

二字熟語意味推定のためのカテゴリー対応規則

2B-7

矢野 啓二郎 瀧口 伸雄 小谷 善行
(東京農工大学 工学部 電子情報工学科)

1. はじめに

自然言語処理システムにおける未知語処理の研究が多数なされているが、その多くは形態素解析レベルでの品詞の推定を行うものである。しかし、意味解析を行う上では単語の意味属性を推定する必要がある。

我々は漢字二字熟語に限定した未知語の処理の研究を行っている。その方法は、形態素解析中の二字熟語の未知語を意味解析処理のための意味属性を推定し、その結果をフレーム構造によって表現するものである。

これにしたいが、[1]では、漢字に対して品詞属性を与えておき、二字熟語中の漢字の品詞属性の組み合わせでその熟語の品詞や性質を推定する規則を設計した。

本稿では、さらに次のことを行う。

- ① 体系づけた意味分類によって漢字1字の意味を分類する
- ② 二字熟語において漢字の品詞の他に①の意味分類を加えた規則を作成する

ここで②の規則をカテゴリー対応規則と呼ぶ。これはまた熟語の品詞だけでなく、熟語の意味属性および、二字の意味関係を与える。

2. 熟語推定のモデル

本稿では熟語を構成する漢字1字1字の意味属性から熟語の意味属性を推定できる二字熟語を対象とする。例えば「文法」とは「文の法則」と考えても間違いではない。このような仮説から以下のようなモデルを提案した。ただし、省略語や固有名詞などは扱わない。

- ・漢字は1字で、いくつかの品詞、意味、格支配構造などの属性を持っている。本稿で使用する品詞を意味的品詞の名付けた。
 - ・二字熟語の属性は、それらを構成する漢字1字ずつの属性によって推定することができる。
 - ・二字熟語はどちらかの漢字の属性を持ち、もう一つの漢字によってその属性を修飾すると仮定する。
- 例) 「美人」は「美しい人」である。
- ・上記の仮定により熟語の属性を推定する規則を作ることができる。これがカテゴリー対応規則である。
 - ・上記の仮定により意味が転じて使用されているような熟語の属性は、推定できない。
 - ・カテゴリー対応規則は、意味的品詞・意味分類・格支配構造によって作ることができる。

3. 漢字の属性

漢字はもともと中国から伝えられたものであり、熟語もそこに起源がある。漢語においては漢字1文字1文字で品詞と意味をもっている。そこで、漢語をもとにした属性を定義した。

表1に意味的品詞について示した。ここに載せた以外の品詞は扱わない。

意味属性は、名詞、動詞、形容詞の語の意味に対して、分類を施して体系づけた。この分類を意味分類と呼ぶ。これらの分類を行うことにより、漢字のもつ意味を分類の集合によって表

現した。名詞の意味分類は、[2]によった。動詞、形容詞の意味分類はIPAL[3][4]のものを使用した。格支配構造もIPALのものを利用した。

表1 意味的品詞

| 品詞名 | |
|-----|-----|
| 名詞 | 代名詞 |
| 動詞 | 助数詞 |
| 形容詞 | 数詞 |
| 疑問詞 | 副詞 |
| 指示詞 | 数量詞 |

名詞の意味分類は、まず「自然」「人事」「文化」の3つに分類し(意味階層1と呼ぶ)、その中をそれぞれ3, 4, 3で10の分類分けをしている(意味階層2と呼ぶ)。その10の分類の中をそれぞれ10の分類で分けられている(意味階層3と呼ぶ)。したがって10×10の100の項目で名詞は分類分けされている。本稿では名詞のみを扱う。

4. カテゴリー対応規則

[1]では意味的品詞による組み合わせで、熟語の意味属性を推定している。ここでは、意味的品詞による組み合わせを基本とし、更に漢字の持つ意味分類による組み合わせと格支配情報を用いて熟語の意味属性を推定する。

カテゴリー対応規則を以下のように定義した。

$$\begin{aligned}
 H_3(Z_3, r) &\rightarrow H_1(J Z_1), H_2(J Z_2) \\
 H_3(Z_3, r) &\rightarrow H_1(Z_1), H_2(J Z_2) \\
 H_3(Z_3, r) &\rightarrow H_1(J Z_1), H_2(Z_2) \\
 H_3(Z_3, r) &\rightarrow H_1(Z_1), H_2(Z_2) \\
 H_1, H_2, H_3 &\in H : \text{意味的品詞の集合} \\
 J Z_1, J Z_2, J Z_3 &\in J Z : \text{意味階層2の集合} \\
 Z_1, Z_2, Z_3 &\in Z : \text{意味階層3の集合} \\
 Z &\in J Z \\
 r \in R & : \text{2漢字の意味関係の関数の集合} \\
 \text{一字漢字辞書}
 \end{aligned}$$

$$H(J Z) \rightarrow H(Z) \rightarrow K : \text{一字漢字}$$

数字の添え字で1, 2はそれぞれ熟語を構成する前の漢字と後ろの漢字。3は熟語を示している。

ここで、熟語の品詞および意味属性は、その二つの漢字の内の一つから受け継ぐ原則がある。すなわち、

$$H_3 = H_1 \text{ かつ } Z_3 = Z_1 (\leftarrow J Z_1)$$

または

$$H_3 = H_2 \text{ かつ } Z_3 = Z_2 (\leftarrow J Z_2)$$

であり、ここではこれを仮定している。

2漢字の意味関係は表2に示した。本稿では2漢字の意味的品詞の組み合わせを「名詞+名詞」に限定して述べる。

以下に二字熟語推定の例を示す。

- ・「駅長」
- カテゴリー対応規則
- 名詞(地位, 所属) → 名詞(上社会), 名詞(上人物)
- 一字漢字辞書
- 名詞(上人物) → 名詞(地位) → 長
- 名詞(上社会) → 名詞(施設) → 駅
- 以上から「駅長」は「名詞(地位, 所属)」に解析される。
- ここで「所属」はそこにつながる二つの意味構造を結合する。

表2 名詞に与えられる2漢字の意味関係

| 項目名 | 意味 |
|-----|-----------------|
| 用途 | 道具の使い道を限定する |
| 所有 | あるもの・概念の持ち主を示す |
| 時間 | 特定の時間を示す |
| 所属 | 特定の機関に限定する |
| 様子 | あるもの・概念の様子を示す |
| 動作 | あるもの・概念の動作を示す |
| 領域 | あるもの・概念の範囲を限定する |
| 程度 | あるもの・概念の程度を示す |
| 対象 | 特定の対象を限定する |

・「外国」

カテゴリ対応規則

名詞(集団, 領域) → 名詞(上性状), 名詞(上社会)

一字漢字辞書

名詞(上性状) → 名詞(位置) → 外

名詞(上社会) → 名詞(集団) → 国

以上から「外国」は「名詞(集団, 領域)」に解析される。

5. カテゴリ対応規則獲得実験

新聞から抽出した110語の熟語の意味分類の組み合わせについて筆者自身が調査した。これにより得られたカテゴリ対応規則を表3に示す。調査は、[2]をもとに行った。

「上人物」など「上」と付くものは意味階層2を表し、付かないものは意味階層3を表す。

6. カテゴリ対応規則と人間の推定結果との差

本学の院生2人に、[7]から抽出した74語の二字熟語の意味属性を推定する実験を行った。学生には意味分類での入力してもらい、辞書はいっさい引かずに行ってもらった。その結果とカテゴリ対応規則によって推定した結果とを比較した。

表4・実験結果

| 調査項目 | | 結果 |
|--------|-----------|-----|
| 完全一致 | 被験者A | 33% |
| | 被験者B | 33% |
| 品詞規則基準 | 被験者A | 92% |
| | 被験者B | 90% |
| 一致 | 人間基準 被験者A | 52% |
| | 被験者B | 43% |

結果を表4に示した。表4の「規則基準」とは、カテゴリ対応規則が選んだ品詞の組み合わせに対して人間が同じ品詞の組み合わせを選んだ割合を示している。

品詞の組み合わせだけに注目すると実験対象には「名詞+名詞」という組み合わせ以外の熟語も含まれていたが、「名詞+名詞」としたのは、カテゴリ対応規則は熟語の58%、被験者Aは33%、被験者Bは28%であった。

また、カテゴリ対応規則では4語が推定不可、被験者Aが2語、被験者Bが9語推定不可であった。

推定結果のくいちがいの原因としては、

- ①人間が意味分類を理解しにくい
- ②規則が不足している
- ③推定に個人差がある

の三つが考えられる。②は調査漏れから生じた問題である。①については表記の問題であり、計算機で扱う上での弊害であると考えられる。③は人間の推定結果が必ずしも正しいものであるとは限らず、被験者の感じ方によって変わってくるものであると考えられる。実際に筆者の考えでは正しいと思われる結果と被験者の考えが異なることが多くあった。

7. おわりに

熟語の意味属性を推定する方法としてカテゴリ対応規則を

提案した。

一般に漢字1字で複数の意味属性を持つため、熟語の意味属性を多数推定することになる。現在、本稿で紹介した以外の規則と、多義性を解消する方法を調査中である。

8. 謝辞

実験に協力して頂いた、本学の大学院生の久保田聡氏、島田志保氏に心から感謝する。

9. 参考文献

- [1] 瀧口ほか：フレーム構造による漢字熟語の表現，電子情報通信学会技術研究報告NLC92-37(1992)
- [2] 大野ほか：類語国語辞典，角川書店(1990)
- [3] 情報処理振興事業協会技術センター：計算機用日本語基本動詞辞書IPAL(Basic Verbs)解説編，情報処理進行事業協会(1988)
- [4] 情報処理振興事業協会技術センター：計算機用日本語基本形容詞辞書IPAL(Basic Adjectives)解説編，情報処理進行事業協会(1990)
- [5] 藤堂ほか：漢字源，学研(1992)
- [6] 吉村ほか：単一化による2漢字未登録語の処理，電子情報通信学会技術研究報告 NLC91-37(1991)
- [7] 一橋出版編集部：演習常用漢字，一橋出版株式会社(1989)

表3 意味分類によるカテゴリ対応規則

| 前の漢字の意味分類 | 後の漢字の意味分類 | 主要な語の位置 | 意味関係 |
|-----------|-----------|---------|------|
| 生理 | 上人物 | 後 | 様子 |
| 天文 | 上人物 | 後 | 所属 |
| 暦日 | 上人物 | 後 | 所属 |
| 気象 | 上人物 | 後 | 所属 |
| 地勢 | 上人物 | 後 | 所属 |
| 景観 | 上人物 | 後 | 所属 |
| 植物 | 上人物 | 後 | 所属 |
| 動物 | 上人物 | 後 | 所属 |
| 物質 | 上人物 | 後 | 所属 |
| 物象 | 上人物 | 後 | 所属 |
| 数量 | 位置 | 後 | 領域 |
| 地勢 | 役割 | 後 | 所属 |
| 時間 | 数量 | 後 | 領域 |
| 位置 | 実質 | 後 | 領域 |
| 天文 | 位置 | 前 | 領域 |
| 上自然 | 上自然 | 後 | 所属 |
| 上自然 | 上性状 | 前 | 領域 |
| 上自然 | 上学芸 | 後 | 所有 |
| 上性状 | 上自然 | 後 | 領域 |
| 上性状 | 上社会 | 後 | 領域 |
| 上性状 | 上人物 | 後 | 領域 |
| 上行動 | 上人物 | 後 | 所属 |
| 上心情 | 上性状 | 後 | 領域 |
| 上人物 | 物象 | 後 | 所有 |
| 上人物 | 位置 | 後 | 対象 |
| 上人物 | 物質 | 前 | 所属 |
| 上人物 | 生理 | 後 | 所有 |
| 上人物 | 上行動 | 後 | 所有 |
| 上人物 | 上性状 | 後 | 対象 |
| 上人物 | 上心情 | 後 | 所有 |
| 上人物 | 上人物 | 後 | 領域 |
| 上人物 | 上性向 | 後 | 所有 |
| 上人物 | 上社会 | 後 | 所有 |
| 上人物 | 上学芸 | 後 | 所有 |
| 上人物 | 上物品 | 前 | 用途 |
| 上社会 | 上性状 | 後 | 領域 |
| 上社会 | 上人物 | 後 | 所属 |
| 上社会 | 上社会 | 後 | 所属 |
| 上社会 | 上学芸 | 後 | 領域 |
| 上物品 | 上人物 | 後 | 所属 |
| 上物品 | 上自然 | 後 | 用途 |