

一問一答型質問応答を利用した関連質問群に対する質問応答

川口 晋平

横浜国立大学 大学院 環境情報学部

森 辰則

横浜国立大学 大学院 環境情報研究院

E-mail: {kawaguchi,mori}@forest.eis.ynu.ac.jp

本稿では、既存の一問一答型質問応答システムを用いて、関連する質問群に解答する手法を提案する。関連質問群間には照応詞、ゼロ代名詞等の照応表現により結束関係が存在する。そこで提案手法では、まずこうした照応表現を探索する。次に照応表現が照応する先行詞を、格フレーム辞書、シソーラスを用いて絞り込み、補完された質問文候補群を生成する。このとき、いずれの補完質問文候補を採用するかという決定に際して、一問一答型質問応答システムの解スコアを用いる。

評価実験によれば、一問一答型質問応答の解スコアの利用が、補完された質問文の決定、解の抽出精度向上に有効であることが示された。

Answering Contextual Question Using a Non-Contextual Q/A System

Shinpei KAWAGUCHI Tatsunori MORI

Graduate School of Environment and Information Sciences,
Yokohama National University

In this paper, we propose a question answering system to answer a series of related questions using a non-contextual Q/A system. The related questions usually have cohesive relation to each other by anaphoric expressions like zero pronouns. First of all, the proposed method detects such anaphoric expressions. Second, it searches for antecedent candidates corresponding to the anaphoric expressions using a case-frame dictionary and a thesaurus. In order to select the best candidate, we use an answer-score of each anaphorically-resolved question produced by a non-contextual Q/A system. The experimental results show the use of answer-score is effective to improve the accuracy of answering contextual questions.

1 はじめに

質問応答システムは、利用者による自然言語での質問に対し、電子化された文書、情報より適切な解答を導くシステムである。その中で新たな質問応答システムの形態として対話を理解し、回答できる質問応答が期待されている。ここで質問応答において対話理解できるシステムとは、「東京タワーはどこにありますか。」という質問に続く「その高さは何メートルですか。」のような単独では答えられない関連する質問群に答えられるシステムを想定している。

このような一連の質問に回答するシステムについては、加藤らの提案により、質問応答システムの評価型ワークショップ NTCIR4 QAC2 Subtask3 において各手法の性能評価が行なわれ [1]。引き続き、NTCIR5 QAC3[2] が進行中である。

本稿では、既存の一問一答型質問応答 [3] を用いて、一連の関連した質問に回答する一手法を提案する。その手法としては、まず、照応表現検出、意味属性に基づく先行詞候補の検出を行ない、その候補に基づいて補完された質問文候補群を生成する。最も適切な質問文を選択する手法とし

て、補完された質問文候補群を既存の一問一答型質問応答システムの入力とし、その解のスコアが大きいものを採用する方法を提案する。解のスコアは知識源に対する質問文の整合性を表すと考えられることから、この手法は、知識源に対して最も整合性が高い補完候補を選択するという考え方に基づいている。

2 先行研究

2.1 一連の質問に回答する質問応答

国立情報学研究所主催の評価型ワークショップ NTCIR4 QAC2 Subtask3 では、図 1 に例示されるような一連の質問列 (質問シリーズ) に対し、解を過不足なくリストにして返す質問応答 (リスト型質問応答) の評価が行なわれた。このワークショップに 14 のシステムが参加し、各種提案がなされている。

提案された手法は大別すると二手法に分けられる。

一つは、文書検索時の検索質問生成段階で、以前の質問に現れる語を加えるなどして、対象文書を絞り込む手法である。高木 [4] は、現在の質問文から抽出された語に加えて、一つ前の質問文から抽出された語を、検索質問として

ミュージカル「ライオンキング」の上演は
何年に始まりましたか。
台本を書いたのはだれですか。
音楽はだれが担当しましたか。
初演されたのはどこの舞台ですか。
だれが演出をしましたか。
日本での初演はいつですか。
その劇場はどこですか。
日本の上演ではだれが演出を担当しましたか。
どんな俳優が出演していますか。

図 1: NTCIR4 QAC2 Subtask3 における質問シリーズの例

文書検索に用いることを提案している。

もう一つは、質問文中に現れる照応表現について、その先行詞で補充し、一問一答型の質問応答システムの入力とする手法である。福本ら [5] は質問文を形容詞表現の省略型、主題部の省略型、代名詞の補充型の 3 つの省略タイプに分け、形容詞表現、主題部の省略型はそれぞれ、形容表現、主題を補い、代名詞の補充型は各代名詞に応じて省略語を推定している。

我々の提案手法は後者の手法に分類される。しかし、我々の提案手法は、先行詞を決定する際にも既存の一問一答型質問応答システムを利用し、その解のスコアを用いている点において、先行研究と異なっている。

2.2 照応解析

前述したように、我々の手法は照応表現を先行詞で補い質問応答システムに入力する手法であるため、照応解析が不可欠である。照応解析は照応詞の検出と、検出した照応詞の先行詞を同定する処理（先行詞同定）からなる。

2.2.1 照応詞の検出

照応詞には指示詞、代名詞、ゼロ代名詞（格要素の省略）がある。特にゼロ代名詞の検出は文脈解析、音声対話などでも重要であり、様々な観点から研究されている。その中でも主な手法として格フレーム辞書を用いた検出方法がある。

格フレーム辞書は、日本語の用言に対し、その用言が取る文型パターンを示しており、IPAI 動詞辞書 [6]、日本語語彙大系 [7] などがある。このような格フレーム辞書を用い、文型パターンと入力文の用言に対する係り受け構造を比較し、充足されていない格要素をゼロ代名詞の候補として検出する。関ら [8] は IPAI 動詞辞書を用い、照応詞の検出を行なっている。

2.2.2 先行詞同定

先行詞同定は、検出された照応詞の照応対象となる先行詞の決定を行なう処理である。

河原ら [9] は、コーパスから自動構築した格フレーム辞書を用いた省略解析の研究を行なっている。その先行詞同定処理は、先行詞候補探索後、格フレーム辞書の文型パターン内格要素と候補の間の類似度と、機械学習による分類器によって先行詞を決定している。類似度計算には、日本語語彙大系 [7] を利用している。

飯田ら [10] は、機械学習による分類器に加え、言語学的知見を利用するため、Walker ら [11] のセンタリング理論を用いた先行詞らしさの選好を利用し、先行詞同定を行なう手法を提案している。先行詞らしさの選好は、

主題(ハ格) > 主語(ガ格) > 間接目的(ニ格) >
直接目的(ヲ格) > その他

という順で先行詞をとりやすいという知見に基づいている。

3 提案手法

本節では提案手法について説明する。図 2 にその概略を示す。本手法では、大きく分けて以下の 3 つの処理で、関連質問群の中の 1 つの質問に対する回答を得る。

1. 現在の質問文内の照応詞に対応する先行詞候補を、直前の質問文とその解、「主題句」より収集、列挙する。
2. 現在の質問文内の照応詞を検出するとともに、その先行詞の候補を絞り込み、照応詞の補充された質問文候補群を生成する。
3. 補充された質問文候補群を一問一答型質問応答システムの入力とし、その解のスコアが最も高いものを、補充された質問文と考え、解を得る。

特に、照応解析の結果の精度の低さを考慮し、照応解析の最尤解を 1 つだけ用いるのではなく、先行詞候補により補充された質問文候補を複数生成し、これらに対する一問一答型質問応答システムの解のスコアを利用することにより、精度の向上を図っている。これは一連の対話の中での曖昧性を、知識源となる文書との整合性により解消することと捉えることもできよう。

3.1 処理の流れ

本節で提案手法の処理を、実例として図 1 の質問シリーズを用いて示す。

質問文 1 ミュージカル「ライオンキング」の上演は
何年に始まりましたか。

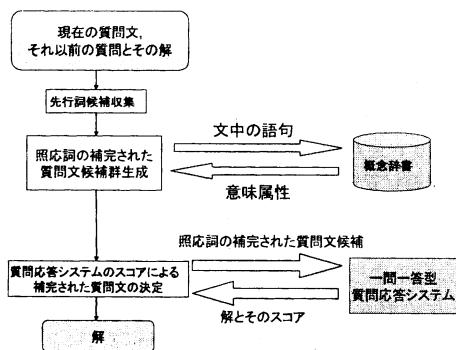


図 2: 提案手法の概略

上記が最初の質問文となる。最初の質問文に対しては、これに続く一連の質問文群で主題となる「主題句」(節 3.2 詳述) を収集し、その意味属性 (意味的な範疇) をシソーラスより得る。この例では、表 1 のように収集される。ここで、意味属性は日本語語彙大系の属性番号と名称により表記している。次に、この最初の質問文を質問応答システム

表 1: 収集された主題句とその意味属性

主題句	意味属性
ミュージカル「ライオンキング」の上演	1992 興行
ミュージカル	1059 見世物 1671 歌劇
「ライオンキング」	1 (未定義)
上演	1992 興行

の入力とし、リスト型質問応答処理 [12] により解のリスト¹を得る。得られた解についても、同様に意味属性を求めておく。その結果を表 2 に示す。

表 2: 解とその意味属性

解	意味属性
2000年	2679年
1995年	2679年

最初の質問の処理が終了したので、次の質問の処理へ移行する。

質問文 台本を書いたのはだれですか。

この質問文に対して、まず、先行詞候補の収集を行なう。その結果、表 1、表 2 にある語句が収集される。なお、この処理は、節 3.2 で詳述する。

次に照応詞の検出を行なう (節 3.3 で詳述)。質問文には指示詞、代名詞がないので、ゼロ代名詞のみを検出する。

¹システムが解であると判定した解。一般には複数である。

すなわち、動詞に対する格フレームを格フレーム辞書から得て、その格要素の対応からゼロ代名詞を検出する。同時にゼロ代名詞に対する意味属性も格フレームから求める。例えば質問文 2 の場合には、動詞「書く」に対する格フレームを調べることで、次のようにデ格の省略が発見される。ただし、N1 は省略されているゼロ代名詞に対応する。

N1 で台本を書いたのはだれですか。

N1:1109 文書 1064 言葉

ここで得られたゼロ代名詞に対応する名詞句の意味属性と、収集した先行詞候補の意味属性について、概念階層上の類似度計算を行なうことにより先行詞候補を絞り込む (節 3.3 詳述)。その結果、先行詞候補は「ライオンキング」「ミュージカル」の 2 つに絞り込まれ、照応詞の補完された質問文候補は、

1. 「ライオンキング」で台本を書いたのはだれですか。
2. ミュージカルで台本を書いたのはだれですか。

の 2 候補になる。この候補群を、補完の整合度順に一問一答型質問応答に入力し、解のスコアの最も高い質問文候補を、最も適切に補完された質問文として決定、その解を、質問に対する解答として決定している。この処理については、節 3.4 で詳述する。

3.2 先行詞候補の収集

先行詞候補として以下のものを収集する。

- 直前の質問文の名詞句
- 直前の質問文の解
- 「主題句」

一連の質問文群 (質問シリーズ) を通じて主題となる名詞句を「主題句」と呼ぶことにする。最初の質問文に含まれる語句が全体の主題となりやすいことより、本稿では、最初の質問文の名詞全てを「主題句」として扱った。これに加えて、「ハ」でマークされる名詞句が最初の質問文の主題を表現するので、「ハ」でマークされる名詞句全体も「主題句」とした。ここで、「ハ」でマークされる名詞句に含まれている個々の名詞も、名詞句全体とは別に収集されている。

なお、当然ながら質問文群において主題が遷移することがある。本来ならば、それを適切に処理しなければならないが、本稿では第一次近似として、照応表現の意味属性と主題句の意味属性の類似度 (後述) が低いときには、主題が遷移したものと判断し、それ以降の質問文では先行詞候補として扱わないこととした。

3.3 補完された質問文候補群の生成

照応表現が補完・解決された質問文候補群生成の過程を図3に示す。質問文中の照応詞の検出と、それに照応可能な先行詞候補の絞り込みを行ない、絞り込まれた先行詞候補を用いて補完された質問文候補群が生成される。

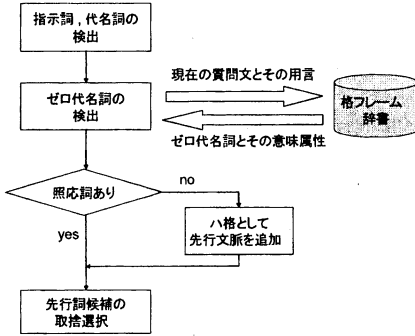


図3: 補完された質問文候補群の生成の流れ

照応詞の検出においては、指示詞、代名詞、ゼロ代名詞を検出している。ゼロ代名詞の検出には節2.2.1で述べた、格フレーム辞書を利用したゼロ代名詞検出手法を用いた。格フレーム辞書として、本稿では、日本語語彙大系[7]を用いた。このときゼロ代名詞に対応する格要素の意味属性を利用するが、同様に指示詞、代名詞についても格フレーム辞書を用いて対応する格要素の意味属性を用いる。また、指示詞、代名詞が存在せず、用言も存在しない(格フレーム辞書を利用できない)場合には、先行文脈を強制的に追加するために「ハ」でマークされた名詞句が省略されると仮定する。

照応詞、ゼロ代名詞の各々について、それに対応する格フレーム中の格要素の意味属性をもとに、先行詞候補との概念階層上の類似度を計算し、候補の絞り込みを行なう。類似度計算には、河原ら[9]の先行詞同定手法と同様に日本語語彙大系を用いる。類似度の計算式を以下に示す。

$$\text{sim}(x, y) = \begin{cases} \frac{2 \times l_{xy}}{l_x + l_y} & \text{if } x \not\subseteq y \\ 1 & \text{if } x \in y \end{cases}$$

ただし、 l_x, l_y は概念 x, y のシソーラス上の深さ、 l_{xy} は、 x, y の共通祖先までのシソーラス上の深さである。この類似度に閾値を設け、閾値以上の先行詞候補を残す絞り込み処理をしている。残った先行詞候補により、質問文候補群を生成する。

3.4 一問一答型質問応答システムによる質問文候補の決定

節3.3に示した処理により照応詞が補完された質問文候補群が生成されるが、これらを既存の一問一答型質問応答システム[3]に入力し、それぞれの解のスコアを得る。このスコアの最も高い質問文候補を、最も適切に補完された質問文として決定している。これは、解のスコアが高い質問文は、知識源の記述と整合性が高い記述であり、知識源から根拠のある確からしい解を導き出したものであるという考えに基づいている。

しかし、補完された質問文候補群が多数生成された場合には、質問応答に要する処理時間が大幅に増加してしまう。そこで補完された質問文に補完の整合度という尺度を設け、整合度の高い候補のみを用いることにする。補完の整合度は、質問文中の照応詞(の意味属性)とそれに対して補われた先行詞(の意味属性)との間の概念階層上の類似度の和とした。これは上記類似度が高いものほど意味的に正しく補完が行なわれていると考えられるからである。また、固有表現は先行詞となりやすいという観察より、先行詞が固有表現であり、かつ、類似度が1の場合(照応詞の意味属性が補われた先行詞の意味属性と一致、もしくは完全な親子関係にある場合)には、補完の整合度を高くしている。文 S の補完の整合度 $C(S)$ の計算式を以下に示す。

$$C(S) = \sum_{\langle r_i, a_i \rangle \in \text{resolv}(S)} \text{sim}_2(a_i, r_i)$$

$$\text{sim}_2(x, y) = \begin{cases} \frac{2 \times l_{xy}}{l_x + l_y} & \text{if } x \not\subseteq y \\ 1 & \text{if } x \in y \text{ かつ } \\ & \text{ } x \text{ が固有表現ではない} \\ 1.5 & \text{if } x \in y \text{ かつ } \\ & \text{ } x \text{ が固有表現である} \end{cases}$$

ただし、 $\text{resolv}(S)$ は文 S 中の照応詞とその先行詞の組の集合である。

整合度の高い補完された質問文候補だけを一問一答型質問応答の入力とすることで、処理時間の短縮と、整合性の低い補完文の選択を防ぐことを可能にしている。

最後に、質問に対し過不足ない解答を求めるために、決定した補完された質問文を、石下ら[12]の提案するリスト型質問応答処理の入力とし解を得ている。

4 実験と評価

4.1 実験目的と評価手法

提案手法における、先行詞同定の精度と質問応答の精度の評価を目的とし、実験を行なった。予備実験より、一問一答型質問応答の入力は補完の整合度上位4件とした。

ベースラインとして解のスコアを利用しない手法と、節2.2.2で述べたセンタリング理論における選好規則を用

いる手法の二つを用意した。

- 解のスコアを利用しない手法
補完された質問文の決定に際して一問一答型質問応答を利用せず、補完の整合度が最も高いものを、補完された質問文とする方法であり、その他の処理の流れは提案手法と同一である。
- センタリング理論を用いる手法
先行詞同定に Walker ら [11] のセンタリング理論を用いた先行詞らしさの選好を用いて、質問文を補完する。この時、選択された候補が疑問詞であれば、その質問の解を先行詞としている。ゼロ代名詞の検出には提案手法と同様に格フレーム辞書、日本語語彙大系 [7] を、先行詞として意味的矛盾が生じないかの判断基準にも提案手法同様に、概念階層上の類似度を用いている。

実験は、テストセットとして NTCIR4 QAC2 Subtask3[1] の質問文 36 シリーズ、251 問を、知識源として 1998 年、99 年の読売新聞、毎日新聞の新聞記事 2 年分を用いて行なった。

4.2 実験結果

先行詞同定精度の評価指標として以下の式で定義される指標を用いる。

$$\begin{aligned} \text{再現率1} &= \frac{\text{先行詞同定に成功した照応詞数}}{\text{正解照応詞数}} \\ \text{適合率1} &= \frac{\text{先行詞同定に成功した照応詞数}}{\text{システムが出力した照応詞数}} \\ F\text{値1} &= \frac{2 \times \text{再現率1} \times \text{適合率1}}{\text{再現率1} + \text{適合率1}} \end{aligned}$$

解の抽出精度の評価指標として以下で定義される「再現率2」「適合率2」「MF 値」を用いる。

質問文全体にわたる解の抽出精度の評価指標は以下の式で定義される。

$$\begin{aligned} \text{再現率2} &= \frac{\text{全ての質問通しての抽出に成功した正答数}}{\text{全ての質問通しての全正答数}} \\ \text{適合率2} &= \frac{\text{全ての質問通しての抽出に成功した正答数}}{\text{全ての質問通してのシステムの回答数}} \end{aligned}$$

個別の質問におけるリスト型質問応答の精度の評価指標は、以下のように定義される。

$$\begin{aligned} \text{再現率3} &= \frac{\text{個別の質問において抽出に成功した正答数}}{\text{個別の質問における全正答数}} \\ \text{適合率3} &= \frac{\text{個別の質問において抽出に成功した正答数}}{\text{個別の質問におけるシステムの回答数}} \\ F\text{値3} &= \frac{2 \times \text{再現率3} \times \text{適合率3}}{\text{再現率3} + \text{適合率3}} \end{aligned}$$

「修正 F 値」は、「F 値3」において、同じ解答もしくは同じものを表現する異なる表現を複数リストに含めた場

合は、そのうちひとつだけを正解とし、それ以外は誤答としている。また、正解のない質問には空リストを返した時にのみ 1.0 が与えられ、それ以外の場合はずべて 0 としている。「MF 値」は「修正 F 値」の全ての質問通しての平均値である。

表 3 に先行詞同定の精度を示す。

表 3: 三手法における先行詞同定精度の比較

手法	再現率 1	適合率 1	F 値 1
提案手法	0.341	0.271	0.289
QA スコア不使用	0.351	0.292	0.305
センタリング理論	0.300	0.324	0.331

この表によれば、センタリング理論を用いた手法は、先行詞同定精度が最も高いことがわかる。特に、他の二手法に比べて適合率が高くなっている。これは、先行詞候補の切り出し方の違いによる。センタリング理論を用いた手法は、文節ごとに切り出したため、先行詞候補が少なく、文脈中に先行詞が出現しないゼロ代名詞が適切にゼロのままになっている。これに対し、解のスコアを使わない手法、提案手法は先行詞候補が多く(節 3.2)、文脈中に先行詞が出現しないゼロ代名詞も誤って補完してしまうため、適合率が低くなっている。

表 4 は QAC2 Subtask3[1] の正解基準に基づいた解の抽出精度であり、表 5 は QAC2 Subtask3 の正解基準とは表層は異なるが記事内容より正解と考えられるものを含めた解の抽出精度である。

表 4,5 によれば、解の抽出精度は提案手法が最も高く、解のスコアを使わない手法もセンタリング理論を用いた手法を上回っている。その理由としては、提案手法の先行詞同定が再現率を重視する手法であることと、採用している一問一答型質問応答システムにおける文照合過程の頑健性が考えられる。同システムにおいては、質問文と知識源に現れる文の間の照合をする際に、対応する格助詞の一致についても考慮をして解スコアを計算している。しかし、この時、仮に格助詞が一致しなかつたとしても、解スコアは完全に 0 になるわけではなく、若干の減点に留まる。一方で、質問応答の初期段階では知識源内での文書検索・パッセージ抽出が行なわれ文書が絞り込まれるが、その際には質問文の中に現れる格助詞等の機能語は捨て去られ、内容語のみが検索質問として利用される。そして、この段階において答を含む関連文書が全く検索されないと、それ以降の過程で答を見つけることは当然できない。

この状況において、適合率重視の先行詞同定手法においては、質問文としての補完は適切に行なえていても、上記の文書検索・パッセージ抽出の段階で重要なパッセージを取りこぼしてしまう可能性が高くなる。一方で、再現率重視の手法においては、文照合の過程でスコアの減点が行なわれるものの、初期段階での取りこぼしが少な

表 4: 解の抽出精度の比較 (QAC2 Subtask3 解答セット参照)

手法	再現率 2	適合率 2	MF 値
提案手法	0.172	0.157	0.236
QA スコア不使用	0.170	0.157	0.224
センタリング理論	0.167	0.142	0.208

いと考えられる。

解のスコアを利用しない手法と提案手法とを比較すると、表 3によれば、前者のほうが再現率、適合率ともに高い。一方で、表 4.5によれば、提案手法のほうが求解精度が高い。これは、先行詞同定の精度を多少犠牲にしても、知識源との整合性(一致の度合)の高い補完質問候補を選ぶほうが結果として求解精度が高くなるということを示唆する結果である。

ここで、提案手法において、ゼロ代名詞をどのように誤ったり過剰に補完しているのかを調べてみる。その原因には主に以下の 2 項目があった。

- 文脈中に先行詞が出現しないゼロ代名詞を誤って補完している
- 他のゼロ代名詞の先行詞が誤って別のゼロ代名詞に補完されている

いずれも質問文の補完としては誤りであるにもかかわらず、解の抽出に成功した例が多く見られた。それぞれ具体的に述べる。

文脈中に先行詞が出現しないゼロ代名詞を誤って補完している

質問文 どこから連れてこられましたか。

という質問文において、ゼロ代名詞が正しく補完されたとき質問文は、「佐渡トキ保護センターで 9 年 5 月に生まれたトキの優優の両親、「友友」と「洋洋」はいつ日本へ連れてこられましたか。」となる。しかし、提案手法は以下のように補完されてしまった。

友友が日本に佐渡島をどこから連れてこられましたか。

「ヲ格」のゼロ代名詞は文脈中には存在しないため省略されるはずが、「佐渡島を」と、誤って検出、補完されている。しかし、この「佐渡島」という語が文書検索等において解を導く手がかりとなり、正しく答えることができていく。

他のゼロ代名詞の先行詞が誤って別のゼロ代名詞に補完されている

質問文 どこで生まれましたか。

表 5: 解の抽出精度 (表層は異なるが記事内容より正解と考えられるものを含む)

手法	再現率 2	適合率 2	MF 値
提案手法	0.180	0.164	0.247
QA スコア不使用	0.176	0.162	0.233
センタリング理論	0.172	0.147	0.218

という質問文において、ゼロ代名詞が正しく補完されたとき質問文は、「棋士・佐藤康光はどこで生まれましたか。」となる。しかし、提案手法は以下のように補完されてしまった。

生年月日が棋士・佐藤康光にどこで生まれましたか。

「生年月日」は誤って補完されている。²本来は「ヲ格」に「棋士・佐藤康光」が選択され、「二格」は省略されなければならない。こうしたゼロ代名詞の格助詞の誤りがあるにもかかわらず、「棋士・佐藤康光」を手がかりに正解を導いている。

4.3 考察

本節では、実験結果に基づき、提案手法と他の二手法をより詳細に比較し、考察を行なう。

4.3.1 提案手法と解のスコアを使わない手法との比較

表 6 に二手法の比較から、補完された質問文の決定に、解のスコアを使うことにより正解を取得できるようになった問題の数を「取得成功数」として、取得できなくなった問題の数を「取得失敗数」として示す。ここではリストの中に 1 つでも正解が含まれるようになれば成功、リストの中から正解が 1 つでも外れれば失敗としている。

表 6: 一問一答型 QA 利用による解取得問題数の差

	問題数 (/251 問)
取得成功数	15 問
取得失敗数	13 問

表 4.5 においては全ての質問シリーズを総合すると提案手法の MF 値の方が大きく、一問一答型質問応答を利用することの有用性が示されている。一方で、個別の質問について見てみると、表 6 に示すように、解のスコアを用いることで解を取得できるようになった問題数が純増

²ここではゼロ代名詞「ヲ格」の意味属性「2443 関連」と「生年月日」の意味属性「2678 年月日」の類似度が高く「生年月日」を選択、同様に「二格」の意味属性「4 人」と「棋士・佐藤康光」の意味属性「48 男」の類似度より、「棋士・佐藤康光」が選択されている。

しているわけではなく、逆に取得できなくなった問題もあるという点に注意する必要がある。取得できなくなった例を以下に示す。

取得できなくなった例

質問文 98年で何回目を迎えますか。

上記の質問に対し、正しい補完文は「「お水取り」は98年で何回目を迎えますか。」である。この時、以下に示す補完の整合度順に補完された質問文候補群4件が一問一答型質問応答の入力となった。

1. 「お水取り」の行が98年で何回目を迎えますか。
2. 「お水取り」が98年で何回目を迎えますか。
3. お水取りが98年で何回目を迎えますか。
4. 奈良が98年で何回目を迎えますか。

解のスコアは「奈良が98年で何回目を迎えますか。」が最も高くなり、正しい解は取得できない。しかし、一問一答型質問応答を用いない手法では、補完の整合度が最も高い「「お水取り」の行が98年で何回目を迎えますか。」が補完された質問文となり、先行詞として整合性がとれているとともに、正しい解も導き出している。一問一答型質問応答の解のスコアにより正しい補完文が選択されなかった。

このように、一問一答型質問応答を利用することで取得できなくなってしまう、または、補完された質問候補群に正しい補完文があっても選択されないことがあるため、補完質問文の決定を一問一答型質問応答システムの解スコアのみで行なう手法には、改良の余地がある。例えば、現在は決定の際に、1位の解のスコアしか用いていないが、これを1件だけではなく候補集合のスコアの分布を調べ決定する方法が考えられる。また、言語学的知見も考慮してみることも考えられる。センタリング理論を用いて先行詞の絞り込みをより厳密に行なうことや、最終決定にも解のスコアだけでなく、センタリング理論において優先度の高いものより重みをつけ、この重みと解のスコアの両方を考慮して決定する方法などである。

4.3.2 提案手法とセンタリング理論を用いた手法の比較

提案手法とセンタリング理論を用いた手法との比較については、節4.2でゼロ代名詞の誤検出、過剰補完と、それによる解の抽出精度の向上について述べた。センタリング理論の解の抽出精度が低くなる原因には、先行詞同定に失敗し、手がかりとなる語を失うと、一連の質問文群ではそれ以降の質問文においても、その語を失うため連続して解を取得できないという点が挙げられる。これに対しても提案手法では、節3.2で述べた主題句を設けることで

対処しており、解の抽出精度の高さより、その有効性が証明された。

5 おわりに

本稿では、既存の一問一答型質問応答システムを利用した関連する質問群に解答するシステムを提案した。一問一答型質問応答システムの解のスコアを用いて、質問文の補完候補を選択することにより、解の抽出精度が向上することを示した。

今後の課題としては、一問一答型質問応答システムを用いた補完文の決定にも曖昧性が残ることから、解集合のスコア分布を調べる方法や、センタリング理論のような、照応解析における言語学的知見を用いるなどの方法で、曖昧性を解消し、精度の向上を図ることなどが挙げられる。

参考文献

- [1] T.Kato, J.Fukumoto and F.Masui. Question Answering Challenge for Information Access Dialogue. In *Working Notes of the Fourth NTCIR Workshop Meeting*, pages 291-297, 2004.
- [2] NTCIR5 Question Answering Challenge(QAC-3) <http://www.nlp.is.ritsumei.ac.jp/qac/index-e.htm>
- [3] Tatsunori Mori. Japanese Q/A System using Λ^* Search and Its Improvement: Yokohama National University at QAC2. In *Working Notes of the Fourth NTCIR Workshop Meeting*, pages 345-352, 2004.
- [4] Toru Takaki. NTT DATA Question-Answering Experiment at the NTCIR-4 QAC2. In *Working Notes of the Fourth NTCIR Workshop Meeting*, pages 402-405, 2004.
- [5] J.Fukumoto, T.Niwa, M.Itoigawa and M.Matsuda. Rits-QA: List answer detection and Context task with ellipses handling. In *Working Notes of the Fourth NTCIR Workshop Meeting*, pages 310-314, 2004.
- [6] 情報処理振興事業協会技術センター. 計算機用日本語基本動詞辞書 IPAL(Basic Verbs) 説明書. 1987.
- [7] NTT コミュニケーション科学研究所. 日本語語彙大系. 岩波書店, 1997.
- [8] 関和広, 藤井敦, 石川徹也. ゼロ代名詞の検出と補完を統合した確率的照応解消モデル. 言語処理学会第8回年次大会発表論文集, pages 591-594, 2002.

- [9] 河原大輔, 黒橋禎夫. 自動構築した格フレーム辞書と先行詞の位置選好順序を用いた省略解析. 自然言語処理, Vol.11, No.3, pages 3-19, 2004.
- [10] 飯田 龍, 乾 健太郎, 松本 裕治. 文脈の手がかりを考慮した機械学習による日本語ゼロ代名詞の先行詞同定. 情報処理学会論文誌, Vol 45, No.3, 2004.
- [11] Marilyn Walker, Masayo Iida, and Sharon Cote. Japanese Discourse and the Process of Centering. *Computational Linguistics*, Vol.20, No.2, pages 193-232, 1994.
- [12] 石下円香, 森辰則. 解候補スコアの分布を利用したリスト型質問応答. 言語処理学会第11回年次大会, pages 1088-1091, 2005.