

日本語書き言葉を対象とした参照表現の自動省略 -人間と機械処理の省略傾向の比較-

飯田 龍^{1,a)} 徳永 健伸^{1,b)}

概要: 本稿では、日本語書き言葉を対象とした参照表現の自動生成課題、特に参照表現を省略するか否かを分類する問題を対象に、既存研究で利用されている談話的特徴を考慮した自動分類モデルを提案する。このモデルの性能を調査するために新聞記事にアノテーションされた結果を正解として評価を行い、提案モデルがセンタリング理論 [4] に基づくヒューリスティックな手法と比較して有効であることを示す。提案モデルは、元の新聞記事の省略箇所を正解として F 値で 0.550 の分類性能で省略する場合を同定できた。評価用データとして利用する新聞記事コーパスにアノテーションされた共参照・ゼロ照応関係は新聞固有の省略や言い回しのため、必ずしも人間の直感的な判断とは一致しない。そこで、3 人の専門家に新聞記事 50 記事中の参照表現を対象に省略すべきか否かを判断させ、その結果を正解として再評価を行った。この結果、人間の判断が一致する箇所のみを対象に評価した結果、50 記事全体を対象に分類した結果を上回る結果を得た。また、人間が一致した結果を分析し、省略生成の問題に関する今後の課題についても議論する。

キーワード: 自然言語生成, 参照表現, アノテーション, ゼロ照応

Generating Zero Anaphors in Japanese Texts - Human-machine Comparison on Zero Anaphor Generation -

RYU IIDA^{1,a)} TAKENOBU TOKUNAGA^{1,b)}

Abstract: This paper focuses on generating referring expressions in Japanese written texts, especially the task of deciding whether a referring expression is realised or not in a text. We propose a model for generating zero-anaphors, taking into account discourse information utilised in the previous work. The proposed model was evaluated using an annotated corpus which contains coreference and zero-anaphoric relations. Our empirical evaluation demonstrated that our model achieved 0.550 in f-score, which is better than the results of a heuristic method based on Centering Theory [4]. Our error analysis revealed the problem that the test data set has a skewed distribution of zero-anaphors due to the specific writing style of newspaper articles. We prepared an alternative gold standard which was made based on the agreement among three experts who judged each referring expression to be realised or not in coreference chains selected from 50 newspaper articles of the data set. We re-evaluated performance of our model with this human-made gold standard. The model performed better with the human-made gold standard than with the tags annotated in the annotated corpus. We also manually analysed the consistent annotation by the three experts and discuss the future directions of generating zero-anaphors.

Keywords: natural language generation, referring expressions, annotation, zero-anaphora

¹ 東京工業大学 大学院情報理工学研究所
Graduate School of Information Science and Engineering,
Tokyo Institute of Technology

a) ryu-i@cl.cs.titech.ac.jp

b) take@cl.cs.titech.ac.jp

1. はじめに

参照表現の自動生成の課題は談話研究における重要な課題であり、主に自然言語処理生成の研究分野で研究が進められてきた [11]. 特に, Winograd[18] の積み木の世界に代表されるような, ある世界と事物を指す参照表現の相互作用, 例えば, ある世界に配置された物体を弁別するための適切な参照表現を生成する問題が主な研究対象とされており, 現在では動的に変化する場面における物体を指す参照表現の利用や視線などのマルチモーダル情報と生成される参照表現の関係の分析など多岐に渡る研究が進められている [3], [17].

また, 参照表現の適切な生成を行うためには, 表現が生成される談話の文脈を反映した問題も扱う必要があり, この問題は談話の特徴をどのように導入すべきかを考えるべき興味深い問題である. 特に, 日本語では頻りに参照表現が省略されるため, 省略の生成も視野に入れた参照表現の生成が課題となる. この問題を自然言語処理の応用という観点から見ると, 機械翻訳や複数文書要約, 自動推敲など, 出力の冗長性を排除するために横断的に利用可能な重要な技術であるといえる.

前文脈の情報を考慮した参照表現生成では, 主に談話要素間の遷移の良さを扱うセンタリング理論 [4] に基づいた生成手法が提案されている [2], [9], [10], [12], [16]. 例えば, Kibble らの手法 [9] では「Continue transition の後の後ろ向き中心 (Cb) は代名詞化する」という規則を導入することで代名詞化の説明を行おうとしているが, そもそもセンタリング理論で定義された中心 (ある談話単位内で最も顕現性の高い談話要素) の遷移は隣接する文に限定される, また中心を一つしか扱えないという問題がある. また, センタリング理論で利用される談話の顕現性などの特徴を個別に利用した参照表現生成のモデルが提案されている (e.g. 橋本ら [22]) が, このモデルは小規模な分析用事例を対象に構築されたモデルであり, かつ参照表現生成の規則に人手による判断が必要な箇所が含まれているため, (ゼロ) 代名詞化の問題に関して潜在的にどの程度有効かはわからない.

そこで, 本研究では日本語書き言葉を対象に共参照・ゼロ照応関係がアノテーションされた大規模な評価用データを利用し, さまざまな文脈を考慮した参照表現の生成課題を解く. 問題とすべき参照表現にはさまざまなものがあるが, 本研究では機械翻訳や複数文書要約などへの応用を考え, 特に参照表現を省略すべきか否かという 2 値問題 (以降ではこの問題を省略生成課題と呼ぶ) を解く. 既存研究で考慮されてきた談話の手がかりを素性として学習することで 2 値分類器を作成し, 新聞記事データについて評価を行った結果を報告する.

まず, 2 節で省略生成に関する関連研究を説明し, 3 節で自動省略のモデルについて説明する. 次に 4 節で新聞記事の評価用データを用いてモデルの評価を行う. また, 自動省略の課題は正解を一意に決定することが困難であり, 評価が困難であるため, 5 節では複数人がそれぞれ参照表現の使用について判定した評価用データも利用して自動省略のモデルの性能を見積る. さらに, 複数の評価者で判定した結果が一致した箇所について人手で分析することで今後どのような点をモデルに反映すべきかについても議論する. 最後に 6 節でまとめと今後の課題について述べる.

2. 関連研究

本研究で対象とする省略生成の問題は英語における参照表現の代名詞化に相当する問題であるため, 英語の代名詞化の研究が参考になると考えられる. 1 節でも述べたように, 代名詞化の処理については談話要素の顕現性を遷移を扱うセンタリング理論 [4] の概念を用いてどのような場合に代名詞化を許すかを考える. センタリング理論では談話単位 (節や文など) における顕現性の高い談話要素 (中心) の遷移に関して, 予測と実際の振舞の差分を分析することで結束性のつながりの良さを説明する. 現在着目している談話単位内の談話要素集合を統語的な情報などで順序付けしたリスト (前向き中心; Cf) の中で最も順位の高い談話要素を優先中心 (Cp) と呼び, 直前の談話単位で最も順位が高くかつ現在着目している談話単位内に出現している談話要素を後ろ向き中心 (Cb) と呼ぶ. この Cb と Cp が同じか否か, 直前の談話単位の後ろ向き中心 Cb_{i-1} と今着目している談話単位の後ろ向き中心 Cb_i が同じか否かという特徴により 4 通りの遷移を考え, その 4 通りのうち Cb_{i-1} と Cb_i が同じであり, かつ Cb_i と Cp_i が同じである場合に最も結束性が良いという説明をする. つまり, この遷移の場合には人間の理解に負荷をかけないために, Cb_i を代名詞化することを許すという規則が多くのだ名詞生成の手法の基礎的な考え方となっている [2], [9], [10], [12], [16].

ただし, Henschel ら [6] も議論しているように, センタリング理論を使って代名詞化の処理を説明する場合以下の点が問題となる.

- 文章中の Cb が必ずしも代名詞化されるとは限らない (前方文脈に先行詞を持つ談話要素が代名詞化されない)
- Cb ではない談話要素が代名詞化される (1 つの発話単位で 1 つ以上の談話要素が代名詞化される)

この問題を解決するために, McCoy ら [13] は文境界や共参照連鎖中の 1 つ前の談話要素からの距離, 同じ談話スレッド (談話の構造や談話における時間経過によって分割された談話の一部) に含まれるか否か, 潜在的な曖昧性に着目して代名詞化するかを判断するモデルを提案している. 一方, Henschel らの手法 [6] では距離情報に加えて談話にお

ける新情報・旧情報の区別や主語の位置に現れているかという手がかりをもとに代名詞化するか否かを決定する。またこれらの手がかりに加え、並列構造などの文の構造や競合する談話要素が存在するか否かで分類規則を詳細化している。この研究以降も代名詞化のためのさまざまな手法が提案されているが、基本的には談話の顕現性を捉えるための特徴とセンタリング理論から派生した特徴を利用する点が酷似している。

この代名詞化の研究に対し、日本語を対象にした省略生成の研究では滑川ら [19] や宮ら [21] がヒューリスティックな規則に基づく省略生成の手法を提案している。また、吉見 [20] は機械翻訳の際に生成される不適切な代名詞 (e.g. それ) や指示連体詞 (e.g. その) に着目し、その代名詞を「そのまま残す」「省略する」「置き換える」のいずれかに分類する問題を扱っており、特に「置き換える」場合には「私」「彼ら」「我々」「自分」の4種類のいずれかを選択する6値分類問題を解いている。吉見の手法では、代名詞の表記や従属節の接続形など代名詞の省略に影響すると考えられる情報を素性として利用し、決定木学習 [15] を用いて分類を行っている。機械翻訳システムが出力した代名詞を対象に上記6種類のどれに分類すべきかを人手で分類することで評価用データを作成し、そのデータに基づいた評価を行っている。この結果、代名詞を省略すべきか否かの2値分類で評価した場合、79.9%の精度を得ている。ただし、評価対象は代名詞のみであるため、本研究で分類対象とする名詞句は評価対象に含まれておらず、また問題の選別に機械翻訳システムが出力した日本語文内の代名詞を対象としているため、問題が翻訳システムの品質に影響されるという問題がある。

また、橋本ら [22] では、多様な参照表現の生成の問題を考えている。彼女らはGDA [5] のタグを付与された55記事の新聞記事を対象に人手で分析を行い、それぞれの参照表現の生成にどのような特徴が関係をするかを検討している。さらに参照表現を選択するモデルを人手で作成し、モデル作成時に参照した記事とは異なる10記事を対象に評価を行い、約96%の精度で適切な参照表現を生成できたと報告している。ただし、人手作成された決定木には「N2が現在の主題と強く関連をもつ」など、計算機による実装がどの程度可能であるかが不明瞭な条件を含むため、それぞれの条件を実際に実装可能な条件に置き換えた上で実装、その後評価を行う必要がある。

3. 自動省略モデル

2節で示した代名詞化・省略生成の手法は、基本的には分類対象となる談話単位の先行詞、つまり同じ実体を指す談話要素の連続 (以降、**共参照連鎖**) における一つ前の談話要素との関係を判断の対象としているが、本研究ではコーパスにアノテーションされた共参照・ゼロ照応の関係をも

高知県 $e_{2,1}$ の橋本大二郎知事 $e_{1,1}$ は三十一日、都道府県で初めて一般事務職の採用資格から国籍条項を ($\phi_{1,2}$ ガ) 撤廃する方針を明らかにした。高知県 $e_{2,2}$ の現在の一般事務職の採用は日本国籍が要件。国は「公権力の行使や公の意思形成に携わる公務員は日本国籍が必要」との姿勢だが、地方公務員法では日本国籍がない人の任用を禁じる規定はない。橋本知事 $e_{1,3}$ は「地方行政の運営上、国籍条項の必要性は ($\phi_{1,4}$ ガ) 感じない。少なくとも日本に生まれ育った在日韓国・朝鮮人を地方公務員として排斥する理由はない。戦後五十年を契機に ($\phi_{1,5}$ ガ) 実現させたい」と話している。

談話要素 $e_{i,j}$ において、 i は同じ要素が同一の共参照連鎖に含まれることを表し、 j は共参照連鎖における出現の順序を表す。 $\phi_{n,m}$ はゼロ代名詞を表す。

図1 モデル構築のための例文

とに共参照連鎖を復元し、この連鎖から分類に有効だと考えられる情報を抽出し、その情報を素性として利用することで効果的に分類を行うことを考える。

提案する自動生成のモデルを図1の文章中の共参照連鎖を例に説明する。この文章中には最初に導入された「橋本大二郎知事 $e_{1,1}$ 」が以降で省略される、もしくは明示的に記述される場合のいずれかで出現している。この談話要素 $e_{1,i}$ ($1 \leq i \leq 5$)*1の集合が共参照連鎖となるが、ここで省略生成の問題は $e_{1,i}$ ($1 < i$) のそれぞれに対して「省略する」「省略しない」の2値分類問題となる。

ただし、任意の共参照関係を対象としてしまうと、「高知県 $e_{2,2}$ 」のようにAのBの関係で出現している表現を省略するか否かという問題を扱う必要が出てくる。もしこの現象まで対象とすると、対応する「AのB」のAの省略がアノテーションされていなければならないが、一般的に間接照応に分類される関係のアノテーションは困難であり、使用可能なコーパスも限られる。このため、分類対象とする談話要素 $e_{i,j}$ は述語の項、特にガ格、ヲ格、ニ格の項に限定してモデルを構築する。

分類対象となる談話要素 $e_{i,j}$ を省略すべきか否かという2値分類問題を考えた場合、2節に示した内容からも分かるように、共参照連鎖内の直前の談話要素 $e_{i,j-1}$ との関係や、共参照連鎖の最初に出現する談話要素 $e_{i,1}$ が文章中でどのように導入されたかが重要であると考えられる。例えば、図1の $e_{1,3}$ が省略されるか否かを考える際には、直前の談話要素 $e_{1,2}$ ($= \phi_{1,2}$) との間に別の主題が導入されている、また最初の談話要素 $e_{1,1}$ ($=$ 橋本大二郎知事 $e_{1,1}$) が主題化されて導入されている、などの情報を参照することで適切に省略するか否かの判断が可能になると考えられる。

特に、センタリング理論 [4] からわかる談話の顕現性の手がかりは非常に重要であり、ここではセンタリング理論を拡張したNariyamaの手法 [14] を用いて、談話の顕現性の特徴を導入する。具体的には、前向き中心を拡張したり

*1 以降では共参照連鎖 i の j 番目の要素を $e_{i,j}$ と表記する。

表 1 図 1 の例文に対応する SRL の作成

	は	が	に	を	その他
$e_{1,2}$	橋本大二郎知事	-	-	国籍条項	採用資格
$e_{1,3}$	規定	日本国籍	意思形成	任用	人
$e_{1,4}$	必要性	日本国籍	意思形成	任用	国籍条項
$e_{1,5}$	理由	日本国籍	契機	戦後五十年	地方公務員

表内の $e_{i,j}$ の行はそこまで読まれた場合の SRL の内容を表す。

ストである salient reference list (SRL) からわかる情報を素性として利用する。これにより、既存のセンタリング理論に基づく手法と比較して柔軟に主題となる要素を参照することができる。この SRL には「は」、「が」、「に」、「を」、「その他」の 5 つのスポットが用意されており、文章の出現順にそれぞれの談話要素が該当するスポットに上書きされる。この SRL の各段階と $e_{i,j}$ が省略されるか否かの対応を図 1 の例を使って説明する。まず、文章の最初から $e_{1,2}$ に到達するまでに、SRL に保持された談話要素の集合が表 1 の $e_{1,2}$ に相当する。この段階では、 $e_{1,2}$ と共参照関係にある談話要素 $e_{1,1}$ 「橋本大二郎知事」が最も顕現性の高いスポットである「は」の箇所に保持されている。つまり、ある実体を指す談話要素が顕現性の高い状態であらためてその実体を指す談話要素が出現しているため、ここでは省略可能となる。一方、 $e_{1,3}$ まで読み進めると「は」のスポットは別の談話要素「規定」に上書きされるため、この箇所では明示的に参照表現を出現させる必要がある。このように、Nariyama の提案する SRL を使った談話要素の扱いは省略生成と相関があるため、分類対象となる談話要素まで読み進めた場合にこのリスト内でどのスポットに保持されているかを素性として利用する*2。この SRL に基づく素性を含め分類に利用する素性を表 2 にまとめる。

4. 評価実験 1: NAIST テキストコーパスを用いた評価実験

3 節の自動省略のモデルを評価するために評価実験を行った。自動省略のモデル作成と評価には NAIST テキストコーパス [24] にアノテーションされた共参照・ゼロ照応関係を利用する。ただし、NAIST テキストコーパスにアノテーションされたゼロ照応の関係は述語の原形に対して付与されているため、そのまま利用すると受け身など格交替をともなう述語については談話要素の顕現性を反映できないという問題がある。そこで、受け身・使役をともなう述語については述語の出現形に対するガ格、ヲ格、ニ格をアノテーションし、その結果を評価に利用する。また、NAIST テキストコーパスには省略された格要素の出現位置に関する情報が含まれていないため、ゼロ代名詞の生成位置を述語の直前と仮定して評価を行う。また、表 2 では NAIST テキストコーパスに含まれていない段落の情報を

*2 センタリング素性は前述の情報が必要となるため、以降の実験では前方の解析はすべて正しくできていると仮定する。

利用しているが、これについては毎日新聞 95 年のデータに含まれていた段落の情報を参照して素性を抽出した。

評価に利用するデータの詳細を表 3 にまとめる。表 3 の省略の総数と非省略の総数が分類対象となる談話要素 $e_{i,j}$ ($j > 1$) のうち省略すべき要素とそれ以外の要素の総数である。この割合からわかるように、この問題設定では談話要素の非省略の場合が少ないため、非省略を検出する問題として再現率、精度、F 値を求める。また、自動分類モデルの学習・分類には最大エントロピーモデル*3を利用した。

4.1 ベースラインモデル

既存研究の多くは、センタリング理論に基づく遷移の良さを指標に（ゼロ）代名詞化するか否かを決定している。そこで、本研究でも Nariyama[14] の SRL を用いたヒューリスティックな省略生成の手法をベースラインモデルとする。既存のセンタリング理論に基づく手法と同様に現在着目している談話要素を生成する際に同一実体を指す前方の談話要素の顕現性が高い場合に省略し、それ以外の場合は省略しないという戦略をとる。具体的には 3 節で示した SRL を談話の進行にともなって管理し、分類対象となる談話要素に到達した時点で「は」のスポットに入っている談話要素と分類対象の談話要素が同一の共参照連鎖に含まれる場合は「省略する」、それ以外の場合は「省略しない」と分類する。

4.2 実験結果

ベースラインモデルと 3 節の提案する省略生成モデルの評価結果を表 4 に示す。この結果から、SRL に基づく手法ではほとんどの談話要素を「省略しない」と分類してしまうため、再現率は高いが精度が低くなり、結果的に F 値が低くなっている。一方、提案モデルでは再現率は低下するものの、精度が 7 割近くまで上がっているため、F 値で 0.550 という結果を得た。この結果より、センタリング理論に基づく特徴だけではなく、共参照連鎖の最初の要素や分類対象となる談話要素の共参照連鎖における出現位置などの特徴を考慮して省略するか否かを考えることが重要であることがわかる。

ただし、この結果は分類対象となる談話要素の直前までの正解を仮定しているため、そのまま実装して利用可能で

*3 実験では最大エントロピーモデルの実装の一つである MegaM (<http://www.cs.utah.edu/~hal/megam/>) を使用した。

表 2 自動省略モデルで利用する素性

ID	素性タイプ	素性	説明
1	E.FIRST	head_word	$e_{i,1}$ の主辞の文字列
2		ne	$e_{i,1}$ の固有名ラベル (e.g. LOCATION, PERSON)
3		case	$e_{i,1}$ を含む文節内の格助詞 (e.g. は, が, を)
4	E.CURRENT	first_sent	$e_{i,j}$ が文章の最初の文に出現しているか否か
5		last_sent	$e_{i,j}$ が文章の最後の文に出現しているか否か
6		first_para	$e_{i,j}$ が文章の最初の段落に出現しているか否か
8		pred	$e_{i,j}$ の係り先の述語の文字列 (基本形)
9		passive	$e_{i,j}$ の述語の態 (受け身か否か)
10		in_quote	$e_{i,j}$ が括弧 (「」) の中に出現しているか否か
11		arg_label	$e_{i,j}$ が何格を生成しようとしているか: ガ格, ヲ格, ニ格
12		func_word_in_path	$e_{i,j}$ から文末までに出現する機能語の基本形の文字列
13		func_pos_in_path	$e_{i,j}$ から文末までに出現する機能語の基本形の文字列
14		E.PREV	para_diff
15	sent_diff		$e_{i,j-1}$ と $e_{i,j}$ で何文離れているか
16	chain_last		$e_{i,j}$ が共参照連鎖内の最後の要素か否か
17	CENTERING	srl_order	$e_{i,j}$ の SRL 内のどのスロットに入っているか
18		srl_rank	$e_{i,j}$ の SRL 内の順位

素性タイプのうち、E.FIRST は共参照連鎖の最初の談話要素から抽出する素性であり、一方 E.CURRENT は分類対象となっている談話要素から抽出する素性を表す。E.PREV は共参照連鎖内の一つ前の談話要素との関係を表す素性であり、また CENTERING は Nariyama[14] の考えを反映した談話の顕現性を捉える素性である。

表 3 NAIST テキストコーパスのデータの内訳

	記事数	共参照連鎖の総数	分類対象の総数	省略の総数	非省略の総数
訓練用データ	1,753	12,290	21,965	18,278 (83.2%)	3,687 (16.8%)
評価用データ	696	4,975	9,563	7,918 (82.8%)	1,645 (17.2%)

表 4 自動省略の評価結果

	再現率	精度	F 値
ベースライン	0.953	0.172	0.291
提案モデル	0.464	0.678	0.550

はないという問題がある。また、文章中に出現する他の共参照連鎖の影響については SRL を参照することで間接的に考慮できるが、他に同じ意味カテゴリに属する談話要素が出現しているかといった情報は明示的に利用しておらず、この情報を顕現性の強さと組み合わせでどう素性に加えるかについては今後の課題である。

また、この評価では共参照連鎖中の談話要素を省略するか否かという観点で問題を考えたが、そもそも新聞記事における参照表現の省略を正解としてよいかという問題がある。新聞記事では、文章を決められた文字数で埋めるために過剰に省略する、もしくは報道記事においてまとめに相当する一文にはあらためて談話要素を記述するという特徴が見られた。このため、5 節では、より人間の談話に関する直観を反映した評価課題について考える。

5. 評価実験 2: 人手による参照表現の評価

4 節で述べたように新聞記事では記事を短くするために過剰に参照表現を省略したり、最後の一文に記事のまとめ

が書かれる際に明示的に名詞句が書かれるなど、新聞記事固有の参照表現の利用法があると考えられる。そこで、より自然な参照表現の利用について調査を行うために、複数の評価者に各文脈における参照表現をどう生成すべきかを選択させ、その結果を分析する。また、3 節で提案した自動省略モデルを適用し、人間の参照表現の利用と機械の振舞いの違いを見る。

4 節では NAIST テキストコーパスにアノテーションされた関係を正解とみなして評価を行ったが、NAIST テキストコーパスには述語の省略された項の出現位置の情報が含まれないため、評価実験 1 では談話要素の文章中における出現順序を反映した評価ができていなかった。この出現順序に関する問題を例 (1) を用いて説明する。例 (1) では、 $e_{4,j}$ ($2 \leq j \leq 5$) が参照表現として生成される談話要素を表し、各 $e_{4,j}$ の先行詞は見出し中の「新銀行」である。また、各談話要素の後の下付き数字はその談話要素が対応する数字を持つ下線部に係ることを表す。この例では、 $e_{4,j}$ は文法役割や統語的な位置を考慮して適切な位置に配置されているが、NAIST テキストコーパスには各述語「取得する」「増資する」「四百億円とする」「開業する」のガ格が省略されており、かつその先行詞は最初に出現している「新銀行_{4,1}」であることしかアノテーションされてない。各述語のガ格をどの位置に生成すべきかを決定することは容易

でないため、4節のように近似的に述語の直前にガ格を生成すると仮定すると、 $e_{4,2}$ はその係り先である「開業する」の直前に出現することになり、“ $e_{4,3} \rightarrow e_{4,4} \rightarrow e_{4,5} \rightarrow e_{4,2}$ ”という例(1)に示した談話要素の出現順序とは異なる解析順序で問題を考えることになる。この結果、実際はまず $e_{4,2}$ の位置でまず省略すべきか否かを考え、それ以降はその $e_{4,2}$ の顕現性を反映した生成の課題を考える問題であるにもかかわらず、それを無視した解析を行うことになる。

(1) ◆新銀行_{4,1}の中身

($e_{4,2}$:ガ)₁₃ ($e_{4,3}$:ガ)₁₄ 十七日に大蔵省の銀行免許を取得、₁₄ ($e_{4,4}$:ガ)₁₅ 二月上旬までに増資して₁₅ ($e_{4,5}$:ガ)₁₆ 資本金を四百億円とし、₁₆ 三月二十日をめぐり、開業する。₁₃

この点を考慮するために、参照表現の出現位置を手で修正し、その結果に対して手で参照表現を選択する。ただし、任意の表現を許してしまうと問題が発散するため、今回の評価では「参照表現を省略する」「参照表現を主題化して明示的に記述する(つまり、助詞“は”もしくは“には”をともなって表現する)」「主題化せずに明示的に記述する(つまり、助詞“が/を/に”をともなって表現する)」の3種類から選択してもらい、人手生成の傾向を分析する。

5.1 人手評価用データの作成

表3に示した評価用データのうち、記事内に含まれる1つの共参照連鎖の要素数が10以上、20以下である記事集合からランダムに50記事を選択し*4、この記事集合を対象に人手評価を行う。上述のように共参照連鎖中の談話要素の出現順序を考慮するために、NAISTテキストコーパスにアノテーションされた関係のうち、ゼロ照応の関係については出現順序の情報が付与されていないため、著者の一人が人手で参照表現の出現位置を修正して作業用データを作成した。具体的には例(1)に示したように、京都大学テキストコーパスの「形態素・構文タグ付きコーパス作成の作業基準」[23]に示された係り受け構造に合致するように談話要素を配置する。つまり、もっとも文末に近い要素に係る談話要素が主題化されて生成される可能性がある場合は文頭に配置し、それ以外の談話要素は節内で「が」「に」「を」の順序を保持するように配置する。また、どの表現が同じ共参照連鎖に属するかを概観しやすいように同一の色で表示するとともにゼロ照応の関係が把握しやすいように参照表現の係り先の情報も提示した。この問題では4節の問題と同様に述語のガ格、ヲ格、ニ格のみを対象とし、それ以外は文章中でどのように出現しているかはあらかじめ表示した上で「主題化する」「主題化しない」「省略する」の

*4 共参照連鎖の要素数が過剰に多い記事では、書き手の「私」が文章の最初に出現し、その表現が以降で頻繁に省略されるという記事が多いため、この偏りを排除するため20以下の要素数という制約を追加した。

表5 人手による作業結果

	t	o	x	保留
作業員 A	145	234	1048	15
作業員 B	219	210	1011	2
作業員 C	195	140	1107	0

3分類を行うことになる。

図2に人手評価用の問題の例を示す。問題中で上付き文字が同じ参照表現が同一の共参照連鎖に含まれることを表し、下付き文字が同じ下付き文字を持つ下線部に関して述語と項の関係にあることを表す。このような問題が与えられた状況で作業員は括弧内の“ ”の箇所を主題化して記述する場合には“t”、主題化せずに記述する場合には“o”、省略する場合には“x”に置き換えるという作業を文章全体に対して漏れなく行う。ただし、そもそもNAISTテキストコーパスにアノテーションされている関係の不備や、データ作成に関する問題などで作業が困難な場合は該当箇所を作業せず保留することを許した。

5.2 人手作業結果の分析

問題作成をしなかった著者1名を含む自然言語処理研究者3名に作業を依頼し、5.1に示した50記事に対して判断をさせた。作業結果を表5にまとめる。表5からわかるように、作業結果は作業員ごとにかなりばらついており、省略生成の問題は専門的な作業員を雇用しても、作業結果にかなりの揺れが生じることがわかる。

次に、3人のうちそれぞれ2人の間でどのくらい作業結果が一致しているのかを調査した。2者間の比較結果を表6に示す。これらの結果からわかる傾向としては、旧情報に相当する「は/には(t)」と新情報に相当する「が/を/に(o)」の間は揺れが少ないのに対し、「は/には(t)」と「省略する(x)」の間の揺れが大きいことがわかる。人間は新・旧情報の分類については比較的揺れが少ないため、機械で処理する際も同様の傾向になる可能性がある。そこで、まずは新・旧情報の分類を行い、その結果、旧情報に分類された結果のみを省略するか否かという問題で考える方が捉えるべき特徴が明確になることが考えられる。この点を踏まえた段階的な処理については今後検討したい。

次に作業員間の一致率と新聞記事にアノテーションされた結果と各作業員の一致率を調査した。t, o, xそれぞれの一致数と全体の判断の一致率、Cohen[1]のKappa値を計算した値を表7に示す。一致率(strict)はtとoを区別して一致率を調べた結果であり、一致率(lenient)はtとoを区別せず、省略するか否かで一致率を計算した結果を表す。

まず、Kappa値で比較すると作業員Cが比較的新聞に近い選択を行っているため、Cと新聞のKappa値がAとCのKappa値より高くなっているが、それ以外の場合は人の

950115-0038

神奈川県平塚市の平塚農業高校⁵⁵の生徒たちが、(:ガ)¹ てんぶらなど台所から出る使い古しの“廃食油”を 使った¹ 肥料作りに三年がかりで成功した。

(:ガ)² 学校⁵⁵の畑でダイコンやハクサイなどを栽培した² 結果、市販の肥料と大差ない成果も得られた。

廃食油の再利用はこれまでせっけん作りが中心だったが、最近ではディーゼル車の燃料¹⁰も注目を浴びている。

廃食油を肥料に変身させる試みに挑んだのは平塚農高⁵⁵の園芸科学班十八人と教諭の松山明彦さん¹⁵。

「水質汚染の一因になっている廃食油を家庭でもっと簡単に再利用する方法はないのか」。

(:ガ)¹⁵ (:ガ)¹⁵ 生ごみに含まれる油²³も、たい肥にして土に (:フ)²³ 返せば、3 (:ガ)²³ 肥料になる⁴ ことにヒントを得て、5 肥料化の実験を始めた⁶。

(:ガ)¹⁵ (:ガ)¹⁵ 最初は廃食油²⁹と米ぬかを (:ニ)²⁹ 混ぜた⁷⁸ あと (:ガ)¹⁵ イースト菌で (:フ)²⁹ 分解させる⁹ 試みを行ったが、10 温度管理が難しくて失敗¹¹。

(:ガ)¹⁵ 米ぬかと油³⁵の混合比率を変えながら、12 (:ガ)¹⁵ 市販の生ごみ分解微生物などを (:ニ)³⁵ 加えて¹³¹⁴

(:ガ)¹⁵ 実験した¹⁵ 結果、(:ガ)¹⁵ 油³⁸百ミリリットル³⁹に米ぬか二百五十グラムと少量の重曹を混ぜて¹⁶ (:フ)³⁹ 発酵させた¹⁷ あと、(:ガ)¹⁵ 砂を (:ニ)³⁹ 加えて¹⁸¹⁹ (:ガ)¹⁵ 生ごみ分解用の微生物を (:ニ)³⁹ 入れてやると、2021 (:ガ)³⁹ さらさらした肥料になる²² ことが分かった。

図 2 人手評価用の問題の例

表 6 作業員 2 人の間の比較

a) 作業員 A と作業員 B の比較				b) 作業員 A と作業員 C の比較				c) 作業員 B と作業員 C の比較			
A \ B	t	o	x	A \ C	t	o	x	B \ C	t	o	x
t	113	7	25	t	102	2	41	t	144	3	72
o	31	145	57	o	28	114	92	o	12	111	87
x	74	57	917	x	65	21	962	x	39	24	948

表 8 50 記事に関する自動評価

	再現率	精度	F 値
50 記事全体を対象に分類	0.460	0.641	0.536
3 人一致のみ (作業員一致を正解)	0.482	0.684	0.565
3 人一致のみ (新聞記事を正解)	0.507	0.690	0.585
3 人一致以外 (新聞記事を正解)	0.337	0.492	0.400

表 9 分析結果: 主題化する場合 (t) の特徴

特徴	事例数	正解数
主題の遷移を捉える	43	32
顕現性が高いが言い直す	21	8
対比的な記述	6	2
特徴的な言い回し	4	2
その他	15	5

判断の一致率のほうが新聞との一致率よりも高くなるという傾向が見られた。作業員間の一致率は Kappa 値としては中程度の一致だが、新聞固有の言い回しや省略などの影響で作業員間の一致率のほうが新聞との一致率と比べて高くなったと考えられる。また、作業員 3 人がすべて同じラベルを付与した結果は 75.0%であり、このデータに対しては 4 回に 3 回の頻度で判断が一致していることがわかる。

次に、この作業員 3 人が一致した結果を正解データとみなし、3 節で導入した省略分類モデルを適用する。つまり、50 記事全体に出現する共参照連鎖の各談話要素を分類した場合と比較して、3 人が一致した箇所のみを評価することで評価結果にどのような影響があるのかを調査する。この際、3 人が一致した箇所については 3 人が一致した判断を正解とする場合と、その箇所でもどのように新聞記事で参照表現が出現しているかを正解とした場合も比較する。さらに、3 人が一致しない場合の結果とも比較する。

結果を表 8 に示す。この結果からまず 3 人が一致した部

分のみを解析した場合には作業員の判断と新聞記事を正解とした場合ともに全体を解いた場合と比較して F 値が向上することがわかる。一方、一致しなかった箇所のみを解いた場合は逆に結果が悪くなるということがわかる。つまり、人の判断が揺れる箇所は表 2 に示した単純な素性では捉えられないような現象が起きており、結果として自動分類の結果も相対的に悪くなったと考えられる。また、3 人が一致した箇所を問題とした場合、作業員が一致して判断したラベルよりも新聞記事を正解とした場合のほうが F 値が高くなっている。これは、そもそも分類モデルを新聞記事にアノテーションされた関係に基づいて学習しているため、人の判断よりも新聞の出現傾向を反映して、このような違いがでたと考えられる。

5.3 作業員の判断が一致した事例の分析

最後に 3 人で作業した結果判断が一致した箇所につい

表 7 作業の一致率

	A-B	A-C	B-C	A-新聞	B-新聞	C-新聞	3人一致
t の一致数	113	101	142	89	120	117	89
o の一致数	145	112	109	107	99	92	99
x の一致数	917	959	944	971	952	1,025	881
一致率 (strict)	82.4%	82.5%	83.5%	81.8%	81.3%	85.6%	75.0%
一致率 (lenient)	85.1%	84.7%	84.6%	83.5%	82.8%	86.3%	77.2%
Kappa 値	0.605	0.571	0.650	0.535	0.548	0.608	-

て、どのような特徴に起因してそれぞれのラベルを選択したのかを事例を参照しながら分析した。まず3人がtに分類した事例、つまり「は/には」で主題化して記述すべきだと判断した結果を調査し、どのような特徴があるかを分類した結果を表9にまとめる*5。表9からわかるように、tの問題のほとんどは主題の遷移を助詞「は」などを手がかりに利用することで適切に捉えることで判断が可能である。例えば、例(2)*6では一度顕現性が高くなって省略されている「ボル・ポト派」 $\phi_{1,8}$ と分類対象となる $\phi_{1,10}$ の間に「中核部隊」が主題化されて挿入されるため、 $\phi_{1,10}$ の箇所ではあらためて記述される必要がある。

(2) カンボジアの反政府組織ボル・ポト派 $_{1,1}$ を非合法化する法律がきょう十五日、発効する。(中略)($\phi_{1,8}$ ガ) シアヌーク国王派やソン・サン派ともたもとをわかち、国民和解を模索する道から離脱した。中核部隊は依然、一枚岩と見られるが、ボル・ポト氏ら老幹部が実権を堅持する同派への国民の反発は根強く、「ボル・ポト派 $_{1,9}$ の歴史は終局に向かいつつある」との声が強い。($\phi_{1,10}$ ガ) インドシナ共産党の流れを受け、1963年発足したカンボジア共産党を前身とする武装組織。

一方で、例(3)の「米国 $_{1,2}$ は」のようにあらためて言い直さなければつながりかが悪い場合や例(4)の「ディーニ国庫相 $_{1,3}$ は」のように「議会の反対も～見通しだ」の1文で一度話題が切り替わっていることを捉えることは表層的な手がかりだけでは困難だと考えられる。一部のつながりの悪さは談話の話題転換に相当する箇所に関連しているため、談話の構造まで考えた生成の問題を考える必要がある。

(3) (前略)しかしその後、米国 $_{1,1}$ は中ソ対立をにらんで対アジア政策を転換する。米国 $_{1,2}$ は潜水艦発射弾道ミサイルの改善で長距離爆撃機に頼らずとも核の傘を維持できる体制も整えつつあった。

(4) ベルルスコーニ首相辞任で表面化したイタリアの政

*5 表9では各カテゴリに分類された事例数に加えて、それらがどのくらい分類モデルで正しく分類できたかの割合も示してあるが、そもそもこれらの特徴を捉えるための素性はほとんど導入できていないため参考程度に参照していただきたい。

*6 簡略化のため分析結果に関連する共参照・ゼロ照応関係のみを示す。以降の例でも同様の表記を採用する。

治危機解決のため、スカルファロ大統領は十三日午後、大統領府にディーニ国庫相 $_{1,1}$ を呼び、暫定政権の首相就任を要請し、これを受諾した。ディーニ国庫相 $_{1,2}$ はイタリア中央銀行の副総裁から、ベルルスコーニ政権の国庫相になった無党派のテクノクラート。議会の反対も少なく、来週には戦後イタリア五十四代目のディーニ政権が発足する見通しだ。ディーニ国庫相 $_{1,3}$ はフィレンツェ出身で、国際通貨基金イタリア代表の後、中央銀行副総裁など重要な経済・金融ポストを歴任。

次に、主題化せずに記述する場合についても同様の分析を行った。結果を表10に示す。ここでは最も重要となる考え方は「同一文内に他に主題化される要素がある場合にはそれ以外の要素は主題化されない傾向がある」というもので、この考え方に基づいて説明することで多くの事例を説明することができた。例えば、(5)では「党首選挙 $_{2,2}$ 」が主題化されているため、「新進党 $_{1,2}$ 」は主題化されず生成されることになる。

(5) 現職の党首は脱落し、羽田孜副党首と小沢一郎幹事長による事実上の一騎打ちに――。新進党 $_{1,1}$ の党首公選 $_{2,1}$ から目が離せなくなってきた。今回の党首選挙 $_{2,2}$ は、新進党 $_{1,2}$ が昨年十二月十日に結成されて以来、初めての公選である。

本研究では共参照連鎖内の情報を主に参照して省略すべきか否かを考えたが、この例からわかるように、ある共参照連鎖の中の談話要素は他の共参照連鎖内の談話要素と相互に依存して生成すべき参照表現は助詞を選択しているため、この点を踏まえた制約充足問題を解く必要がある。この問題については主題化すべきか否かの分類問題も同時に扱う必要があるため、今後の課題としたい。

最後に参照表現が省略されると3人が判断した事例を分析した。ただし、この場合は他の例と異なり文章中で非常に多く出現するため、3人が省略すると判断し、かつ分類器が省略すべきではないと分類した事例のみを分析対象とした。ガ格について分析した結果を表11に、ヲ格、ニ格について分析した結果を表12にまとめる。

まず、ガ格を省略すべき場合について頻出した現象は文章の主題となる表現、もしくはその主題と密接に関連する表現は頻繁に省略されるというものである。特に、例(6)

表 10 分析結果: 主題化しない場合 (o) の特徴

特徴	事例数	正解数
同一文内に他に主題化された名詞句が存在	12	8
連体節の中でかつ「連体節は」のボタンで出現	6	1
直前に主題化されているが「が」を使う	8	3
従属節内に出現	4	2
ヲ格/ニ格	47	25
引用の中	8	3
$e_{i,j-1}$ と $e_{i,j}$ の間に他の「は」が存在	5	2
その他	18	6

表 11 分析結果: 省略する場合 (x) の特徴 (ガ格)

特徴	事例数
文脈で自明な表現 (大域的テーマなど) は省略される	18
冗長な言い回し	8
省略しなくてもよい	3
引用の中	4
アノテーションのずれ	6
その他	9

の「日本」などの表現に代表されるように、その話題に関して前提とされる表現は特に省略されやすいため、文章の主題と関連した表現をあらかじめ列挙して、それを処理に組み込む必要があると考えられる。

(6) いまアダム・スミスが生きていたら、日本_{1,1}における手放しの「市場原理」礼賛に驚くに違いない。(中略) また、法務に関する分野で「法曹人口の大幅増員」と「外国人弁護士を受け入れに関する規制の緩和」を提言しているが、これでは ($\phi_{1,2}$ ガ) 米国のような訴訟国家になってしまう。

また、「(ϕ_i ニ) 国民_i の理解を求める」などの局所的に表現が重なるような冗長な言い回しについてもほとんどの場合で省略されるため、統語的な近さや特定のボタンでこのような場合を検出する必要がある。

一方、ヲ格、ニ格の場合は作業者が省略するといった事例はおおきく二通りに分類される。一つは Kameyama[8]などで議論された属性共有に関する現象、つまり並列表現などで同一の文法格を共有する場合には後者を省略するという場合である。もう一方は選択選好に適合する表現で、かつ前文脈で主題化されている場合に省略される傾向にあった。前者の属性共有の問題については、Iidaら[7]の手法のような文の統語的な特徴を明示的に分類に利用することで対応できると考えられる。後者の選択選好の適合度については典型的にその動詞の項として埋まりやすいという指標を開発し、その結果を用いた分類を行う必要があると考えられる。

6. おわりに

本稿では、新聞記事を対象とした参照表現の自動生成課題、特に参照表現を省略するか否かという問題を対象に、

表 12 分析結果: 省略する場合 (x) の特徴 (ヲ格、ニ格)

特徴	事例数 (ヲ格)	事例数 (ニ格)
特定のボタンで出現	50	21
主題性 + 選択選好の強さ	12	10
引用の中	9	4
その他	2	0

既存研究で利用されている談話の特徴を考慮した自動分類モデルを提案した。このモデルの性能を調査するために、センタリング理論に基づくヒューリスティックな手法と比較、提案モデルの有効性を示した。新聞記事にアノテーションされた結果を正解として評価を行った結果、F 値で 0.550 という結果を得たが、問題を分析した結果新聞記事固有の省略傾向や言い回しが評価に影響していることがわかった。このため、あらためて人手による評価用データを作成し、複数作業者の判断が一致した箇所のみを対象に評価を行った。この結果、人手の判断が一致した箇所のみを問題とした場合の性能が一致しなかった箇所のみを利用した場合と比較して高くなることが示した。つまり、人の判断が一致するような箇所は談話などの特徴を反映して容易に分類可能である可能性がある。そこで、3人の作業者の判断が一致した箇所を手で分析し、自動省略の性能を向上させるためにどのような問題を扱わなければならないかを分析した。この結果、問題を省略するか否かという2値分類で扱うのではなく、主題化するか否かという特徴も考慮した3値分類の問題を扱い、かつ分類対象ではない共参照連鎖の談話要素の出現も考慮した制約充足問題を解く必要があることを明らかにした。

今後の課題として、上述の3値分類問題を解くモデルをどのように構築するかを考えると同時に、今回分析できなかった人の判断が揺れた箇所についても調査を行い、省略生成の実際の処理として何を問題とすることが妥当なのかについて検討したい。

謝辞

本研究は科研費若手研究 (A) 「談話解析技術に基づいた文章推敲支援」(課題番号: 23680014) の支援を受けた。また、人手作業によるデータ作成に関しては、東京工業大学 笹野遼平氏、国立情報学研究所 横野光氏にご協力いただいた。記して謝意を表す。

参考文献

- [1] J. Cohen. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, Vol. 20, No. 1, pp. 37-46, 1960.
- [2] R. Dale. *Generating referring expressions*. The MIT Press, 1990.
- [3] F. F. Van der Meulen, A. S. Meyer, and W. J. M. Levelt. Eye movements during the production of nouns and pronouns. *Memory & Cognition*, Vol. 29, No. 3, pp. 512-521, 2001.

- [4] B. J. Grosz, A. K. Joshi, and S. Weinstein. Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics*, Vol. 21, No. 2, pp. 203–226, 1995.
- [5] K. Hasida. GDA 日本語アノテーションマニュアル 草稿 第 0.74 版, 2005. <http://i-content.org/gda/tagman.html>.
- [6] R. Henschel, H. Cheng, and M. Poesio. Pronominalization revisited. In *Proceedings of the 18th International Conference on Computational Linguistics*, pp. 306–312, 2000.
- [7] R. Iida, K. Inui, and Y. Matsumoto. Exploiting syntactic patterns as clues in zero-anaphora resolution. In *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics*, pp. 625–632, 2006.
- [8] M. Kameyama. A property-sharing constraint in centering. In *Proceedings of the 24th annual meeting on Association for Computational Linguistics*, pp. 200–206, 1986.
- [9] R. Kibble and R. Power. An integrated framework for text planning and pronominalisation. In *Proceedings of the First International Conference on Natural Language Generation*, pp. 77–84, 2000.
- [10] E. Kraehmer and M. Theune. Efficient context-sensitive generation of referring expressions. In K. van Deemter and R. Kibble, editors, *Information Sharing: Givenness and Newness in Language Processing*. CSLI Publications, 2002.
- [11] E. Kraehmer and K. van Deemter. Computational generation of referring expressions: A survey. *Computational Linguistics*, Vol. 38, No. 1, pp. 173–218, 2012.
- [12] M. Fujiwara M. Yamura-Takei and T. Aizawa. Centering as an anaphora generation algorithm: A language learning aid perspective. In *Proceedings of the 6th Natural Language Processing Pacific Rim Symposium*, pp. 577–562, 2001.
- [13] K. McCoy and M. Strube. Generating anaphoric expressions: Pronoun or definite description? *Proceedings of ACL’ 99 Workshop: Reference and discourse structure*, pp. 63–71, 1999.
- [14] S. Nariyama. Grammar for ellipsis resolution in japanese. In *In Proceedings of the 9th International Conference on Theoretical and Methodological Issues in Machine Translation*, pp. 135–145, 2002.
- [15] J. R. Quinlan. *C4.5: Programs for Machine Learning*. Morgan Kaufmann, 1993.
- [16] J-E. Roh and J-H. Lee. Building a pronominalization model by feature selection and machine learning. In *In Processing of 1st IJCNLP*, pp. 566–575, 2004.
- [17] P. Spanger, Y. Masaaki, R. Iida, and T. Takenobu. Using extra linguistic information for generating demonstrative pronouns in a situated collaboration task. In *Proceedings of Workshop on Production of Referring Expressions: Bridging the gap between computational and empirical approaches to reference*, 2009.
- [18] T. Winograd. *Understanding natural language*. Academic Press, 1972.
- [19] 滑川裕樹, 乾健太郎, 徳永健伸, 田中穂積. 照応・省略を含む日本語論説文生成. 言語処理学会第 5 回年次大会発表論文集, pp. 153–156, 1999.
- [20] 吉見毅彦. 英日機械翻訳における代名詞翻訳の改良. 自然言語処理, Vol. 8, No. 3, pp. 87–106, 2001.
- [21] 宮東城, 永田守男. 形態素を利用した自然な和文訳のためのゼロ代名詞生成規則. 電子情報通信学会技術研究報告 (言語理解とコミュニケーション) NLC2000-5, pp. 33–40, 2000.
- [22] 橋本さち恵, 乾健太郎, 白井清昭, 徳永健伸, 田中穂積. 日本語文生成における照応表現の選択. 情報処理学会自然言語処理研究会 (NL-143-5), pp. 33–40, 2001.
- [23] 黒橋禎夫, 居蔵由衣子, 坂口昌子. 形態素・構文タグ付きコーパス作成の作業基準 version 1.8. http://nlp.ist.i.kyoto-u.ac.jp/nl-resource/corpus/KyotoCorpus4.0/doc/syn_guideline.pdf.
- [24] 飯田龍, 小町守, 井之上直也, 乾健太郎, 松本裕裕. 述語項構造と照応関係のアノテーション: Naist テキストコーパス構築の経験から. 自然言語処理, Vol. 17, No. 2, pp. 25–50, 2010.