

新聞記事文における括弧書き表現の分析とその処理について

2L-9 白井 諭^{*1} 矢部 孝幸^{*2} 松尾 三津恵^{*2} 西垣 万亀子^{*2} 大山 芳史^{*1}

^{*1}NTT コミュニケーション科学研究所

^{*2}NTT アドバンステクノロジー(株)

1 はじめに

日本語の新聞記事では、カギ括弧により引用・強調などが、マル括弧により補足・説明・換言・難読漢字の読みなどの雑多な情報が提示される。すなわち、1種類の記号が複数種類の表現に使用されることになり、それらの用法ごとに扱いが異なるため、日英翻訳等の機械処理を行なう上では、括弧表現の種類を識別することが必要になる。

本稿では、新聞記事や速報文に含まれる括弧を用いた表現を分析し、用法を判定するアルゴリズムを提案する。また、日英翻訳を行なう際に有効となる加工方法についても併せて提案する。

2 検討対象データ

本稿で検討対象としたのは、新聞記事1日分（日本経済新聞、1995年8月1日付、1,137記事）と、市況速報文約1カ月分（1995年7月、1,174記事）である。ともに、日本経済新聞社の有料情報サービスであるテレコンデータベースから抽出した。このテレコンデータベースには、これらの記事の英訳も収録されており、日英記事の対応付け[高橋96]を行なうことにより、日英翻訳のための加工方法を検討する際の参考情報として使用した。

表1に、日本語記事におけるカギ括弧およびマル括弧の出現状況を示す。ただし、カギ括弧の直後に現れるマル括弧は性質が若干異なるため、カギ括弧とマル括弧の連続用法として、それぞれの単独用法とは分けて扱うことにした。

表1 括弧表現の出現状況

種別	連続用法		単独用法		全体記事数	全体文数	平均文長(文字/文)
	「…」(…)	「…」(…)	「…」	(…)			
新聞	132	1,108	1,809	1,137	5,881	48.6	
速報	462	164	1,493	1,174	6,463	35.7	

Classification of Expressions with Parentheses and/or Quotations in Japanese Newspaper Articles

Satoshi SHIRAI^{*1}, Takayuki YABE^{*2}, Mitsue MATSUO^{*2}, Makiko NISHIGAKI^{*2} and Yoshifumi OOYAMA^{*1}

^{*1}NTT Communication Science Laboratories and ^{*2}NTT Advanced Technology Corporation

3 カギ括弧とマル括弧の連続用法

この用法の内容を分類した結果を表2に示す。

この用法において、カギ括弧が引用を示す場合は、直後のマル括弧はその発言者を示す。発言者としては、人名（肩書きが付加されたものを含む）はまれで、組織名が多い。従って、次の2条件を満たせば引用であると判定することができる。

- ①カギ括弧の外側が引用を取る形式である
- ②マル括弧の内側が組織名か人名である

②に関して、1記事中に同じ発言者の話が何度か引用されると、初出を除き（同）（同社）のように書かれるため、これに対する照応処理が必要となる。

表2 カギ括弧とマル括弧の連続用法の分類

番号	分類	用例	新聞	速報
1	「引用」 (発言者)	…この問題に触れ、「武村氏が改めて提携強化の意向を強調した」(さきがけ幹部)という。	65	388
2	「準引用」 (発言者)	…「冷え込んでいた設備投資が持ち直しつつある」(建設省建設経済局)ことを示した。	10	72
3	「強調」 (補足)	最新作「初ものがたり」(PHP研究所)では、いふし銀のような描写、確かな人間観察…	51	0
4	「強調」 (説明)	…「価格感と買物行動調査」(首都圏四十キロ圏の成人…、回収率六十・九%)で…	4	2
5	「強調」 (換言)	…国際コンソーシアム「朝鮮半島エネルギー開発機構」(KEDO)は一日、…	2	0
合計			132	462

表2では、この表現の直後が句点のもの、ト格で受け取っているものを引用、そうでないものを準引用とした。これは、日英翻訳のための加工の仕方が異なるからである。

引用の場合、基本的には次の形式でよく、たまにト格を取る動詞に応じてsaidを差し替えるなどの調整が必要になる程度である。

「引用」(発言者) → “引用” said 発言者

これに対して、準引用(表2の2ではコト、ほかに、ウエ、タメ、ノデ、など)では、引用のように単純に翻訳することができない。英訳では分割されることが多いようであるが、形式にはかなり揺らぎがあるため、もう少し詳細に検討する必要がある。

なお、カギ括弧が強調を表す場合は、次節以降の処理を準用することにより対処する。

4 カギ括弧の単独用法

この用法の内容を分類した結果を表3に示す。

カギ括弧が引用であるか否かは、前節に準じて次の条件を満たすか否かにより判定する。

①カギ括弧の外側が引用を取る形式である

引用の多くは、カギ括弧の外側に発言者が示されていることが多い。例えば、次の形式である。

... 国防相は... 演説し、「...」と語った。

... 「...」と話すのは... 大菅敏次副会長。

いずれも英訳は次のようになっていることが多い。

“引用” said 発言者

着目した文には発言者が示されていないが、前や後の文には発言者が示されていれば、着目した文は単純に引用形式に訳出されているようである。しかし、記事中にまったく発言者が示されていないこともあり、その場合は次のように処理されるようである。

引用, the sources said (引用がクォートされない)

準引用は、前節同様、今後の課題とする。

強調は、英訳においてもクォートされている場合もあるが件数は少ない。書き方を工夫して、クォートを避けているように思われる。理由として、新聞記事の場合、例えば国際条約に関連する論調では、日本語記事の引用は条約の文言であることを暗示しているため、英語記事も条約の文言を引用しない限り、誤訳の恐れがあるためと考えられる。

表3 カギ括弧の単独用法の分類

番号	分類	用例	新聞	速報
1	「引用」	... 同社の調査では、七三・五%の人が「睡眠に何らかの役に立った」と答えた。	419	142
2	「準引用」 →単純な引用でない	この作業を手がけたのも「...戦争とはどんなものか知りたかった」ためだ。	26	3
3	「強調」 →名詞句の形式のもの	今回の調査で明らかになったシヨッキングな数字の一つに「九二・五%」がある。	663	19
合計			1,108	164

5 マル括弧の単独用法

この用法を分類した結果を表4に示す。

マル括弧の中身は、名詞句(1,2,8,10)、数字のみ(3,4,9)、仮名のみ(5)、文(6)、格要素(7)となっている。英訳がないため、対処の方針が決められないものも少なくないが、[AAMT 96]を参考にすると、マル括弧の中身を別に訳した後、1は名詞句の最末

表4 マル括弧の単独用法の分類

番号	分類	用例	新聞	速報
1	(補足)	六月の景気指数は四五・七%と、五月(四六・一%)に続いて五〇%を下回った。	1,128	242
2	(換言)	提携後に着手したリストラクチャリング(事業の再構成)は一段落したのか。	338	150
3	(項番)	機構改革=(1)経営企画部と営業企画部を統合し、総合企画部とする(2)商品本部に...	108	0
4	(年齢)	国立国会図書館副館長の井門寛さん(58)のライフワークは太平洋戦争に関する文献...	93	0
5	(読み)	... 電動で腕を伸縮できるロボット玩(がん)具「電動パンチアクション...」を発売した。	85	0
6	(説明)	... 予算に計上した給与改善費(国は千五百八十八億円、...)の枠内に収まる見通し。	31	0
7	(補完)	(経営努力で)安くできる余力が生まれたら、その分は従業員の時給アップに回す。	23	1
8	(代替)	... 四千字程度にまとめ、住所、氏名、年齢、勤務先(学校名)、電話番号を明記し、...	3	0
9	(銘柄番号)	一方、東電(1)が利食い売りに押され、リョービ(3)、フジタ(4)も下げた。	-	745
10	(時刻、銘柄コード) →文頭のみ	(14時02分、コード6752)10円高の1330円と比較的底堅い値動き。	-	355
合計			1,809	1,493

尾、2は同じ表現にならないよう、3は節の先頭、4は人名の直後、8は名詞句の最末尾、9は中心名詞の直後、10は文頭にそれぞれ埋め戻せばよい。

5では、読みが付与される位置は必ずしも単語境界ではないこと、“スタッドレスタイヤ「OBSERVE(オブザーブ)」”の例もあることから、この処理は若干複雑なものになる。

6は、マル括弧の中身が独立した文として訳されている場合が多いため、今後の課題とする。

7は、マル括弧を除去してマル括弧の外側と一体にして訳した後、該当個所に括弧を付加すればよい。

6 おわりに

本稿では、括弧表現の分類と処理について検討した。今後は、処理系を実現するとともに、未解決の課題に取り組む予定である。

参考文献

- [AAMT 96] アジア太平洋機械翻訳協会 例文評価研究会: 機械翻訳における記号の取り扱いについて, *AAMT Journal*, No.15, pp.11-17
 [高橋 96] 高橋, 白井, 藤波, 池原, 上田, 松島: DB から抽出した日英新聞記事の自動対応付け, 言語処理学会第2回年次大会 B3-3, pp.201-204