

丸カッコの分類とその機械処理のための考察*

横山 晶一 高木 美紀

(山形大学工学部)

1. はじめに

日本語の文には、文字だけでなく、さまざまな記号が使われる。文を区切る句点や、語句を区切る読点はもちろんであるが、その他にも多くの記号が用いられている。

句点は、一文の区切りを表すので、多くの自然言語処理システムにおいては、句点を認識してから文の処理を行うものが多い。読点は、係り受けの曖昧性などを解消するために用いられることが多いが、その正書法的な使用方法が確立していないために、必ずしも曖昧性解消に役立つものではない[1,2]。

自然言語処理システムでは、句読点以外の記号を取り扱わずに、記号の部分を見逃したり、記号で区切られた部分を未知語の名詞として扱ったりする場面が多く見られる。しかしながら、このような取り扱いをすると、非文が生じたり、文の形が損なわれたりする場合も多い。また、記号やそれに付随した部分に重要な情報が含まれている場合には、それらの情報を落としてしまう場合もある。

我々は以上のような観点から、さまざまな記号について調査を試みている。カギカッコ「」¹については、例外もあるが、そのほとんどが引用であり、符号を含んだ部分を名詞句として扱うことができることを確かめた。

本稿では、丸カッコ“()”を取り扱う。丸カッコは、「語句や文の次に、それについての注記を加えるとき用いる」[3]と定義されているが、注記の範囲が非常に幅広く、また、必ずしも上の定義に沿った形で使用されているとは限らないので、取り扱いの

難しい記号の一つである。

本研究では、主として新聞記事を中心に、丸カッコを含む文を約1,400文抽出して、まずそれらに対して分析を加え、使用法を分類した。その分類に基づき、機械処理のためのアルゴリズムを考察した。その結果、文字種や長さ等を手がかりにすれば、ある程度の機械処理が可能であることが判明した[4]。また、丸カッコを含む部分を別に処理すべきか、あるいはそのままカッコをはずすべきかについても分析を加えて考察した。その結果についても述べる。

2. 丸カッコの分類

主として新聞記事の中から、丸カッコを含む文を抽出し、その性質を分類した。文の数は現在約1,400文で、さらにデータ収集を継続中である。以下に例文とともに分類結果を述べる。

2. 1. カッコ内が記号

慣用的で決まった使われ方をするものである。この場合には、カッコを含んだ部分を一つの記号または語として登録しておく必要がある。例文中で下線をほどこした箇所が該当するところである。また、長い文の場合には省略箇所を“…”で示す。

- (1) 人が乗ったまま作業ができる高級品(?)ですと1000万程度、寿命は10年程度。
- (2) (有)花器屋萬右衛門の繁栄を祈りたいと思います。
- (3) …はたまた外国勢力の日本を分裂させるための陰謀! (笑)と妄想にふけてみたり。
- (4) 6月30日(金)に起きたニュースである。
- (5) …入れ換え作業中の客車(4両編成)が停車中の客車(同)に衝突しそうになり、…
- (6) 東京銀行(当時)の知り合いにこの話をしたら…

* Classification of Parentheses and Consideration for their Processing by YOKOYAMA Shoichi and TAKAGI Miki (Yamagata University)

¹ 丸カッコとまぎれないように、記号は“と”でくって表す。

上の例文で、(1)~(3)、(6)は、明らかにこれでま
まった記号として扱うことができる。(2)には、類似
のものとして、“(株)”や“(財)”などがある。(4)
は曜日を表すものだが、これもすべてをつくしても
7通りしかないので、記号と同等の扱いをすること
が可能である。(5)の“(同)”もこれ自体は記号と見
なせるが、一種の照応解析が必要になる。(1)、(3)は
カッコの部分を取り除いても問題はないが、(2)では
カッコの部分の情報が必要になる場合がある。また
(4)も、この情報だけから年月日までを推定するこ
とがある程度可能なので、取り除くと問題が生じる。
(6)の類似表現として、“社会党(現在社民党)”とい
ったものがあるが、機械処理の観点からは、これら
は同等に扱えない。

2. 2. カッコ内が固有名詞

さまざまなバリエーションがある。下記に例を示
す。

- (7) 首位に5打差でスタートしたアニカ・ソレンス
タム(スウェーデン)は、…
- (8) …アルマトイ(カザフスタン)やタシケント(ウ
ズベキスタン)など隣接する国々にも…
- (9) …その後、同じ代表のフランシスコ・パレンシ
ア(エスパニョール)には直行便のチケット…
- (10) A.L.ウィリアムズ(IBM)を委員長に、F.J.ポーチ
(GE)、R.E.ガーステンバーク(GM)等と…

上記の例は、いずれもカッコ内が固有名詞になっ
ている。(7)、(8)は国名であるが、その前の人名に対す
る所属国、その国にある地名となっている。また、
(9)、(10)は所属する組織を表す。カッコ内が所在地
を表す場合(後の文(15)内の下線をしていないカッ
コ)もこれに類似する。

- (11) そこで合意ができて社会党の連立復帰が決まれば「政局は終わる」(自民党幹部)
- (12) 「サッチャー回顧録」(日本経済新聞社)の冒頭
に書かれていた首相就任の風景は興味深い。

これらは、いずれもカギカッコによる引用の直後に

置かれており、(11)は発言者、(12)は引用部の本の出
版社を示す。(11)では、発言者の氏名や特徴が書か
れる場合がある。たとえば、“(〇〇氏)”、“(43歳男
性)”といったたぐいである。また、(12)と同様のもの
として、“(〇〇研究所調査)”、“(「××白書」によ
る)”という形で引用元を明示する場合がある。

2. 3. カッコ内が数字や単位

カッコ内が数字で最も単純なのは箇条書きである
が、これは箇条書き自体を一つの単位として扱うべ
きである。その他に、次のような例がある。

- (13) 女優、広末涼子(21)と仏俳優、ジャン・レノ(53)
が共演した仏映画「WASABI」(ジェラルド・ク
ラウジック監督)の完成披露試写会が…
- (14) …同州の裁判所に10億ドル(約1230億円)の
懲罰的な損害賠償を求める訴訟を起こした。
- (15) 整理回収機構は19日、昨年12月に経営が破綻
した瑞浪商工信用組合(岐阜県瑞浪市)の資産
(元本15億円)を3億円で買い取ると発表した。
- (16) 日本とオランダの交流は、西暦1600年4月(慶
長5年3月)、オランダ交易船の「デ・リーフデ
(慈愛)号」が日本に漂着した時から始まる。

文(13)では年齢、文(14)、(16)ではレートや年号の換
算、文(15)では資本金や資産の額を示している。文
(13)のような例は比較的簡単に扱えるが、その他は
カッコの内外の関係を考慮する必要がある。

2. 4. カッコ内が読み仮名

これは同義語や略語と似ているが、ルビをふらな
い場合には多く見られる。

- (17) 町長にインタビューし、「町の生業(なりわい)
は何ですか」と質問した。
- (18) 金泳三(キムヨンサム)元大統領と会談。

この場合どこまでが読み仮名かが問題となる。

2. 5. カッコの前と中のどちらかが略語

新聞記事等では多くの略語や訳語が出てくる。こ
れをカッコの内外に併記する場合が多い。

(19)日本など先進十八カ国が政府開発援助(ODA)を
抛出する。

(20)…インターネットと携帯電話はワイヤレスウェ
ップ(Wireless Web)というキーワードで結ばれ、
WAP(Wireless Application Protocol)も…

(21)日本でなぜCC(クローズドキャプション)が普
及しないのか。

(22)…当然代表取締役である日本の本社社長の署名
(サイン)が必要になります。

これらの例は、次の同義語や上位、下位概念と見る
こともできるが、カッコの前のどこまでがカッコ内
と等しいかが問題となる。

2. 6. カッコの内外が同義語や上下概念

前節と類似しているが、これにも多くのバリエー
ションがある。

(23)朝鮮民主主義人民共和国(北朝鮮)訪問を終え
たカーター元大統領は十八日、青瓦台(韓国大
統領官邸)を訪れ…

(24)…ブーチン大統領は東アジア3カ国(中国、北
朝鮮、日本)歴訪の最初の日程をクリアする…

(25)5000億円あった蓄え(保険金)は既に使い果た
し、あるのは不良債権だけ。

これらは、カッコの中とカッコの前とが、同義語(文
(23),(25))、上下の概念関係(文(24))になっている。
この特殊な例として、次のようなものがある。

(26)IRE が最近もっとも力を入れているのは、CAR
(Computer-Assisted Reporting = コンピュー
ター補助報道)だ。

この文では、カッコ内に“=”が入っている。カッ
コの中ですぐに“=”が来て、同義であることを明
示する場合もある。

2. 7. 注記の意味での使用

これがカッコの最も基本的な使用法のはずである
が、文例としては意外に少ない。前節の訳語も注記
と見れば、逆に多くのものがこれに含まれる。

(27)「たとえばクー・クラックス・クラン(通称 KKK
といわれている白人の優越性を説くアメリカの
テロリストの秘密組織)の旗を揚げたい…」

2. 8. 置き換え

同義語や上下概念との区別がしにくい場合がある
が、カッコの直前部分とカッコ内とが交換可能であ
るような例である。

(28)…私たちは遠慮深く(礼儀正しく)なり、時に
は卑屈になるのである。

(29)私も今まで選挙に行くたびに無効票(該当者無
し)と投票していました。

これらは、カッコの前と中とで同じ品詞句になって
いるので、そのまま置き換えが可能であるが、その
ような置き換えが不可能な場合もある。

(30)…ふつうのサラリーマンなら 1 万円から 1 万
5000 円の税金を取り戻せる(還付)はずだ。

この文では、下線部の入れ換えをすると、文として
成立しなくなる。このような例も多い。

2. 9. 情報補足

注記の一種と見られなくもないが、情報を補足し
ていると考えた方がよい文は多い。

(31)…多くの都市で意外な(といっちは失礼だが)人
物が現地社会に食い込んでいた。

(32)…に言い(タリバンを懲らしめるだけでなく、
テロリストの基地をたたける)、費用対効果の…

(33)実際、憎悪は日々拡大している。(イスラム諸国
では反米デモは激化し、ビンラディンTシャツ
が飛ぶように売れている)。

(34)そして日本が 8 月 1 5 日に降伏し武器を捨てる
ことで、確かに「世界平和」が実現した。(同時
に日本国民は「一億玉砕」をまぬがれた。)

いずれの文も情報を補足しているが、(31)はカッ
コ内が文や句になっていない。(34)では、本来丸カッ
コの中に句点を入れるべきではないが、このような
例は多い。

3. 機械処理のための考察

前節には、比較的数が多く、分類しやすい例を示した。実際には、分類が不明なパターンを有する文が非常に多い。この分類は、人間が見るには役立つが、機械処理を考慮した場合には、カッコの取り扱いに対していくつかの手順を考慮しておく必要がある。以下その手順について考察する。

(a) 記号や語として登録する

すでに示したように、“(笑)”など、定型的に現れるパターンに対しては、丸ごと辞書登録をして、名詞句的な扱いをすればよい。

(b) カッコの前や中の文字種を手がかりにする

訳語や略語の場合、一部の例外を除いて、文字種が同一であることが多い。略語では、アルファベットの大文字、あるいは、長い漢字の列の一部を取ったものなどが多い。また、読み仮名の場合は、ひらがなやカタカナで書かれる場合が多い。問題は、カッコの前の部分のどこまでをカッコ内とマッチングさせたらよいかということである。その場合でも、文字種が同じところを取ると、一部の例外を除いては大体マッチングが取れる。これは後述の、辞書引きしても得られない固有名詞や新語の場合には特に有効である。

年齢や為替の換算レートのような場合には、カッコの内外に数字が含まれている。ただ、この場合には種々のバリエーションがあって、数字が含まれているから換算だといちがいには言えない。

(c) 文中の位置を手がかりにする

箇条書きは、必ず文頭に位置する。また、出版社や発言者、情報提供者を示す場合には、多くの場合、カギカッコの後に丸カッコが来る。文中の位置を手がかりにできる場合はそれほど多くないが、パターンとして決定することはできる。

(d) 辞書やシソーラスを引く

同義語や上下概念は、シソーラスを引くことによって決定できる場合が多い。また、辞書を引くことによって、訳語や読み方が判定できる場合もある。シソーラスにのっていない語で同義語の可能性があ

るものについては、ユーザに問い合わせることも考えられる。

(e) カッコ内が句か文かを手がかりにする

カッコ内の長さが短い場合には、置き換えや訳語などの場合がある。また、名詞句になっている場合には、注記や情報補足の場合が多い。ただし、例にも示したように、カッコの中が完全な文になっていない場合には、取り扱いが難しい。

(f) カッコの取り外し

カギカッコの場合には、その部分を名詞句または一つの名詞として扱おうと、全体として文の解析に支障をきたさないが、丸カッコの場合には、その部分を名詞句として扱えるのは、文として完全な形でカッコの中に入っている場合のみである。一般的には、カッコの部分をそのまま取り除いて別に解析した方がよい場合が多い。文例には示さなかったが、カッコを取り除いてしまうと非文になる場合もある。

4. おわりに

すでに述べたように、まだ分類できていないさまざまな丸カッコの使用方法がある。今後はさらに文例を集めて分析し、機械処理のための考察を深めたい。

さらに、前節の考察をアルゴリズム化して、実際にシステムを作成し、丸カッコの扱い方に対する方法を研究する予定である。

丸カッコ以外の、二重カッコや、角カッコなどの記号についても、文例を集めて解析する予定である。

参考文献

- [1] 横山晶一：読点の使用法とその分類、情報処理学会第22回全国大会講演論文集 1M-4 (1981)
- [2] 横山晶一：読点の使用法とその分類(2)、情報処理学会第23回全国大会講演論文集 4M-4 (1981)
- [3] 金田一春彦他編：日本語百科大事典 縮刷版、大修館書店(1995) p.355
- [4] 高木美紀：機械処理のための丸カッコの分類、山形大学工学部卒業論文(2002)