世界史用語集の語釈文における 見出し語に照応するゼロ代名詞とその表層格の推定

大矢 康介[†]・阪本浩太郎[†]・渋谷 英潔^{††}・森 辰則^{†††}

本稿では、世界史に関する大学入試論述問題に対して自動要約手法に基づき解答を自動生成する際の知識源の一つとして世界史用語集に注目し、見出し語と語釈部に分かれている文書データから解答となる文章を作成するために、語釈文における見出し語に照応するゼロ代名詞とその表層格を推定する手法を提案する。本稿の扱うタスクは、先行詞候補が見出し語に限られている一方でそれに照応するゼロ代名詞を複数の候補から一つ選ぶという点、および先行詞である見出し語が文中に存在しないため、照応解析において有効な手掛かりとなる先行詞の文脈情報が全く使えないという点で、従来のゼロ代名詞照応解析とは異なる。世界史用語集を対象とした評価実験を行った結果、KNPを用いた既存のゼロ照応解析を使用した手法に比べ、提案手法が有効であることが確認された。さらに、出現頻度の低い表層格で埋め込まれる場合の精度低下が観察されたため、通常の文から擬似訓練事例を生成する手法を検討した。同事例を使用した結果、ヲ格、ニ格の推定のF値を改善できることが確認された。

キーワード：ゼロ代名詞照応解析, 擬似訓練事例, 表層格推定, 照応解析, 世界史用語集

A New Method to Find Zero Pronouns Referring to Entry Words and Estimate their Surface Cases in the Context of a World History Glossary and its Descriptions

KOSUKE OYA[†], KOTARO SAKAMOTO[†], HIDEYUKI SHIBUKI^{††} and TATSUNORI MORI^{†††}

In this paper, we focus on the usage of a world history glossary as one of the knowledge sources for automated answer generation of essay-type questions. The questions we use were derived the University of Tokyo's entrance examinations on world history, and the answer generation uses a multi-document summarization methodology. In the automated answer generation, the glossary's descriptions we used as part of the answers. However, entry words were often omitted from the descriptions. To make complete sentences from entry words and their description, we propose a method to

[†] 横浜国立大学大学院環境情報学府, Graduate School of Environment and Information Sciences, Yokohama National University

^{††} 国立情報学研究所, National Institute of Informatics

^{†††} 横浜国立大学大学院環境情報研究院, Graduate School of Environment and Information Sciences, Yokohama National University

find zero pronouns referring to entry words inside the descriptions and estimate their surface cases. This paper's task differs from conventional zero anaphora resolution in the following two ways. First, with this method the entry word is the only candidate for the antecedent, as opposed to having to select one zero pronoun among several zero pronouns. Second, context information of the antecedent, which may be a useful clue in anaphora resolution, does not exist for the entry words, because the entry word appears alone and is not embedded in a sentence. Evaluation results based on a world history glossary revealed that the proposed method would be more effective than the existing method using zero anaphora analysis with Kurohashi-Nagao Parser. Furthermore, we attempted a method to generate pseudo training data from ordinary sentences in a textbook because we observed low accuracy when the entry word was embedded with low-frequency surface cases. Additionally, the results demonstrated that the introduction of the data improves the estimation of "o"-case and "ni"-case in terms of F-measure.

Key Words: *Zero Anaphora Resolution, Pseudo Training Data, Surface Cases Estimation, Reference Resolution, World History Glossary*

1 はじめに

近年、文書情報に対する情報要求は複雑化、高度化しており、そのような要求を満たすアクセス技術として質問応答が注目されている。質問応答とは、利用者の自然言語による質問に対して情報源となる文書集合から解答そのものを抽出する技術であり、複雑高度な情報要求を自然言語で表現できる点に特徴がある。

しかしながら、従来の質問応答に関する研究では、「アメリカの大統領は誰ですか?」といった比較的簡単な形式の質問を扱うものが多く、質問の確信に至るまでの背景や経緯を複数文にわたって説明したりする現実世界の質問状況とは異なる場合がある。そのような現実世界における質問に対する質問応答を目的とした取り組みは、TREC の Live QA (Agichtein, Carmel, Harman, Pelleg, and Pinter 2015) や NTCIR の QA Lab (Shibuki, Sakamoto, Yoshinobu Kano, Ishioroshi, Itakura, Wang, Mori, and Kando 2014; Shibuki, Sakamoto, Ishioroshi, Fujita, Kano, Mitamura, Mori, and Kando 2016; Shibuki, Sakamoto, Ishioroshi, Kano, Mitamura, Mori, and Kando 2017), 「ロボットは東大に入れるか」プロジェクト (以下, 「東ロボプロジェクト」) (新井, 松崎 2012) などで盛んに行われている。

現実世界における質問のように、質問の背景を複数文にわたって説明する記述や、解答が複数文を含む文章となるような質問の例として、大学入試問題が挙げられる。大学入試問題には大学入試センター試験と二次試験があり、二次試験の世界史分野には数十字から数百字以内で解答を記述する論述問題が含まれていることがある。QA Lab では、世界史の大学入試問題を対象としており、特に二次試験の論述問題への自動解答が挑戦的な課題として設定されている。

先行研究(阪本, 中山, 渋谷, 石下, 森, 神門 2016)では, 情報要求の存在する抽出型の複数文書要約としてこの課題を位置づけ, 教科書や用語集等の知識源から句点区切りの単位でテキストを抽出・整列して論述問題に解答する質問応答システムを提案している。知識源に使用される用語集は見出し語と語釈部に分かれて構成されており, 語釈部には見出し語が明示されていないため, 語釈部を句点で区切った文(以下, 「語釈文」という)だけをそのまま解答文に含めると, 何について述べているかわからない文になってしまう。また, 論述問題において解答に含めなければならない指定語句が見出し語となっている場合, 語釈文だけから解答を構成すると大きく減点されてしまう。

このような背景から, 上述の既存システムでは, 用語集の語釈文を解答の材料として抽出した際には, 見出し語を文頭に主題として付け加えた文を生成し, これを解答の一部とする手法を提案している。しかしながら, この手法によって生成された文は, 文法的に誤りがある場合や, 解答文に適していない場合がある¹。

これらの問題を解消するためには,

- (i) 見出し語を語釈文に埋め込むことができるか, 否か, すなわち, 語釈文の述語の省略された項をゼロ代名詞とみなした場合, 見出し語がその先行詞となるか, 否かを判定する。見出し語を埋め込むことができる, すなわち, 見出し語が先行詞となるのであれば, 見出し語の表層格を推定する。
- (ii) 問題文ならびに論述文章の前後の文等から何を主題にするかを決定し, それに応じて格交替などを行い論述問題の解答の一部とする。

ことが必要である。本稿では, (i)に掲げた課題を解決するために, 語釈文中の各動詞に着目し, それが見出し語に照応するゼロ代名詞を持つか否かを判定するとともに, 持つ場合にはその表層格を推定する手法を検討する。(ii)については, 今後の課題とする。

また, 提案手法は, 教師あり学習に基づく手法となっているため, 訓練事例を必要とするが, 用語集の形式をした事例は数に限りがあり, 特定の表層格で埋め込む場合の事例が限られていることが観察された。そのため, 訓練事例数が少ないという問題を解消するために, 擬似訓練事例の自動生成を行う。

本稿の以降の章では次の内容を述べる。2章では既存の世界史論述解答システムの概要とその問題点を述べ, 本研究で提案する解決策を述べる。3章では世界史用語集に関して予備調査を行った結果について述べる。4章では本研究の関連研究について述べる。5章では見出し語に照応するゼロ代名詞とその表層格を推定する手法を提案する。6章では実験結果を報告し, 7章で考察, 8章でまとめとする。

¹ 2.2 節に後述する。

2 世界史論述問題解答システム

2.1 大学入試における世界史論述問題の構成

例えば、東京大学入試二次試験における世界史は3問の大問から成っており、大問毎に、大論述、小論述、語句と解答の形式が異なっている。本稿では、問1の大論述問題を対象とする。問1の質問文は、図1のようにコンテキスト部と制約指定部から成っており、コンテキスト部を読ませ、制約指定部で指定される字数制約450字～600字、7個～9個の指定語句、「17世紀から18世紀末まで」のような時間制約などを満たすコンテキスト部に関連する解答が求められる。この問題に対する解答例(佐藤 2017)を図2に示す。

<p>1871年から73年にかけて、岩倉具視を特命全権大使とする日本政府の使節団は、合衆国とヨーロッパ諸国を歴訪し、アジアの海港都市に寄航しながら帰国した。その記録『米欧回覧実記』のうち、イギリスにあてられた巻は、「この連邦王国の……形勢、位置、広狭、および人口はほとんどわが邦と比較す。ゆえにこの国の人は、日本を東洋の英国と言う。しかれども営業力をもって論ずれば、隔たりもはなはだし」と述べている。その帰路、アジア各地の人々の状態をみた著者は、「ここに感慨すること少なからず」と記している。(引用は久米邦武『米欧回覧実記』による。現代的表記に改めた所もある。)</p>	<p>コン テ キ ス ト 部</p>
<p>世界の諸地域はこのころ重要な転機にあった。世界史が大きくなうねりをみせた1850年ころから70年代までの間に、日本をふくむ諸地域がどのようにパクス・ブリタニカに組み込まれ、また対抗したのかについて解答欄(イ)に18行(540字)以内で論述しなさい。その際に、以下の9つの語句を必ず一度は用い、その語句に下線を付しなさい。</p> <p>インド大反乱 クリミア戦争 江華島事件 総理衙門 第1回万国博覧会 日米修好通商条約 ビスマルク ミドハト憲法 綿花プランテーション</p>	

図1 東京大学 2008 年度入試二次試験における世界史の問1

イギリスは第1回万国博覧会を開催して圧倒的な国力を誇示し、ヨーロッパではクリミア戦争に介入してロシアの南下を阻止するなど外交の主導権を握ったが、これに対しロシアでは戦後農奴解放令を發布して近代化を模索し、ドイツはビスマルクのもとで国家統一を推進するとともに資本主義体制を確立していった。アメリカは南部の綿花プランテーション地域がイギリスの原料供給地として経済的に従属していたが、南北戦争後は北部主導のもとで急速な工業化を進めていった。オスマン帝国ではタンジマートにつけ込んだイギリスの経済的進出が強まったが、クリミア戦争後立憲政への要求が高まり、アジア最初の憲法であるミドハト憲法が發布された。イギリス産業の原料供給地・商品市場に転落していたインドではインド大反乱が起きたが、鎮圧後は直轄植民地となって従属が強まり、さらにヴィクトリア女王を皇帝とするインド帝国が成立した。アヘン戦争・アロー戦争に敗れた清は対等外交を強いられ総理衙門を設置したが、中体西用をスローガンとする洋務運動が展開されて近代化の動きが始まった。そして日本は日米修好通商条約によって世界の資本主義体制に組み込まれたが、明治維新によって国内の近代化をはかり、対外的には江華島事件起こして朝鮮へ進出を開始した。

図2 東京大学 2008 年度入試二次試験における世界史の問1に対する解答例

2.2 解答システムの事例とその問題点

Sakamoto et al. (Sakamoto, Ishioroshi, Matsui, Jin, Wada, Nakayama, Shibuki, Mori, and Kando 2016) では, 東京大学入試二次試験の世界史分野の指定語句ありの論述問題を対象とした質問応答システムの構築を検討している. 論述問題を情報要求の存在する抽出型の複数文書要約として位置づけ, 論述問題を解く際には, 知識源として教科書 4 冊と用語集 1 冊を利用している.

同研究で提案されている大学入試の論述問題における質問応答システムの概要を図 3 に示す. 同システムが論述問題を解く手順は次の通りである.

まず最初に, 質問文として東京大学入試二次試験世界史問 1 の問題文を入力する. 第一段階の「質問文解析」では, 質問文から, 問題文, 指定語句, 制限文字数, 指定された時間情報を抽出する. 第二段階の「関連文書の検索」では, 文書検索モジュールを用いて知識源から指定語句を含む文書を検索する. 第三段階の「文の抽出」では, 知識源から指定語句を含む文を句点区切りで抽出する. この際, 用語集から抽出した文に対しては文頭に見出し語を主題として埋め込む. 第四段階の「文の組の生成」では, 同じ指定語句を含む文をまとめて指定語句ごとに

図 3 世界史論述問題解答システムの概要

組を作成する。第五段階の「文の順位付け」では、解候補としての相応しさを表すスコアを文に付与し、そのスコアを基に文の組の中で順位付けを行う。スコアが等しい場合は文の文字数が多い順に順位づける。第六段階の「解候補生成」では、それぞれの文の組の上位 5 文から 1 文ずつ選択し組み合わせることで複数の解候補を生成する。第七段階の「文の並べ替え」では、解候補中の文を文の時間情報を基に古い順に並べる。第八段階の「解候補の削減」では、字数制約に反する解候補は削除する。もし字数制約を満たす解候補が存在しない場合は最小字数の解候補のみ残す。第九段階の「解候補順位付け」では、それぞれの解候補に解答らしさのスコアをつけて、スコアを基に順位付けを行う。解候補の解答らしさのスコアは解候補に含まれる文のスコアの合計で計測する。第十段階の「解候補選択」では、解答らしさのスコアが最大の解候補を選択する。以上のように、この研究では指定語句で検索して、知識源の中から解答と思われる記述を文単位で抽出している。

知識源のうちの 1 つである世界史用語集は表 1 のように見出し語と語釈部に分かれて構成されている。上記第三段落において、論述問題に解答する際に、この語釈部を句点で区切った文だけをそのまま解候補の一部に含めてしまうと、何について述べているかわからない文になってしまう。また、論述問題において解答に含めなければならない指定語句が見出し語となっている場合、見出し語を含まない語釈文だけから解答を構成すると大きく減点されてしまう。このため上述のシステムでは、用語集の語釈文を解答の材料として抽出した際には、助詞「は」を添えた見出し語を文頭に付け加えた文を生成する手法を提案している。図 4 の下線部はそれぞれ用語集から語釈文を抽出し、この手法を適用した文である。

この手法は、図 4 下線部 (1) のようにコピュラ文となる場合等には文法的に問題ないが、語釈文によっては図 4 下線部 (2) のように文法的に誤りのある文が生成され得るという問題点がある。また、この手法によって得た文の主題は図 4 下線部 (3) のようにならず見出し語になるため、解答文に適していない場合がある。例えば図 1 に示す 2008 年度の問 1 は、各国の動き

表 1 用語集の見出し語と語釈部の例

見出し語	語釈部
青年トルコ	ミドハト憲法の復活による立憲体制の確立をめざして、青年将校や知識人によって形成された諸組織の総称。1908 年青年トルコ革命をおこし、政権を担当したが、しだいに独裁化して、第一次世界大戦後に消滅した。
クリミア戦争	イェルサレムの聖地管理権をロシアが要求してオスマン帝国と開戦した戦争。東地中海へのロシアの南下を阻止するためイギリスとフランスなどが、1854 年オスマン帝国側について参戦した。ロシアは敗れ、56 年のパリ講和条約でロシアの南下政策がくじかれたため、ナポレオン 3 世の人气が上昇した。

を述べよという問題であり、図2に示す模範解答²ではイギリスを主題として文章が展開されているが、上記手法による解答では、国ではない「クリミア戦争」という出来事が主題となっており、解答に適している文であるとは言えない。

2.3 本研究で提案する解決策

2.2節で述べたような問題を解消するためには、

- (i) 見出し語を語釈文に埋め込むことができるか、否かを判定する。埋め込めるのであれば、見出し語の表層格を推定する。
- (ii) 問題文ならびに論述文章の前後の文等から何を主題にするかを決定し、それに応じて格交替などを行い論述問題の解答の一部とする。

ことが必要である。これらを行うことにより、例えば図4下線部(2)のような文を図5のように解答に適した文にすることができる。具体的には、図3の第三段階の「文の抽出」により語釈文を抽出した後、(i)を行い、見出し語である「クリミア戦争」が「参戦する」の「二格」に埋め込むことができることを推定する。さらに図3の第七段階の「文の並び替え」を行った後に(ii)を行い、問題文から国名「イギリスとフランスなど」を主題にすることを決定し、助詞「は」を添えて文頭に置く。

(1)青年トルコは、ミドハト憲法の復活による立憲体制の確立をめざして、青年将校や知識人によって結成された諸組織の総称。プロイセンの国家と軍事の力を背景にして進められたドイツの統一と、その後ビスマルクのもとでとられた国民統合の政策は、その一つの極を示すものであった。清朝は、外国使節の北京駐在を許し、外務事務をあつかう総理各国事務衙門(総理衙門)を設置して、対等な外国の存在を認める外交をはじめた。鎖国の方針を守る朝鮮(李朝)に対して、日本は開国を求めて江華島事件をおこし、不平等な日朝修好条規(江華条約)をおしつけた。(3)クリミア戦争は、東地中海へのロシアの南下を阻止するためイギリスとフランスなどが、1845年オスマン帝国側について参戦した。

(2)クリミア戦争は、ロシアは敗れ、56年のパリ講和条約でロシアの南下政策がくじかれたため、ナポレオン3世の人气が上昇した。この反乱には、旧支配層や旧地主層、農民、手工業者など、イギリスの支配によって没落した多くの層の人々が参加し、ムガル皇帝を盟主とする巨大な反英反乱となった(インド大反乱)。

図4 東京大学 2008 年度入試二次試験における世界史の問1に対する、先行研究のシステムによる解答

(2)イギリスとフランスなどは、東地中海へのロシアの南下を阻止するため、1854年オスマン帝国側について(1)クリミア戦争に参戦した。

図5 本研究の提案する解決策により期待される文例

² <https://akahon.net/>

ビスマルクは、外交ではフランスを孤立させた。

図 6 先行研究により出力されたものを簡略化した、格交替が必要な文例

フランスは、ビスマルクにより外交で孤立させられた。

図 7 本研究の提案する解決策により期待される文例 2

次に、格交替が必要な例を挙げる。図 6 の文例は、図 1 の問題に対して、先行研究のシステムから出力された解候補中の一文を簡略化したものである。この文に関しても、国名「フランス」を主題にすることが望ましいが、この場合、「ビスマルク」の格が何であるのかを推定する必要がある。さらに、「フランス」はヲ格の項となっているためそのまま主題にすると、不自然な文になってしまう。この例では、見出し語である「ビスマルク」が「孤立させる」の「ガ格」に埋め込むことができることを推定した後、「フランス」を主題にすることを決定し、「孤立させる」を「孤立させられる」のように受動態に書き換え格交替をすることが望ましい（図 7）。

1 章でも述べたように、本稿では (i) に掲げた課題を解決するために、語釈文中の各動詞に着目し、それが見出し語に照応するゼロ代名詞を持つか否かを判定するとともに、持つ場合にはその表層格を推定する手法を検討する。(ii) については、今後の課題とする。

3 世界史用語集の見出し語と語釈文の関係に関する予備調査

3.1 調査の目的と方法

2.3 節で提案した解決策を適用する必要がある文が世界史用語集の語釈部にどの程度存在するかを明らかにし、具体的なタスクの定義をするために、株式会社山川出版社・世界史 B 用語集改訂版 (全国歴史教育研究協議会 2008) を対象に、見出し語と語釈文との間の照応に関する以下の調査を行った。分析対象には用語集の見出し語 7,037 語の中から 70 語間隔で抽出した見出し語 100 語とその語釈部に現れる全文を用いた。また、以下の調査は全て単文単位で行う。すなわち語釈文が複文、重文になっている場合には、述語単位に切り分けたものに対して各調査における分類を行う。

調査 1

見出し語と語釈文中の単文の関係を「1. 単文においてゼロ代名詞化された格要素の一つが、見出し語に照応する」、「2. 単文が、主題がゼロ代名詞化されたコピュラ文で、ゼロ代名詞が見出し語に照応する」、「3. 単文中に見出し語そのものが現れている」、「4. 単文に現れる代名詞が見出し語に照応する」、「5. 1~4 のいずれでもない」のいずれかに人手で分類し、出現数を調

べる。それぞれに分類される語釈文中の単文および見出し語の例を表2に示す。分類1の例では、「行われた」の「ガ格」の格要素がゼロ代名詞化しており、見出し語の「聖職売買」に照応している。分類2の例では、「エピクロス＝サモス島出身の哲学者」の関係が成り立っており、コピュラ文である。この場合先行研究の手法を適用できる。分類3の例では見出し語「四書」が語釈文に含まれている。分類4の例では代名詞「この（危機）」が見出し語「金融恐慌」に照応している。分類5の例は分類1～分類4に該当せず、単文中に見出し語を埋め込むことはできない。

調査2

表3の分類1にはゼロ代名詞化された見出し語に照応する格要素が「動詞」「形容詞・形容動詞」「名詞」のうちどの品詞に関わる格要素であるかという観点で人手で分類を行い、その出現数について調べる。ここで用言のとり格は「が・を・に・で・と・から・へ・まで・より」格であり、名詞のとり格は「の」格のみである。

調査3

動詞のいずれかの格要素になっている場合の表層格毎の出現数についても人手で調べる。

調査4

動詞のいずれかの格要素になっている場合について、表層格毎、節の種類毎の分布についてNPを用いて調べる。

表2 語釈文中の単文と見出し語の関係の分類毎の例

分類	見出し語	語釈文中の単文
1.	聖職売買	9～10世紀にさかんに行われた。
2.	エピクロス	サモス島出身の哲学者。
3.	四書	五経よりも四書を高く評価した。
4.	金融恐慌	この危機のなか、対中国不干渉の若槻内閣が倒れ、
5.	メディチ家	その結果、フィレンツェはルネサンスの中心となった。

表3 語釈文中の単文と見出し語との間の関係の分類ならびにその出現数（調査1の結果）

分類	出現数
1. 単文においてゼロ代名詞化された格要素の一つが、見出し語に照応する	207
2. 単文が、主題がゼロ代名詞化されたコピュラ文で、ゼロ代名詞が見出し語に照応する	89
3. 単文中に見出し語そのものが現れている	6
4. 単文に現れる代名詞が見出し語に照応する	3
5. 1～4のいずれでもない	156

3.2 調査結果および本研究のタスク定義

調査1の結果を表3に示す。「3. 単文中に見出し語そのものが現れている」「4. 単文に現れる代名詞が見出し語に照応する」場合は非常に少なく、「1. 単文においてゼロ代名詞化された格要素の一つが、見出し語に照応する」場合が多いことがわかった。

調査2の結果を表4に示す。ゼロ代名詞化された格要素が見出し語に照応する場合、動詞の格要素になっている場合が非常に多いが、名詞の格要素になっている事例、形容詞・動詞の格要素になっている事例も見られた。

調査3の結果を表5に示す。動詞の格要素になっている場合（表4のA）、表層格は「ガ格」「ヲ格」「ニ格」「デ格」「ト格」「カラ格」となる場合があった。表5を見ると「ガ格」の事例数が最も多く、「ヲ格」「ニ格」「デ格」の事例数は少なく、「ト格」「カラ格」の事例数は非常に少ない。用語集の語釈文に対して先行研究の手法³を適用した時に意味がわかりにくくなってしまふ事例は特に、見出し語が「ヲ格」「ニ格」「デ格」など「ガ格」以外の格に埋め込むことができる場合である。例えば、図8は「聖職売買」が「行われた」の「ガ格」に埋め込める場合であるが、この場合先行研究の手法においても、生成された文単体としては、適格な文となっている⁴。しかし、図9のように「クリミア戦争」が「参戦した」の「ニ格」に埋め込める場合は、先行研究の手法を適用すると文単体として意味がわかりにくい文になってしまう。そのため「ヲ格」「ニ格」「デ格」に埋め込めることを正しく推定することは重要であると言える。

調査4の結果を表6に示す。この表を見ると、見出し語が動詞の格要素になっている場合に

表4 語釈文中の単文においてゼロ代名詞化された格要素と見出し語との間の分類ならびにその出現数（調査2の結果）

分類	出現数
A. 動詞のいずれかの格要素になっている	189
B. 形容詞・形容動詞のガ格になっている	7
C. 名詞のノ格になっている	11

表5 動詞の格要素になっている場合の表層格毎の出現数（調査3の結果）

表層格	ガ格	ヲ格	ニ格	デ格	ト格	カラ格
出現数	165	5	4	13	1	1

聖職売買は、教会が腐敗した9～10世紀にさかんに行われた。

図8 見出し語が下線部動詞の「ガ格」に埋め込める時に Sakamoto et al. (2016) の手法を適用した場合

³ 「見出し語+は、」を文頭に付け加える。

⁴ ただし、格の推定を行っていないので、主題の変更などは行えない。

クリミア戦争は、東地中海へのロシアの南下を阻止するためイギリスとフランスなどが、1854年オスマン帝国側について参戦した。

図 9 見出し語が下線部動詞「二格」に埋め込める時に Sakamoto et al. (2016) の手法を適用した場合

表 6 動詞の格要素になっている場合の表層格毎、節の種類毎の出現数（調査 4 の結果）

節の種類	主節	連用節	連体節
ガ格	78	81	6
ヲ格	2	1	1
ニ格	0	4	0
デ格	7	6	0
ト格	0	1	0
カラ格	0	1	0

その動詞が連体節の一部であるという事例は極めて少ない。

これらの調査結果をふまえ、本研究のタスク定義を以下に述べる。本研究では、語釈文中の各動詞について、ゼロ代名詞化された格要素の一つが、見出し語に照応しているか否か、すなわち、表 4 の A であるか否かを判定し、表 4 の A である場合はさらにその表層格を推定する。埋め込める場合の表層格は、「ガ格」「ヲ格」「ニ格」「デ格」「ト格」「カラ格」となるケースがあったが、「ト格」「カラ格」となるケースは非常に少なかった。そのため本研究では「埋め込めない」「ガ格」「ヲ格」「ニ格」「デ格」「それ以外の格」のうちのどれであるかを推定する。また、表 3 の分類 2 に属すコピュラ文については、体言止めの文でありその体言がサ変名詞以外であるか否か、という簡単な条件で判定できる⁵。その場合は、助詞「は」を添えて見出し語を文頭に置くだけで良いので、本研究の扱うタスクの対象としない。よって本研究の扱うタスクは、用語集の見出し語と語釈文を入力とし、語釈文中の各動詞句⁶に対して「埋め込めない」、「ガ格」、「デ格」、「ヲ格」、「ニ格」、「それ以外の格」のいずれかのラベルを出力するという多クラス分類問題とする。その際、対象となる動詞句が受身形、および使役形の場合であっても原型に戻さず、それぞれの態のままで表層格の推定を行う。入力と出力の例を図 10 に示す。

4 関連研究

日本語では格要素の省略が頻繁に起きることから、日本語ゼロ代名詞照応解析に関する研究

⁵ 例えばコピュラ文は名詞＋判定詞（だ）で終わるが、世界史用語集においては判定詞（だ）が省略されている。例、見出し語）シュメール人、語釈文）メソポタミア南部で最古の文明を築いた民族。また、サ変名詞の場合はコピュラ文にならない。例えば、見出し語）ボンディシェリ、語釈文）1672～74 年フランスが獲得。

⁶ KNP の解析の結果、＜ 用言:動 ＞ タグがふられた句を動詞句としている。

<p>【入力】 見出し語) 聖職売買 語釈文) 教会が腐敗した9~10世紀にさかんに行われた。1075年, 教皇グレゴリウス7世が禁止したが, 以後もつづいた。</p> <p>【出力】</p> <ol style="list-style-type: none"> 1. 腐敗する => 埋め込めない 2. 行われる => ガ格 3. 禁止する => フ格 4. 続く => ガ格

図 10 本研究が扱うタスクにおける入出力例

が多く行われている。大規模格フレーム (河原, 黒橋 2005, 2006) を使用した手法としては, 文内・文間両方のゼロ照応を対象とし, 構文の手掛かりおよび大規模格フレームによる語彙の手掛かりを素性とした対数線形モデル (Sasano and Kurohashi 2011) や, 格フレームの汎化を利用したニューラルネットゼロ照応解析モデル (山城, 西川, 徳永 2018), 外界照応および著者・読者表現を考慮したゼロ照応解析 (Hangyo, Kawahara, and Kurohashi 2013) などがある。機械学習を用いた手法としては, 文内ゼロ照応を対象とし, 文の構造を素性として取り入れた手法 (飯田, 乾, 松本 2006) や, フィードフォワードニューラルネットワーク (FNN) とリカレントニューラルネットワーク (RNN) を組み合わせて使用した手法 (Matsubayashi and Inui 2017) などがある。

これらの関連研究はいずれも Web テキストや新聞記事などの文章中のゼロ代名詞照応解析を行っているのに対し, 本研究では世界史用語集を対象とし, 質問応答における解答の流暢性の向上を目的としている。一般的なゼロ代名詞照応解析のタスクでは, ある格にあるゼロ代名詞に注目したうえで, 先行する名詞句群から高々一つの候補を選択することを行っている。一方で, 本研究の扱うタスクでは先行詞を用語集の見出し語に限定した上で, それに照応できるゼロ代名詞 (とその格) を高々一つ選択することを行っている。さらに, 先行詞である見出し語は, 単独で表示され, 文中に存在するわけではないため, 照応解析において有効な手掛かりとなる先行詞の文脈情報がまったく使えないという点が特徴的である。

また, 擬似訓練事例を利用した研究としては, Q&A サイトの質問分類において, Q&A に対して擬似的な分類ラベルを付与したデータを利用して特徴表現の収集を行った研究 (大森, 森田, 泓田, 青江 2015) や, 単語分散表現において, 学習を高速化させるために擬似的に負例をいくつかサンプリングし損失を算出する手法である Negative Sampling (Goldberg and Levy 2014) などが挙げられる。ゼロ代名詞照応解析のタスクにおいて擬似訓練事例を利用した研究としては, 中国語ゼロ代名詞照応解析の研究 (Liu, Cui, Yin, Zhang, Wang, and Hu 2017) がある。この研究では文章中に 2 回以上出てくる名詞に着目し, より後の方に出現する名詞を取り除いた

ものを, ゼロ代名詞に見立てて擬似訓練事例としている。この研究では, 文章中に2回以上出現する名詞の一方を取り除いているのに対し, 本研究では1回しか出現していなくても, それが世界史固有表現である場合には取り除き, 見出し語としている点が異なる。

5 機械学習を用いた見出し語に照応するゼロ代名詞とその表層格の推定手法

5.1 全体の処理の流れ

本節では3.2節で述べたタスクを実現する提案手法の概要を述べる。照応解析には大きく規則ベースの手法と学習ベースの手法があるが, 照応関係を包括的に捉える規則を書き尽くすことは困難であると考えたため, 本研究では学習ベースの手法を用いることとした。図11に提案手法の処理の流れを示す。提案手法により見出し語に照応するゼロ代名詞の表層格の推定を行う手順は次の通りである。まず最初に, 見出し語とそれに対応する語釈文の組を入力とする。第一段階の「構文・格解析」ではKNP⁷により構文・格解析を行う。第二段階の「動詞句の抽出」では, 構文・格解析結果をもとに, <用言:動>タグがふられた句を抽出する。第三段階の「動詞句毎に素性を抽出」では, 抽出した動詞句それぞれに対して, KNPの構文・格解析結果, その際に使用された京大格フレーム(河原, 黒橋 2006), 日本語語彙大系(池原 1997)などの言語資源を用いて素性抽出を行う。第四段階の「動詞句毎に表層格を推定」では, 第三段階で抽出した素性を入力とし, Support Vector Machine を one-versus-the-rest 法により多クラス分類に拡張したモデルによりそれぞれの表層格を推定する。最後に, 第四段階で推定された動詞句毎の表層格を出力とする。

図 11 提案手法の概要

⁷ <http://nlp.ist.i.kyoto-u.ac.jp/?KNP>

5.2 使用する素性

飯田ら (2006) によると、ゼロ代名詞照応解析に使用する素性は一般的に以下の3種類に分けられる。

- 1) 対象となるゼロ代名詞を持つ述語の語彙, 統語情報に関する素性
- 2) 先行詞候補に関する語彙, 統語, 意味, 位置情報に関する素性
- 3) 対象となるゼロ代名詞を持つ述語と先行詞候補の対から抽出可能な情報に関する素性

語釈文におけるゼロ代名詞照応の場合, 先行詞である見出し語に文脈がないため, これらのうち2)に含まれる先行詞候補に関する統語, 位置情報に関する素性を利用することはできない。本研究では1)に分類される素性として「節の種類」, 「態」, 「格の埋まりやすさ」, 「すでに埋まっている格」, 「動詞の直前に現れやすい格 (直前格)」, 「複合格助詞」, 2)に分類される素性として「見出し語の意味カテゴリ」, 3)に分類される素性として「意味クラス PMI」, 「意味的な類似度」の合計6種類の素性を使用する。「格の埋まりやすさ」, 「意味クラス PMI」, 「意味的な類似度」については, KNP の解析時に使用された格フレームをもとに, 「ガ格」「デ格」「ヲ格」「ニ格」それぞれに設定する。上記の素性は, いずれも先行研究で使われている素性である。

また, 上記の素性では, 格助詞を付与した見出し語を語釈文に埋め込んで作られる文 (以下, 「見出し語を埋め込んだ文」という。) 全体の適格性を考慮せず, 埋め込む前のそのままの語釈文を解析した際に適用された格フレームを素性抽出に利用している。そのため, 本研究で新たに使用する素性として, 埋め込み後の適格性を判断することを目的として, 見出し語を埋め込んだ文を考慮した「KNP スコア差分」と「再選定格フレーム素性」の2つの素性を提案する。

5.2.1 ベースライン素性

ゼロ代名詞照応解析における先行研究で活用されてきた以下の素性をベースラインとする。

節の種類

ゼロ代名詞を持つ述語の統語的選好を考慮するため, 節の種類を素性として使用する (Hangyo et al. (2013), 3 節)。節の種類とは, 注目する述語がどの種類の節に含まれているかであり, 「主節」「連用節」「連体節」のいずれかになる。KNP の解析結果から得られる。

態

KNP の解析結果から得られる, 動詞の態を使用する (Hangyo et al. (2013), 3 節)。「能動形」「受身形」「使役形」のいずれかであり, KNP の解析結果から素性抽出を行う。

格の埋まりやすさ

格フレームのそれぞれの格の埋まりやすさは, その格が明示されていない場合に, ゼロ代名詞として省略されているのか, 単にその格が考慮されていないのかの判断の手掛かりになると考えられる。そこで, 格の埋まりやすさの素性として格スロット生成確率を用いる (笹野ら (2011), 5 節)。格フレーム CF_i の格スロット s_j に格要素が対応付けられている確率 $P(A(s_j) = 1 | CF_i, s_j)$

は, CF_i の直前格スロットを s_k , 格スロット s_j の用例数の合計を n_j とすると,

$$P(A(s_j) = 1 | CF_i, s_j) = \frac{n_j}{n_k} \quad (1)$$

と計算する. ここで, $A(s_j)$ は格スロット s_j が埋まっている時 1 をとる関数である.

既に埋まっている格

すでに対応づけられている項がある格に見出し語を埋め込むことができる事例は極めて少ないと考えられる. それぞれの格について, すでにその格に対応づけられた, 直接係り受けを持つ項がある場合は 1, ない場合は 0 とする (Hangyo et al. (2013), 3 節). 格の対応づけの判定は KNP により行い, 格が明示されているか否かの区別はしていない.

用言の直前に現れやすい格である直前格

直前格はその用言の用法の決定に対して重要な役割を果たす (河原, 黒橋 2005) ため, 直前格が埋まっていない場合には見出し語を埋め込みやすいと考えられる. そのためそれぞれの格に関して直前格であれば 1, そうでなければ 0 とする素性を使用する.

複合格助詞

「～として」の「する」や「～によって」の「よる」などの動詞句は複合格助詞の一部であるが, これらの動詞の格要素に見出し語を埋め込む場合は極めて少ないと考えられる. そのため動詞句が複合格助詞の一部である場合は 1, そうでなければ 0 とする素性を使用する.

見出し語の意味カテゴリ

どのような意味カテゴリに属する見出し語に対する語釈文であるかによって, 語釈文中のある格での省略されやすさに偏りがあると考えられる. 先行研究では IREX で定義された 8 種類の固有表現を使用することが多いが, 本研究では世界史用の固有表現辞書 (石下, 阪本, 中山, 渋谷, 森, 神門 2016) の 17 種類の固有表現クラスのうちどのクラスに属するかをバイナリ素性として使用する. 世界史固有表現辞書を登録した MeCab により見出し語を形態素解析し, 固有表現クラスを獲得する. 表 7 に, 世界史固有表現辞書で定義されている 17 種類の固有表現クラスと, その固有表現クラスに属するサブクラスおよび語の例を示す. サブクラスについては「意味クラス PMI」や「意味的な類似度」の計算に用いる.

意味クラス PMI

ある格スロットにどのくらい先行詞候補が入りやすいかを考慮するため, 見出し語の意味クラスと対象の格スロットの意味クラス情報との間の自己相互情報量 (PMI) を素性として使用する (笹野ら (2011), 5 節). ここで, 意味クラスには日本語語彙大系 (池原 1997) において深さ 5 の位置にある意味属性を使用し, さらに深い位置の意味属性に属する語はその上位の深さ 5 の位置にある意味属性に含める. また, 世界史用語集の見出し語が日本語語彙大系にそのまま登録されていることは稀であるため, 見出し語の汎化を行う必要がある. 見出し語の汎化には世界史固有表現辞書のサブクラスにあたる語を使用する. ただし, 見出し語が世界史固有表現辞

表 7 世界史固有表現辞書で定義される固有表現クラスとその固有表現クラスに属するサブクラスおよび語の例

固有表現クラス	サブクラス	例
場所	国家, 地域, ...	タイ, チベット
文明	文化, 文明, ...	エジプト文化, ナスカ文明
出来事	文化運動, 紛争, ...	ルネサンス, 島原の乱
言語	言語, 語族, ...	ペルシア語, セム語族
技術	文字, 技術, ...	アラビア数字, 製紙法
時代	時代, 年代, ...	新石器時代, 戦国時代
人物	皇帝, 政治家, ...	アウラングゼブ, ケネディ
建造物	遺跡, 聖堂, ...	アンコール, ミラノ大聖堂
民族	民族, 人種, ...	アーリヤ人, ユダヤ人
作品	科学書, 物語, ...	医学典範, 金瓶梅
制度	条約, 法令, 社会制度, ...	ドイツ憲法, ミラノ勅令
組織	国際組織, 会議, ...	EC, アフリカ民族会議
社会概念	権利, 通貨, ...	タバコ利権, 独占販売権
宗教	宗教	イスラーム教, シク教
宗教概念	宗教概念, 神, ...	巡礼, 輪廻
様式	建築様式, 美術様式, ...	コリント式, ロマン主義
その他	動物, 土, ...	人類, 黄土

書に登録されていない場合には見出し語の末尾の名詞を使用する。対象の格フレーム CF_i の格スロット s_j の用例における見出し語と同じ意味クラス CL_k に属する語の占める割合を、一般のテキストにおける CL_k に属する語の占める割合で割ることにより計算する (式 (2))。

$$\log\{P(CL_k|CF_i, s_j)/P(CL_k)\} \quad (2)$$

格スロット s_j の用例に見出し語と同じ意味クラスに属する語が無い場合は一定の小さな値 $\log(0.000001)$ を与える。

意味的な類似度

ある格スロットにどのくらい先行詞候補が入りやすいかを考慮するため、世界史固有表現辞書により得られた見出し語のサブクラスにあたる語と、対象の格スロットの用例との意味的な類似度を日本語語彙大系により計算し (笹野ら (2005) 付録, 式 (1) を使用), 最も値の大きいものを素性として使用する。意味クラス PMI と異なり, 用例の出現頻度を考慮せず, 1 つでも見出し語と意味が類似している語があれば大きな値をとる。

5.2.2 本研究で提案する素性

本研究で新たに導入する素性を以下に示す。

表 8 KNP スコア差分算出例

見出し語を埋め込んだ文	KNP スコア	KNP スコア差分 (提案素性)
教会が腐敗した 9~10 世紀にさかんに <u>聖職売買</u> が行われた	-58.5	$-58.5 - (-43.2) = -15.3$
教会が腐敗した 9~10 世紀にさかんに <u>聖職売買</u> を行われた	-61.4	$-61.4 - (-43.2) = -18.2$
教会が腐敗した 9~10 世紀にさかんに <u>聖職売買</u> に行われた	-61.1	$-61.1 - (-43.2) = -17.9$
教会が腐敗した 9~10 世紀にさかんに <u>聖職売買</u> で行われた	-59.9	$-59.9 - (-43.2) = -16.7$

KNP スコア差分

見出し語を埋め込んだ文を KNP で解析した時のスコアから、元の語釈文を KNP で解析した時のスコアを引いたものを「KNP スコア差分」とする。KNP スコアは KNP による構文・格解析結果のもっともらしさであり、このスコアが高いほど日本語として出現しやすい文であると言える⁸。そのため、埋め込みの前後でのスコアの差は、埋め込みの良さを表すと考える。表 8 に KNP スコア差分の算出の例を示す。この例では、KNP スコア差分の値が最も大きい（絶対値が一番小さい）ガ格が適切であると判断される。

再選定格フレーム素性

見出し語を埋め込んだ文を KNP により解析し、その時に使用された格フレームを使用し、京大格フレームを使用して抽出する素性「格の埋まりやすさ」「意味クラス PMI」「意味的な類似度」「直前格」を再計算・再抽出したものを「再選定格フレーム」素性とする。

5.3 擬似訓練事例の利用

3.1 節で述べたとおり、「ヲ格」「ニ格」「デ格」に関する埋め込みの推定精度は重要である。しかし、これらの格について学習に用いることのできる訓練事例が非常に少ないため、上手く学習できないことが予想される。また、相対的に割合の低い事例の数を増やすためには、非常に多くの語釈文を正解事例として人手で解析する必要があるが、世界史の分野で見出し語と語釈部に分かれている文書データの数に限られており、正解事例を大量に増やすことは現実的ではない。

この問題を解消するため、比較的潤沢に入手可能な世界史教科書の記述から擬似訓練事例を自動生成し、これを活用することを提案する。生成手順としては、教科書中の文を句点区切りで分割し、分割されたテキストごとに以下の処理を行う。

- (1) 「を」「に」「で」のいずれかでマークされた世界史固有表現に注目し、2, 3 の処理を行う。
- (2) その世界史固有表現を文中から取り除き、見出し語とする。

⁸ 文の出現のしやすさの指標を算出するためには、厳密には、適用可能な全ての格フレームにおける出現確率の総和を取る必要がある。一方で、本論文では KNP の解析である最終的な 1 つの格フレームしか使用していないため、ここで計算された値は文の出現のしやすさの近似的な指標であると考えられる。

- (3) 直後の「を」「に」「で」のいずれかの格助詞を取り除き、埋め込みされる表層格の正解ラベルとする。

図 12 に擬似訓練事例の自動生成の例を示す。

この手法によって生成された擬似訓練事例は、大きく分けて 3 種類に分類される。各分類の定義は以下の通りである。分類 A は取り除きたい文節の係先が連体節となっておらず、取り除きたい文節にかかる文節が存在しない事例である。分類 B は取り除きたい文節の係先が連体節となっている事例で、分類 C は取り除きたい文節にかかる文節が存在する事例である。各々の例を図 13、図 14、図 15 に示す。ここで、着目する固有表現および格助詞で構成される文節を <> で囲み、着目した固有表現が 1 つの項となっている述語に下線を引いている。分類 C の文は、文節を取り除いた結果、係り受け関係が崩れてしまい元の文の意味を維持しなくなってしまうために排除すべき事例であると考えられる。分類 C に属さない文においても不自然な文となっている場合があるが、自動生成された文が自然であるか不自然であるかを機械的に正確に判別することは難しい。そのため、不自然な文を機械的に取り除く近似的な方法として分類 B を導入している。分類 B には多くの場合不自然な文が分類されるが、一部自然な文も含まれて

図 12 擬似訓練事例の自動生成の例

図 13 分類 A

図 14 分類 B

シーア派に対し、ムハンマドのスナを尊重し、正統カリフ、ウマイヤ朝、
 アッパース朝と続いた<カリフ>を正統な指導者と肯定する党派をスナ派
 という

図 15 分類 C

図 16 擬似訓練事例を自動で分類する手順

おり、一方で分類 A には多くの場合自然な文が分類されるが、一部不自然な文も含まれている。そこで本研究では自動生成された擬似訓練事例のうち、分類 C を使用せず、分類 A、B のみを使用する。また、自動生成された訓練事例が分類 A、B、C のどれに該当するかの分類については、KNP の解析結果を利用し図 16 の手順により自動で行う。文節の係り受け関係およびある文節が主節・連用節・連体節のうちどれであるかは KNP の解析結果よりわかる。

6 評価実験

見出し語が語釈文に埋め込めるか否かを判定し、埋め込める場合に見出し語の表層格を推定するという本稿の提案する解析手法の有効性を調べるために、ベースライン素性（5.2.1 節）のみを使用した手法と、従来型のゼロ照応解析に基づく手法による推定精度をそれぞれ求め比較検討する。また、本稿で提案する素性（5.2.2 節）の有効性を調べるために、ベースライン素性に提案素性も加えた手法の推定精度を求め、ベースライン素性のみの手法と比較検討する。さらに、擬似訓練事例の有効性を確認するために、擬似訓練事例の追加の前後における推定精度について比較検討を行う。

なお、従来型のゼロ照応解析に基づく手法としては、見出し語を 1 文目、語釈文を 2 文目と

した文章を KNP に入力し、得られた述語項構造解析の結果（これにはゼロ照応解析結果も含まれる）から見出し語を項としている結果のみを取り出して得られたものを出力とする、簡便なものを採用した。

6.1 使用するデータと実験設定

評価実験では、世界史用語集 (全国歴史教育研究協議会 2008) 中の見出し語 7,037 語から 3 節で使用した語を除いたもののうち、ランダムに抽出した 900 語とそれに対応する語釈部を使用する。これらの語釈文中のすべての動詞句に対し、見出し語がどの表層格に埋め込むことができるか、もしくは埋め込むことができないかの正解事例を、第一著者が作成したものをデータセットとして使用する。正解ラベル毎の事例数は表 9 の通りである。また、正解ラベル毎、節の種類毎の事例数は表 10 の通りである。全ての実験において、Support Vector Machine の実装には python の機械学習ライブラリである scikit-learn を使用した。カーネル関数には RBF を使用し、各パラメータの値はデフォルト値である。また、評価実験は 10 分割交差検定により行い、精度は F 値によって評価する。提案手法で用いる擬似訓練事例の生成には世界史 B 教科書 (東京書籍 2007) を使用する。この教科書から生成された分類毎の擬似訓練事例の数は表 11 に示す通りであった。なお、擬似訓練事例は訓練事例のみに含まれており評価事例には含まれていない。

表 9 見出し語 900 語文の語釈文における正解ラベル毎の事例数

正解ラベル	埋め込めない	ガ格	ヲ格	ニ格	デ格	その他の格
事例数	1399	1432	82	44	134	31

表 10 見出し語 900 語文の語釈文における正解ラベル毎、節の種類毎の事例数

節の種類	主節	連用節	連体節
埋め込めない	327	448	312
ガ格	688	666	21
ヲ格	50	22	3
ニ格	22	0	21
デ格	76	65	0
その他の格	22	3	1

表 11 擬似訓練事例の分類毎の数

	分類 A	分類 B	分類 C
ヲ格	201	114	234
ニ格	186	70	143
デ格	40	19	156

最後に、提案手法で得られた解析結果を論述問題解答作成手法に組み込んだ場合にどのような解答が作成されるかについて、事例を調査し、既存手法との比較を行う。実験に使用する論述問題を図 17 に示す。この実験をする際に、対象となる用語（「バグダード鉄道」「仇教運動」「ゴールドラッシュ」）については訓練事例から除いて学習した。

6.2 実験結果

6 節の冒頭で述べた、KNP の解析結果に基づく従来型のゼロ照応解析手法を「KNP ゼロ照応解析手法」、ベースライン素性（5.2.1 節）のみを使用した手法を「ベースライン素性手法」、ベースライン素性に加え、本稿で提案する素性（5.2.2 節）を使用した手法を「提案手法」として、実験結果を表 12 に示す。

また、使用する素性を 1 つずつ除くことにより素性の有効性を検証する実験を行った結果を

<p>私たちは、情報革命の時代に生きており、世界の一体化は、ますます急速に進行している。人や物がひんぱんに往きかうだけでなく、情報はほとんど瞬時に全世界へ伝えられる。この背後には、運輸・通信技術の飛躍的な進歩があると言えよう。</p> <p>歴史を振り返ると、運輸・通信手段の新展開が、大きな役割を果たした例は少なくない。特に、19世紀半ばから20世紀初頭にかけて、有線・無線の電信、電話、写真機、映画などの実用化がもたらされ、視聴覚メディアの革命も起こった。またこれらの技術革新は、欧米諸国がアジア・アフリカに侵略の手を伸ばしていく背景としても注目される。例えばロイター通信社は、世界の情報をイギリスに集め、大英帝国の海外発展を支えることになった。一方で、世界中で共有される情報や、交通手段の発展によって加速された人の移動は、各地の民族意識を刺激する要因ともなった。</p> <p>運輸・通信手段の発展が、アジア・アフリカの植民地化をうながし、各地の民族意識を高めたことについて、下記の9つの語句を必ず1回は用いながら、解答欄（イ）を用いて17行以内で論述しなさい。</p> <p>スエズ運河 汽船 バグダード鉄道 モールス信号 マルコーニ 義和団 日露戦争 イラン立憲革命 ガンディー</p>
--

図 17 東京大学 2003 年度入試二次試験における世界史の問 1

表 12 世界史用語集を用いた実験結果

	KNP ゼロ照応解析手法			ベースライン素性手法			提案手法		
	適合率	再現率	F 値	適合率	再現率	F 値	適合率	再現率	F 値
埋め込めない	0.443	0.945	0.603	0.754	0.821	0.786	0.760	0.820	0.789
ガ格	0.164	0.007	0.014	0.815	0.866	0.840	0.821	0.872	0.846
ヲ格	—	0.000	—	0.483	0.354	0.408	0.525	0.390	0.448
ニ格	—	0.000	—	—	0.000	—	—	0.000	—
デ格	0.030	0.007	0.011	0.286	0.014	0.026	0.176	0.021	0.037
その他の格	—	0.000	—	1.000	0.615	0.762	0.944	0.654	0.773
全体結果	0.430	0.430	0.430	0.778	0.778	0.778	0.782	0.782	0.782

表 13 に示す.

次に, 擬似訓練事例を訓練事例に追加して学習を行ったモデルにより, 評価実験を行った結果を表 14 に示す. また, 今回は事例数の少ない「ヲ格」「ニ格」「デ格」の精度も重視したいため, 全体結果を「Micro F」「Macro F」で評価する.

さらに擬似訓練事例を使用したモデルとの比較のために, クラス頻度に偏りがある学習において, その頻度を考慮した学習を行うオプションとして `class_weight="balanced"` を指定する. このパラメータ指定により, 訓練データのクラスの頻度に反比例する重みを, パラメータ C に乗じた状況で学習したモデルによる実験結果を表 15 に示す. なお, このモデルでは擬似訓練事例を使用していない.

最後に, 提案手法で得られた解析結果を論述問題解答作成手法に組み込んだ場合にどのような解答が作成されるのかについて, 事例を調査し, 既存手法との比較を行った. 既存手法を用い

表 13 素性を 1 つずつ除いた場合の F 値

使用しない素性	埋め込めない	ガ格	ヲ格	ニ格	デ格	その他の格	F 値
全ての素性を使用	0.789	0.846	0.448	—	0.037	0.778	0.782
節の種類	0.759	0.823	0.393	—	0.014	0.683	0.759
態	0.785	0.841	0.425	—	0.013	0.711	0.777
格の埋まりやすさ	0.787	0.843	0.426	—	0.025	0.714	0.779
既に埋まっている格	0.749	0.793	0.155	—	0.118	0.583	0.735
直前格	0.786	0.845	0.318	—	0.013	0.683	0.778
複合格助詞	0.789	0.846	0.448	—	0.037	0.773	0.782
見出し語の意味カテゴリ	0.761	0.818	0.433	—	—	0.732	0.759
意味クラス PMI	0.784	0.842	0.448	—	0.026	0.773	0.779
意味的な類似度	0.784	0.843	0.438	—	0.013	0.773	0.779
KNP スコア差分	0.788	0.846	0.414	0.045	0.013	0.773	0.782
再選定格フレーム素性	0.786	0.840	0.408	—	0.026	0.762	0.778

表 14 擬似訓練事例の分類ごとの実験結果

	擬似訓練事例なし	分類 A のみ			分類 A+B		
	F 値	適合率	再現率	F 値	適合率	再現率	F 値
埋め込めない	0.789	0.797	0.777	0.787	0.787	0.761	0.774
ガ格	0.846	0.810	0.869	0.839	0.815	0.869	0.842
ヲ格	0.448	0.392	0.622	0.481	0.351	0.634	0.452
ニ格	—	0.113	0.186	0.140	0.127	0.182	0.150
デ格	0.037	0.250	0.014	0.026	0.231	0.022	0.041
その他の格	0.773	0.944	0.654	0.773	1.000	0.548	0.708
Micro F	0.782	0.770			0.766		
Macro F	0.482	0.508			0.494		

表 15 SVM に class_weight=“balanced” を指定したモデルによる実験結果

	適合率	再現率	F 値
埋め込めない	0.778	0.785	0.781
ガ格	0.830	0.869	0.849
ヲ格	0.514	0.439	0.474
ニ格	0.188	0.070	0.102
デ格	0.297	0.229	0.259
その他の格	0.900	0.692	0.783
Micro F			0.777
Macro F			0.541

(1)バグダード鉄道は、トルコのコンスタントノポリスからバグダードをへて、ペルシア湾に通じる鉄道(計画)。日露戦争後、中国東北地方については、北はロシア、南は日本の勢力下とされ(1907年、第1回日露協約)、モンゴルをめぐっても両国の勢力範囲を確定した。義和団事件への出兵にあたって中国東北地方を占領したロシアは、その後も撤兵することなく、鴨緑江にまで進出した。(2)ペルシアは、イラン立憲革命の旧称。(3)北京議定書は、1901年9月7日に北京で清朝と11カ国(出兵8カ国とベルギー・オランダ・スペイン)との間で調印された義和団事件の講和に関する最終議定書。1869年のスエズ運河の開通は列強のアフリカへの進出を刺激し、アフリカは分割競争の対象となった。(4)バグダード鉄道は、諸列強がねらったが、ドイツが1899年に敷設権を獲得し、1903年に会社を設立した。(5)バグダード鉄道は、最終的には一部分だけの開通に留まった。(6)仇教運動は、とくに97年の山東におけるドイツ人宣教師殺害事件を口実とするドイツの進出は、98年から義和団による山東教案となって爆発し、大規模な反帝国主義・排外運動に発展した。(7)ゴールドラッシュは、鉄道や汽船による交通網の発展が背景にあり、以後、西部開拓が進展した。

図 18 東京大学 2003 年度入試二次試験における世界史の問 1 に対する、先行研究のシステムによる解答

(1)バグダード鉄道は、トルコのコンスタントノポリスからバグダードをへて、ペルシア湾に通じる鉄道(計画)。日露戦争後、中国東北地方については、北はロシア、南は日本の勢力下とされ(1907年、第1回日露協約)、モンゴルをめぐっても両国の勢力範囲を確定した。義和団事件への出兵にあたって中国東北地方を占領したロシアは、その後も撤兵することなく、鴨緑江にまで進出した。(2)ペルシアは、イラン立憲革命の旧称。(3)北京議定書は、1901年9月7日に北京で清朝と11カ国(出兵8カ国とベルギー・オランダ・スペイン)との間で調印された義和団事件の講和に関する最終議定書。1869年のスエズ運河の開通は列強のアフリカへの進出を刺激し、アフリカは分割競争の対象となった。(4)諸列強がバグダード鉄道をねらったが、ドイツが1899年に敷設権を獲得し、1903年に会社を設立した。(5)最終的には一部分だけの開通に留まった。(6)とくに97年の山東におけるドイツ人宣教師殺害事件を口実とするドイツの進出は、98年から義和団による山東教案となって仇教運動が爆発し、大規模な反帝国主義・排外運動に発展した。(7)鉄道や汽船による交通網の発展が背景にあり、以後、西部開拓が進展した。

図 19 東京大学 2003 年度入試二次試験における世界史の問 1 に対する、提案手法を組み込んだ場合の解答

て作成された解答を図 18 に、提案手法を用いて作成された解答を図 19 に示す。それぞれの図で、下線部は用語集から生成された文を示している。

7 考察

7.1 実験毎の考察

本節では各実験毎に考察を述べる。表 12 をみると、「埋め込めない」、ならびに、各格の推定の全てにおいて F 値の値が、KNP ゼロ照応解析手法、ベースライン素性手法、提案手法の順に向上していることが見て取れる。各手法の推定精度の間に統計的な有意差があるかを確認するために McNemar 検定を行った結果、KNP ゼロ照応解析手法とベースライン素性手法との比較、ベースライン素性手法と提案手法との比較においては、P 値がそれぞれ、 2.270×10^{-161} 、ならびに、0.179 となり、前者の間には有意差があるものの、後者の間には有意差があるとは言えないことがわかった。表 13 の、素性を 1 つずつ除いて実験を行った結果を見ると、「節の種類」「見出し語の意味カテゴリ」「埋まっているか」「再選定格フレーム素性」が有効であることがわかる。動詞句が連体節に含まれる場合、その動詞句のいずれかの格に見出し語を埋め込むことができる場合は少ない。そのため「埋め込めない」の推定の際には「節の種類」の素性が有効である。また、世界史用語集では、図 20 のように見出し語が場所を表す語のときには「ヲ格」に入りやすいなどの傾向があるため、「見出し語の意味カテゴリ」が有効であると考えられる。すでに対応づけられている項がある格に見出し語を埋め込むことができる事例は極めて少ないと考えられるため、「埋まっているか」の素性は非常に重要である。「意味クラス PMI」や「意味的な類似度」の素性はあまり有効ではなかったが、これには 2 つの原因が考えられる。1 つ目は 5.2 節でも述べたように、見出し語を埋め込んだ文を考慮せず格フレームを選定している点である。また 2 つ目は、日本語語彙大系を使用してこれらの素性を計算している点である。日本語語彙大系に登録されている語は京大格フレームの用例として出現している語と比較すると少なく、世界史用語集に出現するような語はほとんど登録されていない。そのため、分散表現を用いて格フレームの用例を汎化して利用する (山城 他 2018) ことで、さらに F 値を向上させることができる可能性がある。

擬似訓練事例を利用した実験の結果 (表 14)、分類 A の擬似訓練事例のみを利用した場合に Macro F が向上していることがわかる。一方で、分類 B の事例も利用した場合に Macro F は分

<p>見出し語) ボンディシェリ 語釈文) 1672~74年フランスが獲得。1954年インドに返還した。</p>
--

図 20 見出し語が「場所」を表す場合の見出し語と語釈文の例

類 A のみを利用した場合に比べ下がり, Micro F は大きく下がってしまった. 分類 B は質の低い擬似訓練事例であると言えるため, 質の高い擬似訓練事例を利用することは重要であると考えられる. また, 擬似訓練事例を利用して事例数の少ない格の訓練事例数を増やした場合に, 元々事例数の多い「埋め込めない」「ガ格」の F 値が下がってしまったため, バランスよく訓練事例数を増やす必要があると考えられる. SVM の `class_weight` パラメータを “balanced” にした場合の実験では, Micro F, Macro F のどちらも擬似訓練事例を用いた場合よりも高かったが, ヲ格, ニ格の F 値については擬似訓練事例の分類 A のみを使用した場合の方が高かった. デ格の F 値は擬似訓練事例を用いた場合の方が低いが, これはデ格の擬似訓練事例がヲ格, ニ格と比較して少ないことが起因していると考えられる. 本研究では教科書の文から世界史固有表現を取り除くことで擬似訓練事例を生成したが, 「埋め込めない」事例の場合にはこの手法が適用できないため, 「埋め込めない」の擬似訓練事例の自動生成手法については今後検討する必要がある. また, 教科書の文から生成できる擬似訓練事例の数には限りがあるため, NAIST テキストコーパス, 京大ウェブ文書リードコーパスなどの, 述語項関係のアノテーションがされたコーパスを使用して, 学習事例を増やすことが考えられる. これについては, 今後の課題としたい.

最後に, 既存のシステムの解答と, 既存のシステムに提案手法を組み込んだ手法で作成された解答との比較を行った結果を考察する. 下線部 (1), (2), (3) はコピュラ文であるため, 既存のシステムと提案手法を組み込んだ手法で作成された文は同じである. 下線部 (4) については「バグダード鉄道」が「ねらう」のヲ格に埋め込めることを推定できている. 下線部 (5) については「留まる」に対して「バグダード鉄道」が埋め込めないと判定されてしまった. 下線部 (6) では「爆発する」のガ格に「仇教運動」が埋め込めることを推定できている. 下線部 (7) では, どの述語の格にも「ゴールドラッシュ」が埋め込めないと判定された. 「鉄道や汽船による交通網の発展がゴールドラッシュの背景にあり, ...」と続くのが自然と考えられる. これらの結果を見ると, 述語の項に埋め込められない場合にも文中の名詞のノ格になっている場合などがあるため, それも含めて推定することが今後の課題であると考えられる.

7.2 失敗分析

本節では推定に失敗した具体的な事例に基づき考察を行う. 7.2.1 節では擬似訓練事例を用いない場合における失敗事例を分析し, 7.2.2 節では分類 A の擬似訓練事例を追加したことにより, 推定が成功から失敗に転じてしまった事例について分析する. 分析対象には, 6 章の実験で 10 分割にしたうちの 1 セットにあたる見出し語 90 語とその語釈部からなるデータセットを使用する.

7.2.1 擬似訓練事例を用いない場合における失敗分析

擬似訓練事例を用いない場合に推定に失敗した事例を調査した結果、失敗の原因と考えられるものには、

- (1) 格フレームの選択誤り
- (2) 見出し語が世界史固有表現辞書に登録されておらず、見出し語の意味カテゴリの素性が抽出できないことによる分類誤り
- (3) KNP による格解析・形態素解析の誤り
- (4) 見出し語に照応する名詞が語釈文中に存在するにも関わらず、重ねてゼロ照応をしてしまうという誤り
- (5) 見出し語以外に照応するゼロ代名詞が存在し、格が埋まっているにも関わらず、格を埋めようとしてしまう誤り

の5つがあった。失敗原因の内訳を示した混同行列を表16に示す。表中の(1)～(5)は上述の原因に対応している。分析対象のデータセットには正解ラベルが「その他の格」となる事例は存在せず、システムが「その他の格」と予測した事例も存在しなかったため「その他の格」については省略している。また、明確な原因が明らかではなく、おそらく複合的な要因による失敗であると考えられる事例を表中では「他」としている。「他」に該当する失敗事例の数には特に偏りはなく、誤りの傾向はつかめなかった。また、正解ラベルが「埋め込めない」である事例を「ガ格」と予測したもの、正解ラベルが「ガ格」である事例を「埋め込めない」と予測したものが多いことがわかった。さらに、正解ラベルが「デ格」となる事例を「埋め込めない」と予

表 16 失敗原因の内訳を示した混同行列

		予測されたラベル				
		埋め込めない	ガ格	ヲ格	ニ格	デ格
正解ラベル	埋め込めない	97	(1) 0 (2) 4 17 (3) 1 (4) 0 (5) 0 他 12	(1) 0 (2) 0 3 (3) 0 (4) 0 (5) 0 他 3	(1) 0 (2) 0 0 (3) 0 (4) 0 (5) 0 他 0	(1) 0 (2) 0 0 (3) 0 (4) 0 (5) 0 他 0
	ガ格	(1) 2 (2) 0 18 (3) 4 (4) 2 (5) 2 他 8	137	(1) 0 (2) 0 0 (3) 0 (4) 0 (5) 0 他 0	(1) 0 (2) 0 0 (3) 0 (4) 0 (5) 0 他 0	(1) 0 (2) 0 0 (3) 0 (4) 0 (5) 0 他 0
	ヲ格	(1) 1 (2) 0 2 (3) 0 (4) 0 (5) 0 他 1	(1) 1 (2) 0 3 (3) 0 (4) 0 (5) 0 他 2	1	(1) 0 (2) 0 0 (3) 0 (4) 0 (5) 0 他 0	(1) 0 (2) 0 0 (3) 0 (4) 0 (5) 0 他 0
	ニ格	(1) 1 (2) 1 3 (3) 0 (4) 0 (5) 0 他 1	(1) 0 (2) 0 0 (3) 0 (4) 0 (5) 0 他 0	(1) 0 (2) 0 0 (3) 0 (4) 0 (5) 0 他 0	0	(1) 0 (2) 0 0 (3) 0 (4) 0 (5) 0 他 0
	デ格	(1) 6 (2) 2 18 (3) 2 (4) 0 (5) 0 他 8	(1) 1 (2) 1 3 (3) 0 (4) 0 (5) 0 他 1	(1) 0 (2) 1 1 (3) 0 (4) 0 (5) 0 他 0	(1) 0 (2) 0 0 (3) 0 (4) 0 (5) 0 他 0	0

測する事例が非常に多く存在しており, これらの事例の判別が難しいということがわかった。

(1) の格フレームの選択誤りの例を図 21 に示す。見出し語を埋め込まない状態で格解析した際に選定された格フレームと, 見出し語を埋め込んだ状態で格解析した際に選定された格フレームのいずれの必須格にも正解となる格が存在しないため, 「格の埋まりやすさ」, 「意味的な類似度」などの素性値が 0 となってしまう, 上手く分類できないと考えられる。また, この場合は必須格に「デ格」が含まれる格フレームが選定されることで上手く分類される可能性がある。しかし, 図 21 に示した必須格に「デ格」が含まれる「収まる」の格フレーム 6 では, 「ガ格」, 「ニ格」, 「デ格」のいずれの用例も語釈文中の「収まる」の格要素と意味的に遠いものであった。この場合, 格フレーム生成確率が高い「収まる」の格フレーム 1, 2 などが選定されやすく, 「収まる」の格フレーム 6 が選定されなかったと考えられる。ここで, 格フレーム生成確率 $P(F_i|v_i)$ は用言 v_i の頻度を $C(v_i)$, v_i が格フレーム F_l をとる頻度を $C(F_i, v_i)$ として

$$P(F_i|v_i) = \frac{C(F_i, v_i)}{C(v_i)} \quad (3)$$

と計算される (河原, 黒橋 2007)。

(2) については, 特に見出し語が助詞を含んでいる場合などに, 世界史固有表現辞書に登録されていないことがある。例えば「アルバニアの対ソ外交」や「ドイツの軍備制限」などである。これらの場合には「見出し語の意味カテゴリ」の素性が全て 0 となってしまう, 分類に有効な素性が使用できないため失敗してしまう場合があると考えられる。

(3) の KNP による格解析・形態素解析の誤りが分類の失敗の原因となっていると考えられる

<p>【用語集の記述】 見出し語) ナントの王令 語釈文) ユグノー戦争は急速に収まり, フランスの国家統一の出発点となった</p> <p>【期待される出力】 収まる => デ格</p> <p>【選定された「収まる」の格フレーム 2】 <ガ格> 本:39, C D:36, ... <ニ格> ケース:812, 箱:357, ... <ノ格> たばこ:10, 専用:9, ...</p> <p>【再選定された「収まる」の格フレーム 1】 <ガ格> タイヤ:29, 足:18, ... <ニ格> 中:3644, 下:349, ... <ノ格> 手:286, 腕:228, ...</p> <p>【必須格に「デ格」が含まれる「収まる」の格フレーム 6】 <ガ格> 全体:4, 顔:3, ... <ニ格> 写真:1253, 画像:44, <デ格> 笑顔:26, ポーズ:11,</p>

図 21 格フレームの選択誤りが起きた事例

事例を図 22 に示す。この例では、「行った」を動詞「行く」の活用形と誤解析してしまっており、適切な素性の抽出ができていないと考えられる。

(4) の見出し語に照応する名詞が語釈文中に存在するにも関わらず、重ねてゼロ照応をしてしまうという誤りが起きている事例を図 23 に示す。この例では、「組織」が「ミール」を照応しており、この文に対して「ミール」を埋め込むことはできない。見出し語に照応するゼロ代名詞の推定を行う前に、見出し語に照応する名詞の推定を事前に行うことができればこのような事例においても正しく推定できると考えられる。

(5) の見出し語以外に照応するゼロ代名詞が存在し、格が埋まっているにも関わらず、格を埋めようとしてしまう誤りの事例を図 24 に示す。この例では「建国する」の「ガ格」の格要素は「オスマン」であるが、「オスマン」は「建国する」と直接係り受け関係になく、ゼロ代名詞化している。この場合、「ガ格」は埋まっていないと解析されるため、いずれかの格で埋め込めると誤分類してしまう場合がある。見出し語に照応するゼロ代名詞の推定を行うと同時に、文内ゼロ代名詞照応解析も行うことでこのような事例においても正しく推定できると考えられる。

【用語集の記述】
見出し語) イタリア政策
語釈文) 教皇や教皇領の保護を理由に行ったが、帝国教会政策の維持をはかる
ことこそが目的であった
【期待される出力】
残される => ヲ格

図 22 KNP による解析誤りが起きた事例

【用語集の記述】
見出し語) ミール
語釈文) 1861年の農奴解放の際にも組織は残され、地主から土地を購入し、
その購入代金の支払い、土地の共同利用などの責任を負うこととなった
【期待される出力】
残される => 埋め込めない

図 23 見出し語に照応する名詞が語釈文中に存在する事例

【用語集の記述】
見出し語) ルーム = セルジューク朝
語釈文) 13世紀モンゴルの侵攻を受けて弱体化していたが、この王朝に仕えていたオスマンが自立してオスマン帝国を建国した
【期待される出力】
残される => 埋め込めない

図 24 見出し語以外に照応する代名詞がゼロ化している事例

7.2.2 擬似訓練事例の分類 A を利用した際の失敗分析

表 12 および表 14 からわかるように、擬似訓練事例を利用した場合には、利用していない場合に比べ、「埋め込めない」、「ガ格」の再現率が下がり「ヲ格」、「ニ格」、「テ格」の再現率が上がっている。実際に失敗した事例を分析すると、図 25 のように、見出し語を語釈文中に「埋め込めない」場合に「ニ格」や「ヲ格」などに埋め込めると推定してしまう場合がいくつか見られた。しかしこれだけでは、何が失敗の原因となっているかまでは考察できない。そこで、世界史教科書から自動生成された擬似訓練事例の中に、分類の失敗の原因になるような事例がないかの調査を行った。調査対象は評価実験に使用した分類 A の擬似訓練事例 427 個とした。

調査の結果、自動生成された擬似訓練事例の中には「元の文において、注目している動詞について表出している格要素がただ一つしかないために、それを取り除くと動詞しか残らない事例」があった。図 26 の例では、「除く」の唯一の格要素である「稲作」が含まれる文節を取り除いてしまうことで、「除く」に対応づけられる格要素が存在しなくなってしまう。このような擬似訓練事例では、着目する動詞がどのような用法で使われているか、すなわちどの格フレームを使用するかを決定づけるための手掛かりが少ないため、適切な格フレームを選定できていないと考えられる。その結果、質の低い擬似訓練事例となってしまう、分類に失敗してしまう場合があると考えられる。このような事例を擬似訓練事例から取り除くことで、さらに質の高い擬似訓練事例のみを学習に使うことができるようになり、分類精度が向上する可能性がある。

また、自動生成された擬似訓練事例の中には見出し語とみなした語の意味カテゴリに偏りがあり、特に「場所」を意味する語の割合が多いことがわかった。用語集の事例および擬似訓練事例における正解ラベル毎の、見出し語の意味カテゴリが「場所」である割合を表 17 に示す。用語集の事例中の、見出し語の意味カテゴリが「場所」である割合と比較して、擬似訓練事例中に含まれる、見出し語の意味カテゴリが「場所」である割合は非常に高くなっている。このような場合、擬似訓練事例を含む訓練事例の分布と、評価事例の分布が大きく異なってしまう、

<p>【用語集の記述】 見出し語) 真番郡 語釈文) その位置について、楽浪郡の北方説と南方説の二説があるが、南方説が有力とされる 【期待される出力】 残される => 埋め込めない</p>

図 25 「埋め込めない」事例に対して「ニ格」と推定してしまった例

<p><稲作を> 除けば、しばしば耕地をかえなければならなかった</p>
--

図 26 元の文において、注目している動詞について表出している格要素がただ一つしかないために、それを取り除くと動詞しか残らない事例

表 17 用語集の事例および擬似訓練事例における正解ラベル毎の、見出し語の意味カテゴリが「場所」である割合

正解ラベル	用語集の事例	擬似訓練事例
埋め込めない	0.13(182/1,399)	
ガ格	0.18(252/1,432)	
ヲ格	0.34(28/82)	0.57(115/201)
ニ格	0.30(13/44)	0.81(151/186)
デ格	0.07(9/134)	0.90(36/40)
その他の格	0.16(5/31)	
全体	0.16(489/3,123)	0.71(302/427)

学習が上手くいかないと考えられる。この結果、見出し語を語釈文中に「埋め込めない」場合に「ニ格」や「ヲ格」などに埋め込めると誤分類してしまう場合があると考えられる。擬似訓練事例の分布が、訓練事例の分布と近くなるように、擬似訓練事例を間引くなどして調整することで、このような誤りを解消することができると考えられる。

8 まとめ

本研究では、文章中のゼロ代名詞照応解析と異なり、見出し語と語釈文に分かれている文書データに対し、語釈文における見出し語に照応するゼロ代名詞の表層格の推定を行った。評価実験を行った結果、KNPを用いた既存のゼロ照応解析を使用した手法に比べ、提案手法が有効であることを示すことができた。見出し語を埋め込んだ文を考慮した素性を使用することにより、全体的にF値を向上させることができたが、最終的なF値は0.782であり改善の余地があると考えられる。また、事例数の少ない格の推定精度を向上させるために擬似訓練事例を使用し、ヲ格、ニ格のF値を向上させることができた。今後の課題としては、「埋め込めない」事例の擬似訓練事例の自動生成や、擬似訓練事例の分布の調整、さらに質の良い擬似訓練事例の選定、ニューラルネットワークに基づくゼロ照応解析モデルとの比較などが挙げられる。

謝 辞

本研究の一部は、JSPS 科研費 JP16K00296、ならびに、横浜国立大学大学院環境情報研究院共同研究推進プログラムの助成を受けたものである。

参考文献

- Agichtein, E., Carmel, D., Harman, D., Pelleg, D., and Pinter, Y. (2015). “Overview of the TREC 2015 LiveQA Track.” In *Proceedings of The 24th Text REtrieval Conference*.
- 新井紀子, 松崎拓也 (2012). ロボットは東大に入れるか?: 国立情報学研究所「人工頭脳」プロジェクト. 人工知能学会誌, **27** (5), pp. 463–469.
- Goldberg, Y. and Levy, O. (2014). “word2vec Explained: Deriving Mikolov et al.’s Negative-Sampling Word-Embedding Method.” In *arXiv preprint arXiv:1402.3722*.
- Hangyo, M., Kawahara, D., and Kurohashi, S. (2013). “Japanese Zero Reference Resolution Considering Exophora and Author/Reader Mentions.” In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pp. 924–934, Seattle, Washington, USA. Association for Computational Linguistics.
- 飯田龍, 乾健太郎, 松本裕治 (2006). 文の構造を利用した文内ゼロ照応解析. 言語処理学会第 12 回年次大会発表論文集, pp. 488–491.
- 池原悟 (1997). 日本語語彙大系. 岩波書店.
- 石下円香, 阪本浩太郎, 中山周, 渋谷英潔, 森辰則, 神門典子 (2016). 東大入試世界史第 2 問 (小論述問題) 及び第 3 問 (語句問題) を解く質問応答システムの検討. 言語処理学会第 22 回年次大会発表論文集, pp. 557–560.
- 河原大輔, 黒橋禎夫 (2005). 格フレーム辞書の漸次的自動構築. 自然言語処理, **12** (2), pp. 109–131.
- 河原大輔, 黒橋禎夫 (2006). 高性能計算環境を用いた Web からの大規模格フレーム構築. 情報処理学会自然言語処理研究会 171-12, pp. 67–73.
- 河原大輔, 黒橋禎夫 (2007). 大規模格フレームに基づく構文・格解析の統合的確率モデル. 自然言語処理, **14** (4), pp. 68–81.
- Liu, T., Cui, Y., Yin, Q., Zhang, W.-N., Wang, S., and Hu, G. (2017). “Generating and Exploiting Large-scale Pseudo Training Data for Zero Pronoun Resolution.” In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 102–111, Vancouver, Canada. Association for Computational Linguistics.
- Matsubayashi, Y. and Inui, K. (2017). “Revisiting the Design Issues of Local Models for Japanese Predicate-Argument Structure Analysis.” In *Proceedings of the 8th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pp. 128–133, Taipei, Taiwan. Asian Federation of Natural Language Processing.
- 大森勇輔, 森田和宏, 泓田正雄, 青江順一 (2015). 擬似訓練データを用いた Q&A サイトの質問分類. 言語処理学会第 21 回年次大会発表論文集, pp. 489–492.
- 阪本浩太郎, 中山周, 渋谷英潔, 石下円香, 森辰則, 神門典子 (2016). 東大入試世界史第 1 問

(大論述問題) を解く質問応答システムの検討. 言語処理学会第 22 回年次大会発表論文集, pp. 553–556.

Sakamoto, K., Ishioroshi, M., Matsui, H., Jin, T., Wada, F., Nakayama, S., Shibuki, H., Mori, T., and Kando, N. (2016). “Forst: Question Answering System for Secondstage Examinations at NTCIR-12 QA Lab-2 Task.” In *Proceedings of The NTCIR-12 Conference*.

Sasano, R. and Kurohashi, S. (2011). “A Discriminative Approach to Japanese Zero Anaphora Resolution with Large-scale Lexicalized Case Frames.” In *Proceedings of 5th International Joint Conference on Natural Language Processing*, pp. 758–766, Chiang Mai, Thailand. Asian Federation of Natural Language Processing.

佐藤貢 (2017). 東大世界史 27 ヲ年 [第 6 版]. 教学社.

Shibuki, H., Sakamoto, K., Ishioroshi, M., Fujita, A., Kano, Y., Mitamura, T., Mori, T., and Kando, N. (2016). “Overview of the NTCIR-12 QA Lab-2 Task.” In *Proceedings of The NTCIR-12 Conference*.

Shibuki, H., Sakamoto, K., Ishioroshi, M., Kano, Y., Mitamura, T., Mori, T., and Kando, N. (2017). “Overview of the NTCIR-13 QA Lab-3 Task.” In *Proceedings of The NTCIR-13 Conference*.

Shibuki, H., Sakamoto, K., Yoshinobu Kano, T. M., Ishioroshi, M., Itakura, K. Y., Wang, D., Mori, T., and Kando, N. (2014). “Overview of the NTCIR-11 QA-Lab Task.” In *Proceedings of the 11th NTCIR Conference*.

東京書籍 (2007). 世界史 B. 東京書籍株式会社.

山城颯太, 西川仁, 徳永健伸 (2018). 分散表現による大規模格フレームの汎化を利用した統合的ニューラルゼロ照応解析. 言語処理学会第 24 回年次大会発表論文集, pp. 588–591.

全国歴史教育研究協議会 (2008). 世界史 B 用語集 改訂版. 株式会社山川出版社.

略歴

大矢 康介: 2018 年横浜国立大学理工学部数物・電子情報系学科卒業. 2018 年より横浜国立大学環境情報学府情報環境専攻博士課程前期在籍中. 自然言語処理の研究に従事.

阪本浩太郎: 2014 年横浜国立大学大学院環境情報学府情報メディア環境学専攻博士課程前期修了. 2014 年より横浜国立大学大学院環境情報学府博士課程後期在学. この間, 2015 年 8 月より 2016 年 8 月まで米国カーネギーメロン大学言語技術研究所外来研究員. 自然言語処理, 質問応答などの研究に従事. 言語処理学会会員.

渋谷 英潔: 1997 年小樽商科大学商学部商業教育養成課程卒業. 1999 年同大学

大学院商学研究科修士課程修了。2002年北海道大学大学院工学研究科博士後期課程修了。博士（工学）。2006年北海学園大学大学院経営学研究科博士後期課程修了。博士（経営学）。

森 辰則：1986年横浜国立大学工学部情報工学科卒業。1991年同大学大学院工学研究科博士課程後期修了。工学博士。同年、同大学工学部助手着任。同講師、同助教授を経て、現在、同大学大学院環境情報研究院教授。この間、1998年2月より11月までStanford大学CSLI客員研究員。自然言語処理、情報検索、情報抽出などの研究に従事。言語処理学会、情報処理学会、電子情報通信学会、人工知能学会、ACM各会員。

(2019年7月26日 受付)

(2019年11月8日 再受付)

(2019年12月24日 採録)