

2K-4

単語間の共出現度数によるかな漢字一括変換

齊藤 哲 江澤 義典
(関西大学大学院) (関西大学)

1. はじめに

最近のかな漢字変換方式は入力ストローク数が減少するという利点から全文一括処理型の変換方式が主流となりつつある。一般に、全文一括型の処理においては文節単位への分解をした後で文節中の単語ごとに latest match 方式で変換を行っている。つまり、同音異義語の処理において文脈情報が全く無視されていた。そこで、同一文章における単語間の共出現度数を意味ネット構造で表現し、同音異義語の選択において文脈情報を活用するかな漢字一括変換方式を考察したのでその概要を報告する。

2. 共出現度数について

一般的にかな漢字混じり文を生成する段階で複数の同音語から適切な漢字列を決定するとき、当該文章中の他の単語(漢字列)との相互関係を手掛かりにしていると考えられる。

文章中の単語関係に着目することは限定された形で文脈情報を利用していることになる。ここでは、一つの文章中に同時に存在する自立語どうしの組の出現回数をカウントしたものを共出現度数と呼ぶ。例えば、「LISPがマイコンで使える」という文章の場合の部分意味ネットは図1となる。本システムの共出現度数では自立語の前後関係は考慮していない。また複合名詞は一つの自立語として扱う。

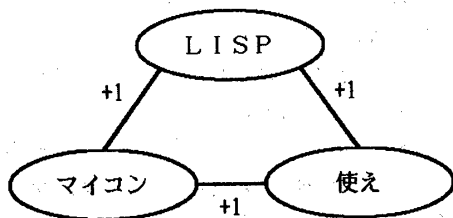


図1 単語間の部分意味ネット作成例

3. 処理の流れ

本システムの処理の概要を図2に示す。

かな文の入力において、本システムでは文節単位に区切り記号を挿入したかな文字列を入力することにより文節単位への分解は入力段階に手作業で行っている。次に自立語辞書と付属語辞書により各文節を可能な自立語または自立語と付属語の組み合わせ候補を作成する。ここで複数の候補が存在する場合は各候補別に文章中で注目している文節より前に確定された自立語からの共出現度数を合計し、最も高い度数を示したものを選定する。ここで選定できない場合(度数合計の最大値を示した候補が複数個あるかまたは文章の始めの文節である)は最終使用の単語を用いる。

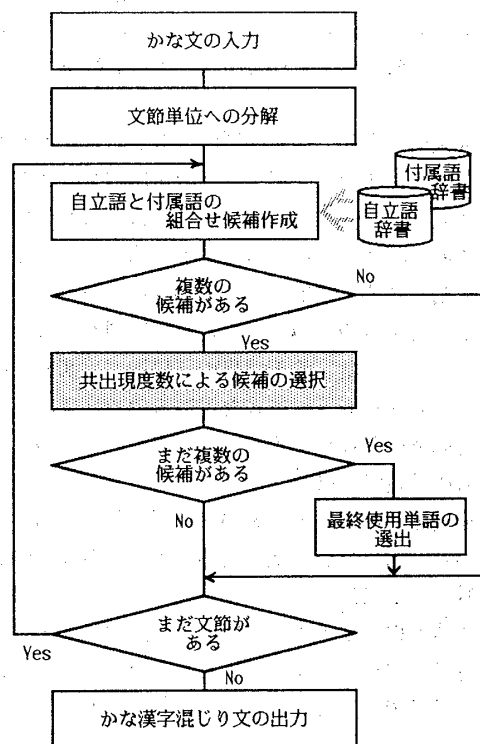


図2 本システムの処理の流れ

One stroke translation of Kana words into Kanji-Kana words using the frequencies of co-existence.

Akira SAITO Yoshinori EZAWA
Kansai University

4. 実行例

以下に本システムと従来の連文節一括変換型のシステムの比較例を図3に示す。

第1行：帰納による推論もできる。
第2行：プログラムにこの機能がある。

図3(a) 目的とする文章

キウコリスイロシキ。
→ 帰納による推論もできる。
プログラムニキウカアル。
→ プログラムにこの帰納がある。

図3(b) 連文節一括変換型による実行例

キウコリスイロシキ。
→ 帰納による推論もできる。
プログラムニキウカアル。
→ プログラムにこの機能がある。

図3(c) 本システムによる実行例

従来の連文節一括変換型のシステムによる変換例
図3(b)においては第1行目の「帰納」は手作業による選択を行っている。次に第2行目を入力したところ latest match 方式により第1行目で使った「帰納」が第一候補として出力されている。

本システムの変換方式による実行例を図3(c)に示す。ここでは、第1行目の「帰納」は第一文節のため latest match 方式を用いるので従来のシステムと同様の手作業をする必要がある。しかし、第2行目の「機能」は「プログラム」と「この」からの共出現度数を用いて自動的に選択されている。すなわち、第2行目の変換処理を実行する直前の部分意味ネット構造が図4のようになっていて、漢字列「機能」に対する共出現度数の合計値 $(a + c)$ が「帰納」に対する値 $(b + d)$ よりも大であるために、ここでは「機能」が選択されている。

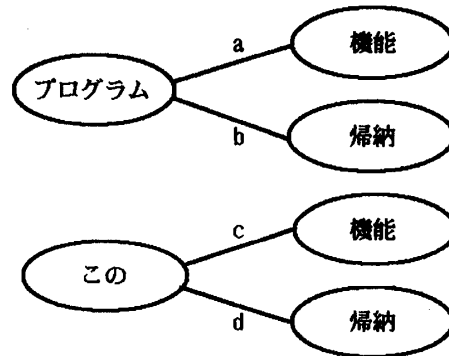


図4 「機能」と「帰納」に関する部分意味ネット

5. あとがき

本システムにおいて、複合名詞はそのまま自立語として登録したが単名詞に分解して登録し単名詞間の結合度合による複合名詞の生成が可能となるように拡張する予定である。

また、かな文入力において文章は一括して入力するが文節に分解する作業は手作業にたよっている。この点も自立語辞書と付属語辞書により組み合わせ候補を作成し、共出現度数を応用した文節処理が可能となるように拡張する予定である。

謝辞

色々と貴重なご意見を賜った関西大学の久保井不二男教授ならびに伊藤都男教授に深く感謝申し上げます。

参考文献

- 1) 本間・山階・小橋：「連語解析を用いたべた書きかな漢字変換」, 日本語文書処理研究会資料 21-2, pp.1-7, 1985年.
- 2) 溝口・北沢：「知識工学入門」, 講談社, pp.28~66, 1982年.
- 3) 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, 22-1, pp.59-67, 1981年.