ＬＴＢ文生成部の構成

4C-6

幡野浩司*　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[2]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（デバッガ）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。

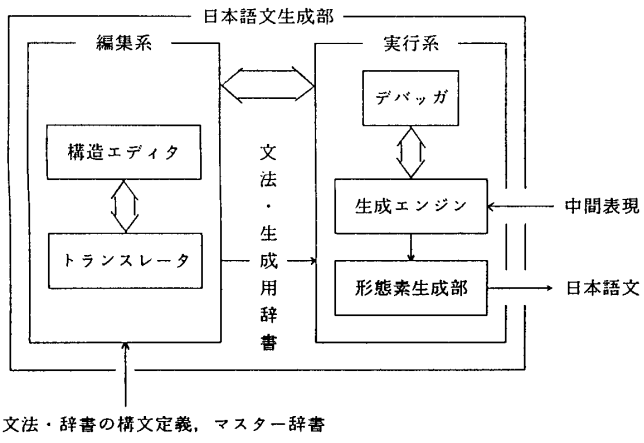


図１　文生成部の全体構成

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[3]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）[4]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) の過程をデバッガで追跡，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[5]にゆずることとし，本稿ではデバッガのもつ機能について説明する。

デバッガ

文生成用デバッガは，ＣＩＬデバッガ[3]の機能を用いたものである。ＣＩＬデバッガ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デバッガ上に文法規則，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。デバッガ

"Configuration of the Sentence Generator in LTB"
HATANO, Kozi*　FUKUSHIMA, Hideaki*　OJIMA, Ryo**　IKEDA, Teruo***　SHIGENAGA, Shin-ichi***
*Institute for New Generation Computer Technology　**Kanri Kogaku, Ltd.
***Tokyo Research Lab., Matsushita Electric Industrial Co., Ltd.

を用いて生成過程の追跡を行なっている例を図２に示す。図では文法規則の適用過程を追跡している。

2.2 編集系

生成に必要なデータは，次のようにして作成する。

(1) 文法規則：ユーザがエディタを用いて作成する。

(2) 生成用辞書：マスター辞書を変換して作成するか，生成用辞書を直接編集する。

上記 (1)，(2) の作業を編集系が支援するのである。編集系は，文法の内容をそれらの構文定義にしたがって構造的に編集する構造エディタ，マスター辞書を生成用辞書に変換するトランスレータからなる。以下に構造エディタ，トランスレータについて説明する。

構造エディタ

文法規則など，定まった構文を持つデータの編集のために，構造エディタをツールとして用意した。本エディタの操作は，基本的にＳＩＭＰＯＳ上のＰＭＡＣＳの操作体系に基づいているが，ユーザがＢＮＦ記法による編集対象データの構文定義ファイルを作成することにより，次のような構造的編集を行なうことを可能にしている。

(a) データの構文チェック

(b) データのプリティプリント

(c) 節単位の移動，複写，削除，

(d) データのエントリ名検索

(e) データ中の各種パラメタ記述の変更

本エディタによれば，データをテキストとして編集したり，構造的に編集したりという切り換えをコマンドにより編集中に行なうことができる。図２に本エディタを用いて文法規則を編集している例をあげた。図では，文法規則の修正を行なっている。

トランスレータ

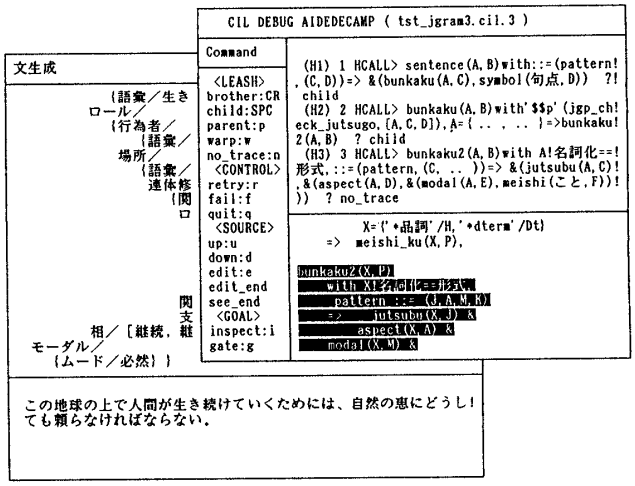
トランスレータは，マスター辞書から，生成に必要な情報をとりだして生成用辞書を自動作成する。変換は，マスター辞書が格納されているデータベースをアクセスしながら行なう。本トランスレータは，前述の構造エディタからコマンドによって呼び出すことができる。

3．開発環境の使用例

以上に述べた開発環境を用いて，我々は 4000 単語規模の辞書および文法を作成した。現在，これらのデータは文生成部に組み込まれ，全体として汎用的な日本語文生成部として稼働している。

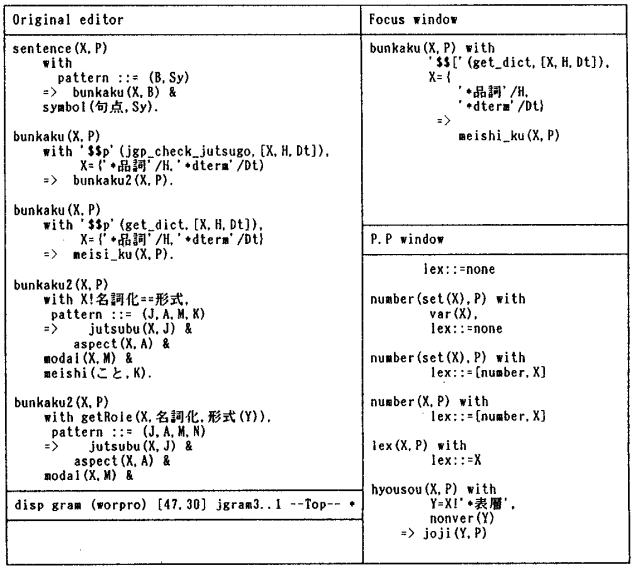
4．おわりに

中間表現から日本語文を生成する文生成機構，および文生成用文法規則開発者用の開発環境をＳＩＭＰＯＳ上に構築し，汎用的な日本語文生成部を開発した。現在我々は，この文生成部を利用しながら談話理解システムＤＵＡＬＳ第３版[6]の文生成プラニング部の設計を行なっている。



左上：生成エンジンに入力した中間表現　左下：出力した日本語文
右上：文法規則の追跡経過　右下：文法規則のソースファイル（現在適用されている規則節が反転表示されている。）

図２ デバッガの使用例



左：編集中のデータ（文法規則）のソースファイル内容　右上：現在編集中の節　右下：プリティプリントされたデータ

図３ 構造エディタの使用例

【参考文献】

[1] 杉村他：汎用日本語処理系ＬＴＢの構成，情報処理学会第３７回全国大会予稿集，1988年9月

[2] 杉村他：論理型形態素解析ＬＡＸ，Proceedings ofthe Logic Programming Conference，1988年4月

[3] ＣＩＬ言語マニュアル（第３版第２刷），ICOT-TM-242，1988年4月

[4] 瀧塚他：ＬＴＢマスター辞書の構成，ソフトウェア科学会，論理と自然言語研究会，1987年12月

[5] 池田他：ＬＴＢ文生成部の生成方式，情報処理学会第３７回全国大会予稿集，1988年9月

[6] 橋田他：ＤＵＡＬＳにおける談話処理，情報処理学会第３７回全国大会予稿集，1988年9月