

階層的な文章構造表現に基づく接続表現の生成と省略の処理

6G-2

池田光生

ikedai@icot.or.jp

新世代コンピュータ技術開発機構

1 はじめに

文章の生成においては、生成する文章の意味内容の生成、意味内容の文章構造への変換、文章構造に基づいた表出の三つの処理を行う。本稿ではこのうち表出について述べる。本稿で述べる日本語表出モジュール Sancho は (1) 隣接する文をいかに適切な接続表現を用いて結合するか、(2) 意味表現中のどのオブジェクトを省略するか、等を考慮することによって自然なつながりを持った結束性の高い文章を生成することができる。本稿ではこれらの二点について述べる。

Sancho への入力には木構造を用いて階層的に表現された文章構造データである。階層を用いて文の接続の強さを表すことにより、適切な接続表現を選択することができる。また、文章中で話題として取り上げられるオブジェクトの継続範囲を木構造を用いて表し、助詞「は」の使用やオブジェクトの省略に活用している。ここで言う階層的な文章構造表現とは文献 [2] で述べた、統語的な制約を反映した意味表現の階層的な考え方に基づいて設定した表現形式である。

本モジュールは、ICOTで開発した立論テキスト生成実験システム [1] の表出部であり、並列論理型言語 KLI を用いて記述した。KLI の各プロセスを意味表現の各階層に割り当ててモジュール性を高めた。本モジュールを含む立論テキスト生成実験システムを用いて約60個の文章の生成を行い、本モジュールの有効性を確かめた。なお、立論テキスト生成実験システムはICOTの無償公開ソフトウェアの一つである。

2 階層的な文章構造表現

本モジュールへの入力には立論テキスト生成実験システムの文章構造生成部によって作られた文章構造表現である。この入力データは文章の構造を木構造で表現したもので、

- 文の接続の強さを意味表現の階層を用いて表現している
- 話題となるオブジェクトの継続範囲を簡潔に表現している

の二点を特徴として持つ。階層的な文章構造表現を用いることにより、本モジュールにおける適切な接続表現の選択と助詞「は」の使用やオブジェクトの省略が可能になった。

木構造の葉に当たる項を命題項と呼び、一つの命題項または接続表現で結合した複数の命題項によって一文を表現する。後続する命題項に結合された命題項は従属節として表出される。命題項は表出の際の述語に相当する関係オブジェクトとその引数 (agent, object, goal など)、およびテンス・アスペクト、モダリティなどに関する情報からなる。本モジュールではこれらの命題項を適切な接続表現で結合していくことによって文章を表出する。文章構造表現の例を図1に示す。ただし、命題項の記述は簡略化してある。

「危険」「実施」「歩行者」のような個々のオブジェクトを基本オブジェクトと言う。基本オブジェクトから結合した複数の命題項までのオブジェクト間の関係によって意味表現の階層を分類する。基本オブジェクトを意味表現の第一層とし、基本オブジェクト間の関係を表す層を意味表現の第二層とする。第二層のオブジェクトにテンス・アスペクト、モダリティなどの情報を与えたものが命題項で、命題項は意味表現の第三層に相当する。

木構造中の各ノードは子ノード間の接続関係を記述し、隣接する命

Selecting connectives and ellipses based on a semantic representation hierarchy

IKEDA, Teruo

Institute for New Generation Computer Technology

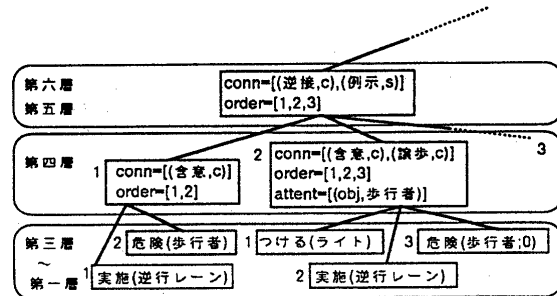


図1: 文章構造表現例

題項や部分文章構造の接続関係を規定する。各ノードは接続関係の強さに応じて異なる意味表現の階層に属する。助詞「は」の支配が文末にまで及ぶような従属節として命題項を表出する接続関係を持つノードは意味表現の第四層に属する。また、助詞「は」の支配が従属節を越えない接続関係を持つノードは意味表現の第五層に属する。隣接する命題項を重文化せず、接続詞を用いて二文として表す接続関係を持つノードは意味表現の第六層に属する。

木構造の各ノードは以下に示す3つの情報を持っている。

1. 子ノードの表出順序

表出の順に並べられた子ノードのリスト。

2. 子ノード間の接続関係

子ノード間の接続関係を表すデータのリスト。リストの*i*番目の要素が*i*番目に表出される子ノードと*i+1*番目に表出される子ノードとの間の関係を表す。接続関係は接続関係名と接続型の組で表される。接続関係名は、因果、理由、並列、逆接などの値をとり、接続型は、eop (段落を変える)、s (二文に分ける)、c (重文化する)のいずれかの値をとる。接続関係名と接続型の組み合わせから子ノードを結合するための接続表現を求める。

3. そのノード以下の表出における着目対象オブジェクト

ある部分文章構造において、システムがあるオブジェクトに着目して文章を生成することがある。その際、着目対象オブジェクトは、助詞「は」を伴って導入され、以後、その部分文章構造内では省略される。木構造の各ノードはそのノード以下の表出における着目対象オブジェクトが存在する場合、着目したオブジェクト名と着目時のルール名 (agent, object, goal など) の組をノード内に記述する。

子ノード間の接続の強さは木構造のルートに近いほど弱く、葉に近いほど強い。例えば、木構造の葉と葉は強い接続関係で結ばれているため、「~れば」「~とき」などの結び付きの強い従属節を用いて結合する。木構造のルートに近づくにしたがって接続関係が弱くなるため、結び付きの弱い従属節を用いて文を生成する。さらにルートに近いノードでは接続詞を用いて二文に分けたり、あるいは段落を変える処理を行う。

このように、入力となる文章構造データは接続の度合いに関する統語的な制約に基づいて文章の意味内容を構造化しているため、適切な接続表現を用いて結束性の高い文章を生成することができる。また、木構造を用いることによって、オブジェクトへの着目の継続範囲を簡潔に表現できる。

3 表出処理

本モジュールは並列論理型言語 KL1 を用いてインプリメントした。表出処理は前章で述べた意味表現の各階層ごとに割り当てたプロセスとそれらを管理するデータサーバ、および辞書サーバ、文脈管理サーバ、接続表現サーバ、活用処理プロセスの各プロセスを用いて行う。

各階層プロセスのうち第一層～第三層プロセスでは命題項の表出処理を行う。第三層プロセスでテンス・アスペクト、モダリティの処理を行い、第二層プロセスで基本オブジェクトを適切な語順に並べ、格助詞などの接辞を付加する。第一層プロセスでは基本オブジェクトを適切な語彙データに変換し、語彙データ列を活用処理プロセスへ送る。

第四層～第六層プロセスでは隣接する命題項を接続するための接続表現を選択する。第四層～第六層プロセスに入力された木構造データのルートノードは子ノード間の接続関係および表出順序に関する情報を持っている。第四層～第六層プロセスでは、まず、子ノード間の接続関係に関する情報を接続表現サーバに送ることにより接続表現を得る。次に、子ノードと接続表現を表出順に並べたリストをデータサーバに送る。

データサーバは各階層プロセスをコントロールする。表出順に並んだ文章構造データのリストが入力されると、データサーバはそれぞれの文章構造データの意味表現の階層を判定し、適切な階層プロセスにデータを送る。各階層プロセスでは前述のような処理を行い、データを細分化して接続表現と共に表出順に並べたリストにして返す。返された文章構造データは一般に意味表現の階層が下がる。

このようにして、文章構造データは部分構造データに分割され命題項のリストとなり、第三層プロセスに送られて処理される。各階層プロセスは並列に動作する。データサーバは表出処理の同期を取って適切な表出順序を保つ機能を持ち、各階層プロセスは個別に処理を進められる。表出処理のうち接続表現の生成とオブジェクトの省略処理について以下で述べる。

3.1 接続表現の生成

接続表現は木構造中の各ノードに記述された接続関係名と接続型から決める。接続表現サーバは入力された接続関係名と接続型に対応する接続表現を返す機能を持っている。対応する接続表現がいくつか考えられる場合はそれらを適当な順序で用いる。接続関係名、接続型と対応する接続表現を表 1 に示す。表中で nil とあるのは接続詞を使わないことを示す。また、今回の実験では日本語の接続表現の一部を抜くにとどまった。

3.2 オブジェクトの省略処理

各階層プロセスでは表出しようとするすべてのオブジェクトについて省略すべきかどうかを文脈管理サーバに問い合わせる。文脈管理サーバは表出した命題項の履歴を用いて省略処理を行う。省略処理の内容を以下に示す。

- データサーバは階層プロセスから返された文章構造データを処理するときに、データのルートノードに着目対象オブジェクトと着目時のロール名の組が記述されていれば、それを文脈管理サーバに送り、現在の着目対象リストとして登録する。その文章構造データの処理がすべて終了した後にはリストから削除する。
- 文脈管理サーバは送られてきた各命題項および命題項を構成するオブジェクトとロール名の組に対して以下の順で処理を行う。
 - 命題項またはオブジェクトとロール名の組が現在の主題リストの中のどれかと同じならば省略する。
 - 命題項が直前の命題項と同じならばその命題項全体を省略し、現在の主題リストに登録する。
 - 命題項またはオブジェクトとロール名の組が現在の着目対象リストの中のどれかと同じならば助詞「は」を用いて表出し、現在の主題リストに登録する。
 - 直前の命題項の中に同じオブジェクトとロール名の組があれば省略し、現在の主題リストに登録する。
 - 一つの命題項を表出し終わったときに、現在の主題リストの中からその命題項に含まれていないオブジェクトとロール名の組を削除する。

表 1: 接続関係名, 接続型と接続表現との対応

階層	接続関係	接続型	接続表現
第六層	演繹	s, eop	したがって, よって, だから, ゆえに
	因果	s, eop	だから, その結果, そして
	理由	s, eop	なぜなら, というのは
	展開	s, eop	すると
	逆接	s, eop	しかし, だが
	並列	s	また
	並列	eop	nil
	演繹	c	ので, から, ために
	転換	s, eop	ところで, 一方
	例示	s, eop	例えば
	一般化	s, eop	このように
	提示	s, eop	nil
	目的	s, eop	そのために
第五層	因果	c	ので, から, ために
	展開	c	と, たら, とすると, てみたら
	逆接	c	が, けれど, のに
	並列	c	れば,
	提示	c	が
第四層	目的	c	ために
	含意	c	れば (仮定的条件句), ので (事実的条件句)
	条件	c	ときに (仮定的条件句), ので (事実的条件句)
	累加	c	れば
	譲歩	c	ても
前提	c	ときに	

4 生成結果と評価

生成した文章の一つを以下に示す。

御堂筋では警察によって逆行レーンが実施されなくて、一方通行が実施された。だから、バスの会社がルートを変えた。その結果、40% 乗客を失った。このように、警察によって一方通行が実施されるときに逆行レーンが実施されなければ警察がルートを変えて会社に交えて、会社が乗客を失う。また、バスを廃止してしまう。

警察によって逆行レーンが実施されれば歩行者が危険のように見えるが、歩行者はバスがライトをつければ警察によって逆行レーンが実施されても危険でない。

したがって、逆行レーンが実施されなければならない

本モジュールの入力として設定した階層的な文章構造表現によって適切に接続表現の選択とオブジェクトの省略が行える。より質の高い文章を生成するためには、連体修飾句や埋め込み文などを生成するためのより豊かな文章構造を生成する必要がある。意味表現の階層は統語的な制約を反映したものであるから、本モジュールの前段階である文章構造生成部の処理内容を充実させなければならない。

参考文献

- IKEDA Teruo, KOTANI Akira, HAGIWARA Kaoru, and KUBO Yukihiro. Argument text generation system (dulcinea). In *International Conference on Fifth Generation Computer Systems 1992*, volume 1, pages 385-394. ICOT, 1992.
- 池田先生. 意味表現の階層に基づく自然言語生成システム. コンピュータソフトウェア, 8(6):28-37, 1991.