

経験的優先規則による古典文の係り受け解析

2C-5

齋藤 大輔[†] 石隈 博史^{*} 小川 泰幸[†] 江畑 秀規[†] 上原 徹三[†] 石川 知雄[†]武蔵工業大学大学院工学研究科電気工学専攻[†] 日立ソフトウェアエンジニアリング(株)^{*}

1 はじめに

本研究では、平安時代の作品である伊勢物語を処理の対象として、意味解析に入りこまない範囲で、経験的優先規則による古典文の係り受け解析を行う。

2 本研究での係り受け解析

本研究での係り受け解析は、前もって伊勢物語の解釈文 [1] を形態素解析したものを入力情報として与えて、代表品詞と下接特性の設定処理、係り受けの決定処理、優先度の設定処理の3つの処理を行うことから成り立っている。

2.1 入力情報

入力情報は、まず伊勢物語の解釈文の文を文節ごとに区切り、文節ごとに代表品詞と下接特性を設定するために、文献 [2] を参考に文節の単語ごとにその単語の読みと表記と品詞と活用形を示したものである。入力情報で用いている品詞の種類は学校文法で用いられている品詞を中心としたものである。但し、名詞、副詞、助詞については複数に分類している。

2.2 代表品詞と下接特性の設定処理

ここでは、与えられた入力情報を基にして、文節ごとにどのような受け特性を持つかを表す代表品詞とどのような受け特性を持つ文節に係り受け可能かを表す下接特性を設定する [3]。

2.2.1 代表品詞の設定

ある文節の代表品詞の設定は、自立語の品詞により設定する。代表品詞の種類は、それぞれ、「体」「用」「副」「連体」「無し」の5種類である。

2.2.2 下接特性の設定

ある文節の下接特性の設定は、その文節の最後の単語の品詞と活用形により設定する。単語の品詞によっては複数の下接特性をもつ場合もある。下接特性の種類は「体」「用」「終」の3種類である。

2.3 係り受けの決定処理

ここでは、文節ごとに定めた代表品詞と下接特性を用いて係り受けの決定を行う。すなわち、下接特性と代表品詞が一致した文節同士は係り受けが可能と判断される。しかし、それだけでは候補が膨大なものとなるため、係り受けの交差はない、文節の係り先は一つ、後ろの文節が前の文節に係ることはないという3つを絶対条件として、以下のような規則により係り受けの関係に制約を加えた。

- a: 下接特性が「終」の文節より前にある文節は、下接特性が「終」である文節を越えては係らない。
- b: ある文節に関して、その文節の直後の文節に係り受け可能であるならば、一部の文節を除いて他に係り先のない文節とする(直結文節と呼ぶことにする)。

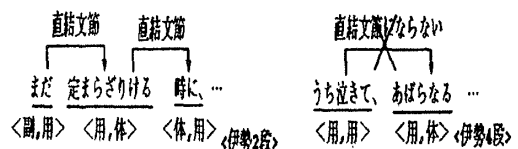
Dependency Analysis of Japanese Classics Sentence based on the Heuristic Priority Rules

[†]D.Saitou, Y.Ogawa, H.Ebata, T.Uehara, T.Ishikawa

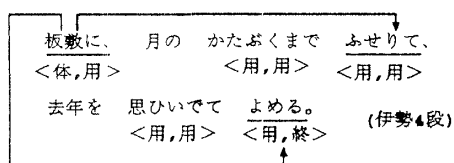
^{*}H.Ishiguma

[†]Electrical Engineering, Graduate School of Research Division in Engineering, Musashi Institute of Technology

^{*}Hitachi Software Engineering, Ltd



- c: 「～ば、～と」などの一部の接続助詞が文節末の場合には、それより前にある文節はそれより後の文節には係らない。
- d: 読点がある文節はそこで一端意味がまとまり、その文節の係り先は句読点がある文節ということが多いため、一部の例を除いてその係り先は句読点がある文節に限ることにする。逆に、読点がない文節は一部の例外を除いて読点がある文節を越えて係ることはないものとする。



- e: 係り結びの関係の関係を満たす文節同士の場合は、それ以外の係り受けのパターンは存在しないものとする。

2.4 優先度の設定処理

係り受けの決定処理を終えた段階でも、文候補の数は複数存在する。それで、選出されたすべての文候補の間で、係り受けの妥当な文を判定するために、優先度を設定する。設定方法は、各文節の係り先すべてに対して、ある規則を設定し、それに従って評点を与えることによって行なう。そして、その合計点により各文候補の評価点が決定し、文候補の間で優先度が設定される。

3 解析結果

伊勢物語の1段から39段までの歌を除いた108文について解析を行った。解析の結果と予め作成した正解を比較した結果は次のようになった。但し、予め作成した正解では、人手によってもどの文節に係るかを限定できない場合にはどちらの文節に係っても正解としてある。正しい係り受けの文が選出されたのは95文で、88%となっており正解率は高いものとなった。そのうち、優先度の設定をして順位が1位となったものは、77文で72%となった。108文のうち13文では、解析に失敗している。また、解析に成功した95文のうち18文

では優先度の設定処理での順位が2位以下となっている。それらの理由と今後の方針を次で述べる。

4 問題点と今後の方針

係り受けの決定に失敗した理由としては以下のようなことが上げられる。

1. 規則Cが成り立たない場合がある。
2. 正解の係り受け関係の中に下接特性と代表品詞が一致しない例があり、例外の規則でも正解の文節に係ることができず解析失敗になる。
3. 規則dに反する係り受けが正解となる場合がある。

優先度の設定処理で正解の順位が悪くなってしまいう理由としては以下のようなことが上げられる。

1. 読点のある文節での優先度の設定では、直後の読点のある文節か、句点の文節に係る場合に点を加算したが、そうすると正解の順位が悪くなってしまいう場合がある。
2. 読点が無い文節では単純に遠くの文節に係るよりも近くの文節に係る場合の点を高くしたが、そうすると正解の順位が悪くなってしまいう場合がある。

今後の方針としては、先に述べた問題点の解決策の検討を随時進めていく。また、文候補の数の増大の理由の1つとして代表品詞と下接特性の種類の少なさを考えているので、今後種類を増やしていく予定である。なお、本研究は一部、平成7年度科学研究費補助金（一般研究C）「日本語文語文のコンピュータによる処理」によって進めたものである。

参考文献

- [1] 木村雅則: 伊勢物語翻刻・解釈ファイル, 勉誠データセンター, 1993
- [2] 三ツ木 徳彦: 伊勢物語精訳, 中道館, 1994
- [3] 兵藤 安昭 池田 尚志: 表層的情報とN近傍ブロック化手法による日本語長文の骨格構造解析, 情報処理学会論文誌, Vol36, No9, pp2091-2101 (1995).