

構文規則の変更可能な古文係り受け解析機能の試作

田中 成美* 上原 徹三** 石川 知雄**

*武蔵工業大学大学院工学研究科 **武蔵工業大学工学部

1 はじめに

我々は、日本語の古文の文法処理を検討している。しかし、現在まで公開されている日本語の文法解析ツールには文法規則の変更機能が用意されていない（英語の解析ツールにはこの機能を持つものがある [1]）。

そこで、古文の言語処理の環境を充実させるために、解析法や文法規則を変更できる構文解析プログラムを試作した。また、この解析プログラムを用いて、平安時代の仮名書き古文を対象とした解析実験を行った。本稿では、解析プログラムの機能とその適用性を検討する。

2 解析機能

本解析プログラムの特徴を以下に挙げる。

- 変更の多い構文規則をデータとして規則ファイルに記述し、解析プログラムに与える。
- 解析で使用可能な構文規則の種類を規則ファイル名指定ファイルに記述し、解析に用いる構文規則の種類やそれらを解析に適用する順番を設定する。
- 解析対象文の品詞体系は、解析プログラムに依存していない。

これらの特徴を持つ構文解析プログラムの構成を図1に示す。この図の右上に規則ファイル名指定ファイルがある。解析プログラムの外側にあるそれ以外のファイルは規則ファイルである。このファイルに構文規則を記述して解析プログラムに与えると、ファイルの構文規則を用いて解析が行なわれる。この機能には、解析対象の文の2文節間の係り受け可能性を求める処理とその処理の後に係り受け可能性のある2文節間や文候補に優先度を付与する処理がある。前者で用いる構文規則を制約規則、後者で用いる構文規則を優先規則とする。また、解析では、「係り先は一つ」「係り受けの交差はしない」「文節はその文節より後に現れる文節に係る」の3つを絶対規則とし、他の規則よりも優先する。ここで、文候補とは、絶対規則を守りつつ制約規則により求めた係り受け可能性のある2つの文節を組み合わせた文全体の係り受けパターンのことである。

解析プログラムの入力情報は、平安時代の仮名書きの古文の例文集とする。これには、単語分けされた文や単語情報（表記、読み、品詞、活用

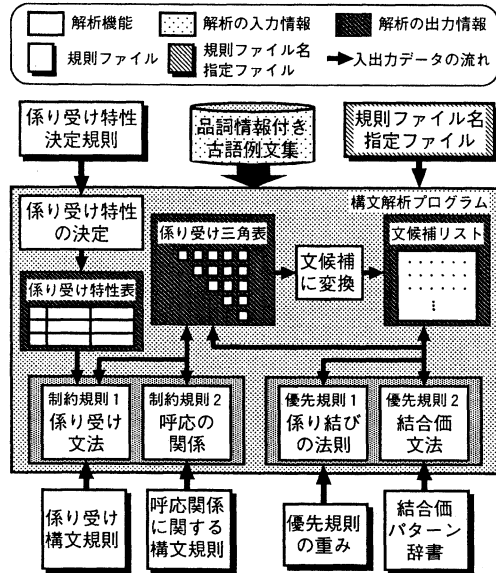


図1: 構文解析機能の構成

形), 係り受け情報(文節間の係り受け関係)等が記載されている。出力情報は、各文節に対する係り受け特性の設定結果, 係り受け三角表, 優先規則によって付与した優先度の合計により順位付けした文候補とする。また、解析で使用した規則の種類と番号を別のファイルに出力する。ここで、係り受け三角表とは、解析対象の文の2文節間の係り受け関係の可能性を記録した表のことである。

3 使用規則およびそれらの適用順の指定法

解析で使用する構文規則やそれらの適用順は実行時に指定する。この機能により、解析プログラムを変更せずに使用可能な複数の構文規則を組み立てることができ、組み立てた複数通りの中で目的に最適な解析法を実際に実験して求めることができる。

使用する構文規則やそれらの適用順は、使用可能な構文規則の名前、その規則に対応するコード名、その規則を記述したファイルの名前を予め規則ファイル名指定ファイルに記述し、この中から使用する規則のコード名とそれらの適用順を対話形式で指定するという方法で行う。

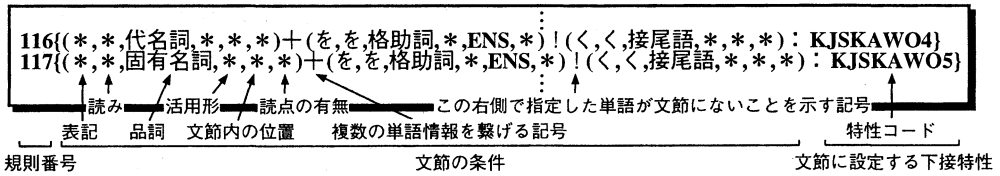


図 2: 下接特性の決定規則ファイル

4 係り受け可能性の制約規則

制約規則は、「係り受け文法に関する構文規則」、「呼応関係に関する構文規則」とする。以下では、構文規則の指定法と解析への適用法を説明する。

4.1 係り受け構文規則

係り受け文法を用いて係り受け可能性を求めるときには、文節ごとに与えられる係り受け特性を用いる。係り受け特性は、当該文節が他の文節に係る特性（下接特性）と他の文節を受ける特性（代表品詞）から成る。係り受け可能性は、まず、文節の係り受け特性を求めて、係り側の下接特性と受け側の代表品詞の係り受け可能な組合せを規定した構文規則により求める。

本解析プログラムでは、係り受け特性の決定規則と係り受け文法に関する構文規則を外部から読み込む方法で係り受け関係を求める。以下で、両規則の指定法と解析への適用法を記す。

4.1.1 係り受け特性の決定規則

係り受け特性の決定規則の指定は、文節の条件と係り受け特性のコード名を図 2 に示す形式でファイルに記述するという方法で行う。これにより、品詞や活用形の体系を変更した際に、係り受け特性の決定規則を容易に書き換えることができる。ここで、図 2 では下接特性の決定規則ファイルを示しているが、代表品詞の決定規則ファイルもこれと同じ形式で作成する。

以上の方法を解析に適用して文節「富士の山を」の係り受け特性を求める例を図 3 に示す。この文節は、固有名詞「富士の山」に格助詞「を」が付属しているため、下接特性の決定規則ファイルの規則により下接特性のコードは KJSKAW05 となる。また、この文節の自立語の品詞が固有名詞であるため、代表品詞の決定規則ファイルの規則により代表品詞のコードは DTAKO となる。

4.1.2 係り受け文法に関する構文規則

係り受け文法に関する構文規則は、係り側の下接特性と受け側の代表品詞のペアを 1 つの規則として、係り受け構文規則ファイルに指定する。

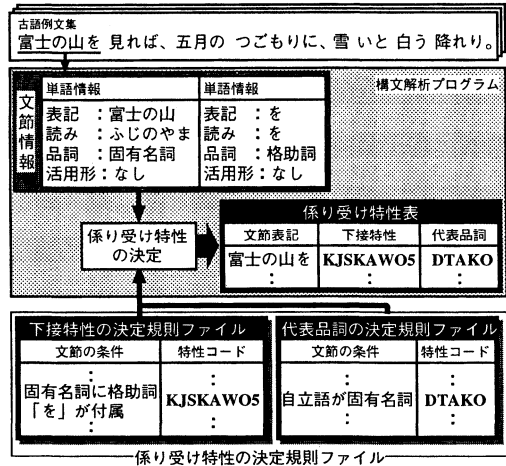


図 3: 係り受け特性決定規則の適用例

これにより、係り受け特性の分類の変更に応じてそれに従う構文規則を作成する作業を容易にする。

以上の方法を解析に適用して 2 文節間の係り受け可能性を求める例を図 4 に示す。係り受け

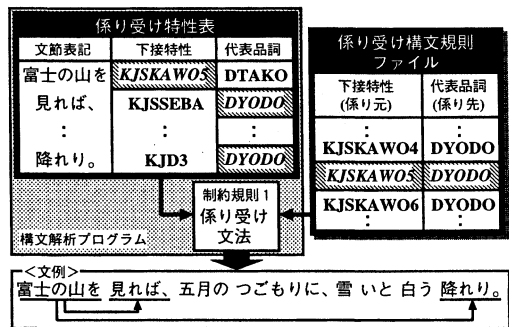


図 4: 係り受け文法に関する構文規則の適用例

特性表から、文節「富士の山を」の下接特性は KJSKAW05 であることがわかり、構文規則ファイルにある規則によりこの文節は代表品詞 DOYODO の文節に係る可能性があることがわかる。

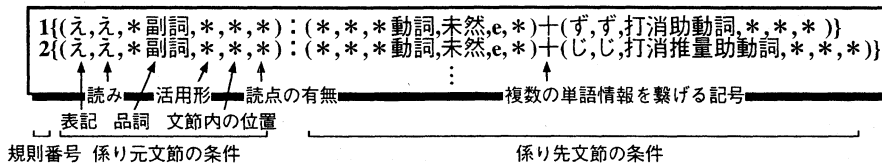


図 5: 呼応関係に関する構文規則の規則ファイル

この代表品詞に設定された文節は「見れば、」と「降り。」であるので、「富士の山を」はこの2つの文節に係る可能性があると判断される。

4.2 呼応関係に関する構文規則

呼応関係とは、例えば現代文での「全く～ない」のように「全く」という単語が文に現れた場合は、必ずその後に「ない」のような否定を表す単語があり、これらが係り受け関係を持つというものである。呼応関係は古文においても多数あり、作品の文体によって特徴がある。このことから、規則の変更が多いと考え、その作業を容易にするために呼応関係に関する構文規則を外部から与えることとした。

規則の指定は、係り側と受け側の文節の条件を図5に示す形式でファイルに記述するという方法で行う。例えば、図5の規則番号1のように、表記と読みが共に「え」で品詞が副詞の単語は、表記と読みが共に「ず」で品詞が打消助動詞の単語と呼応の関係にあることを示すことができる。その規則を指定すると、解析対象の文にこの関係を持つ2つの文節がある場合、この文節間には係り受けの可能性があると判断される。

5 優先規則による優先度付け

優先規則は、「係り結びの法則に関する構文規則」、「結合価文法に関する構文規則」とする。解析では、優先規則によって係り受け可能性のある2文節間や文候補に優先度を付与し、その合計で文候補の優先順位を決めるが、正解候補を全文候補の1位あるいは上位にするために、優先規則の優先度の調整が多いと考えられる。

このような理由から、優先規則の優先度を解析プログラムの外部から与え、優先度の調整作業を容易にする。規則ファイルには、優先度付けの種類(どのような条件で優先度を与えるか)を表すコード名と優先度の値を記述する。以下では、各々の優先規則を用いて係り受け可能性のある2文節間や文候補に優先度を付与する方法を説明する。

5.1 係り結びの法則に関する構文規則

係り結びの法則とは、古文に特有な文法で、節や文の終りの形が終止形、連体形、已然形のいずれかであることが助詞により決まっており、この助詞と節または文の終りの形の呼応の法則のことである。しかし、この法則は成立しない場合もあるので、優先規則として解析に適用する。

5.2 結合価文法に関する構文規則

結合価文法は、「動詞に係る名詞文節は動詞ごとに異なり、動詞はそれを反映した固有の文構造を持つ」という特徴に着目した文法である。例えば、「紹介する」は「～ガ・～ニ・～ヲ」の3項と結合し得る結合価3の動詞であり、「紹介する」が出現する文は、この動詞を中心とした動詞に固有な文構造を持つと捉える。ここで、この例の「～ガ」などは名詞に格助詞が付属した名詞文節であり、これを格要素という。また、動詞と格要素の格助詞の組合せを結合価パターンという。

本研究では、動詞の結合価パターンを優先規則として解析に適用する。解析に適用する際に、優先度を与える条件を以下に挙げる。

- 動詞とそれに係る格要素の格助詞との組合せが結合価パターン辞書にある場合、動詞を含む文節と格助詞が付属する文節の間に優先度を与える。
- 1つの格要素が複数の動詞に係る可能性があり、格要素の格助詞とそれぞれの動詞の組合せが結合価パターン辞書にある場合、文節間に与える優先度に差をつける。例えば、文「格要素A 動詞1 動詞2」に「格要素と動詞との距離が遠い程、その2文節間の優先度を大きくする」という規則を適用する場合は、「格要素A-動詞1」間よりも「格要素A-動詞2」間の優先度の方を大きくする。
- 複数の格要素が同時に同じ動詞に係る文候補があり、その組合せがその動詞の結合価パターンとして辞書に登録されている場合、その文候補に優先度を加える。

6 解析機能の適用実験

試作した解析機能の適用性を検討するために、本稿で述べた構文規則を適用した解析プログラムを用いて実験を行なった。ここで、実験で用いた結合価パターン辞書は、古語例文集のレコード番号が奇数の文をデータとして作成した。

図6, 7に古語例文集のある文を解析した結果を示す。図6の解析結果は、制約規則である呼応関係に関する構文規則が解析に反映していることを示す解析例である。この図の文にある副詞

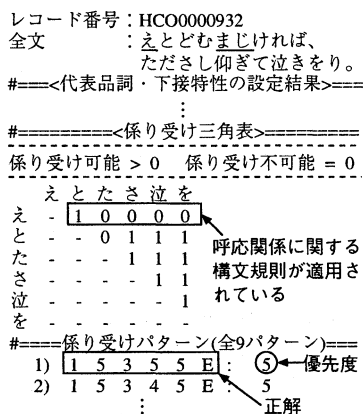


図 6: 解析結果 (出力ファイル)

「え」と打消推量助動詞「まじ」の間には、呼応の関係がある。この関係は、呼応関係に関する構文規則ファイルに記述した規則 $\{(え, え, *副詞, *, *, *) : (*, *, *動詞, 終止, e, *) + (まじ, まじ, 打消推量助動詞, *, *, *)\}$ を満たすので、2つの単語を含む文節の間に係り受け可能性があると判断されている。このことは、図6の係り受け三角表の1行目(係り元文節が「え」の行)が文節「とどむまじければ、」の列と交わるところのみ1となっていることからわかる。これによって、係り先候補を絞り込むことができ、文候補を削減することができている。

図7の解析結果は、優先規則である係り結びの法則に関する構文規則と結合価文法に関する構文規則が解析に反映していることを示す解析例である。この図の文には、格要素「こと人に」と動詞文節「あはせむと」、「いひけるを、」があり、制約規則によってそれぞれ係り受け可能性があると判断されている。この格要素とそれぞれの動詞の組合せは、結合価パターン辞書にあるため、辞書のパターンとマッチして優先規則により優先度が加算されている。このことは、図7

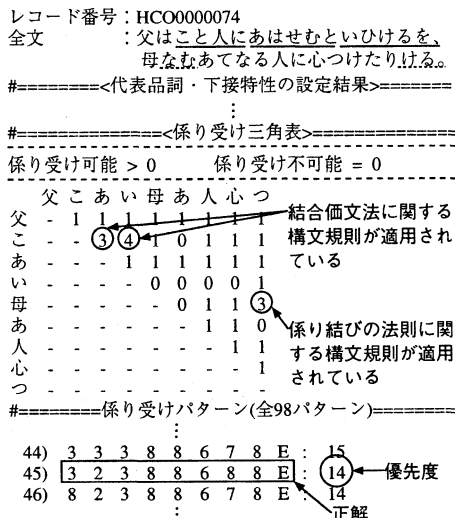


図 7: 解析結果 (出力ファイル)

の係り受け三角表の格要素の行とそれぞれの動詞文節の列が交わるところが1より大きい数値となっていることからわかる。また、この文には「母なむ」と「つけたりける。」があり、この間に係り結びの法則が適用されている。このことは、同図のこの2文節間にあたる係り受け三角表の優先度が1より大きい数値となっていることからわかる。

7 おわりに

本研究では、日本語の古文の研究環境を充実させるために、古文を対象として構文規則の適用順の指定や構文規則の変更が容易にできる構文解析機能を提案し、この機能を持つ解析プログラムを試作した。解析機能には、対象文の2文節間の係り受け可能性を求める機能と係り受け可能性のある2文節間や文候補に優先度を与える機能がある。これらの機能は、実験により解析に反映されていることが確認できた。

今後は、この解析機能を利用して実際に古文の解析を行ない、また、それを通して解析機能の向上を図る予定である。

なお、本研究の一部は文部省科学研究費補助金(基盤研究C2 No.11680422)によって実施したものである。

参考文献

- 1) 関根 聡. 英語構文解析システム「Apple Pie Parser」. 情報処理学会 学会誌 pp1221-1226 (2000.11).
- 2) 中村 幸弘, 碁石 雅利. 「古典文の構造」. 右文書店 (1996.4).