

LTB 文生成部の構成

梶野浩司^{*} 福島秀顕^{*} 小島量^{**} 池田光生^{***} 重永信一^{***}

^{*}新世代コンピュータ技術開発機構 ^{**}管理工学研究所 ^{***}松下電器東京研究所

1. はじめに

我々は、汎用日本語処理系 (LTB)^[1] に組み込まれる日本語文生成部の設計開発を行なっている。文生成部は、日本語文を生成する生成エンジン、形態素処理機構のほかに、編集機能、デバッグ機能など、文生成に必要なデータをユーザが作成するためのツールを提供するものである。本稿では、SIMPOS^[2]上に設計した文生成部の全体構成についての説明をするとともに、開発用ツール群を使用して文法規則および生成用辞書を作成した例について述べる。開発した文生成部は次のような特徴を持つ。

- ・日本語の構文的な情報のみを用いて生成を行なう。
- ・形態素生成に用いる活用表は、LAX^[3]の形態素文法に準拠する。
- ・生成する語彙や語順を必要に応じて指定できる。
- ・文法と辞書が完全に分離されている。

2. 文生成部の構成

文生成部の全体構成を図1に示す。文生成部は、次のふたつの処理系からなる。

実行系: 入力した中間表現をもとに表層文を生成するための処理機構 (生成エンジン、形態素生成部) と、生成過程を追跡、デバッグするためのツール (デバッグ) からなる。

編集系: 文生成に必要な各種データを作成、変更するためのツール (構造エディタ) と、データ変換のためのツール (トランスレータ) からなる。

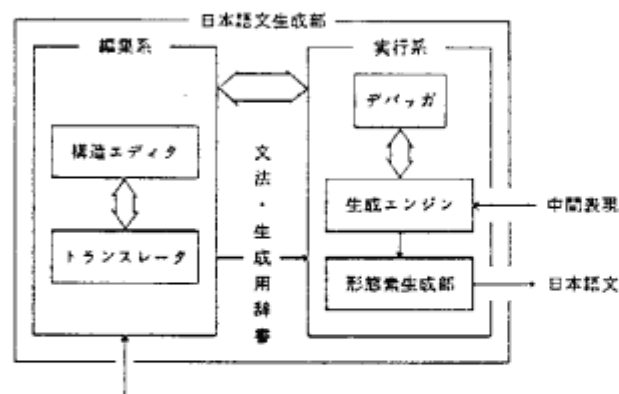


図1 文生成部の全体構成

"Configuration of the Sentence Generator in LTB"

HATANO, Kozi^{*} FUKUSHIMA, Hideaki^{*} OJIMA, Ryo^{**} IKEDA, Mitsuo^{***} SHIGENAGA, Shin-ichi^{***}

^{*}Institute for New Generation Computer Technology ^{**}Kanri Kogaku, Ltd.

^{***}Tokyo Research Lab., Matsushita Electric Industrial Co., Ltd.

からなる。

ここでいう中間表現とは、生成すべき日本語文の構文構造を記述したものであり、CIL^[4]の部分項で表現されている。また、辞書については現在、解析/生成共用辞書 (マスター辞書)^[5]を生成用辞書に変換したものを利用している。しかし、マスター辞書にない項目を一時的に追加したり、直接生成用辞書に変更を加えることも可能な構成になっている。実行系/編集系の利用は、メニューによるコマンド選択方式によって統合的に行なうことができる。つぎに、実行系および編集系について説明する。

2.1 実行系

表層文の生成は次のようにして行なう。

- (1) あらかじめ用意した文法規則および生成用辞書を用いることにより、入力した中間表現から表層構造を生成する。規則の適用および辞書の参照は生成エンジンが制御する。
- (2) (1) で生成した表層構造に形態素処理を施すことにより、表層文を生成する。この処理は、形態素生成部が行なう。

上記の (1) および (2) の過程をデバッグで追跡、デバッグするわけである。生成方式の詳細の説明については他稿^[6]にゆずることとし、本稿ではデバッグのもつ機能について説明する。

デバッグ

文生成用デバッグは、CILデバッグ^[7]の機能を用いたものである。CILデバッグの種々の機能のうち、文生成部のデバッグには特に次の機能が有効である。

- (a) 規則適用過程を追跡する機能
- (b) 適用中の規則をソースファイル上で反転表示する機能。
- (c) 実行中に各種データの内容を表示、変更する機能 (インスペクタ)
- (d) 実行中の規則に一時的な変更を加える機能。

文生成の際、文法規則および辞書は内部的にCIL述語に変換して用いているが、CILデバッグ上に文法規則、辞書それぞれの構文定義を行なうことによってユーザが内部表現を意識せずにデバッグすることが可能になっている。また、上記 (a)~(d) の機能の他に、生成用の独自の機能として、文生成部に入力した中間表現をウィンドウにプリティプリントする機能などがある。デバッグ

- 14 -