

例文による品詞指定を用いた単語登録方式

3 J-3

富山幸男、森川好美、†菅井勝

日本電気マイコンテクノロジー(株) ソフトウェア生産技術部、†日本電気(株) ソフトウェア生産技術開発本部

1 はじめに

日本語ワードプロセッサやワープロソフトなどの発達にともない、日本語入力システムにおける 仮名漢字変換効率に対する要望も高まってきている。通常、日本語入力システムの操作中に 変換できない単語がある場合、単語登録を行う。この単語登録において、登録を行う単語に対する品詞の選択が重要になる。

一般の日本語入力システムの代表例として VJE[1] の単語登録における品詞選択の例を示す。単語登録のモードに移ると、名詞系を中心とした五種類ぐらいの品詞が並ぶ。登録する品詞がない場合は、たくさんの単語登録が出来るモードに移る。このモードでは、だいたい 30 種類の品詞が並び、この中から該当すると思う品詞を探すようになっている。また、他の日本語入力システムでは、登録できる品詞の数を名詞を中心とした五種類ほどにしているものもある。上記に示す方法では、名詞系の品詞を除く単語(品詞分類の難しい語)を正しく登録する場合に、(1) たくさんの品詞の中から、選択しなければならない。(2) 日本語文法を熟知していないと難しい。(3) 選択した品詞が合っているかどうかの確認が出来ない。(4) 登録できない品詞がある。といった問題が感じられる。

上記の問題を解決する方法として、日本語入力システムにおける単語登録に 例文を使用して、品詞選択の情報を提供していく例文による品詞指定方式について述べる。

2 例文による品詞指定システム

図1に例文による品詞指定方式の流れを示す。

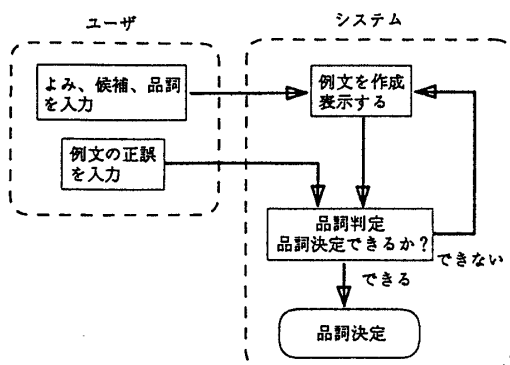


図1例文による品詞指定の流れ

ユーザはよみ、候補、品詞を入力する。この情報をもとにシステムは、例文を作成し、表示する。ユーザは、システムが表示した例文の正誤を判定し、入力する。システムは、ユーザの入力した、よみ、候補、品詞と、例文の正誤の情報をもとに、品詞判定を行い品詞を決定する。品詞決定できない場合は、次の例文を作成し、表示する。

図2に岩波国語辞典[2]の品詞分類を用いた品詞を示す。木構造の末端の品詞37種類が登録できる品詞である。図2に示す品詞をVJEの方法を用いて単語登録を行うと、この37種類の中からユーザが直接、品詞を指定しなければならない。一方、図2に示す品詞を本論文の方式を用いた場合、点線部分は、例文を用いて品詞決定していくので、ユーザは、四角で囲まれた部分の11種類の中から品詞指定を行えば良い。

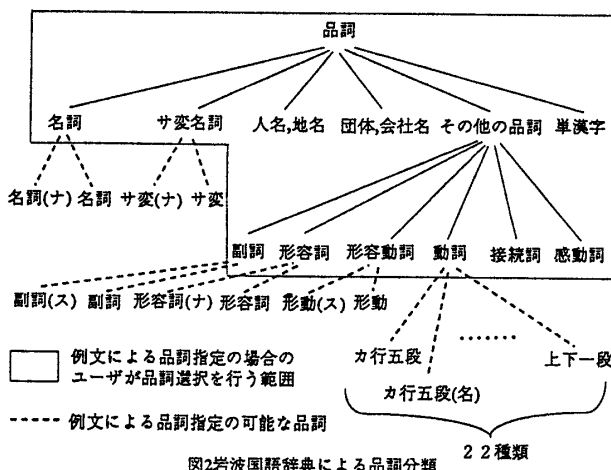


図2岩波国語辞典による品詞分類

3 品詞判定のアルゴリズム

例文による品詞指定登録は、図2に示すように 名詞、サ変名詞、動詞、形容詞、形容動詞、副詞の時に使用される。以下にそれぞれの場合の例文の作成方法を示す。

名詞、サ変名詞

よみに(かいしゃ)、候補に(会社)、品詞に(名詞)を入力したとする。システムは、入力された候補の「会社」に「+」な「+名詞」をつなげて、「会社+ +名詞」という文を作る。つなげる名詞には、もの、とき、を用いる。そして、「会社+ +名詞」は正しいかどうかをユーザに聞く。図3に画面例を示す。

よみ	かいしゃ	候補	会社
品詞	名詞(ナ)	人名	地名
		団体	名詞
		サ変	單漢
			その他
例文は文法的に正しいですか？			
☒ 会社なとき			
登録		取消	
		終了	

図3 名詞の登録例

ユーザの正誤入力：NO

名詞の選択

正誤入力 YES の場合 名詞(ナ)

正誤入力 NO の場合 名詞

サ変名詞の選択

正誤入力 YES の場合 サ変名詞(ナ)

正誤入力 NO の場合 サ変名詞

動詞

よみに(よむ)、候補に(読む)、品詞に(動詞)を入力したとする。ユーザは、動詞を登録する時は よみ、候補を、終止形で入力する。システムは、入力された候補の「読む」をもとに、変格活用であるかどうか調べ、変格活用でなければ、五段活用を作成する。作成した五段活用が正しいかどうかユーザに聞く。

システムの出力する例文：読まない

ユーザの正誤入力：YES

YES: 五段活用 と判定できる。

NO: 一段活用 と判定できる。

五段活用の場合は「読む」をもとに五段活用連用形を作成し、「+ が、を + 動詞」をつなげ、「読み + が、を + 動詞」は正しいかどうかユーザに聞く。下に画面例を示す。

よみ	よむ	候補	読む
品詞	ま行五段(ナ)	人名	地名
		団体	名詞
		サ変	單漢
			その他
文法的に正しいものを選択して下さい。			
☒ 読まない		☐ 読ない	
☒ 読みがいい		☐ 読がいい	
登録		取消	
		終了	

図4 ま行五段活用(名)の登録例

ユーザの正誤入力：YES

YES: 連用形が名詞と扱える 判定できる。

NO: 連用形が名詞と扱えない と判定できる。

上下一段活用と判定した場合は「読む」をもとに上下一段活用連用形を作成し、「+ が、を + 動詞」をつなげ、「読 + が、を + 動詞」は正しいかどうかユーザに聞く。

システムの出力する例文：読がいい

ユーザの正誤入力：NO

YES: 連用形が名詞と扱えると判定できる。

NO: 連用形が名詞と扱えない と判定できる。

形容詞

よみに(しかくい)、候補に(四角い)、品詞に(形容詞)を入力したとする。ユーザは、形容詞を登録する時は よみ、候補を、終止形(〜い)で入力する。

システムの出力する例文：まるなもの

ユーザの正誤入力：YES

YES: 形容詞(ナ) と判定できる。

NO: 形容詞 と判定できる。

形容動詞

よみに(ゆうのうだ)、候補に(有能だ)、品詞に、形容動詞を入力したとする。形容動詞を登録する時は よみ、候補を、終止形(〜だ)で入力する。

システムの出力する例文：有能する場合

ユーザの正誤入力：NO

YES: 形容動詞(ス) と判定できる。

NO: 形容動詞 と判定できる。

副詞

よみに(きわめて)、候補に(きわめて)品詞に(副詞)を入力したとする。

システムの出力する例文：きわめてするとき

ユーザの正誤入力：NO

YES: 副詞(ス) と判定できる。

NO: 副詞 と判定できる。

なお、図3,4の登録例は、ユーザインタフェース構築環境鼎[3]を用いて実現したものである。

4 まとめ

単語登録において、例文を提示することにより、品詞の選択項目を少なくすることが出来た。また、日本語文法について熟知していないユーザでも容易に正しい品詞を指定することのできる方式を提案した。

参考文献

- [1] 田中亘、他「VJE-β 活用ハンドブック」技術評論社
- [2] 西尾実、他：「岩波国語辞典第四版」岩波書店
- [3] 暦本、菅井、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。