

動的な重要度を用いた固有名詞同形語判別機構

4M-03

高橋 充彦

宮崎 正弘

新潟大学大学院自然科学研究科

1 はじめに

名詞を「一般名詞」「固有名詞」「数詞」の3つに大別すると、図1に示すようにこれらの名詞間またはそれぞれの名詞において「同形語」と呼ばれる、表記が同じで意味が異なる単語があり、特に、固有名詞間、一般名詞-固有名詞間に多くの同形語があることが分かる[1]。

重要度を用いて、一般名詞と固有名詞間の同形語判別、固有名詞間の同形語判別の方法を提案し、その有効性を示す。

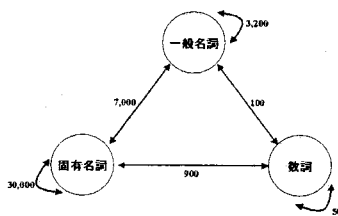


図 1: 各名詞間の同形語の数

2 重要度を動的に変動させる方法

固有名詞は地名、人名、組織名などその数は膨大であるが、出現頻度が高い比較的少数の語と出現頻度が低い多くの語がある。前者を重要度の高い固有名詞、後者を重要度の低い固有名詞とする。そしてこのような静的な重要度によって同形語を判別することが考えられる。しかし、重要度が低

い固有名詞であっても重要度が高い固有名詞と共起したり、特定の一般名詞と共起する場合、その重要度が高くなるものと考えられる。以下、重要度を動的に変動させることによって同形語を判別する方法について提案する。

2.1 重要度が高い固有名詞と共起してる場合

重要度が高い固有名詞が文章中にあったとき、それをキーワードとしてリンクされた固有名詞の重要度を上げて同形語を判別する。例えば図2に示すように、「新潟」は重要度が高い固有名詞で、これをキーワードとしてその地理的上下関係の下位語である重要度が低い固有名詞「内野」「寺尾」「古町」といった単語を動的に重要度が高い固有名詞に変動させる。「内野」は一般名詞「内野（ないや）」と同形語になっているが、重要度が高い固有名詞の方を優先的に選択し、固有名詞「内野（うちの）」と判別する。

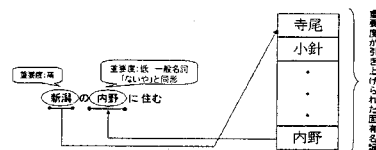


図 2: 同形語判別法 1

2.2 一般名詞と共起してる場合

重要度が低い固有名詞が特定の一般名詞と共起しやすい場合、特定の一般名詞が出現したらそれ

Disambiguation Mechanism of Japanese

Homographs using dynamic word-cost.

Mitsuhiko Takahashi, Masahiro Miyazaki

Niigata University

をキーワードとしてリンクされた固有名詞を動的に重要度が高い固有名詞に変動させる。例えば図4のように「敬称」として使用される一般名詞「さん」「様」「君」といった一般名詞（接尾語）と重要度が低い固有名詞「名」・「姓」が共起しやすいため、それらの間に共起関係のリンクをはっておく。そのようにすると「平野」という単語が出

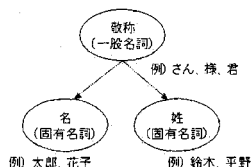


図3: 敬称と姓・名の共起関係

現し、単独では一般名詞「へいや」になってしまうが、「平野さん」のように「姓（固有名詞）」+「敬称（一般名詞）」のような複合名詞で構成されれば、敬称「さん」をキーワードとして共起されている姓の固有名詞の重要度が動的に高く変動するため、「平野（ひらの）」が導き出される。

3 本手法の問題点

一定の条件を満たした状況において、本稿で提案した方法を適用すると失敗してしまう場合がある。

3.1 名詞句のとき

・「新潟 の 球場 の 内野」
という名詞句において、2.1の同形語判別法を使うと「新潟」をキーワードとしてその下位語である「内野」「寺尾」「古町」等を重要度が高い固有名詞に動的に変動させてしまう。そして「内野」を固有名詞「うちの」と判定する。しかし、本例文では「内野」は一般名詞「ないや」である。つまり重要度が低い固有名詞を動的に変動させてしまったことにより、一般名詞が正解であったが強制的に変動させた固有名詞を選んでしまうという結果になる。このような場合には名詞句の構造を考慮して選択すればよい。

3.2 重要度を動的に変動できないとき

次に重要度が低い固有名詞が単独で出現してしまったために動的に変動させることができずに失敗してしまった例をあげてみる。図4の例文で姓の固有名詞「平野（ひらの）」が単独で出ている。これは一般名詞「平野（へいや）」と同形語である。重要度が低い固有名詞と一般名詞では一般名詞の方が優先的に選択されてしまうため、図4のように一般名詞「平野（へいや）」と判定ミスを起こしてしまう。このような場合には格パターンチェックの利用により解決できる。

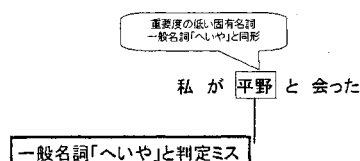


図4: 重要度を動的に変動できないときの失敗例

4 おわりに

固有名詞が特定の固有名詞や一般名詞と共起しやすいことに着目して固有名詞の重要度を動的に変動させることによって同形語を判別する機構を提案しそれを用いた。また本手法の問題点を示した。

今後、3の問題点の解決策を図ると共に、重要度の高い固有名詞が出現すると、現段階ではそれが最優先に判断されてしまうため、重要度の高い固有名詞の重要度を動的に下げる場合の検討も必要である。

参考文献

- [1] 宮崎, 大山: 階層的単語属性を用いた同形語の自動読み分け法
電子通信学会論文誌, Vol.J68-D, No.3,
pp.392-399 (1985-3)
- [2] 池原, 宮崎, 白井, 横尾, 中岩, 小倉, 大山, 林:
日本語語彙大系 (全5巻)
岩波書店 (1997)