2004－ＦＩ－ 76　(22)
2004－ＮＬ－163　(22)
2004／9／17

「新幹線要約」のための文末の整形

池田 諭史 , 大橋 一輝 , 山本 和英
長岡技術科学大学 電気系
E-mail:{ikeda,ohashi,ykaz}@nlp.nagaokaut.ac.jp

新幹線の電光掲示板で使用されるニュース記事は冗長度の少ない高密度表現となっている。また、体言や格助詞で終わるといった独特の文末表現をしている。そこで本稿ではこのような高密度文の文末に着目し文末が体言や格助詞で終わる形に整形することによる要約を試みた。整形はパターンマッチを用いて行ない、結果として文末の要約率は 52%であり 1 文当たり 2.50 文字の削除ができた。また、人が判断した正解率は 95%であった。

Processing sentence end reduction of a newsflash

Satoshi Ikeda , Kazuteru Ohashi , Kazuhide Yamamoto
Department of Electrical Engineering,Nagaoka University of Technology
E-mail:{ikeda,ohashi,ykaz}@nlp.nagaokaut.ac.jp

The electrical bulletin board news consists of high density expressions. The end of the sentence is unique shape that is nouns or case particles. This paper focuses on expressions of the sentence end, and attempt to summarize them by forming them into nouns or case particles. We summarize the news sentence by pattern matching approach. Our evaluation illustrates that our summarizer reduces 2.50 characters on average; the summarization ratio of sentence ends is 52%. We also show that the correctness of reduction is 95%.

1　はじめに

新幹線の電光掲示板やニュースの字幕のような文章では例 1 に示すような体言や格助詞で終わる文が多く見られる。

例 1) 日本人拉致疑惑などは与党の動向見極め 判断。
　　対ロ支援は 継続へ。

電光掲示板や字幕ニュースは文字数制限がある文章なので冗長度が少ない。すなわち高密度表現であるといえる。

これらの文は一部が省略されているが人が文意を補完することで意味をとることが可能になる。我々はこのような文の補完を意識せずに行なうことができる。さらに人によって違う意味になるのではなく同じ意味に補完される。省略前と比較すれば表現された内容は減っているが人が見れば文意を容易に補完することができるので、省略後も文意の保持された文といってもよい。これらの文は文意を保持する上では必要ないと思われる文末を削除し、省略内容の補完は人が行なうという形で文意を保ち、その上で文を短くしているので冗長度が低い。

我々は、電光掲示板のニュース文のような文への要約を目指し、その第一段階として体言や格助詞で終わる文への要約を行なった。文末は不必要な部分が多いと考えられるので、最初に削除しておくことにより、要約の第一段階としても使えると考えている。

佐藤ら [4] はパソコンで閲覧する目的で作成された新聞記事と携帯端末で閲覧する目的で作成された新聞記事から、文末の言い換えパターンの抽出を行なっている。しかし、そのパターンには複数の語への言い換えがある語もあり、このままでは言い換えることができない。

若尾ら [6] はテレビニュース番組の音声によるニュースと字幕によるニュースの調査を行なっている。調査は文末の調査も行なっており、本稿で行なったような要約方法の可能性を示している。短縮可能な表現の列挙も行なっており、またニュース字幕におけるこれらの短縮表現の使用頻度の調査している。

新幹線の電光掲示板と同様に、文字数制限のある文章のテレビニュースの字幕や、携帯端末用の新聞記事への要約はいくつか行なわれている [1,3,5]。石ざこら [1] は、重複部を削除することによる要約を行なっており、大森ら [3], 三上ら [5] は、文章全体の要約を行なっている。しかし、[1,3,5] は文末については触れられていない。

2　新幹線要約とその特徴

実際に新幹線の電光掲示板のニュースで使われている記事と同じ記事を配信しているサービスがある。これは日経ニュースメール (1) という NIKKEI-goo が行なっているサービスで、平日に 1 日 3 回メールでニュースが配信される。日経ニュースメールの中で、「主なニュース」の欄にある記事の本文が実際に新幹線の電光掲示板のニュースで使われている文であり、以下に例を示す。

例 2) ◆国土交通省、外国船の監視を強化
国交省はテロ対策の一環として外国船の監視を強化する。航行情報を一元管理するほか、港湾近くには監視カメラを設置へ。

日経ニュースメールは例 2 のように記事の見出しがあり、その下に数文でその内容が示され、その多くが 2 文である。1 記事当たりの文字数は 60 文字以内であり、その多くが 56 文字である。我々はこのメールを 1999 年 12 月〜現在までの約 4 年半分を収集した。収集したニュースメールの統計を表 1 に示す。

表 1: 収集したニュースメールの統計

メール数	3365
記事数	21127
文数	40374

新幹線の電光掲示板に配信されるニュースメールの記事は、一般の新聞記事の要約と考えられる。以降ではこれを新幹線要約と呼ぶ。新幹線要約の記事は、新聞等の一般的な記事に比べると文末に特徴がある。表 2 に新幹線要約と一般的な文として日本経済新聞 (2) の文末の品詞の使用頻度を示す。

表 2: 文末における品詞の出現比率

品詞	出現比率 [%]	
	日経	新幹線要約
名詞	23.7	55.92
(うちサ変名詞)	(5.00)	(39.90)
動詞	28.66	15.91
形容詞	1.80	0.19
副詞	0.20	0.22
助詞	1.56	8.83
(うち格助詞)	(0.34)	(6.41)
助動詞	38.59	18.52
記号	5.42	0.40

表 2 から一般的な記事である日本経済新聞は動詞や助動詞といった用言で終わっているものが大半を占めていることが分かる。一方新幹線要約ではサ変名詞で終わっている文、つまり体言止めになっている文が多い。これより、体言止めを行なった方が高密度表現になっているのではないかと推測できる。

また、新幹線要約では文字数が少なくて情報が多い単語が使われていると考えられる。一般的に和語よりも漢語の方が情報が多いとされるので、ニュースメールと新聞において漢語と和語の対応がとれる主な単語の出現比率を調べた。この結果を表 3 に示す。表 3 から、実際に新幹線要約では和語に比べて漢語が多く使われており、また一般的な記事と比較すると約 3 倍の比率で和語よりも漢語が使われていることが分かる。

以上より、新幹線要約は同じ意味をもつ表現ならばより短い表現を好んで使っているということが分かる。すなわち、新幹線要約は新聞記事に比べて高密度表現になっているといえる。また人間の補完能力を利用することにより、意味をとる上では不必要だが、一般的な文としては必要な助動詞の削除を行なうことで、文字数の削減を行なっている。

表 3: 新幹線要約と日本経済新聞の和語と漢語の比

和語	漢語	和語に対する漢語の比率		a/b
		日経 (a)	新幹線要約 (b)	
見つかる	発見	1.059	3.126	0.335
決める	決定	0.622	2.147	0.290
選ぶ	選出	0.342	2.522	0.136
分かる	判明	0.188	2.633	0.072
命じる	命令	1.356	3.605	0.376
述べる	発言	0.446	0.191	2.432
調べる	調査	5.914	50.882	0.116
以上 7 語の統計		2.712	1.011	0.373

3　要約手法

本稿では以下に示す手順で要約を行なった。

1. 断定表現及び敬語表現の削除
2. 「示す」の削除
3. サ変動詞の換言
4. 「なる」の削除
5. 「明らかに」の後の削除
6. 和語の換言
7. 「しまう」の削除
8. 「立つ」の削除
9. 未来の行動を示唆する表現
10. 文末の複合名詞への換言

本稿ではこの順番で要約を行なったが、手順 3.〜9. は任意の順序で処理が可能である。

なお、終助詞「か」で終わっている疑問表現は新幹線要約では使われないので処理の対象としない。

3.1　断定表現及び敬語表現の削除

以下に示すものを断定表現もしくは敬語表現とし、この表現で終わっている文についてその表現を削除する。

・ 断定表現: 「だった」 「である」 「だ」

・ 敬語表現: 「ます」 「です」

例 3) 幸い２０００年問題は解決しても、油断は禁物だ。
→幸い２０００年問題は解決しても、油断は禁物。

3.2　「示す」の削除

文が「を示す」や「を示した」で終わっている時、それを削除する。

例 4) 自民党との最終調整に入る意向を示した。
→自民党との最終調整に入る意向。

3.3　サ変動詞の換言

サ変動詞を換言する際にサ変動詞以降の部分を削除することで、サ変名詞で終わる体言止めの文にする。この際、サ変動詞より後に自立語が存在する場合にそれを削除すると意味が変わることや意味がとれなくなることが

ある。

例 5) 気がつくと、いつの間にかコンピューターに包囲されているのに驚く。
→*(1)気がつくと、いつの間にかコンピューターに包囲。

そこで自立語がサ変動詞の後にある場合は、処理を行なわないこととする。

Step 1 サ変動詞の「する」以降、つまりサ変名詞よりも後の部分を削除する。これ以降、サ変動詞の換言処理中の「サ変名詞」とは「する」の部分が削除されて名詞化したサ変動詞を指す。

Step 2 削除した部分に「見られる」「だろう」といった推定を表す表現が含まれていた時、文末に「か」を付加する。その後の処理は行なわずにサ変動詞の換言を終了する。

例 6) 逃走資金に困って自首したとみられる。
→逃走資金に困って自首か。

Step 3 削除した部分に「ない」や「ぬ」といった否定の表現が入っている時、「せず」を付加して処理を終了する。同時に「れる」といった受動を表す表現が含まれる時は、「されず」を付加する。その後の処理は行なわずにサ変動詞の換言を終了する。

例 7) 今年は牝馬が出走しない。
→今年は牝馬が出走せず。

Step 4 文末が「名詞＋を＋サ変名詞」であれば、「を」を取り除き、「名詞＋サ変名詞」の複合語にする。

例 8) 今月から、奇数月に各駅と車内のつり広告に詰め将棋を掲載する。
→今月から、奇数月に各駅と車内のつり広告に詰め将棋掲載。

Step 5 残った部分が「助詞$_1$＋名詞＋すること＋助詞$_2$＋名詞」で終わっている時、「すること」を取り除く。その時、助詞$_1$が「を」「か」の場合は、それを「の」に置き換える。

例 9) 両国政府の実務者協議を２月にも開催することで合意した。
→両国政府の実務者協議を２月にも開催で合意。

例 10) 知事選に中馬弘毅・同府連会長を擁立することを正式決定した。
→知事選に中馬弘毅・同府連会長の擁立を正式決定。

「初めて」が削除した部分に含まれている時、Step2 以降は、上記の処理ではなく以下の処理を行なう。

Step 2 削除した部分に「するのは」もしくは「したのは」が含まれる時、サ変名詞の前に「初めて」を付加する。削除した部分に「みられる」が含まれていれば、「か」を付加する。その後の処理は行なわずにサ変動詞の換言を終了する。

例 11) 人口関連活動を日本企業が直接援助するのは初めて。
→人口関連活動を日本企業が初めて直接援助。

例 12) 同条約に基づき日本政府が海外からごみを撤収したのは初めて。
→同条約に基づき日本政府が海外からごみを初めて撤収。

Step 3 削除した部分に「して」が含まれる時、「後初」を付加する。サ変名詞の直前が助詞「が」ならば、その助詞「が」を「の」に置き換える。

例 13) 会見に応じたのはカルマパ１７世が中国から出国して初めて。
→会見に応じたのはカルマパ１７世が中国から出国後初。

例 14) イスラエルに和平推進派のバラク政権が発足して初めて。
→イスラエルに和平推進派のバラク政権の発足後初。

Step 4.1 サ変名詞が「発言」「言及」の時、サ変名詞の前に「初めて」を付加する。

例 15) 主要閣僚が、６月選挙を前提に公式に発言したのは初めて。
→主要閣僚が、６月選挙を前提に公式に初めて発言。

例 16) ロシア軍幹部が撤退に言及したのは初めて。
→ロシア軍幹部が撤退に初めて言及。

Step 4.2 サ変名詞が「発言」「言及」以外の時、サ変名詞の直前の単語を調べ、その単語により以下の処理を行なう

- 助詞「の」「が」＋サ変名詞
  → 助詞「の」＋サ変名詞＋「は初」
- 助詞「を」「も」＋サ変名詞
  → 助詞は変化せず＋「初めて」＋サ変名詞
- 上記以外の時、
  ～サ変名詞 →～のサ変名詞＋「は初」

Step 5 削除した部分に「みられる」が含まれる時、文末に「か」を付加する。

3.4 「なる」の削除

「助詞＋なる」という表現が文中に存在する場合に、「助詞＋なる」以降の部分を削除する。この際、削除される部分に自立語が含まれていると意味が変わることや意味がとれなくなることがある。

例 17) １月末まで入札趣意書を受け付け、３月中旬に優先的交渉の対象となる企業を決める。
→*１月末まで入札趣意書を受け付け、３月中旬に優先的交渉の対象に。

そこで「助詞＋なる」より後に自立語が含まれる場合は処理を行なわない。推定を表す「だろう」が「助詞＋なる」の後に含まれる場合は、これを削除すると意味が変わるので、処理を行なわない。

(1)記号 '*' はその表現が不自然であることを表す。

「助詞＋なる」も含めてそれ以降を削除し、以下の処理を行なう。

助詞が「に」「と」の場合は、「に」を付け足す。

例 18) 総選挙投票３カ月半後の合意で、ぎりぎりの選択 となった。
→総選挙投票３カ月半後の合意で、ぎりぎりの選択 に。

削除した部分に、否定を表す「ない」や「ぬ」を含む時は、「ならず」を付け足す。

例 19) 火薬が湿っていたのか、ほとんどが起爆剤にならなかった。
→火薬が湿っていたのか、ほとんどが起爆剤にならず。

3.5 「明らかに」の後の削除

「明らかに」という表現が文中に存在する場合に、「明らかに」よりも後の部分を削除して「〜明らかに」という文つまり、助詞「に」で終わる文にする。この際、「明らかに」の後に自立語が存在する場合、自立語を削除すると意味が変わることや意味がとれなくなることがある。

例 20) 香港の人権擁護団体が３日明らかにしたところによると、北京 の…。
→*香港の人権擁護団体が３日明らかに。

そこで、「明らかに」より後に自立語を含む場合は処理を行なわない。

Step 1 「明らかに」の後の部分を削除する。

Step 2 削除した部分を調べ、削除された部分により以下の処理を行なう。

- 否定の表現「ない」か「ぬ」と受身の表現「れる」が同時に存在する場合
  →文末に「されず」を付加する。
- 否定の表現「ない」か「ぬ」が存在する場合
  →文末に「せず」を付加する。

例 21) 特別損失額は明らかに していない。
→特別損失額は明らかに せず。

例 22) 費用は明らかに されていない。
→費用は明らかに されず。

Step 3 「明らかに」の直前が「することを」ならば、「すること」を削除する。「すること」の直前が「助詞「に」＋サ変名詞」ならば、「に」を「へ」に置き換える。「すること」の直前が「助詞「を」＋サ変名詞」ならば、「を」を「の」に置き換える。

例 23) ４月に北区の西友赤羽店 に移転することを明らかにした。
→４月に北区の西友赤羽店 へ移転を明らかに。

例 24) 宗教法人の代表役員 を辞任することを明らかにした。
→宗教法人の代表役員 の辞任を明らかに。

3.6 和語の換言

表３に示した和語が文中に存在する場合に、和語より後を削除して和語を漢語に置き換える。この際、和語より後に自立語が存在する場合、自立語を削除してしまうと意味が変わることや意味が通らないことがある。

例 25) ヨットレースのアメリカズカップの挑戦艇を決める ルイ・ヴィトン杯 は 延期 となった。
→*ヨットレースのアメリカズカップの挑戦艇を決定。

そこで、和語の後に自立語が存在する場合は、処理を行なわない。表３の和語について漢語に換言した。

Step 1 和語も含めて和語以降の部分を削除する。

Step 2 削除によって、文末が「することを」になった場合、「すること」を削除し、対応する漢語を付加する。和語の換言の処理を終了する。

例 26)「災害広域支援マニュアル」の作成に着手 することを決めた。
→「災害広域支援マニュアル」の作成に着手を決定。

Step 3 以下の条件の場合、処理を行なう。

- 削除後の文末が助詞「が」で和語が「分かる」
  →漢語「判明」を付加して、処理を終了する。
- 削除後の文末が助詞「が」で和語が「調べる」以外
  →助詞「が」を助詞「を」に置き換える。
- 削除後の文末が「が＋名詞＋で」
  → 「の＋名詞＋を」
- 削除語の文末が助詞「は」で和語が「分かる」
  → 処理を行なわない。
- 和語が「調べる」で削除した部分に「している」
  →「〜調査中」。処理終了

Step 4 和語に対応する漢語を付加する。

例 27) 昨年１２月の大学理事会で解任されていたことが分かった。
→昨年１２月の大学理事会で解任されていたことが判明。

例 28) 変造硬貨計３５９枚 が見つかった。
→変造硬貨計３５９枚 を発見。

例 29) 寝室から中根さん が焼死体で見つかった。
→寝室から中根さん の焼死体を発見。

例 30) 西淀川署で出火原因を 調べている。
→西淀川署で出火原因を 調査中。

3.7 「しまう」の削除

この処理は文末以外にも適用できるので、文末以外にも適用している。「しまう」が文中に存在する場合に、「しまう」を削除する。「しまう」の後ろに「ば」が付いて「しまえば」になっている時、処理は行なわない。

「しまう」で文が終わっている時、「しまう」を削除して、その前の単語を原形にして、処理を終了する。

例 31) 難しい人名は振り仮名がないとまごついてしまう。
→難しい人名は振り仮名がないとまごつく。

「しまう」が文の途中にある時、「しまう」の直前の一語と「しまう」を取り除く。

例 32) 日本の宇宙開発は金縛り状態に陥って しまったのだろうか。
→日本の宇宙開発は金縛り状態に陥ったのだろうか。

3.8 「立つ」の削除

「立つ」が文中に存在する場合に、「立つ」を含めてそれ以降を削除する。この際、自立語が「立つ」以降に含まれている場合、この自立語を削除すると意味が変わることや意味が通らなくなることがある。

例 33) 行司の「ハッケヨイ」の掛け声で 立つ よう 競技規則 を 改正 した。
→*行司の「ハッケヨイ」の掛け声で。

そのため「立つ」より後に自立語を含む場合は、処理を行なわない。

「立つ」が慣用表現の一部になっている場合は、処理を行なわない。例外として「めどが立つ」については行なうこととする。

Step 1 「立つ」を含む「立つ」以降の部分を削除する。

例 34)「トップボーイ」はＴＶゲームの専門小売店の頂点に 立つ。
→「トップボーイ」はＴＶゲームの専門小売店の頂点に。

Step 2 削除した部分に否定の表現である「ない」や「ぬ」を含む場合は、文末に「立たず」を付加する。

例 35) 零式装備品が何になるのか、まだ見通しは 立っていない。
→零式装備品が何になるのか、まだ見通しは 立たず。

慣用表現「めどが立つ」の場合は、「が立つ」を削除し名詞「めど」で体言止めの文にする。削除部に「ない」や「ぬ」を含む場合は「めどが立たず」とする。

例 36) 来月、義足をつける めどが立った。
→来月、義足をつける めど。

例 37) 環境問題で二転三転し、建設の めどは立っていない。
→環境問題で二転三転し、建設の めどは立たず。

3.9 未来の行動を示唆する表現

以下に示す語を未来の行動を示唆する表現として、「する＋未来の行動を示唆する表現」が文中に存在する場合に、これらの表現を削除して助詞「へ」を付加する[(2)]。

・「予定」　・「計画」　・「方針」　・「方向」

「する＋未来の行動を示唆する表現」が文中に存在しても、これより後に「ない」「ぬ」のような否定を表す表現が入っていると意味が逆になるので、処理を行なわない。また、「という」や読点「、」が含まれる場合も処理を行なわない。複数の「する＋未来の行動を示唆する表現」が文に含まれる時は、一番最後に出現したものについて処理を行なう。

[(2)]大橋ら [2] では動詞的意味の弱い名詞として「予定」「計画」「方針」の 3 語を挙げている。

「する＋未来の行動を示唆する表現」を含むそれ以降の部分を削除する。削除後の文末が助詞ならば、この助詞も削除する。その後、文末に「へ」を付加する。

例 38) 来月の通常国会に提出し、２００１年度から実施 する予定だ。
→来月の通常国会に提出し、２００１年度から実施 へ。

3.10 文末の複合名詞への換言

上記の処理を行なった後に「名詞＋助詞＋サ変名詞」で終わっている場合に、助詞を削除し「名詞＋サ変名詞」の複合名詞で終わる文にする。ただし、名詞の種類が茶筌 (3) の分類で、代名詞、人名、固有名詞、接尾のいずれかの時、もしくは自立語でない時は処理を行なわない。助詞が「から」「で」「も」の時についても処理を行なわない。

複合名詞の妥当性を見るために、毎日新聞 (4) の記事の見出しと新幹線要約文から、複合名詞を取り出して辞書を作成した。助詞が「に」の時、「名詞＋サ変名詞」の複合名詞が辞書にあれば、その複合名詞は妥当であるとして、「名詞＋助詞＋サ変名詞」を「名詞＋サ変名詞」の複合名詞に置き換える。上記以外の助詞の時は、「名詞＋助詞＋サ変名詞」を「名詞＋サ変名詞」の複合名詞に置き換える。

例 39) ３階の焼け跡から男性の遺体 が見つかった。
→３階の焼け跡から男性の 遺体を発見。
→３階の焼け跡から男性の 遺体発見。

この換言を行なう際の複合名詞は、体言止めの時のみ使われる表現が多く、一般的な文では妥当性の判断は難しい。助詞「に」の時は一般的な文で使われる複合名詞が多く、また助詞「に」を削除してしまうと意味が保持されない場合が多い。従って作成した辞書より妥当性を判断を行なった。「に」以外の助詞の場合は、例 39 のように意味がとれる形になる。この例の場合「遺体発見」を「遺体 が 発見した」や「遺体 に 発見した」ではなく「遺体 を 発見した」という意味の文であると解釈できる。

4 要約実験

本稿の要約手法の妥当性をはかるために、毎日新聞 2000 年度版の全記事を入力として、要約を行なった。新聞記事は句点「。」を文の区切りとして、1 文づつ入力した。232,038 文を入力し、73,512 文について要約された出力を得た。

4.1 文末の要約率

原文と文末整形したものを前方から一致させて、違う表現になった文字以降を文末と定義して、文末部分の要約率を求めた。

例 40) 日本交通公社などが 1 月 1 日、社名 を変更する。
→日本交通公社などが 1 月 1 日、社名 変更。

例 40 の場合「日本交通公社などが 1 月 1 日、社名」は前方から一致すると同じなので、下線を引いた「を変更する」(5 文字) →「変更」(2 文字) の要約率を求める。

$$\text{例 40 の文末の要約率} = \frac{2}{5} = 0.40$$

となる。

3.7 節の「しまう」については、文末の整形ではないので除いている。表 4 に各手法による要約率を示す。ここでは、一つの処理しか行なってない文のみを示し、複数の処理が行なわれた文については省いている。要約手法の数字は節番号を示す。提案手法全体で文末の要約率 52%を実現した。

表 4: 各手法による文末の要約率

要約手法	3.1	3.2	3.3	3.4	
要約率	0.60	0.33	0.49	0.45	
文数	16825	1313	37995	7510	
要約手法	3.5	3.6	3.8	3.9	全体
要約率	0.62	0.66	0.41	0.36	0.52
文数	199	7194	197	848	72727

4.2 人手による正解評価

要約された 73,512 文から無作為に 1,000 文を取り出し、3 人の被験者が正しい文になっているかを個別に評価して、3 人の評価の多数決によって正解の評価を行なった。評価の基準は意味を保持していることと、違和感が許せる範囲内であるという 2 点である。結果を表 5 に示す。要約手法の数字は節番号を示す。

表 5: 本手法における正解率

要約手法	3.1	3.2	3.3	3.4	3.5
文数	231	19	492	107	9
正解数	205	18	481	106	8
正解率	0.89	0.95	0.98	0.99	0.89
要約手法	3.6	3.7	3.8	3.9	全体
文数	116	21	3	13	1000
正解数	113	17	3	12	952
正解率	0.97	0.81	1	0.92	0.95

またそれぞれの人による正解についてのゆれをみるために、一人以上が正解とした時と 3 人が正解とした時の正解率をそれぞれ求めた。それを表 6 に示す。全ての場合において正解率 90%以上という結果が得られた。

表 6: 正解の人数を変えた時の正解率

	1 人以上が正解	2 人以上が正解	3 人が正解
正解率	0.98	0.95	0.91

4.3 人間の文末整形との比較

要約文の出力された 73,512 文から無作為に 100 文を取り出しその入力文について人手で文末の整形を行った。その文について文末の要約率を求めた。表 7 にその結果を示す。表 7 より、人手による整形と同程度の要約率が得られていることが分かる。

整形後の文の削減文字数を求めた。また上記の人手で整形した 100 文についても、同様に削減文字数を求めた。表 8 に示す。表 8 より本手法では 1 文で 2.5 文字を削除でき、削減文字数においても要約率と同様に、人で同程度の文字数の削減ができている。

5 考察

5.1 不正解の文について

不正解の文についてその例と対処方法について述べる。

表 7: 本手法と人手による整形の文末の要約率

	本手法	人手による整形
文数	72727	100
要約率	0.52	0.51

表 8: 文字の削減数

	本手法	人手で整形
入力文数	73512	100
全削減文字数	183606	290
1 文当たりの削減文字数	2.50	2.90

例 41) 本当の顔は世界中の映画を取引する巨大な見本市なのだ。
→*本当の顔は世界中の映画を取引する巨大な見本市 なの。

例 41 は断定の表現「だ」を削除した時に間違った文となった例である。文末「なのだ」の「だ」だけを削除したために、残った部分が「なの」という文末としてはふさわしくない形になっている。この文は「なのだ」を削除すれば正しい文になる。断定の表現が不足していたことによって起こった間違いであるので、断定の表現を増やすことにより解決が可能である。

例 42) 顔はその人の年輪みたいなもので、喜怒哀楽の表情を積み重ね、人柄を示す。
→*顔はその人の年輪みたいなもので、喜怒哀楽の表情を積み重ね、人柄。

例 42 は「を示す」を削除した時に間違った文となった例である。「を示す」を削除する際に、「を示す」の直前が名詞の時に起こる間違いである。名詞がサ変名詞の時は正しい文になる。またサ変名詞以外でも「考え」「意向」「見通し」といったような名詞の時も正しく削除される。これらの名詞は数が多くないので辞書を作ることにより防ぐことができる。

例 43) スポーツ記事に出てくる『結果を出す』という言葉が 気になります。
→*スポーツ記事に出てくる『結果を出す』という言葉が 気に。

例 43 は、「なる」を削除した時に間違った文となった例である。これは「気になる」が一つの決まった慣用表現であり、この表現を途中で削除したために間違った文になったと考えられる。慣用表現の辞書を作ることで防げる

例 44) 北部ニーダーザクセン州のシュターデ原発（１９７２年に完成）を２００３年までに廃止することを明らかにした。
→*北部ニーダーザクセン州のシュターデ原発（１９７２年に完成）を２００３年まで へ 廃止を明らかに。

例 44 は「明らかに」の後を削除した時に間違った文となった例である。これは「明らかに」の直前が「助詞「に」＋サ変名詞」の時に「に」を「へ」に置き換える処理をした時に間違った文となった。この時、助詞「に」の前が場所を表す単語ならば正しい文になるが、それ以外の単語であれば、間違った文になる。また、場所を表

す場合に助詞「に」を「へ」に言い換えなくても例 45 に示すように正しい文になる。しかし、助詞「に」を「へ」に置き換えられた例 23 の方がより正しい文となる。

例 45) 4 月に北区の西友赤羽店に移転することを明らかにした。
→*4 月に北区の西友赤羽店 に 移転を明らかに。

助詞「に」の前の単語が場所の場合のみ「へ」に置き換えること、もしくは助詞「に」を「へ」に置き換えないことにより防ぐことができる。

例 46) 利用者に過度の使用を警告することを決めた。
→*利用者に過度の使用 を 警告を決定。

例 46 は和語を換言した時に間違った文となった例である。和語を変換する際に、「すること」を削除する時がある。この時、漢語への換言後に、係り受け関係が間違っているために起こった間違いである。「助詞$_1$ ＋名詞＋助詞$_2$ ＋漢語」で文が終わっている時、助詞が両方とも「を」の時に表現は正しくない。これは 助詞$_1$ を「の」に置き換えれば正しい文になる。例 47 のように修正すれば良い。

例 47) 利用者に過度の使用を警告することを決めた。
→利用者に過度の使用 の 警告を決定。

例 48) 母親を殺してしまおうと思っていた。
→*母親を 殺しう と思っていた。

例 48 は「しまう」を削除した時に間違った文となった例である。「しまう」を削除する際に、「しまう」の前の動詞と「しまう」の活用が一致しない時に起こる間違いである。「しまう」を削除する際に、活用形を「しまう」の活用形に一致させる必要がある。

例 49) 遊泳プールの水質基準を厳しくする方向で検討に入った。
→*遊泳プールの水質基準を 厳しくへ。

例 49 は「する＋未来の行動を示唆する表現」以降を削除し「へ」を付加した時に間違った文となった例である。「する」がサ変動詞の一部になっていない時に「する」以降を削除すると、間違った表現の文になる。よって「サ変名詞＋する＋未来の行動を示唆する表現」が文に含まれる時に「する」以降を削除して、「へ」を付加すればよい。

5.2 文末のサ変名詞の動詞性と名詞性

3.3 節のサ変動詞の換言において、「～は初」で終わる文になることがある。この表現は文意は分かるのだが人による違和感のゆれが大きかった。このため元々「～は初」とする処理の多くを、今回は「～初めて～」とする処理に変更した。処理を変更する前の例を例 51～例 52 に示す。

例 50) プーチン大統領がアラブ国家の指導者と会談するのは初めて。
→プーチン大統領がアラブ国家の指導者との 会談は初。

例 51) ＮＨＫと民放が同じ内容で放送するのは、初めて。
→ＮＨＫと民放が同じ内容での 放送は初。

例 52) 参加の背景を、ＡＳＥＡＮ側が公式の場で説明したのは初めて。
→参加の背景を、ＡＳＥＡＮ側が公式の場での 説明は初。

これらの例については違和感を感じたり文が間違っていると感じる場合がある。しかし、元の文に「初めて」という単語がない場合、つまり「は初」がない場合には正しい表現であるといえる。「初めて」という情報を削除した文についてサ変動詞の変換を行なった文を例 53～例 55 に示す。

例 53) プーチン大統領がアラブ国家の指導者と会談する。
→プーチン大統領がアラブ国家の指導者と 会談。

例 54) ＮＨＫと民放が同じ内容で放送する。
→ＮＨＫと民放が同じ内容で 放送。

例 55) 参加の背景を、ＡＳＥＡＮ側が公式の場で説明した。
→参加の背景を、ＡＳＥＡＮ側が公式の場で 説明。

これらについては違和感を感じる場合は少ないと思われる。これは「は初」の前のサ変名詞が大きく関係してくるためと考える。サ変名詞は名詞にも関わらず動作を表すことができる。複合名詞中のサ変名詞の動詞的な働きについては大橋ら [2] も議論している。サ変名詞が表す動作の度合は、サ変名詞の文での出現位置や人の感性によって様々である。

具体的に例 50 と例 53 での「会談」について考えてみる。まず例 53 では「会談する」というサ変動詞と同じ意味で文を補完して読むのが一般的であると思う。日本語において通常述語が最後に来るので、文において述語を見つけられない時は最後の単語を述語的に考える。一方例 50 では「会談」を名詞としてとらえる見方と例 53 と同様に動詞としてとらえる見方が考えられる。「会談」を普通名詞としてとらえた場合は文に違和感を感じ、サ変動詞としてとらえると違和感を感じることはない。例 53 の場合は例 50 とは違いサ変名詞は文末にないということで、このためにサ変名詞「会談」は動詞的な「会談する」だけではなく名詞的な「会談」という意味の 2 通りの見方ができる。

5.3 人手の整形文との比較

4.3 節で作成された人手での整形文と本手法の出力文について考察する。本手法と同じ結果の文と、文意を変えずに助詞が変わっていた文がほとんどであったが、大きく違う文もあった。例は原文、本手法の出力文、人手での整形文の順に示す。

例 56) 今後になお課題を残した 形だ。
→今後になお課題を残した 形。
→今後になお課題を 残す。

例 56 は「形」という単語には情報がほとんど含まれていない。人手で整形を行なう際にはそのことが判断できるので削除されている。

例 57) カラー写真を使ったグラフ面なども あります。
→カラー写真を使ったグラフ面なども ある。
→カラー写真を使ったグラフ面なども。

例 57 は人手での整形では助詞「も」で終わる文にしている。これは、本手法で実現できていない形の文で、新幹線要約ではしばしば見られる。また我々が文意を補

完する際には肯定の意味で補完するので肯定の意味を表す単語の「ある」は削除しても文意が補完される。このことを利用して人手での整形では「ある」が削除されている。同様に、「ある」についてもパターン化できると考える。

例 58) 同改正案は今月中に成立 する見通しだ。
→同改正案は今月中に成立 する見通し。
→同改正案は今月中に成立 へ。

例 58 は「見通し」も 3.9 節で述べた「へ」で終わる文への整形が可能であることを示している。

例 59) クレジットカードの利用 を中止すると発表した。
→クレジットカードの利用 を中止すると発表。
→クレジットカードの利用 の中止を発表。

例 60) 府の職員条例を準用し 1 期目の退職金５６８４万円 を返還させることを決めた。
→府の職員条例を準用し 1 期目の退職金５６８４万円 を返還させることを決定。
→府の職員条例を準用し 1 期目の退職金５６８４万円 の返還を決定。

例 59 では動作「中止する」ということを「発表」している。この時の「中止する」というサ変動詞に対してもサ変名詞化が可能であるという例である。同じような「発表」と同じように使われるサ変名詞は他にも例 60 で示されている「決定」など多くある。これらのサ変名詞が文末にある場合に直前に出現したサ変動詞をサ変名詞化することが可能であることを示している。

例 61) この結果、調査捕鯨問題をめぐる日米間の対立は収束した。
→この結果、調査捕鯨問題をめぐる日米間の対立収束。
→この結果、調査捕鯨問題をめぐる日米間の対立は収束。

例 61 は本手法が短くなった例である。人手の整形は文末を複合名詞化しておらずこのような文はいくつか見られた。

5.4 要約されなかった文

本手法がどの程度実行漏れがあるかを調べるために、本手法で要約されなかった 158,526 文から無作為に 200 文取り出して、本手法で要約する文にも関わらず要約されていない文がないかを調べた。その結果、本手法で処理されるべき文で処理されていない文はなかったが、要約を想定していたにも関わらず要約されていない文が 9 文あった。以下に例を原文、想定していた要約文の順で示す。

例 62) 焼け跡から池本さんが遺体で発見 された。
→焼け跡から池本さんが遺体で発見。

例 63) ３者で会合を持つ考えを 明らかにした。
→３者で会合を持つ考えを 明らかに。

例 62 は自立語の判断を茶筌の解析結果が「非自立」となった単語のみ自立語ではないとしたので、「れる」が自立語であるとして、処理されなかった。

例 63 は「明らかにする」の形では「する」という単語の情報はほとんどないので削除しても良いが、「する」は自立語であるので処理の対象にはならなかった。

6 おわりに

新幹線の電光掲示板で使用されるような文への要約を目的として、文末の整形の手法を提案した。提案手法を実装し、新聞記事を入力として要約したところ、文末の要約率は 52%であり、1 文当たり 2.5 文字削除することができた。これは人間が行なった文末整形の結果とほぼ同じ値となった。人手で出力の評価を行なった結果、正解率は 95%となった。

今後の課題として 5.1 節で示したように、断定の表現や慣用表現が不足しているのでそれに対応すること、係り受け関係が間違う文があったので、その文に対して係受け関係を修復することが挙げられる。新幹線要約で良く見られる表現として文末が助詞「も」で終わる記事があるがこれへの対応も課題である。

謝辞

本研究の一部は、科学研究費補助金 若手 (B) 「高密度表現を利用したまとめ型要約に必要な言語変換技術」 課題番号 16700134、及び科学研究費補助金 基盤 (A) 「円滑な情報伝達を支援する言語規格と言語変換技術」 課題番号 16200009 によって実施した。

使用した言語資源及びツール

(1) 日経ニュースメール,NIKKEI-goo, http://nikkeimail.goo.ne.jp/

(2) 日本経済新聞全記事データベース 2000 年度版, 日本経済新聞社.

(3) 形態素解析器「茶筌」,Ver.2.3.3, 奈良先端科学技術大学院大学 松本研究室, http://chasen.naist.jp/hiki/ChaSen/

(4) 毎日新聞全記事データベース 2000 年版, 毎日新聞社.

参考文献

[1] 石ざこ 友子, 片岡 明, 増山 繁, 中川聖一: テレビニュース番組の字幕作成のための重複部削除による要約, 情報処理学会 研究報告,NL-133-7,pp.45-52,1999.

[2] 大橋 一輝, 山本 和英: 「サ変動詞＋名詞」の複合名詞への換言, 言語処理学会第 10 回年次大会発表論文集,pp.693-696,2004.

[3] 大森 岳史, 増田 英孝, 中川裕志: Web 新聞記事の要約とその携帯端末向け記事による評価, 情報処理学会 研究報告,NL-153-1,pp.1-8,2003.

[4] 佐藤 大, 岩越 守孝, 増田 英孝, 中川 裕志: Web と携帯端末向けの新聞記事の対応コーパスからの言い換え抽出, 情報処理学会 研究報告,NL-159-27,pp.193-200,2004.

[5] 三上 真, 増山 繁, 中川 聖一: ニュース番組における字幕生成のための文内短縮による要約, 言語処理学会論文誌「自然言語処理」,Vol.6,No.6,pp.65-81,1999.

[6] 若尾 孝博, 江原 暉将, 白井 克彦: テレビニュース番組字幕に見られる要約の手法, 情報処理学会 研究報告,NL-122-13,pp.83-89,1997.