がやんだ．そして，日がさしてきた．
雨がやんだ．すると，日がさしてきた．
雨がやんだ．しかし，日はさしてこなかった．

（ｂ）接続表現なし．（句点結合）

（例６）
雨がやんだ．日がさしてきた．
雨がやんだ．日はさしてこなかった．

（ｃ）接続助詞による結合（読点結合）

（例７）
雨がやんで，虹がかかった．
雨がやむと，虹がかかった．
雨はやんだが，虹はかからなかった．

ここで，注意すべきは（ｃ）の読点結合の場合である．これは，文章の意味構造の上では，

２つの述語（＜雨が＞やむ，＜虹が＞かかる）で表されていたものが，接続助詞という言語的手段の使用により表層の表現では１つの文で表現されたことになる．このような場合には文章の意味構造と生成された文章の表層の構造は１対１に対応しない．

本報告では接続表現の生成の制御のうち，特に文間の接続表現の生成の制御方略に焦点をあてて考察を進める．

４．文章生成における接続表現生成

文章生成における我々の出発点は，文章の意味構造である．物語文章，説明文章などの通常の文章には節，段落，文といった意味的なまとまりに基づく階層構造が存在する．我々は文章の意味構造を以下のような構造とした．

文章における意味の最小単位は，述語を中心とした文フレームである．文フレームには，述語，格情報，法情報を含んでおり，通常の単文の意味構造に相当する．この文フレームの並びが段落フレームを構成し，段落フレームの並びが節フレームを構成し，節フレームの並びが文章を構成する．

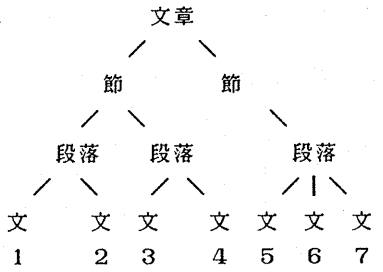


図１．文章の階層構造

文章生成はこの階層構造の木の末端を左から右にたどることにより進められる．また，結束性を考慮する処理を行うために必要な一般的な文脈情報として，以下に示すような情報がこの意味構造には含まれている．

文フレーム中：
　前の文フレームとの連接関係
　ローカルな焦点情報
段落フレーム中：
　文フレームの並び
　段落におけるグローバルな焦点情報
節フレーム中：
　段落フレームの並び
文章フレーム中：
　節フレームの並び
　文章のタイプ
　（説明文，物語文，作文，詩，戯曲等）
　文章のスタイル
　（丁寧，普通，口語調等）

表１．文間の連接関係（例）

TOP	文頭
ATT	添加・単純
SEQ	添加・継起
SUP	補足
TUR	転換
ANS	（疑問に対する答としての）連鎖・補足説明
TRA	（新しい話題への）転換・推移
REV	逆接
SEM	同時

４．１　文章の意味表現と文章の表層表現

文章の意味構造は図１に示すように階層的な構造を持っているが．この章では文フレームの構造が生成文章の表層表現に与える影響についてふれ，「How to say」のレベルでの生成の制御について考察する．

われわれは，段落フレームの構造を文フレームのリストで表している．このことは，発話の内容を決定する過程において，生成される文の並ぶ順序は既にほぼ決定されているということを意味している．従って，次に示す異なる文章の意味構造はそれぞれ発話の内容を決定する過程で異なるものとして定められているものとする．

<4a>（こどものさけは）まだ，力が弱いので，水にながされながら，いく日も，いく日もかかって，川を下ります。

<4b>（こどものさけは）まだ，力が弱いので，水に流されながら，川を下ります．いく日も，いく日もかかります．

<4c>（こどものさけは）まだ，力が弱いので，水に流されながら，川を下るのに，いく日も，いく日もかかります．

ここで，

文フレーム１（述語：弱い）

文フレーム２（述語：流される）

文フレーム３（述語：｛時間が｝かかる）

文フレーム４（述語：下る）

とする．

このような表層の文章表現の違いは，4aの文章の意味構造が，

文フレーム４（文フレーム３
（文フレーム１），
文フレーム２）

という構造をもち，同様に4bは

文フレーム４（文フレーム２
（文フレーム１））
文フレーム３（）

という構造を，また，4cは

文フレーム３（文フレーム４
（文フレーム２
（文フレーム１））

という構造を持つといった構造の違いによって生じると考える．この生成の制御を行うためには，文フレーム１，２，３，４間の関係を一般的なネットワーク構造としてとらえ，そのネットワークのたどりかたの方略を立てなければならないが，その制御要因は発話のゴール・プラン等に深く関わってくる．

本報告で考察の対象とする接続表現の生成の制御は，文フレームのリストの構造についての変更は，複数の文フレームが形式的に１つの文により表現されるといった読点結合の生成の制御に限られる．

４．２　生成の方略

小学校低学年の国語の教科書の分析から得られた接続表現の制御方略（第１版）を以下のように構成する．

まず，一般的な原則として

<r1>接続詞の種類は文のタイプ，スタイルでおおまかに限定される．一度選択された接続詞の種類は同一文章内ではあまり変化しないが，同一の接続表現は連続して使用されない．

<r2>連続した２つの文フレームに，文脈に関する意味のギャップが生じる場合（場所，時間，主題，アスペクト，感情等の変化及び常識や推論を用いてはじめて２つの文フレームの関係が正当化できる場合）には，それに対応する接続詞を生成する．

<r3>２つの文フレーム間の連接関係には，それに対応した接続詞が存在する．

が考えられる．

さらに，小学校低学年の国語の教科書中に現れる接続詞の種類は約６０－７０種であるが，その出現頻度にはかなり特徴がある．（表２）

表２．出現頻度の多い接続詞　（小学校国語教科書１，２年．昭和５８年度用）

	そして	すると	それから
教育出版	３１	１２	５
学校図書	３３	２９	１３
光村図書	３７	３１	１４

また	けれども	でも	しかし
８	８	１２	５
１４	１４	２２	１０
１５	１８	２４	６

（このほかの種類の接続詞の出現頻度は，「ところが」，「それで」，「そこで」等が１０回程度を除いては殆ど５回未満であった．なお，学校図書，光村図書の資料については，文献１３によった．）

この様に，継起の接続詞（そして，すると，それから）や，逆接の接続詞（けれども，でも，しかし）の出現が多く，小学校高学年の国語の文章においてもこの傾向は当てはまる（一方，１度しか用いられないていない接続詞の数が増え，その種類は１５０－１８０種になる）．以下では連接関係が継起，逆接の場合の接続表現の生成に焦点をあてて方略を構成する．

連接関係が「継起」（または「同時」）の場合は文の並び自体がその意味を表すことができるため，接続詞を用いて生成文の間の関係を陽に表現しなくても，文章の結束性はそれほど損なわれない．この場合は，読点結合を用いるか否かは，冗長性の観点から以下のような方略により制御する．

<r4.1>直前の生成文で読点結合が用いられていなければ，２つの文フレーム間の連接関係に対応する読点結合を生成する．

<r4.2>直前の生成文で読点結合が用いられていれば，２つの文フレーム間の連接関係に対応す

る接続詞を生成する．

連接関係が逆接の場合も読点結合が生じる可能性があり，その生成も冗長性の観点からの制約を受けるが，この連接関係は文の並びによって自然に表現できないため，現在のところその生成方略は構成していない．

接続詞の選択は文章の種類によっておおまかな語彙の選択が行われたあと，①連接関係に対して１意に接続詞を決定してよいものと，②更に文脈情報によりその選択が制限されるものに分けることができる．

①のタイプの接続表現としては逆接の接続詞，{しかし}，{けれども}等がある．

（例５）

a．一所懸命勉強した．でも，５点だった．

b．一所懸命勉強した．しかし，５点だった．

c．一所懸命勉強した．けれども，５点だった．

(a)と(b,c)の表現の選択は文章のスタイルで決まるが，bとcのいずれかを選ぶか選択は非決定的になされる．

②のタイプの接続表現として，継起の接続詞を選択する生成方略を以下に示す．

<r5>連接関係「継起」に対して，主題が継続していて主題の省略が生じている場合には{そして}，{て}を，主題が変化している場合には{すると}，{と}を，時間の変化が特に文フレームに示される場合は{それから}を用いる．

（例５）

秋になるころからおとなのさけは，たくさんあつまって，たまごをうみに，海から川へやってきます．そして，いきおいよく川を上ります．

（あつまる－[継起（て）]－やってくる－[継起（そして）]－上がる）{主題継続，アスペクト変化なし}）

（例６）

おかあさん鳥は，えさを見つけると，ココココと，合図をします．すると，ひよこはかけよって　きて，ついばみます．

（みつける－[同時（と）]－合図をする－[継起・主題変化（すると）]－かけよってくる－[継起（て）]－ついばむ）

この分類の境界は曖昧である．よりきめ細かな接続表現の生成の制御のためには，①のタイプの接続表現の数を減らすように，さらに文脈情報による制限を検討していく必要がある．

５．汎用文章生成システム上での実現

前章で構成された接続表現の生成の制御方略を汎用文章生成システムＨｉ－ＧＴＧ[15,16]上で実現してみた．Ｈｉ－ＧＴＧシステムは北海道大学大型計算機センターのUTILISPおよびApollo DomainのCommon-lisp上に作られている．システムの外部仕様はＭＩＴのＮＬＰシステム[1,3,4]（PAUL　Version）の仕様をほぼ満たしており（プリプロセッサ部を除くすべての機能を含んでいる），さらに日本語の活用語尾生成のための機能が拡張されている．本システムは，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[5]，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．

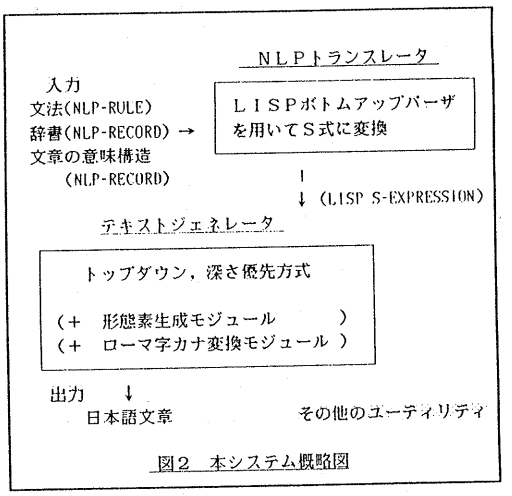


図２　本システム概略図

５．１　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　カテゴリー　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））（意味解析部））

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アルゴリズムの詳細については，文献６に詳しい．

５.２　ＮＬＰ記法

（１）ＮＬＰレコード

以下では，シンボルとは英数字および"_"，"＊"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．

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通りである．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．

（２）ＮＬＰルール

ＮＬＰルールはAPSG（Augmented Phrase Structure Grammars）[4]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機能が備わっている．これらの機能の追加，変更は容易である．主な機能しては，属性の参照に関しは，あるレコードの上位概念に沿った属性の参照や，別のレコードの属性からの間接的な参照が可能である．

また，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．

（例８）

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

現在のレコードのレコード名が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レコードとVPというレコード名のレコードが作られる．レコードNPにはレコードSENTの属性SUBJECTが持つ属性がすべてコピーされ，また，属性REFにレコードSENTの属性SUBJECTの値がコピーされ，属性SUBJECTに「T」がセットされる．

５.３　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，深さ優先方式で文章を生成するというシンプルなも

のとなっている．そのため，現在の文法記述能力では副詞句の生成や，強調による繰り返しなどの，語順の変化を伴う生成をうまく扱うことができない．

5.4　生成規則（の一部）

Ｈｉ－ＧＴＧの生成規則は現在約１１０個のルールで記述されているが，そのうち４章で考察した方略を取り込んだ規則の一部を以下に示す。

（１）文章のスタイルによる接続詞の選択は，ＮＬＰレコードの階層的な意味表現能力による（<r1>）

```
conjunction(
  rev:=list<'shikashi','keredomo'>);
```

（２）２つの文フレームを比べ，接続詞を生成するか否かの決定（<r4>）

```
contcept(slist('next')='seq',
         +slist('current')='seq',
         +te('conj')) -->
  sent(%contcept , te) comma*
contcept(slist('next')='sem',
         +slist('current')='sem',
         +to('conj')) -->
  sent(%contcept, to('conj')) comma*;
contcept(slist('previous')='seq',
         slist('current')='seq',
         te('conj')) -->
  conj(%concept, -te('conj')) comma*
  sent(%contcept);
```

（３）表層の接続表現の選択（<r1>,<r5>）

```
conj(special) -->
 conj(-special,sup:=special);
conj(rev)-->
 conj(-rev,
  sup:=myrandom<rev('conjunction'))>);
conj(seq,aag=aag('previous'),
    +timeflag  )
  --> conj(-seq,sup:='soshite' ) ;
conj(seq,+(aag=aag('previous')),
   +timeflag)
  --> conj(-seq,sup:='suruto' ) ;
conj(seq,timeflag )
  --> conj(-seq,sup:='sorekara' ) ;
conj(seq) --> conj(-seq,sup:='soshite');
```

なお，２つの述語間の因果関係を外部の知識として与え文フレーム間のギャップをチェックすることによる接続詞の挿入は，関数check2-concept,getconjを導入することにより

```
concept(check2concept<sup,
                sup('previous'),s*r>)
 --> conj(special,sup:=getconj
     <sup,sup('previous'),s*r)>) ;
     sent(%concept) ;
```

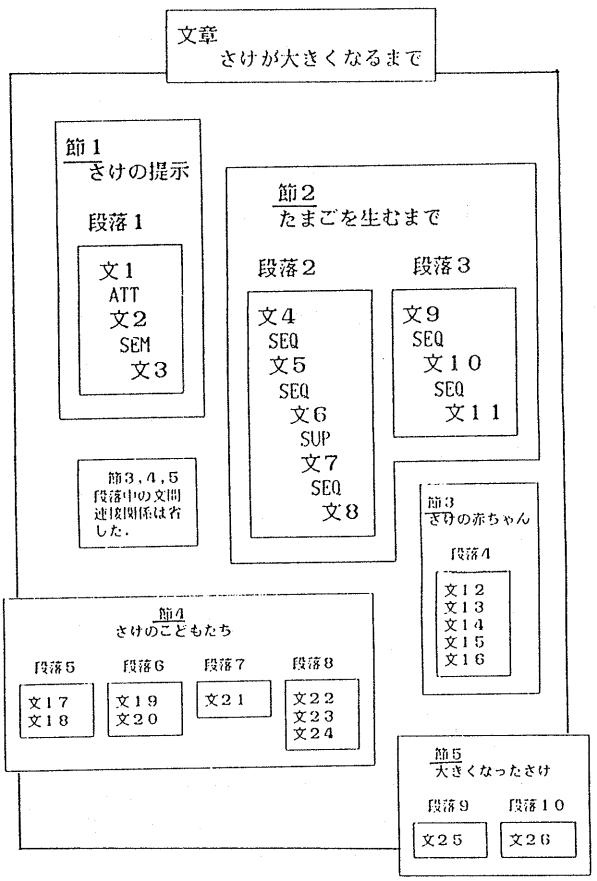
と表現できるが（<r2>），この方略はまだ実現していない．

5.5　生成実験例

しょうがくこくご１下，「さけが大きくなるまで」の文章を生成した例を示す．

まず，文章の意味構造の概略図を示す．

文章は５つの節フレームからなり，その節フレームは１０個の段落フレームに分かれる，全体で２６個の文フレームからなっている．



また，このＮＬＰ記法による表現の一部を以下に示す

```
text (paras := list<'para1','para2',
  'para3','para4','para5'> ,
   title := 'sake_ga_ookiku_narumade' ,
   type := 'setumeibun' ,
   text_style := 'pol') ;
para1 (contents := list <'subp1'> ,
       document := 'sake_no_teiji' );
subp2 (sents := list<'sent4','sent5',
'sent6','sent7','sent8'>,
 gfocus := 'otonano_sake' ,
 para_rel := 'seq' ) ;
sent4 ('atumaru' , prs , stative ,
        declarative , pol ,
        aag := 'sake1' ,
        timef := 'time1' ,
        focus :='sake1' ,
        s*rel :='tur') ;
```

生成実験結果を以下に原文と対比して示す．

（１）原文＜一部＞

秋に　なる　ころから，おとなの　さけは，たくさん　あつまって，たまごを　うみに，海から　川へ　やって　きます．そして，いきおいよく　川を　上ります．三メートルぐらいの　たきでも　のりこえて，川上へ　川上へと　すすんで　いきます．

（２）Ｈｉ－ＧＴＧによる生成文章＜一部＞（出力結果はローマ字であるが，見やすいように漢字かなまじり文に直した．）

秋に　なる　ころから，おとなの　さけは，たくさん　あつまって，たまごを　うみに，海から　川へ　やって　きます．そして，いきおいよく　川を　上ります．三メートルぐらいの　たきを　のりこえて，たいへん　川上へ　すすんで　いきます．

（原文での「川上へ　川上へと」いう強調表現は，生成文章では「たいへん　川上へ」と生成された．）

６．まとめ

本報告では，階層化された文章の意味構造を出発点として「how to say」の過程における接続表現の生成の制御について述べ，小学校低学年程度の文章における基本的な接続表現である経起，逆接等の接続表現の生成方略をＨｉ－ＧＴＧにより実現した．

本報告で構成した接続詞の生成方略は，大変限定されたものであるが，さらに実際の文章中で用いられている接続表現の調査を続け，その生成の制御方略の構成を行っていきたい．

参考文献

(1)Granville,R.A.: "Cohesion in Computer Text Generation:Lexical Substitution",MIT/LCS/TR-310(1983).

(2)Derr,M. and McKeown,K.:"Using Focus to Generate Complex and Simple Sentences", Proc. of COLING-84,pp.319-326(1984).

(3)Heidorn,G.E. :"Natural Language In-puts to a Simulation Programming System", Tech.Rep.NPS-55HD72010A,Naval Postgraduate School(1972).

(4)Heidorn,G.E.:"Augmented Phrase Structure Grammars",IBM Research RC 5787 (#25076)(1975).

(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(198).

(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 (1985).

(10)久野暲：談話の文法，大修館書店(1978).

(11)畠弘巳：接続詞と文章の展開，日本語教育５６号，pp.13-27(1985).

(12)宮地裕　２文の順接・逆接，日本語学，Vol.2，No.18,pp.22-29(1983)

(13)鈴木一彦・林巨樹（編）研究資料日本語文法第４巻，修飾句・独立句編　明治書院，pp.320-325(1984).

(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，情処第３１回全国大会，pp.1363-1364(1985).

(16)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，情処研究報告NL-56-2(1986).