

# 要約条件と単文型知識源に基づく non-factoid 型質問応答手法

木村 輔<sup>†</sup> 田上 諒<sup>††</sup> 宮森 恒<sup>†</sup>

<sup>†</sup> 京都産業大学 大学院 先端情報学研究科 〒603-8555 京都府京都市北区上賀茂本山

<sup>††</sup> 京都産業大学 コンピュータ理工学部 ネットワークメディア学科 〒603-8555 京都府京都市北区上賀茂本山

E-mail: <sup>†</sup>{i1658047,miya}@cc.kyoto-su.ac.jp, <sup>††</sup>gl344775@cse.kyoto-su.ac.jp

**あらまし** 物事が起こった理由やその過程を問う質問への解答を目指す，大学入試問題を想定した non-factoid 型質問応答が近年活発に研究されている。これまでは，主に解答候補文の検索と要約によって解答を生成する手法が提案されている。しかし，質問文中で指示されたまとめ方に応じた解答生成が十分になされているとは言えず，また，解答候補となる文を抽出できない場合があるといった問題のため，現状では，必ずしも適切な解答生成に結びついているとは言い難い。本稿では，大学入試二次試験の世界史 B の論述問題を対象とし，単文を基本単位とした知識源構築，および，まとめ方の指示に応じた，解答候補文の抽出を組み合わせた自動解答手法を提案する。まず，質問文中からまとめ方の指示を推定し，そのまとめ方の指示に応じて，解答候補文の抽出処理の内容を変更することで，より適切な解答文の生成を目指す。また，単文型知識源を利用することで，情報源が異なる知識源間や，2つの事柄間を，元々の文より細かな粒度で比較できる為，従来手法では困難だった，2項間の共通点や相違点の発見に繋がることが期待できる。

**キーワード** non-factoid 型質問応答，大学入試問題，論述，自動解答，自然言語処理

## 1. はじめに

近年，質問応答システムは，情報検索の手段の一つとして，Siri<sup>(注1)</sup>や Google Now<sup>(注2)</sup>などのアプリケーションに搭載されており，徐々に人々の日常生活へ浸透してきている。従来の検索方式では，与えられたクエリに対して，複数の文書を返すため，ユーザは，情報要求を満たす適切な回答を，自身で確認する必要がある。それに対して，質問応答システムは，与えられた自然言語の質問に対して，1つの回答を返す検索方式であるため，ユーザは，効率よく正確に情報へ到達できる点で，より優れた検索方式といえる。

質問応答システムは，大きく分けて2種類あり，人名や日時など単純な事実や出来事を答える factoid 型質問応答と，事物間の関係や方法，理由などを答える non-factoid 型質問応答がある。近年，より現実的な質問への解答を目指した，大学入試問題を対象とする質問応答システムの研究 [3][14] が活発に進められている。国立情報学研究所 (NII<sup>(注3)</sup>) の評価型ワークショップである NTCIR-13 QA Lab-3 Task [2] では，factoid 質問応答を想定した，多肢選択問題や穴埋め問題に解答するタスクと，non-factoid 質問応答を想定した，論述問題に解答するタスクが実施されている。

factoid 質問応答である，大学入試センター試験の世界史 B 問題を解くタスクでは，安定して高得点を取る手法が提案されているが，non-factoid 質問応答である，大学入試二次試験の

世界史 B の論述問題においては，平均点は超えるものの，未だ高い水準には達していない。これまでは，主に論述の解答候補文の検索と要約によって解答を生成する手法が提案されている。しかし，質問文中で指示されたまとめ方に応じた解答生成が十分になされているとは言えず，また，解答候補となる文を抽出できない場合があるといった問題のため，現状では，必ずしも適切な解答生成に結びついているとは言い難い。

本稿では，大学入試二次試験の世界史 B の論述問題を対象とし，単文を基本単位とした知識源構築，および，まとめ方の指示に応じた，解答候補文の抽出を組み合わせた自動解答手法を提案する。まず，質問文中からまとめ方の指示を推定し，そのまとめ方の指示に応じて，解答候補文の抽出処理の内容を変更することで，より適切な解答文の生成を目指す。また，単文型知識源を利用することで，情報源が異なる知識源間や，2つの事柄間を，元々の文より細かな粒度で比較できる為，従来手法では困難だった，2項間の共通点や相違点の発見に繋がることが期待できる。

本稿の構成は以下の通りである。2章では，関連研究と本研究の位置付けについて述べ，3章では，提案手法の目的とシステム構成について詳述する。4章で実験とその結果，5章で考察を述べる。最後に，6章で，まとめと今後の課題を整理する。

## 2. 関連研究

質問応答システムの種類は多岐にわたる。質問応答システムについて，解答できる質問の種類および分野と，提案された手法の種類の，2つの側面から関連研究について述べる。

質問応答システムの種類は，2つの軸によって，それぞれ大きく2種類に分けられる。1つめの軸は，質問応答システムが

(注1) : iOS - Siri - Apple (日本) : <http://www.apple.com/jp/ios/siri/>

(注2) : Google : <https://www.google.com/intl/ja/landing/now/>

(注3) : 国立情報学研究所/National Institute of Informatics : <http://www.nii.ac.jp/>

解答できる質問のタイプで、一方を、人名や日時など単純な事実や出来事を答える factoid 型質問、もう一方を、事物間の関係や方法、理由などを答える non-factoid 型質問と呼ぶ。2 つめの軸は、質問応答システムが対象としている分野で、一方を、質問内容が特定分野に限定されているクローズドドメイン質問応答、もう一方を、質問内容が特定分野に限定されていないオープンドメイン質問応答と呼ぶ。

また、質問応答システムの研究において提案された手法の種類は、大きく 4 種類に分けることができる。1 つめは、特定の質問形式や質問分野を想定したルールベース手法、2 つめは、質問を元言語、解答を翻訳先言語とみなした統計的機械翻訳や、Noisy Channel Model を用いた統計的手法、3 つめは、述語と係り元の意味の関係などを用いた意味的構文解析手法、そして、4 つめは、それぞれを合わせたハイブリッド手法である。

クローズドドメインの factoid 型質問応答システムの一つとして、Green ら [6] によって開発された BASEBALL がある。また、オープンドメインの factoid 型質問応答システムの例としては、Kupiec ら [9] の MURAX が挙げられる。これは、人名や地名など短い語句を回答する、ルールベース手法を用いた質問応答システムで、文書検索を用いた質問応答システムの基本的な構成を示したものと見える。また、別の例として、IBM が開発した Watson [1] が挙げられる。特徴的な点として、回答候補から正答を選び出す手法が挙げられる。この手法では、Web 検索などによって得られた回答候補に対して、クイズ番組の過去問から事前に学習することで生成した観点をを用いて、各回答候補へスコア付けし、回答を選び出す。

Fukumota [5] は、non-factoid 型質問をさらに 3 つの質問タイプに分類し、各質問タイプに応じた解答を生成する手法を提案している。質問タイプは、質問文中の表層パターンによって分類され、質問文中の手がかり語と、検索された文書中の手がかり語を元に、各質問タイプに沿った解答抽出戦略を提案している。また、文書検索時には、全文検索エンジンとして Namazu<sup>(注4)</sup>を用い、Google のスニペット情報を合わせることで、検索精度の改善を目指している。

近年、大学入試の二次試験問題に対する質問応答システムに関する研究も活発に行われている。大学入試の二次試験の世界史 B 問題の場合は、世界史に分野が限定されるという点で必ずしもオープンドメイン質問応答ではないが、人名や地名といった単純な名称の回答だけでなく、適切な内容説明や理由の回答、年代の正しい並びや地図の理解、写真の理解と組み合わせた回答が求められる factoid 型質問や、問題から与えられた文脈を踏まえた上で、質問時に要求される要約指示や要約文字数に応じた、論述を生成する non-factoid 型質問を含んでいる。このような質問に、適切に解答するためには、従来にはない、より現実的な回答生成機能が要求される。すなわち、factoid 型質問や non-factoid 型質問に、柔軟かつ適切な形式で正確に回答することが課題の一つであると考えられる。

Sakamoto ら [13] は、文書検索を用いた質問応答システムを

基礎とした、11 のモジュールで構成された質問応答システムを開発している。質問解析において、大論述では問題中に提供された指定語句をクエリ語として利用し、そのような指定語句が存在しない小論述においては、質問文を語句解答問題とみなし、大学入試の二次試験の世界史 B 語句解答モジュールを用いてクエリ語を獲得している。また、解答候補文に対して、命題を含む文節を 1 つの単位とした文圧縮することで、大論述問題において、ROUGE [10] による評価に改善はなかったものの、小論述問題においては、無回答となる場合を削減したと報告している。

Takada ら [15] は、論述問題を解答するために、文書検索を用いた質問応答システムをベースとし、文書検索を選択する上で重要な、各文へのスコアリングにおいて、「質問文中の後半に現れる語句の重要性が高い」という仮説を元にした項を、スコアの計算式に導入している。また、要約文生成では、質問文と時間表現が一致した解答候補文を優先的に選択するよう処理をしている。さらに、制限文字数を超えるが、スコアが高い解答候補文を取りこぼさないために、特定の条件において、そのような解答候補文を文圧縮をし、生成された各圧縮文に対して、再びスコアリングすることで解決している。彼らは、問題点として、質問文中の語句と知識源中の文の語句間での抽象度の違いによる名詞の不一致により、適切な抽出が困難となる点、また、制限文字数が少ない問題において、知識源中の文の組み合わせによって解答する問題に対応できていない点を挙げている。

我々の提案手法は、これまでの手法で対象とされなかった質問文中のまとめ方の指示に着目し、各まとめ方の指示に沿った文書検索と文要約を目指した手法を提案する。また、先行研究で挙げられた問題点に対して、単文を基本単位とした知識源構築による文書検索によって解決を目指した。

### 3. 提案手法

大学入試の二次試験の論述問題に解答するためには、次の 3 点が重要であると考えられる。

まず、質問文中の焦点カテゴリの活用である。例えば、「明代の長江流域の農業・工業の変化について簡潔に述べよ」という論述問題において、「農業」と「工業」のように、論述対象の焦点となるカテゴリについて明記されていることが多い。そのため、検索ベースによって解答候補文の元となる文書集合を獲得する場合、論述対象の焦点カテゴリを用いてクエリ語を拡張するなど、焦点カテゴリを活用することで、より適切な文書検索ができると考えられる。

次に、論述のまとめ方の指示に応じた解答生成である。論述問題におけるまとめ方は、表 1 に示す 5 種類に分類できる。そのため、それぞれの要約指示に応じて適切に処理を変更することで、より正確な解答候補抽出が生成できると期待される。

最後に、検索対象の知識源の整理である。知識源となる教科書や参考書では、文書全体の構成や書き手の傾向により、同一の事象に対する表現が異なることが多いと考えられる。この表層的な表現を単純化することで、内容の比較がしやすくなり、事柄間の共通点、差異を獲得しやすくなることが期待できる。

(注4) : 全文検索システム Namazu : <http://www.namazu.org>

表 1 論述問題におけるまとめ方の指示の分類

まとめ方の指示	要求される要約内容
まとめ	簡潔な要約
過程/変化	特定時期における概観を軸とした要約
関連/影響	指示された語句間の関係を明示した要約
特色	指示された語句の特徴を明示した要約
比較	指示された語句間の共通点などを明示した要約

そこで我々は、大学入試二次試験の世界史 B の論述問題を対象とし、単文を基本単位とした知識源構築、および、まとめ方の指示に応じた、解答候補文の検索と要約を組み合わせた自動解答手法を提案する。質問文中からまとめ方の指示を推定し、そのまとめ方の指示に応じて、解答候補文の抽出処理を変更することで、より適切な解答文の生成を目指す。また、単文型知識源を利用することで、情報源が異なる知識源間や、2つの事柄間を、元々の文より細かな粒度で比較できる為、従来手法では困難だった、2項間の共通点や相違点の発見に繋がる事が期待できる。

本節では、まず、3.1 節にて、解答の対象となる論述問題の特徴について述べ、3.2 節にて、提案手法を取り入れた自動解答システムの構成、最後に、3.3 節にて、単文型知識源の構築について説明する。

### 3.1 論述問題の基本構成

本節では、対象データとなる大学入試二次試験の世界史 B の論述問題の特徴について説明する。論述問題は、10 数行程度の文で解答する大論述と、1 行から数行程度の文で解答する小論述の 2 種類が存在する。大学入試二次試験の世界史 B の大論述問題の例を図 1 に、小論述問題を図 2 に示す。

文脈部
人類の歴史において、戦争は多くの苦悩と惨禍をもたらすと同時に、それを乗り越えて平和と解放を希求するさまざまな努力を生みだす契機となった。第二次世界大戦は1945年に終結したが、それ以前から連合国側ではさまざまな戦後構想が練られており、これらは国際連合など新しい国際秩序の枠組みに帰結した。しかし、国際連合の成立がただちに世界平和をもたらしたわけではなく、米ソの対立と各地の民族運動などが結びついて新たな紛争が起こっていった。たとえば、中国では抗日戦争を戦っているなかでも国民党と共産党の勢力争いが激化するなど、戦後の冷戦につながる火種が存在していた。
質問部
第二次世界大戦中に生じた出来事が、いかなる形で1950年代までの世界のありかたに影響を与えたのかについて、解答欄に17行以内で説明しなさい。その際に、以下の8つの語句を必ず一度は用い、その語句の部分に下線を付しなさい。なお、EECに付した( )内の語句は解答に記入しなくてもよい。
指定語句
大西洋憲章、日本国憲法、台湾、金日成、東ドイツ、EEC(ヨーロッパ経済共同体)、アウシュヴィッツ、パレスチナ難民

図 1 大学入試二次試験の世界史 B の大論述問題の例 (2005 年 世界史 東京大学 第 1 問)

図 1 に示すように、大論述問題は、基本的に、文脈部、質問部から構成され、問題によっては、指定語句を与えられることがある。正しく解答するためには、文脈部を踏まえた論述の生成が求められる。つまり、質問部において与えられた、論述文字数の制限、論述に含めるように指示された指定語句、論述の

文脈部
ビザンツ世界やイスラーム世界と異なり、中世の西ヨーロッパは古代ギリシアやヘレニズムの文明をほとんど継承しなかった。ギリシア・ヘレニズムの学術文献が西ヨーロッパに広く知られるようになるのは、12世紀以降である。
質問部
これらの学術文献はどのようにして西ヨーロッパに伝わったのか。3行以内で説明しなさい。

図 2 大学入試二次試験の世界史 B の小論述問題の例 (2005 年 世界史 東京大学 第 2 問 問 (3))

まとめ方の指示に沿った文生成が求められる。

それに対し、図 2 に示すように、小論述問題は、文脈部と質問部から構成される。大論述問題と同様に、質問部において、論述文字数の制限、論述のまとめ方の指示が与えられる。

### 3.2 論述問題解答システムの構成

開発した論述問題解答システムの処理手順を、図 3 に示す。

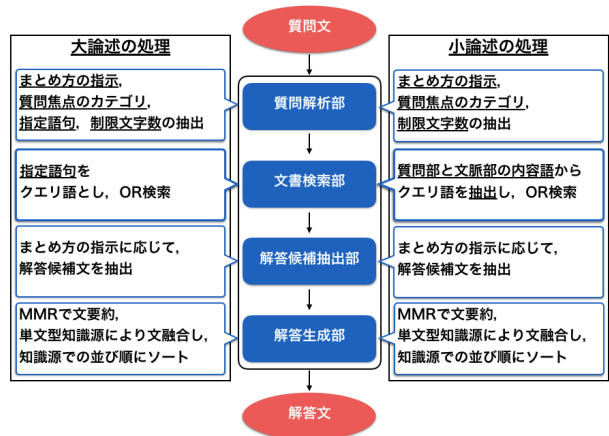


図 3 論述問題解答システムの処理手順

本システムは、Kupiec ら [9] が提案した文書検索を用いた手法をベースとした構成となっており、質問解析部、文書検索部、解答候補抽出部、解答生成部での処理を経て解答を生成する。

図 3 が表すとおり、大論述問題と小論述問題の解答生成手順における主な違いは、クエリ語の獲得方法であり、その他の処理については、基本的には同様の構成になっている。3.2.1 節から 3.2.4 節では、共通する処理手順について説明する。

#### 3.2.1 質問解析部

質問部から、論述をどのようにまとめるかを明示しているまとめ方の指示、論述をどのような事柄についてまとめるかの指示である質問焦点のカテゴリ、および、生成する論述文における制限文字数を、それぞれルールベースを用いて抽出する。ここでは、まとめ方の指示、および、質問焦点のカテゴリの推定方法について述べる。

表 1 に示すまとめ方の指示を推定するために、各まとめ方の指示と、それらを特定する手がかりとなる語句の対応表を人手で作成した。質問部を形態素ごとに分かち書きした後、作成した手がかり語との対応表を用いて、質問部のまとめ方の指示を特定した。なお、質問部中に該当する手がかり語がない場合は、

まとめ方の指示を単に「まとめ」と判断した。まとめ方の指示と手がかり語の対応表の例を表2に示す。

表2 まとめ方の指示と手がかり語の対応例

まとめ方の指示	手がかり語
過程/変化	改変, 経緯, 展開, 展望
関連/影響	結果, 背景, 関係
特色	特徴, 特質
比較	違い, 対比, 相違

質問焦点のカテゴリを推定するために、大学入試センター試験の世界史 B 問題解答用に筆者らが作成した、語句の上位概念ラベルを付与する手法 [8] を利用した。まず、イベントオントロジー EVT<sup>(注5)</sup> で用いられている上位概念ラベルと、QA Lab-2 タスクのオーガナイザから提供された、固有表現アノテーション付東京書籍教科書 (NE\_Tokyoshoseki) で用いられている固有表現タグが、それぞれ対応するようなパターン表を人手により作成した。ここで、上位概念ラベルは、イベントオントロジー EVT 中の固有表現に対して付与されている、「国家・王朝」、「社会制度」、「技術・発明」などのラベルである。また、固有表現タグは、NE\_Tokyoshoseki 中の固有表現に対して付与されている、「personType\_socialRole」、「socialSystem」、「historicalEvent」などの 32 種類のタグである。質問焦点のカテゴリを推定する際は、質問部を形態素解析し、もし、得られた語句が NE\_Tokyoshoseki において固有表現タグが付与されていた場合、作成したパターン表を用いて上位概念ラベルを獲得する。そして、獲得された全ての上位概念ラベルを質問部が持つ質問焦点のカテゴリとした。

### 3.2.2 文書検索部

3.3 節で説明する知識源を対象に、3.2.5 節、および、3.2.6 節で説明する方法により得られたクエリ語で OR 検索し、要約文の候補となる文の集合を取得した。また、世界史に用いられる語句の表記ゆれに対応するために、大学入試センター試験の世界史 B 問題解答用に筆者らが作成した、世界史に特化した表記ゆれ対応表 [8] を利用した。これは、次の 3 つの手法により獲得された同義語から構成されている。まず、Wikipedia<sup>(注6)</sup> の転送機能を用いた手法で、転送記事そのもののタイトルと転送先のタイトルを同義語として獲得した。次に、Wikipedia の段落構成を用いた手法で、Wikipedia の記事の、第 1 段落中に現れる特定のパターンで抽出された語句と、その記事タイトルを同義語として獲得した。そして、イベントオントロジー EVT のタグ情報を用いた手法で、イベントオントロジー EVT の固有表現に付与されている、日本語及び日本語以外の別名を表すタグを、その固有表現の同義語として獲得した。

### 3.2.3 解答候補抽出部

抽出されたまとめ方の指示に応じて、要約文の元となる解答候補文を抽出する。

「関連/影響」および「比較」以外のまとめ方の指示では、文書検索部で出力された文を、そのまま解答候補文とした。

「関連/影響」のまとめ方の指示では、検索された文書のうち、原因・理由表現を持つ文のみを解答候補文とした。

「比較」のまとめ方の指示では、「第二次世界大戦における、アメリカとイギリスの共通点について述べよ」という質問における共通点、「連合国に属した」のように、比較対象語句間で共通する事項を特定し、文書検索部で出力された文のうち、特定された共通する語句を含む文を解答候補文とした。共通する語句を抽出するために、事前に、3.3 節で説明する教科書から構築された知識源を対象に、1 文を 1 文書として、教科書に現れる語句間の自己相互情報量 (Pointwise Mutual Information) を計算し、共起語知識源を構築した。このとき、語句 A と共起する語句 B について、3.2.1 節で説明した手法で上位概念ラベルを推定し、焦点カテゴリとして付与した。これにより、「語句 A と共起する、特定の焦点カテゴリを持つ語句 B」を検索することができるようになる。共通する語句を抽出する際は、3.2.1 節で説明した質問の焦点と、構築した共起語知識源を用いて、各比較対象語句に共起する語句を取得し、その共起語句のうち共通する語句を特定した。この手法によって、共通する語句を特定できなかった場合は、文書検索部で出力された文を、そのまま解答候補文とした。

### 3.2.4 解答生成部

まず、抽出された各解答候補文のスコアとして、文書検索時に付与された Okapi BM25 [11] のスコアを与える。そして、そのスコアを基準とし、Maximal Marginal Relevance (MMR) [4] に基づく文要約手法により解答を生成し出力する。MMR は、生成する要約文における冗長性を低減するためのペナルティとして計算される。スコアリングされている候補となる要約文の集合から、スコアの高い候補文を順に選択し、要約文を作成する過程において、既に選択された候補文と類似した候補文のスコアに MMR に基づくペナルティを与え、生成される要約文内の冗長性を抑える手法である。提案手法においては、既に選択された候補文との類似度を、形態素単位の 3-gram オーバーラップで算出し、ペナルティのスコアとした。また、知識源として、3.3 節で説明する単文型知識源を用い、かつ、元々が同じ文から生成された文が、既に要約文として選択されていた場合、その 2 つの文を再融合した。最後に、生成した要約文を、検索対象の知識源での元々の並び順にソートし出力する。

### 3.2.5 大論述問題におけるクエリ語の取得方法

大論述問題において、図 1 に示すように、解答の論述に含むように指示された指定語句が提示されている場合は、この単語群をクエリ語として扱う。それ以外の場合は、小論述問題と同様の方法でクエリ語を獲得する。

### 3.2.6 小論述問題におけるクエリ語の取得方法

図 2 に示すように、小論述問題では、指定語句が存在しないため、知識源への文書検索には、質問部の内容語をクエリ語と

(注5) : イベントオントロジー EVT - Researchmap : <http://researchmap.jp/zoelai/event-ontology-EVT/>

(注6) : Wikipedia : <https://ja.wikipedia.org/>

して扱う。

また、図2のように、文脈部で明示した内容（ギリシア・ヘレニズムの学術文献）を、質問部において参照する形で指示語と一般的な名詞のみ（これらの学術文献）で表現する場合がある。しかしながら、文脈部の全ての文の内容語をクエリ語とし、解答候補を文書検索によって取得することが適切であるといふ難い。そのため、質問部において、文脈部を参照する表現が存在する場合、文脈部にある参照された語句を含む文のみを獲得し、その文の内容語をクエリ語に追加した。

さらに、文脈部と質問部から推定された焦点カテゴリと、クエリ語のうち、固有表現である語句を用いて、3.2.1節で説明した共起語知識源から共起する語句を獲得し、クエリ語として追加した。このクエリ語の拡張により、ある固有表現と共起する、焦点カテゴリを持つ語を獲得することができる。

### 3.3 単文型知識源の構築

情報源が異なる知識源間において、表層的な表現を単純化することで、元々の文より細かな粒度で比較することを目的とした単文型知識源を作成した。

#### 3.3.1 単文を基本単位とした文

ここで、単文型知識源中の文書である、単文を基本単位とした文の定義を説明する。まず、図4に、節（主部と述部の1対1関係）による文の分類と例を示す。

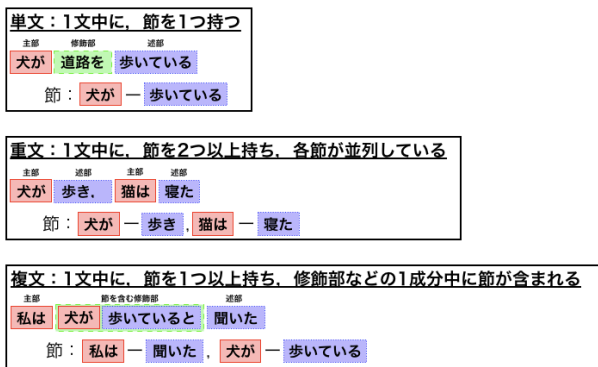


図4 節（主部と述部の1対1関係）による文の分類

通常の文は、「成分を含まない単文」であることは珍しく、「成分を含む単文」、「複文」、および「重文」であることが多い。つまり、「1つの節によって、1つの意味が表現される」と考えたとき、「成分を含む単文」、「複文」、および「重文」は、複数の意味を持った複雑な文であると言える。また、節が1つ以上あり、修飾部などの成分を1つ以上含むということは、ある事柄を表現するとき、同じ節や成分の集合を与えられたとしても、表現者の特性によって、それらの並び順が一意に定まらず、表層上の表現が異なる可能性が高い。

これらの文に対して、「成分を含まない単文」は、節を1つのみ含み、修飾部などの成分を持たないため、表現者の特性によらず、表層的な表現は、単純かつ類似したものになる可能性が高いと予測される。そこで我々は、図5に示すように、一般的な文の表現である、「成分を含む単文」、「複文」、および「重文」を、「修飾部のような成分を含まない単文」へ変換することで、

表層的な表現の単純化を目指した。

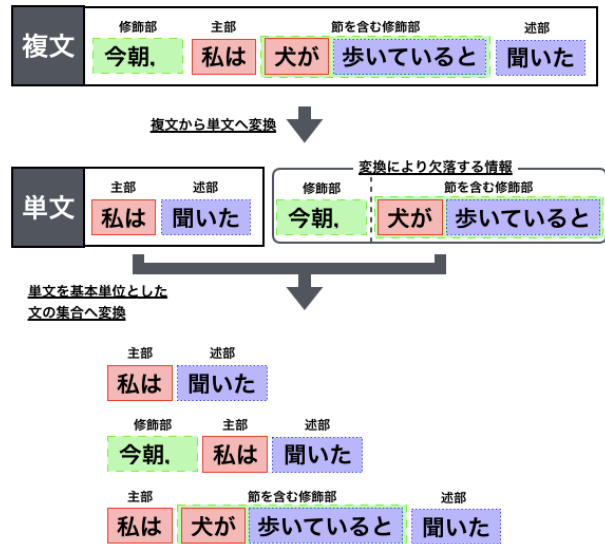


図5 複文から単文への変換例と、複文から単文を基本単位とした文の集合への変換例

しかしながら、この変換によって、成分や修飾部内の節の情報が欠落してしまうことは、知識源として活用する上で好ましくない。そのため、「修飾部のような成分を含まない単文」を基本単位とし、元々の文が持つ成分や修飾部内の節のそれぞれを組み合わせた文を合わせて生成することとした。また、変換によって終端が読点になる文は、読点の前方2つの形態素の品詞によって、助詞を原形に変換するなどし、句点で終わる自然な文へ変換した。

#### 3.3.2 一般的な記述と単文型の記述の比較

知識源における一般的な記述例として、図6に東京書籍の「世界史B」の一文、図7に山川出版社の「詳説世界史」の一文を示す。また、図6、図7の文を再構成することで作成した、それぞれの単文型知識源における記述を、図8、図9に示す。図6と図7の文書間の unigram のオーバーラップによる表層類似度は、0.126であり、同様の事柄についての文書であるが、類似度は比較的低い値を示す。これに対して、図8と図9の各文書間の内容語による表層類似度は、図8の「あ」と図9の「エ」で最大値 0.407 となった。このように、単文型へ変換した知識源を用いることで、異なる知識源に記述された、同一の事象を発見しやすくなることが期待できる。

この王国のハンムラビ王はメソポタミア全土を支配し、各地の法慣習を集大成してハンムラビ法典を發布した。

図6 一般的な知識源における記述（東京書籍）

王は運河の大工事をおこなって治水・灌漑をすすめ、またハンムラビ法典を發布して、法にもとづく強力な政治をおこなった。

図7 一般的な知識源における記述（山川出版社）

あ.ハンムラビ法典を發布した。  
 い.この王国のハンムラビ王はメソポタミア全土を支配し、ハンムラビ法典を發布した。  
 う.各地の法慣習を集大成してハンムラビ法典を發布した。

図 8 変換後の単文型知識源における記述 (東京書籍)

ア.王はおこなった。  
 イ.法にもとづく強力な政治をおこなった。  
 ウ.王は法にもとづく強力な政治をおこなった。  
 エ.王はハンムラビ法典を發布して、おこなった。  
 オ.ハンムラビ法典を發布して、法にもとづく強力な政治をおこなった。  
 カ.王はハンムラビ法典を發布して、法にもとづく強力な政治をおこなった。  
 キ.王はすすめた。  
 ク.治水・灌漑をすすめた。  
 ケ.王は治水・灌漑をすすめた。  
 コ.王は運河の大工事をおこなってすすめた。  
 サ.運河の大工事をおこなって治水・灌漑をすすめた。  
 シ.王は運河の大工事をおこなって治水・灌漑をすすめた。

図 9 変換後の単文型知識源における記述 (山川出版社)

### 3.3.3 単文型知識源の構築手法

単文型知識源作成のための処理の具体例を図 10 に示す。

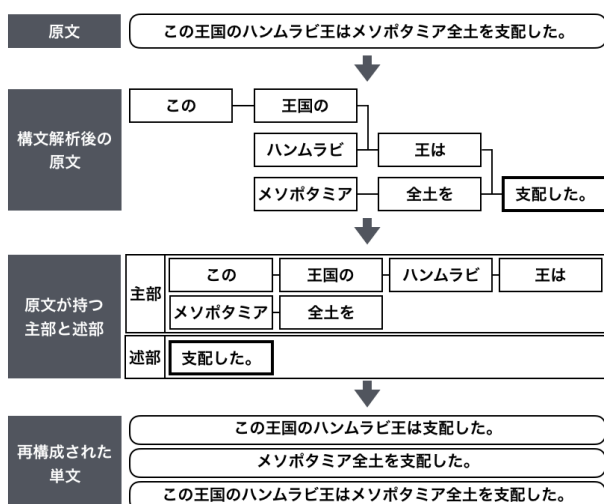


図 10 単文型知識源作成処理の具体例

まず、対象となる文を構文解析し、主格(主題格)にかかる述部を含む文節の一覧を取得する。次に、その他の文節から各述部へかかる一連の文節を統合する。最後に、統合された文節において、取得した述部を含む文節とそれ以外の各文節間の、元々の係り受け関係を用いて、その組み合わせによって単文型の文を作成する。

文書検索する文書集合として、4つの知識源を作成した。QA Lab-2 から提供された東京書籍の世界史教科書データ<sup>(注7)</sup><sup>(注8)</sup><sup>(注9)</sup>と、山川出版社の世界史教科書データ<sup>(注10)</sup>のそ

(注7): 世界史 A, 東京書籍 (2008)  
 (注8): 世界史 B, 東京書籍 (2007)  
 (注9): 新選世界史 B, 東京書籍 (2007)  
 (注10): 詳説世界史, 山川出版社 (2007)

れぞれの1文を1文書として登録した, 2つの知識源を作成した。(以下,T; 東京書籍教科書, Y; 山川出版社教科書と表す)。また, 各教科書の一文に対して, 本提案手法によって単文型へ変換した, 各1文を1文書として登録した, 2つの知識源を作成した(以下,TR; 変換後東京書籍教科書, YR; 変換後山川出版社教科書と表す)。

なお, 3.2.3 節で説明した, まとめ方の指示が「関連/影響」である場合において, 原因・理由表現を持つ文を特定するために, 手がかり語の対応表を手で作成し, 原因・理由表現の有無を示すフラグを各文書に付与した。

## 4. 評価実験

### 4.1 実験目的

本実験は, 大学入試二次試験の世界史 B 論述問題を解答する上で, 提案した単文型知識源の利用が, どのように貢献しているかを明らかにすることを目的とする。またデータセットとして, QA Lab-2 および QA Lab-3 で提供された以下の東京大学の二次試験の世界史 B 論述問題を用いた。システムの開発には, 下記のデータセットを用いた。

- 東京大学 (2005 年度 ~ 2009 年度, 隔年 3 回分)
- 東京大学 (2003, 2011 年度, 計 2 回分)

各データセットにおける, 大論述問題と小論述問題の内訳を表 3 に示す。

表 3 各データセット中の論述問題の内訳

年度	大論述問題	小論述問題
2003	1 問	2 問
2005	1 問	0 問
2007	1 問	6 問
2009	1 問	0 問
2011	1 問	5 問

システム全体における形態素解析器として MeCab<sup>(注11)</sup>を用い, 筆者らが NTCIR-12 QA Lab-2 タスク, 大学入試センター試験の世界史 B 問題解答用に作成した, 世界史 B に特化したユーザ辞書 [8] を利用した。また, 構文解析器とその形態素解析器として, それぞれ, KNP<sup>(注12)</sup>と JUMAN++<sup>(注13)</sup>を利用した。

### 4.2 実験内容

単文型知識源の利用によって, 解答の生成に与える影響を明らかにするため, 構築した自動解答システムに, 大学入試二次試験の世界史 B 論述問題を解答させた。生成されたシステム要約と, 人手で作成された評価要約間の, N-gram のオーバーラップによって計算される評価指標 ROUGE-N [10] (N = 1,2) により評価した。ROUGE-N を算出するため, QA Lab-2 のオーガナイザから提供された評価用要約文のデータセットを用いた。

(注11): MeCab: Yet Another Part-of-Speech and Morphological Analyzer : <http://taku910.github.io/mecab/>

(注12): KNP - KUROHASHI-KAWAHARA LAB : <http://nlp.ist.i.kyoto-u.ac.jp/?KNP>

(注13): JUMAN++ - KUROHASHI-KAWAHARA LAB : <http://nlp.ist.i.kyoto-u.ac.jp/index.php?JUMAN++>

### 4.3 実験結果

大論述問題における実験結果を表4, 表5に示す. また, 小論述問題における実験結果を表6, 表7に示す. なお, 東京大学2003年度大問2-(b) 問7-(b)については, 評価要約が入手できなかったため, 評価指標の値をNAとした.

表4 大学入試大論述問題における各知識源の違いによる ROUGE-1 の比較

知識源	東京大学 2003 年 大問 1	東京大学 2011 年 大問 1
T	0.399	0.448
TR	0.511	0.502
Y	0.390	0.402
YR	0.471	0.482

表5 大学入試大論述問題における各知識源の違いによる ROUGE-2 の比較

知識源	東京大学 2003 年 大問 1	東京大学 2011 年 大問 1
T	0.093	0.132
TR	0.161	0.168
Y	0.092	0.087
YR	0.136	0.133

表4, および, 表5より, 大論述の要約文生成において, TR および YR の単文型知識源を用いた場合, もととの知識源である T および Y と比べて, ROUGE-1, ROUGE-2 共に高い値を示すことがわかる.

また, 表6, および, 表7より, 小論述の要約文生成においても, TR および YR の単文型知識源を用いた場合, 元々の知識源である T および Y と比べて, ROUGE-1, ROUGE-2 共に高い値を示すことがわかる. また, 元々の知識源を用いた場合, ROUGE の値が0, つまり, 制限文字数を満たす解答が生成できなかった問題に対して, 単文型知識源を用いることで, 解答を生成できていることがわかる. しかし, 短い文書を多く含む単文型知識源を用いた場合でも, 制限文字数を満たす単文を要約候補文として取得できず, 要約文を生成できなかった問題も存在することがわかった.

## 5. 考 察

4章の評価実験より, 大論述および小論述のどちらの問題形式においても, 元々の知識源を用いる場合と比べ, 単文型へ変換した知識源を用いた提案手法の評価が, ROUGE-1, ROUGE-2 共に高いことがわかる. これは, 制限文字数のために要約文として選択されなかった, 答えとなる文節を含んだ文を, 文字数が短い単文型へ変換することによって, 要約文として選択されたことが一因と考えられる.

また, 知識源として T および Y を用いて生成された解答の要約文において, 制限文字数に達しない場合が多く確認された. これは, 生成したクエリによって抽出される, 要約候補文の数が少ないことが原因と考えられる. これに対し, 単文型知識源である TR および YR を用いた場合は, 全ての解答要約文にお

いて, 制限文字数に近い文字数を用いて, 文要約をしていることが確認できた. 特に小論述問題では, 元々の知識源では解答を生成できなかった問題に対して, 提案した単文型知識源を用いることで解答できる場合を確認できた. 一般的な知識源と比べ, 単文型の知識源は検索対象文書の数が増加したことが原因と考えられる.

実際に生成された要約文を手で確認したところ, 単文型知識源である TR および YR を用いた場合, 文頭が接続詞で始まる要約文や, 「王は, した.」のようなほとんど意味を持たない文を含む要約文が生成されていた. この問題に対して, 単文型知識源の文書として, 元々, サ変接続の名詞と共に用いられた, サ行変格活用の動詞のみの述部を持つ文を不採用にすることや, 最小の意味を持つ文のみを採用することで解決できると考えられる. 文の持つ意味の定義については様々あるが, ROUGE-N と同じ自動要約の評価指標である, Basic Element [7] で計算に使用される, 「最小意味単位の集合」という考え方などを参考に, 文が持つべき意味的要素を定義することが考えられる. また, 知識源として T および Y を用いた場合においても, 大論述問題における指定語句を, 全て含んでいない要約文を生成する場面があった. この問題については, Takada ら [13] のように, 一度, 指定語句毎に解答文候補を集め, それぞれの指定語句の解答文候補から最低1文毎に, 要約文となる文を選択することで解決できる.

## 6. ま と め

本稿では, まとめ方の指示に応じた要約文生成手法と, 単文型知識源による, 細かい粒度の要約文候補の抽出手法を組み合わせた手法を提案した. 大学入試二次試験の世界史 B 論述問題を対象とした質問応答システムを構築し, 実験によって, 提案した単文型知識源によって, 生成された要約文の ROUGE-N がどのように変化したかを明らかにした. 提案した単文型知識源を用いることで, 元々の知識源では対応できなかった, 短い制限文字数を満たす要約文を生成できた. しかし, 本提案手法で構築した単文型知識源中には, 表現する内容があまりにも少ない不自然な文を含むため, 単文型文書の作成手法を改善する必要がある. また, 現在の提案手法において, 単文型知識源の生成元の文が異なる文同士が要約文として選択された場合, それぞれの文を結合しながら自然な文へ変換することは難しい. そこで今後我々は, まとめ方の指示に沿った要約文の生成のために, Rush ら [12] が提案した, Attention ベースの Encoder-Decoder を用いた生成型の文要約手法をベースとした, 生成元の文が異なる単文型の文同士の結合による, 自然な要約文生成を目指す手法を導入する予定である.

## 謝 辞

本研究は京都産業大学総合学術研究所の研究活動によるものです.

表 6 東京大学入試小論文問題における各知識源の違いによる ROUGE-1 の比較

知識源	2003 年度		2011 年度				
	大問 2-(a) 問 5	大問 2-(b) 問 7-(b)	大問 2-問 1-(a)	大問 2-問 1-(b)	大問 2-問 2	大問 2-問 3-(a)	大問 2-問 3-(b)
T	0	NA	0	0	0.245	0.217	0.314
TR	0	NA	0.244	0.294	0.309	0.314	0.340
Y	0.200	NA	0.197	0	0.296	0.245	0.203
YR	0.321	NA	0.246	0.308	0.417	0.353	0.269

表 7 大学入試小論文問題における各知識源の違いによる ROUGE-2 の比較

知識源	2003 年度		2011 年度				
	大問 2-(a) 問 5	大問 2-(b) 問 7-(b)	大問 2-問 1-(a)	大問 2-問 1-(b)	大問 2-問 2	大問 2-問 3-(a)	大問 2-問 3-(b)
T	0	NA	0	0	0.030	0	0.095
TR	0	NA	0.053	0.065	0.046	0.087	0.090
Y	0.043	NA	0.034	0	0.057	0.043	0.026
YR	0.038	NA	0.077	0.167	0.133	0.043	0.097

## 文 献

- [1] IBM Watson (ワトソン) - Japan. <http://www.ibm.com/smarterplanet/jp/ja/ibmwatson/>.
- [2] NTCIR-13 QA Lab-3 Task. <http://research.nii.ac.jp/qalab/index.html>.
- [3] ロボットは東大に入れるか。Today Robot Project. <http://21robot.org>.
- [4] J. Carbonell and J. Goldstein. The use of mmr, diversity-based reranking for reordering documents and producing summaries. In *Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval*, pages 335–336. ACM, 1998.
- [5] J.-i. Fukumoto. Question answering system for non-factoid type questions and automatic evaluation based on be method. In *NTCIR, 2007*.
- [6] B. F. Green, Jr., A. K. Wolf, C. Chomsky, and K. Laughery. Baseball: An automatic question-answerer. In *Papers Presented at the May 9-11, 1961, Western Joint IRE-AIEE-ACM Computer Conference, IRE-AIEE-ACM '61* (Western), pages 219–224, New York, NY, USA, 1961. ACM.
- [7] E. Hovy, C.-Y. Lin, and L. Zhou. Evaluating duc 2005 using basic elements. In *Proceedings of DUC*, volume 2005. Citeseer, 2005.
- [8] T. Kimura, R. Nakata, and H. Miyamori. KSU team's multiple choice QA system at the NTCIR-12 QALab-2 task.
- [9] J. Kupiec. Murax: A robust linguistic approach for question answering using an on-line encyclopedia. In *Proceedings of the 16th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '93*, pages 181–190, New York, NY, USA, 1993. ACM.
- [10] C.-Y. Lin. Rouge: A package for automatic evaluation of summaries. In S. S. Marie-Francine Moens, editor, *Text Summarization Branches Out: Proceedings of the ACL-04 Workshop*, pages 74–81, Barcelona, Spain, July 2004. Association for Computational Linguistics.
- [11] S. Robertson, H. Zaragoza, et al. The probabilistic relevance framework: Bm25 and beyond. *Foundations and Trends® in Information Retrieval*, 3(4):333–389, 2009.
- [12] A. M. Rush, S. Chopra, and J. Weston. A neural attention model for abstractive sentence summarization. *CoRR*, abs/1509.00685, 2015.
- [13] K. Sakamoto, M. Ishioroshi, H. Matsui, T. Jin, F. Wada, S. Nakayama, H. Shibuki, T. Mori, and N. Kando. Forst: Question answering system for second-stage examinations at ntcir-12 qa lab-2 task.
- [14] H. Shibuki, K. Sakamoto, M. Ishioroshi, A. Fujita, Y. Kano, T. Mitamura, T. Mori, and N. Kando. Task overview for ntcir-12 qa lab-2. In *In Proceeding of the 12th NTCIR Conference*, 2016.
- [15] T. Takada, T. Imagawa, T. Matsuzaki, and S. Sato. Sml question-answering system for world history essay and multiple-choice exams at ntcir-12 qa lab-2.