

カナ漢字変換における同音語選択 に関する一考察

2K-5

清水敏夫 西淳一郎 岡田都

松下電器産業(株)無線研究所

1. まえがき

近年、日本語ワードプロセッサ等の情報機器の急速な普及にしたがって、日本語入力手段としてのかな漢字変換技術の重要性が益々増大している。特に文節分かち書きを必要としないべた書きかな漢字変換の研究が盛んであるが、その問題点としてまず第一に、べた書き入力文からいかに文節を区切るかがある。これはこれまでに2分節最長一致法、文節数最小法が提案されており、実用化にもなっている。第2の問題点として同音異義語処理がある。この問題については現在、次の3通りの提案がなされている。

- (1) 学習辞書を用いた頻度情報
- (2) 複合語・派生語に関する単語固有の情報
- (3) 格文法を用いた意味的語彙情報

本稿では、上記のうち(3)について着目し、意味的情報を用いた同音語処理手法について述べる。

2. 格文法モデルの問題点

これまでに、国立国語研究所の分類語彙表を用いた格フレーム辞書を持ったかな漢字変換システムが提案されてきている。しかしながら分類語彙表は時代と共に意味が変化するという言語の特質を完全に表現できているとは言えない。また分類語彙表の情報のもれが無いとはいえない。しかし、独自に分類語彙表を作成し、格フレーム辞書を構築するのは余りに作業が煩雑であると言えよう。

本稿では、かな漢字変換システムに学習格フレーム辞書を設けることにより、同音語処理をより効率よく行うかな漢字変換システムを提案する。

3. 格文法モデルを用いたかな漢字変換システム

3.1 格フレーム辞書

同音語選択処理に於て、各用言がその固有の格構造を持つことから、体言・用言の意味的、用法上の関係を格構造から判断することにする。ここで問題となるのは、個々の体言、用言の意味体系をいかに正確に記述するかということであるが、

ここでは、比較的厳密に作られた国立国語研究所の分類語彙表を採用した。この分類語彙表を基本に各体言に対して意味体系を設定した。

用言(主に動詞)の格については、個々の用言により異なるため用言に共通性の高い格と個々の用言特有の格に分け、それぞれ必須格、任意格とした。また形容詞、形容動詞に関しては、格として主格のみを用いた。今回我々が用いた格を以下に示す。

主格、対象格、目標格、起点格、終点格、理由格、道具格、場所格、経路格、原因格、方法格、時格。

格フレーム辞書にこれらの格の項に、それに対応する意味体系を記述した。そのさいある用言に於て異なった格であり、かつ同じ意味体系を持つ場合には、後出する格の意味体系に関する情報は省略し、既出の意味体系へのポインターを与える。

格フレーム辞書の記述例を図1に示す。

3.2 学習格フレーム辞書

意味体系の規範として用いた分類語彙表は比較的厳密に作られているとはいえ、言語の性質上、言語の意味が変わって行くことには対応していない。また一度、格フレーム辞書を構築したときにも同様の問題が生じる。そこで、我々は学習格フ

読み：カウ

表記：買う

品詞コード：5段ワ行

必須格フレーム：(主格：人、国・機関・組織名)

(対象格：動物、植物、自然物、

人工物、原材料、

手段・設備、

生成物・生産物

)

任意格フレーム：(起点格：主格意味体系へのポインター

)

(場所格：場所)

(時格：時間)

図1 格フレーム辞書記述例

フレーム辞書と呼ぶ記憶エリアを用いて、格構造を基本とした同音語選択処理に失敗し、使用者による同音語選択が行われた場合、その選択された語（特に体言）の係受け関係を基に学習格フレーム辞書に選択された語の情報を記述することにする。ここで、使用者による同音語選択とは、格構造と関係なく同音語候補から使用者が所望の候補を選択することである。

3.3 動作例

図2に格構造を基本とした同音語選択による失敗例を示す。図2に示された解析変換結果では図1に示された「買う」の格フレーム辞書を用いて解析を行うので、使用者の望む「怒り」が出力されない。従って、使用者は同音語選択作業を行うことにより、所望の「怒り」を選択することになる。

このときに、学習格フレーム辞書更新処理が行われる。図3に図2を基にした学習格フレーム辞書の更新例を示す。

3.4 かな漢字変換システム概要

格フレーム辞書及び学習格フレーム辞書を用いたかな漢字変換システムの処理フローを図4に示す。同音語処理部で格フレーム辞書、学習格フレーム辞書を参照し、同音語選択を行う。使用者による同音語選択が行われると、その語の係受け関係により、学習格フレーム辞書の更新を行う。

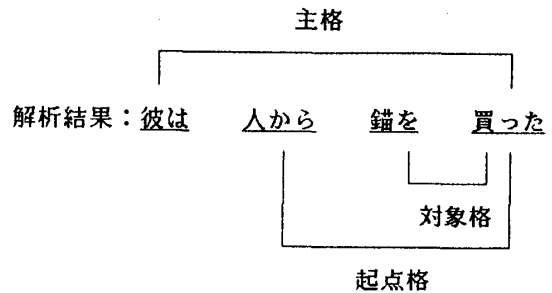
4. むすび

以上、学習格フレーム辞書を備えたかな漢字変換システムについて述べてきた。言語の特質である、意味の変化や新たな慣用句などに対しても本システムは有効に同音語選択を行えるものと考ええる。なお、本システムは現在試作中であり、今後その性能評価等が今後の課題である。

参考文献

- 1) 牧野、木澤：べた書き文の仮名漢字変換システムとその同音語処理、情報処理学会論文誌、Vol. 22, No. 1, pp. 59-67 (1981)。
- 2) 宮崎：係り受け解析を用いた複合語の自動分割、情報処理学会論文誌、Vol. 25, No. 6, pp. 970-979 (1984)。
- 3) 松田、高田、小谷：未知語の属性を修得する自然言語処理系、情報処理学会第31回全国大会論文集、pp. 1201-1202 (1985)。
- 4) 大島、阿部、湯浦、武市：格文法による仮名漢字変換の多義解消、情報処理学会論文誌、Vol. 27, No. 7, pp. 679-706 (1986)。

例文：カレハヒトカライカリヲカッタ



正答： 彼は 人から 怒りを 買った

図2 変換失敗例

読み：カウ

表記：買う

品詞コード：5段ワ行

学習格フレーム：(対象格：(1 読み：イカリ

表記：怒り

図3 学習格フレーム記述例

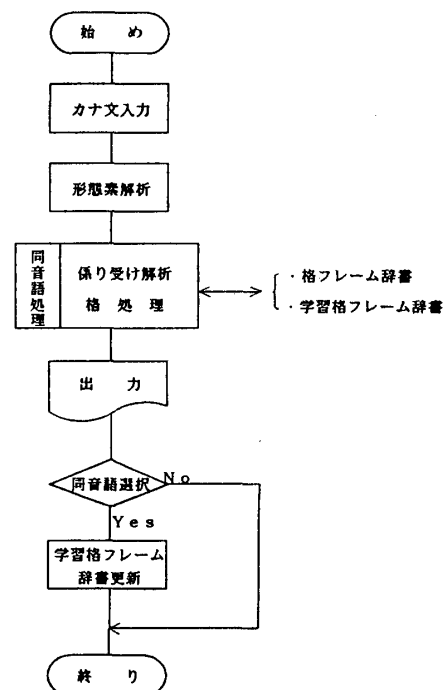


図4 カナ漢字変換処理フロー