

深層格を用いた仮名漢字変換

1 S-8

*後藤 浩文 *横田 浩 **中里 茂美 **大黒 和夫

*東芝ソフトウェアエンジニアリング **(株)東芝 情報通信システム技術研究所

1. はじめに

近年ワープロなどの仮名漢字変換機能は、全文まると変換などの採用により高速で使いやすいものに仕上がっている。

しかし、変換精度という点でみると、第1候補が希望する漢字でなかった場合複数回のキー操作が必要であり、必ずしも高速で使いやすいとはいえない面も残されている。

変換精度を上げるためには、形態素的な解析知識や漢字の出現頻度だけでは十分な効果が得られず、表層格を用いた共起仮名漢字変換でその一部を補っているが、知識情報を収集する労力やその情報量を考えると、すべてを網羅することは難しい。

本報告では、単語の意味や用法に着目した一実現方法として我々が試作を進めている、単語間の意味情報を知識として持ち同音語の絞り込みを行う、深層格を用いた仮名漢字変換について述べる。

2. 概要

単語間の意味情報とは、動詞と共起する単語を、格助詞の種類と名詞の意味的な分類(カテゴリ)で表したものである。同音語を持つ各動詞に記述された知識である(以下、集められた知識を単にAI共起辞書という)。

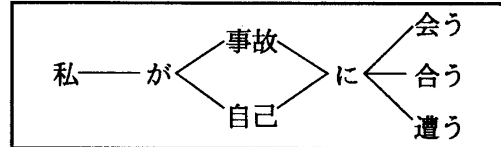
[あう]を例にあげると

(人)	が	(人)	に	会う
(趣味)	が	(人)	に	合う
(人)	が	(災難)	に	遭う

のような意味情報を知識として持っている。

ここで『わたしがじこにあう』という入力を例にあげて同音語の絞り込みを説明する。

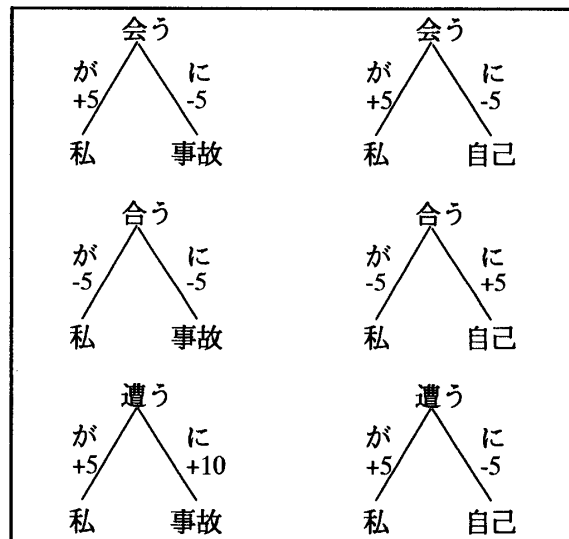
① まず同音語の組み合わせを行う。



② 動詞に着目する。

会う 合う 遭う

③ これに掛かる前方の名詞を探し、tree構造を作成しAI共起辞書の情報により点数付けを行う。



④ 文法的知識、出現頻度、意味情報の総合評価で第1候補の漢字を決定する。

- 1) 私が事故に遭う。
- 2) 私が事故に会う。
- 3) 私が自己に会う。
- 4) 私が自己に合う。
- 5) 私が自己に遭う。
- 6) 私が事故に合う。

The Kana to Kanji conversion using depths case pattern.

* Hirofumi GOTOU, * Hiroshi YOKOTA, ** Shigemi NAKASATO, ** Kazuo OHGURO

* TOSHIBA Software Engineering Corp. ** TOSHIBA Information and Communication System Laboratory.

3. 連体修飾・受動態

表層格を用いた共起変換で、連体修飾・受動態を実現する場合、それぞれ別の共起情報が必要となる。しかし、深層格を用いれば格解析(係り受け解析)を行っているため、複数の共起情報を持たなくてもそれぞれに対応することが可能である。

3.1 連体修飾

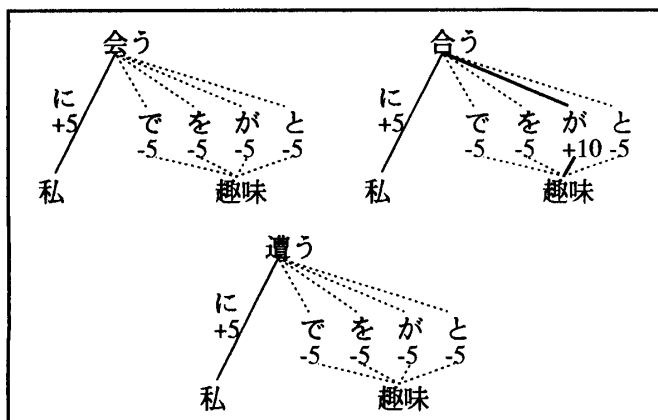
連体修飾への応用を『私に合う趣味を探している。』を例にあげて説明する。

動詞の活用を活用語尾から判断する。動詞の活用が連体形の場合、

① 動詞の後方の名詞を探す。

私に合う趣味を探している。

② 動詞と名詞の間に格を補って tree 構造を作成し、AI 共起辞書の情報により点数付けを行う。



ここで動詞と[趣味]の間に[に]格がないのは、[二重格の否定]のルールがあるためである。

③ 文法的知識、出現頻度、意味情報の総合評価で第1候補の漢字を決定する。

- 1) 私に合う趣味を探している。
- 2) 私に会う趣味を探している。
- 3) 私に遭う趣味を探している。

3.2 受動態

受動態への応用を『犬が彼女に飼われている。』を例にあげて説明する。

動詞の後ろに受身の助動詞(れる)がある場合、

① 動詞の前方の名詞を探し、tree 構造を作成する。

② 受動態と判断された動詞に付く格をルールに従い変換する。

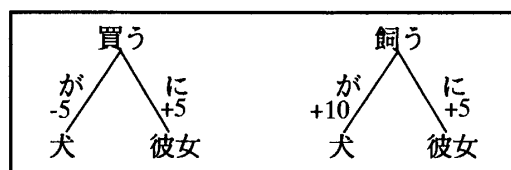
通常の情報

(人)	飼	(物)	養	買う
(人)	飼	(動物)	養	飼う

受動態用に格変換した情報

(人)	飼	(物)	養	買う
(人)	飼	(動物)	養	飼う

③ AI 共起辞書の情報により点数付けを行う。

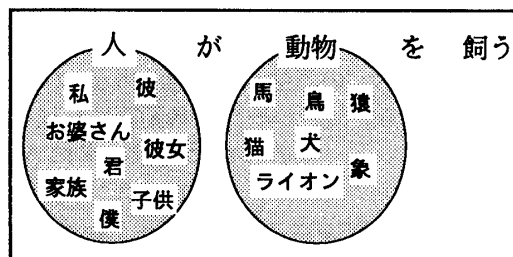


④ 文法的知識、出現頻度、意味情報の総合評価で第1候補の漢字を決定する。

- 1) 彼女に犬が飼われる。
- 2) 彼女に犬が買われる。

4. 効果

以上述べてきたように、一つの共起情報で複数の共起関係を抽出することができる。



深層の共起情報を1つ記述するだけで、表層の共起を数十個記述したのと同じくらいの効果が得られる。また、連体修飾・受動態への応用を考えるとかなりの効果が期待できる。

5. おわりに

単語間の意味情報を知識として持つ、深層格を用いたかな漢字変換について述べた。現在、この方式を組み込んだかな漢字変換システムで評価を行っている。

このシステムをより充実したものとするために、辞書の整備・コンパクション、文節切りの改良、処理の高速化、学習機能などに力を入れていく予定である。