

L T B 文生成部の生成方式

4C-7

池田 光生*、幡野 浩司**、福島 秀顕**、重永 信一*
 松下電器東京研究所*、新世代コンピュータ技術開発機構**

1. はじめに

本文生成部は、属性-属性値の属性構造である中間表現形式 [重永 他 87] によって与えられる文法的な情報を用いて日本語文を生成する。生成される表層文の語彙や語順などは、与えられる中間表現上で指定されるが、直接の指定のない限り、デフォルト処理に従って生成される。

図1に生成部の生成処理の流れを示す。辞書及び文法は、編集系のツール類 [幡野 他 88] を用いて作成される。

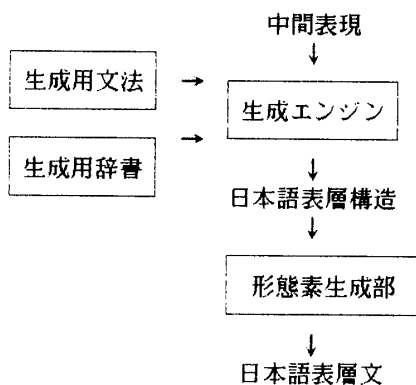


図1

2. 中間表現形式の例

図2に、中間表現形式による記述例をあける。これは、「人間は、その優れた技術を使って、自然から得たものを巧みに加工し、自分たちの生活を豊かにしている。」という日本語文に対応している。

- 1) {関係/並立、
- 2) 関係子/and、
- 3) ロール/{要素 1/
- 4) {関係/{語彙/加工}
- 5) ロール/
- 6) {行為者/{語彙/人間、
- 7) 取り立て詞/は、
- 8) 対象/{語彙/物、
- 9) 道具/{関係/{語彙/使う}、
- 10) ロール/
- 11) {対象/{語彙/技術、
- 12) 連体修飾/
- 13) set([{関係/{語彙/優れる}、
- 14) 支配/対象、
- 15) 時制/過去、
- 16) ロール/
- 17) {対象/{語彙/技術}}、
- 18) {限定詞/その}]]))、

- 19) 様態/{語彙/巧み}}、
- 20) 要素 2/
- 21) {関係/{語彙/する}、
- 22) 相/結果状態、
- 23) ロール/{終点/{語彙/豊か}、
- 24) 対象/{語彙/生活、
- :

図2

例文は「加工する」と「する」という二つの述語が並立してできている。「加工する」という述語で表される関係は”行為者”、”道具”、”対象”、”様態”という格に対応する4つのロールを持っていて、「する」という述語で表される関係は”対象”、”終点”というロールを持っている。「技術」は「その」と「優れた」に連体修飾されている。「する」のアスペクト素性として、「結果状態」という値が与えられていて、その結果「している」という表層表現と対応する。

3. 生成エンジンの処理

3.1 ロールの処理

述語の格成分を中間表現データから取り出し、辞書の格情報を参照しながら、それらの間の語順と、格成分に付加すべき接辞を決定する。表層属性によって表層表現が与えられているときにはそれを用いる。

例6),8)では、「加工」の辞書記述を参照して、行為者格の接辞を「が」に対象格の接辞を「を」にする。また、例23),24)では「する」の辞書記述を参照して、対象格の接辞を「を」に、終点格の接辞を「に」にする。

「道具」、「目的」などの自由格は辞書に書かれていない。自由格については、格成分の品詞などから接辞を決定する。例9)では、「加工」の道具格の成分が従属文なので、接辞として「て」を付加する。また、例19)では様態格の成分は、「巧み」という形容名詞であるから、連接形を指示して、「巧みに」と活用させる。

語順属性を用いて、ロールの語順が指定されている場合にはその語順で生成する。指定のない場合には、語順規則により決定する。本生成部では、格要素を7つのグループに分けて、そのグループ間の順序を定めている。なお、取り立て詞の「は」を持つものは他の格よりも前にする。

例7)では、行為者格が取り立て詞の「は」を持っているため先頭にくる。例8)の対象格は辞書に記述されている必須格であるから他の格よりも後ろにする。

3.2 特殊関係の処理

関係の値は、通常、動詞や形容詞などの述語であるが、本生成部では、述語以外に以下のような特殊関係を用意している。それぞれの処理について述べる。

「並立」並立関係は要素1、要素2というロールを持つ。

- 1) 文並立 文並立は関係子の値として、「and」、「or」、「順接」、「逆接」の4つを持つ。例2)では、「and」となっており、「加工し」、「する」と生成する。「or」の場合は、「する」という動詞を付加して、「たり」、「たりする」と生成する。
- 2) 句並立 関係子の値が「and」のとき、「と」、「or」のとき、「や」と生成する。

以下は連体修飾で用いられる関係である。二つの名詞句A、B間に以下の関係があるものとする。

- 1) の 関係子の値が「テーマ」のとき「AについてのB」それ以外は「AのB」と生成する。
- 2) 同格 「AというB」と生成する。
- 3) 名詞連接「AB」と生成する。例) 情報処理

3.3 連体修飾の処理

例の「その優れた技術」のように連体修飾句が複数ある場合は、連体修飾属性の値をリストにするか、あるいはset()を用いて()内にリストを入れる。リストで与えられたときは、その順番に生成する。set()を用いたときは、連体修飾の語順規則によって生成順を決定する。

本生成部では、連体詞による修飾句を最も被修飾句の近くにおき、次に「その」「あの」などの限定詞による修飾句をおく。「の」「同格」などの特殊関係と、述語による修飾句は他の修飾句よりも前にくるが、述語が形容詞の場合は特殊関係より後、それ以外は特殊関係より前にする。

3.4 動詞化

関係属性の値には述語性の名詞を入れることができる。例4)では、サ変名詞である「加工」が関係属性の値となっている。関係属性の値が名詞であり、かつ派生名詞化の指定がなければ動詞化を行う。形容名詞及び状態性の副詞と連体詞には、「だ」を付加して活用型をダ系にする。サ変名詞には、「する」を付加して活用型をウ系サ変にする。

3.5 名詞化

格成分が動詞である場合には、格要素名と接辞とを参照して名詞化属性の値を付与する。名詞化属性の値によって、名詞化の方法が指定されているときはそれに従う。指定のない場合には、名詞化属性付与規則によって属性値を付加する。

3.6 アスペクト辞の生成

述語の辞書記述には各述語のアスペクト素性の値が書かれている。このアスペクト素性の値と中間表現のアスペクト属性の値とから、アスペクト辞を決定する。例ではアスペクト素性が「瞬間」である「する」という動詞に対して、「結果状態」というアスペクト属性を与えている。「瞬間」と「結果状態」とから「ている」というアスペクト辞を決定する。また、「継続(ていく)」のように直接アスペクト辞を指定された場合は、それを用いる。

3.7 モダリティの処理

次のようなモーダル属性の値を例に説明する。

*)モーダル/{ムード/[許可,説明(のだ),推量(う)],
態度/疑問}

)は「てもよいのだろうか」という表層表現に対応している。複合モダリティの場合はリストを用いて表し、リストに書かれた順番に生成する。)では態度の値が「疑問」であるから、最後に終助詞「か」を付加する。

推量のように「かもしれない」「らしい」「ようだ」「にちがいない」などいくつかの表現が可能なものについては推量(ようだ)のように直接指定する。

4. 形態素生成部の処理

4.1 前処理部

形態素生成部では不適当な語彙データを修正するための前処理を行った後に活用処理を行う。例6)~8)では、行為格である「人間」に格助詞「が」と取り立て詞「は」が付加されて形態素生成部に渡される。形態素生成部では前処理として、格助詞「が」を削除し取り立て詞「は」を残す。

4.2 活用処理部

活用処理部では、語彙データ列から活用を変化させるための後続する最初の語彙データを求め、接続表により活用形を決める。例文の「その優れた技術を使って」の部分の語彙データは「その」「優れる」「た」「技術」「を」「使う」「て」となっている。「優れる」と後続する「た」とから「優れる」の活用形を決めて活用させる。同様に「使う」と後続する「て」とから「使う」の活用形を決めて活用させる。

5. まとめと今後の課題

L T B文生成部の生成方式について述べた。入力データとなる中間表現には生成に必要な文法的な情報が書かれている。直接の指定がされない限り、デフォルト処理に従った表層表現を生成する。

デフォルト処理の部分は現在、文法規則の手続き中に散在していて、ユーザの変更が困難であり、汎用性を欠いている。これを切り分けることが今後の課題である。

また、述語が省略された場合のデフォルト処理、サ変名詞を動詞化せずに普通名詞として使用する生成、並立関係と格関係の整理、中間表現形式の見直しなどについて今後の検討が必要である。

参考文献

- [奥津 他 86] 奥津敬一郎、沼田善子、杉本武
「いわゆる日本語助詞の研究」
にほんごの凡人社 1986
- [金田一 76] 金田一春彦 編 「日本語動詞のアスペクト」 むぎ書房 1976
- [寺村 84] 寺村秀夫 「日本語のシンタクスと意味」
くろしお出版 1984
- [幡野 他 88] 幡野浩司 他 「L T B文生成部の構成」
情報処理学会第37回全国大会 1988
- [重永 他 87] 重永信一 安川 秀樹
「自然言語処理システムのための日本語
文生成ツールの試作」
情報処理学会第35回全国大会 1987