

## ”サ変名詞+読点”の品詞判別処理

6C-1

鎌田海平 湯村武 西田行輝  
三洋電機 情報通信システム研究所

## 1. はじめに

日本語解析において、複数の文章からなるテキストを処理する際には、その文章を関連する文章の中の1文として解析する必要がある。また、複数の事象が接続した長文解析においても、それぞれの事象を抽出し、事象間の関係を解析しなければならない。しかし、新聞記事など簡潔な表現が好まれるテキストには、”サ変名詞+読点”で表現される事象が多く、事象の抽出が困難な場合がある。すなわち、”サ変名詞+読点”出現パターンが単事象として分割できるのか、単文の構成要素であるのか判断する必要がある。我々は、この問題に対し構文情報、辞書情報を活用してこの曖昧さの解消を試みた。

## 2. 単文化処理

単文化処理は複数の事象からなる文をそれぞれの事象に分割する処理のことをいう。テキストに含まれる文章のうち、単文化処理によって複数の事象に分割されるのは、次のような場合である。

〈単事象に分割され得る構文パターン〉

- ・”句点”によって文章が終了。
- ・”接続助詞”によって他の事象と接続。
- ・”連用中止形”によって他の事象と接続。
- ・”体言(サ変名詞)止め”によって文章が終了。
- ・”サ変名詞+読点”パターンによって他の事象と接続。

本稿では、特に”サ変名詞+読点”パターンを対象とした処理について述べる。なお、処理対象として新聞記事1120文から”サ変名詞+読点”パターンを81事例抽出した。

## 3. 処理の概要

”サ変名詞+読点”の品詞判別規則として構文情報に基づいた21の規則を設定した。(名詞判別:9規則、動詞判別:12規則 第4図に例示)

処理は、1次判別処理と2次判別処理から成り立っており、1次判別処理では、近傍の構文的特徴を利用して品詞の絞り込みを行なう。2次判別処理は、1次判別処理で品詞を決定できなかった場合に起動され、動詞の格フレームを参照することによって判別する規則と、1次判別処理で判明した品詞やその他の構文情報などによって品詞を判別するための規則(第4図動詞判別規則④)を設定した。これは、

《事象内に「が(は)」格と「を」格が存在する場合はその事象文内の”サ変”の品詞は動詞である。》

という判別規則であるが、この規則を適用する場合には単文内に存在する「が」格や「を」格が、”サ変”の格フレームと一致しているかどうかを判定しなければならない。

しかし単語の持つ多義性によっては、第1図に示すように「購入する」の動詞格フレームと、「が」格である「林」とが照合を失敗する場合があった。

「林が来て五千株を一株三千円で購入、購入資金は同社から融資をうけたと述べた。」

=== 1次判別処理 ===  
購入→動詞?名詞?

[が] 林  
格フレームとのマッチング(する/しない)

[を] 株  
格フレームとのマッチング(する/しない)

=== 2次判別処理 ===  
購入→動詞

第1図 動詞判別規則の適用例

## 4. 2次判別処理

第2図に見られるように、文章中に”サ変”が複数存在し、その中にどの動詞判別規則でも判別できなかった”サ変”が存在した場合、1次判別処理では品詞を決定せずに、結果を保留しておく。次に2次判別処理では、1次判別処理で判明した品詞やその他の構文情報などを利用することで、”サ変”の品詞を判別する。2次判別処理で処理されるこのような事例は、おもに動詞判別規則(第4図 動詞判別規則④)において、格フレームにマッチングする「が(は)」格、「を」格のいずれかしか事象内に存在しないために、品詞を判別できなかったことによるものである。なお、2次判別処理においては次のような判別規則を設定した。

《品詞判別すべき”サ変”に後続する事象が動詞(句)で結ばれている場合の”サ変”の品詞は動詞である。

(このときの動詞(句)には1次判別処理で、動詞と品詞判別された”サ変”も含まれる。)

第2図は2次判別規則が適用されたときの判別結果である。1次判別処理では格フレームにマッチングする格が「が」格しか存在しないために、「解析、」の品詞判別を2次判別処理まで保留している。ただし、このとき「出力、」に関しては、1次判別処理において、品詞は動詞に決定されている。2次判別処理では、この結果を利用し、2次判別規則を適用することで、「解析、」の品詞を動詞に決定する。

Category Determination in "Sa-Series Noun + comma" Pattern

Kaihei KAWATA, Takeshi YUMURA, Yukiteru NISHIDA  
Information & Communication System Research  
Center, SANYO Electric Co., Ltd.

「このシステムが解析、他のシステムが変換する結果を出力、印刷する。」

=== 1次判別処理 ===

解析→動詞?名詞?

[が] システム

格フレームとのマッチング(する/しない)

出力→動詞

[が] システム

格フレームとのマッチング(する/しない)

[を] 結果

格フレームとのマッチング(する/しない)

=== 2次判別処理 ===

解析→動詞

第2図 2次判別処理の適用例

## 5. 品詞判別結果

本稿で述べた“サ変名詞+読点”の品詞判別規則を、新聞記事から抽出した事例に適用した結果、品詞が正しく判別できたのは81事例中77事例(95.1%)であり、正しく判別できなかったのは4事例であった。この4事例のうち3事例は、動詞と判別されるべき“サ変”が、名詞判別規則によって誤判別されたものである(第3図)。これらは、サ変名詞が複合名詞の要素となっているものであり、これを正しく品詞判別するには、複合語を構成する要素間の概念的な関係を記述した言語外知識などが不可欠だが、この点については今後の研究課題であろう。

「JR東中野駅で電車追突、運転ミスか。」

=== 1次判別処理 ===

追突→名詞

第3図 品詞判別の失敗例

## 6. 考察

以上の実験結果から以下のことが明らかになった。

- ・品詞判別において、“サ変”の直前に付随する機能語(助詞、助詞的名詞)によって、品詞判別できるケースが名詞判別処理で11事例、動詞判別処理で15事例となり全事例に対する割合は32.1%(81事例中26事例)となり、もっとも多い。
- ・動詞判別処理においては、2次判別処理規則によって品詞判別が行なわれるケースがもっとも多く、1次判別処理では品詞を判別できないケースが多いことがわかる。

これは、新聞記事では、主格や目的格などが省略されるケースが多いことを証明していると同時に、本稿の2次判別処理が“サ変+読点”パターンの品詞判別に効果的に作用していることを示している。

## 7. おわりに

本稿では、辞書情報や構文情報から“サ変名詞+読点”パターンの品詞を、動詞と名詞に品詞判別する手法について述べた。

また本稿の単文化処理は、言語外知識を利用することがない点で、自然言語応用システムへの組み込みが容易であるという利点がある。今後は、単文化処理によって分割された事象ごとの情報を内部に記憶し、入力される文章内容の変化に伴ってその内容を変化させ、テキストを解析するための処理モデルの研究につなげていく予定である。

### [名詞判別規則]

- ① “サ変”ノードの直前のノードに連体修飾句タイプの格助詞が付随。  
「ズサンな領収検査を行った日航に対する指導、監督も不十分だった、とした。」
- ② “サ変”ノードの直前のノードの文法カテゴリーが名詞並列句。  
「システムが解析や変換、生成の処理を行なう。」
- ③ “サ変名詞+読点”パターンが文頭に出現。  
「消費税紛糾、マル優廃止反対の連絡会が開かれた。」

### [動詞判別規則]

- ① “サ変”ノードの直前が接続助詞的名詞。  
「堤さんが会話するため左側を向いたまま運転、中央分離帯の立ち木に衝突して横転した。」
- ② “サ変”ノードの直前が格助詞的名詞。  
「購入金額が高すぎるなどと指摘、その不自然さを追求した。」
- ③ “サ変”ノードの直前が副詞(副詞的名詞)。  
「29万人の職員を2万人削減し、組織構造のうち総支社を来春廃止、組織を簡素化する。」
- ④ “サ変”を含むスコープ内に「が(は)」格と「を」格が存在する。  
「政府は、総務庁を中心に規制緩和推進要綱を策定、今月中に閣議決定した。」

第4図 品詞判別規則と判別例

### [参考文献]

- (1)小松、安原、日比谷、加藤; “要約支援システム COGITO” 情報処理学会 vol.30 no.10 1989
- (2)中岩; “接続関係情報・用言意味情報を用いた省略格補完処理” 人工知能学会第3回全国大会