

2G-6

日本文インタフェース構築ツール「ゆい」における例文からの辞書登録ツール

吉田宗弘 町田和浩 三枝克広 杉山高弘* 和田孝*
 日本電気技術情報システム開発(株) *日本電気(株)ソフトウェア生産技術開発本部

1 はじめに

計算機のコマンドになれ親しんでない人でも、日常の会話文に近い日本語の命令文を入力することによって計算機を容易に使いこなせるような日本文インタフェースが必要となってきた。しかし、日本文インタフェースを応用システム毎に作ることは、通常、自然言語処理に関する難しい知識と膨大な辞書を開発する期間を必要とする。そこで、汎用な日本文インタフェース構築ツール「ゆい(YUI: Yet another User Interface)」[1]を開発した。構築ツールの構成を図1に示す。本稿では特に構築ツールの主機能である辞書登録ツールについて説明する。

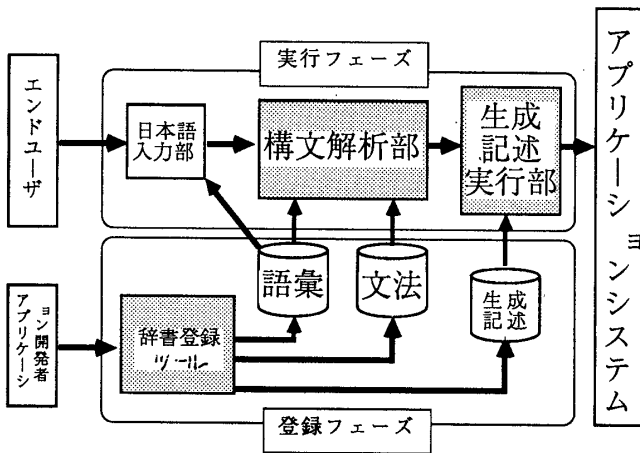


図1: 日本文インタフェース構築ツールのシステム構成

2 辞書登録ツール

辞書登録ツールは、難しい言語処理知識を必要とせずアプリケーションの分野の典型的な例文から直接、文法・語いを作成し、それぞれの辞書に登録する支援を行うため、辞書をコンパクトにすることができ、日本文

A dictionary development tool from examples in the Japanese language interface toolkit YUI

Munehiro Yoshida, Kazuhiro Machida, Katuhiro Saegusa, Takahiro Sugiyama*, Takasi Wada*
 NEC Scientific Information System Development, Ltd, * NEC Corporation.

インタフェースの移植性を高めるとともに、分野依存した表現を可能にしている。また、既存の文法知識にとらわれない登録者独自の日本文の認識単位の導入により口語文なども表現することができる。日本文インタフェースの運用時に受理できなかった入力文を保存しているログファイルから、本ツールにより文法を生成することにより受理される確率を高めていくことができる。このツールは文法登録部、語い登録部、生成記述登録部からなり、以下で詳しく述べる。

2.1 文法辞書登録機能

辞書登録ツールの最も重要な機能である文法辞書登録機能は、構文解析したい日本文の例文を入力することによって文法を登録する。ユニフィケーション文法(LFG等)で採用されているSBJ(主語)、OBJ(目的語)等の構文的に決まる機能を表す文法機能や、キーワードとなる述語と文中に出現する助詞の種類によって決定する述語項構造は本ツールが自動的に登録する。

図2で示すように本ツールの文法作成支援として、まず形態素解析ツールを用いて自動的に入力文を語いレベルで分かち書きし、入力文の名詞部分をマークし文節間に区切り記号を挿入する。この解析結果に誤りが生じていたり、登録者独自の文節認知単位に修正したいときは適宜グラフィック画面上で例文を直接操作することによって修正が可能である。ただし、ここでいう文節とは以下の形式に従うものとする。

文節 := [接頭語] 変数項 [接尾語] | 定数項
 変数項 := <マークされた名詞>

マークされた名詞部分を抜き出し必要な部分だけ(主に用言と助詞からなる格構造)を残した文法のひな型を作る。このひな型をスケルトンと呼ぶことにする。スケルトンは、定数項(文字列)と変数項からなる。定数項は、スケルトンに残った文字列を表し、構文解析時には解析文と文字列マッチングされる。変数項はスケルトンの抜き出された部分を表し、構文解析時に入力文中の文字列で、かつ、語辞書に登録されている文字列がバインディングされる。登録画面上では\$変数に1番から通し番号が自動的に割り当てられる。

最初に入力した例文以外の入力文もスケルトンによって解析できるように、以下の拡張を文法登録ツール

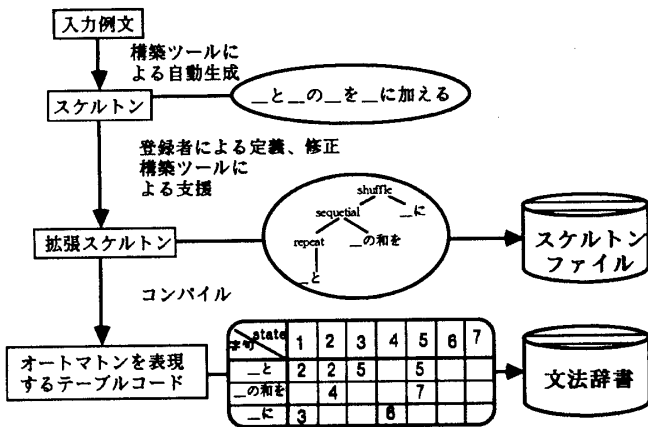


図 2: 文法の作成

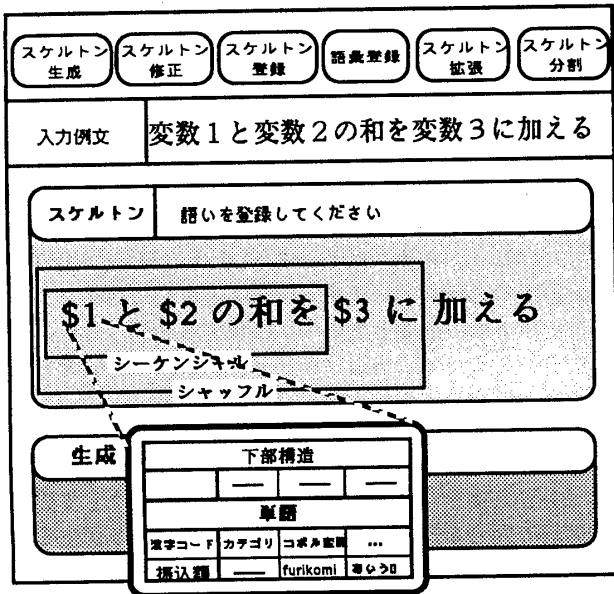


図 3: スケルトン拡張

ルは支援する。実際に文法登録ツールでは、図3で示されるように指定する。

- 助詞の出現順の順不同 (シャッフル)
- 助詞の出現順を固定 (シーケンシャル)
- 省略可能 (オプション)
- 同一句の繰り返し (繰り返し)
- 同等な言い回し (並列)
- ネスト構造に分割

2.2 語い登録機能

構文解析時に参照する語い辞書の登録機能は、アプリケーションシステムへのカスタマイズ機能として重要である。語い登録については、大きく2つの手段があ

る。1つは、前節で示したスケルトン生成時に入力文中に表れる語いをシステムが自動的に登録する。もう1つは、スケルトンのふさわしい変数項に直接語いを追加登録する。語い辞書に登録する際に、以下で示す3つ組の登録内容とその語いが出現するスケルトンと変数項の位置情報をシステムが自動的に付け加える。

- 入力文中に出現する漢字コード
- 語いをアプリケーション対象分野で用いられる形式に変換した生成オブジェクト
- 生成記述で意味解析を助ける意味カテゴリ

語い登録エディタは、図3の右下にある表入力形式のエディタである。登録したい変数項の位置をマウスで指定することによって登録エディタがポップアップする。

2.3 生成記述登録部

本生成記述を用いれば、登録者はアプリケーションシステムが必要とするコマンドや構文解析結果を既存のプログラム言語で書くより簡単に記述できる。特に、複雑な条件式や制御文を記述せず、あらかじめ用意された生成記述の機能を用いればよい。生成記述と構文解析文法が一体となった登録画面のため、記述がローカルにとじて解析の流れが追いやすく保守性も向上する。現在、リスブ関数の他、Xlib、XToolkit 関数、その上で開発したグラフィック画面操作関数「鼎ライブラリ」[2]等が記述できる。

3 まとめ

本ツールは EWS4800 (UNIX SYSTEM V) および SUN3 (UNIX BSD3.4) の X ウィンドウ (X11R2) 上で C 言語により開発した。グラフィック画面作成ツールは「ユーザインタフェース構築環境：鼎」を利用して開発した。本ツールを用いることによって難しい自然言語処理知識を必要とせず、例文から文法・語いを獲得することによって日本語インタフェースを構築し、今まで困難とされていた分野のアプリケーションシステムを日本語を用いることによって操作できる。

参考文献

[1] 杉山他「例文からの文法獲得に基づく日本語インタフェース構築ツール『ゆい』」情報処理学会、自然言語処理研究会 73-2 (1989).

[2] 暦本、菅井、杉山、他「X ウィンドウ上のマルチメディアユーザインタフェース構築環境: 鼎」情報処理学会第 30 回プログラミングシンポジウム予稿集、(1989).