

照応解析に知識源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 環境情報学府

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

森辰則[‡]

[‡] 横浜国立大学 大学院 環境情報研究院

1 はじめに

質問応答システムは、自然言語で与えた質問に対し、文書群より適切な解答を導くシステムである。その中でシステムの新たな形態として対話を理解し、回答ができる文脈依存型質問応答が期待されている。ここで文脈依存型質問応答とは、「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。」関連する質問群に答えられるシステムである。

本稿では、既存の文脈処理のない一問一答型質問応答システムを用いて、一連の関連した質問に回答する一手法を提案する。一般に後続する質問には照応表現が含まれるが、その先行詞を検出し、適切に補完することができれば、一問一答型の質問応答システムでも後続質問に答えられるであろう。上記の照応補完は通常、i) 省略や指示表現を格フレーム辞書等の言語資源を用いて検出する、ii) 各照応表現の先行詞を同定する、という手順となるが、各段階において曖昧さが生じる。例えば、一つの動詞が複数の格フレームを持ち得るため、ゼロ代名詞の認定に複数の可能性があるし、照応表現の先行詞候補が複数存在することが普通である。この曖昧性を解消するために、従来、文脈との結束性に注目した照応解析の研究が数多くなされてきた。すなわち、新しい文に対して、先行文脈との結束性が高くなる解釈を与える。この種の方法論は有効に働くことが多いが、省略検出における格フレーム選択の問題などすべての曖昧性を解消するものではない。そこで、本稿では知識源との結束性という新しい概念に基づき照応解析の曖昧性を解消する手法を提案する。質問応答システムは、ある質問を解釈する際に、それ以前の質問群はもちろんの事、内蔵する知識源も参照可能である。また、質問応答処理それ自身が、質問文と最も首尾一貫した文脈を持つ語句を知識源から見つけることである。そこで、質問文の解釈と質問応答処理を並行して行なって、答が適切に見つかる質問文の解釈を選び、その時の答を最終的な答とする方法を検討する。

2 先行研究

関連質問群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 NTCIR-5 QAC3[1] などにおいて各手法の評価が行なわれている。そこで提案されている手法は大別すると二つに分けられる。一つは、文書やパッセージ(文書の断片)を検索する際に、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。例えば、Murata et al.[2] は、過去の質問文の一部や質問の回答を「履歴」と考え、答を探索する範囲となるパッセージを、その履歴との類似度が高くなるように検索する手法を提案している。

二つめは、文脈との結束性の観点から質問文を補完し、通常の質問応答処理を行なう手法である。Matsuda et al.[3] は質問文における参照関係を、代名詞によるもの、ゼロ代名詞によるもの、修飾表現もしくは被修飾表現の省略によるものの3つに分類し、選択制限や文内での登

場位置等の情報に基づき先行詞を一意に決定している。そして、得られた解釈に基づき一つの補完質問文を生成、質問応答を行なっている。

我々の手法は後者の方法論に分類されるが、知識源との結束性を考慮している点が大きく異なる。

3 提案手法

提案手法は、以下の四つの処理を行ない、関連質問群の個々の質問に回答する。

1. 照応表現の検出
2. 照応表現の補完された質問文候補群の生成
3. 知識源による補完された質問文の決定、解の取得
4. 先行詞候補リストの更新

先行文脈から取得している先行詞候補リストと、現在の質問文を本手続きの入力とする。先行詞候補リストには先行文脈から抽出した名詞や名詞句が保存されており、以下で述べる照応表現の補完手法により内容が異なる。また、最初は空である。

まず、格フレームを用いて、質問文より、ゼロ代名詞、指示詞といった照応表現を検出する。一般に、一つの動詞に対して格フレームが複数存在するので、照応表現の検出結果を付与した質問文解釈の候補は複数になる。

次に、上記質問文解釈候補の各々について、検出された照応表現を補完する語句を先行詞候補リストより選択し、照応表現が補完された質問文候補群を生成する。我々の方法では、この処理の後に知識源との結束性に基づく判定が存在するので、この段階では複数の可能性に対応する複数の質問文候補を生成してもよい。そのため、先行詞の補完候補の選定には自由度が存在する。そこで、知識源との結束性の判断の前処理として、文脈に基づいてどのような制約を適用することが適しているかを調べるために、a) 言語学的知見を用いる手法、b) 関連質問群の特性を利用する手法、c) 最低限の選択制限のみを用いる手法の三つの補完質問文生成手法を検討する。また、補完候補が非常に多くなる場合が存在するので、格フレームから得られた意味属性と先行詞候補の意味属性から計算される文補完の整合度を定義し、その値により可能性の低い候補を削除することも検討する。

照応表現の補完された質問文候補について、各々、知識源との結束の度合を測り、最良の補完質問文候補の一つを選択する。そして、これを一問一答型の質問応答システム[4]の入力とし、リスト型質問応答処理[6]により解のリスト¹を得る。ここで、我々は、与えられた質問文に対する知識源に対する結束性の度合がその質問に対して求めた解の「良さ」と相関があると仮定し、前者の尺度として後者を利用する。これは、質問応答処理それ自身が、質問文と最も首尾一貫した文脈を持つ語句を知識源から見つけることであり、各解に付随するスコアがその

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。

結束性の指標であるべきであるからである。解の「良さ」を測る尺度としては、a) 解のスコアの最大値を用いる手法と、b) 解候補群のスコア分布を考慮する手法[6]を検討する。

上記手続きの最後の段階においては、次の質問のために、そこまでの文脈から得られた先行詞候補のリストを作成し、次の質問の処理へと移行する。

上記の処理の流れを、照応表現が補完された質問文候補群の生成に言語学的知見を用いる手法を、知識源との結束性に基づく補完質問文の決定に解候補群のスコア分布を考慮する手法のそれぞれを用いた場合を例にして、以下に説明する。なお、詳細な説明は節 3.2以降で述べる。

3.1 例による処理の説明

以下の質問文 1 ~ 3 が与えられたとする。

質問文 1 チャールズ皇太子はどこ皇太子ですか。

最初の質問文は、照応表現を補完する必要がないので、そのまま一問一答型質問応答システムの入力とし、リスト型質問応答処理により、{ 英国, 英 } という解のリストが得られる。

次に最初の質問文より次の質問に対する先行詞候補リストを作成する。言語学的知見を用いる手法では、センタリング理論を基礎とした Salient Referent List (SRL)[5] に基づく照応解析を行なっている。そのため、先行詞候補リストは SRL と同等の内容を保持するリストとなる(節 3.3で詳述)。² ただし、SRL に保存すべき要素が疑問詞であればそれを解に置き換えてから保存している。また、保存する際に、それぞれの要素の意味属性もシソーラスより獲得し、付加する。以下にこの時点での先行詞候補リストを示す。括弧の中はそれぞれの意味属性である。

主題：チャールズ皇太子 [男][息子][貴人]

質問文 2 ダイアナ妃といつ結婚しましたか。

二問目であるので、照応表現の検出(節 3.2で詳述)を行なう。ここでは指示詞等の参照表現がないので、ゼロ代名詞の検出のみが行なわれる。ゼロ代名詞は、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。質問文 2 では、動詞「結婚する」に対する格フレームを調べると、以下のが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 がダイアナ妃といつ結婚しましたか。
N1 : [人][動物]

得られた照応表現の各々について、その先行詞を選定し、補完された質問文候補の生成を行なう(節 3.3で詳述)。その結果、ここでは先行詞候補リストに基づき「チャールズ皇太子」がゼロ代名詞の補完候補となるので、照応表現の補完された質問文候補は、

チャールズ皇太子がダイアナ妃といつ結婚しましたか。

になる。ここでは、補完質問文候補が 1 つしか生成されなかったため、これを質問として一問一答型質問応答により解のリストを得る。解として { 81年 } が得られる。

次に先行詞候補リストを作る。質問文 2 より、「いつ」に対する解と、「ダイアナ妃」を「その他」のスロットに入れ、それぞれ意味属性を取得する。結果のリストは以下ようになる。

² 節 3.3の脚注に示すように、判定詞「だ」については特別に扱っている。

主題：チャールズ皇太子 [男][息子][貴人]
その他：ダイアナ妃 [妻][女][貴人], 81年 [年][期間]

質問文 3 子供は誰ですか。

この質問文に対して、照応表現の検出を行なうが、指示詞や代名詞、ゼロ代名詞も存在しない。よって先行文脈を引き継ぐために「ゼロ主題」(‘ハ’でマークされたゼロ代名詞)を質問文に付加し次文を得る。

N1 は 子供は誰ですか。

ゼロ主題を強制的に追加するとき、ゼロ代名詞に対応する名詞句の意味属性がわからず、対応する格もない。このとき、意味属性が抽象関係以外の先行詞候補の要素全てを補完候補とする³。照応表現の補完された質問文候補は、

1. チャールズ皇太子は子供は誰ですか。
2. ダイアナ妃は子供は誰ですか。

の 2 候補になる。この候補群を一問一答型質問応答の入力とし、知識源との結束性に基づき補完文の候補を決定する(節 3.4で詳述)。解の良さを測る指標をスコア分布に基づいて定義し、この指標が最大となる質問文候補を、最も適切に補完された質問文として決定する。節 3.4で定義する指標に基づき、表 1 に解の良さを値として示す。

表 1: 質問文候補とその解の良さの指標の値

補完質問文候補	解の良さの指標
チャールズ皇太子は子供は誰ですか。	0.068
ダイアナ妃は子供は誰ですか。	0.046

解の良さの指標が最大の補完質問文候補は、「チャールズ皇太子は子供は誰ですか。」であり、これを最も適切に補完された質問文として決定、その解を質問応答処理により求め、{ ウィリアム, ダイアナ, ラトビア } が得られる。

この後に質問文から先行詞候補リストを作り、次の質問に備えるという処理が繰り返される。

3.2 照応表現の検出

照応表現は、表層に現れる指示詞・代名詞の検出、ゼロ代名詞の検出、ゼロ連体格代名詞(名詞句に係る接続助詞‘ノ’)の検出と、三つの処理に分けて検出している。ゼロ代名詞の検出には格フレーム辞書を利用している。本稿では、日本語語彙大系 [8] を用いた。後の処理で利用するために、ゼロ代名詞の検出の際に、対応する格要素の意味属性も保存する。ゼロ連体格代名詞の検出は、磯江ら [7] の手法を用いている。また、照応表現が何も検出されない場合は、先行文脈を引き継ぐためにゼロ主題(‘ハ’でマークされたゼロ代名詞)を付加する。

3.3 照応表現の補完された質問文候補群の生成 先行詞候補リストの作成

言語学的知見を用いる手法

言語学的知見を用いる手法では、日本語センタリング理論を拡張した Nariyama [5] の手法を利用した。先行詞候補リストに Salience Referent List (SRL) を用いている。SRL は文脈を保持しておく記憶領域である。リストは主題のなりやすさに基づいて以下のように優先順位のあるスロットをもつ。

³ 抽象関係の意味属性をもつ名詞句は、補完しても質問文の曖昧性が大きく、「ハ’でマークされた格として補うには不適切であると考えられるため、除外している。

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

スロットに入る新しい要素が文脈中に現れれば上書きされる。文中の照応表現は上記と同じ順番で処理され、SRL中の未選択の最上位要素が先行詞となる。SRLを質問応答の文脈に対応させるため、疑問詞は求まった解に置き換えるとともに、複数解の場合には、1つのスロットに複数要素を持つことを認めるなど拡張した。⁴なお、ゼロ主題については、強制的に追加されたものであるので例外的に扱っており、SRL中の全ての要素を先行詞候補として複数の質問文を生成している。また、SRLでの優先順位に加え、意味属性に基づく照応表現と先行詞の間の意味的整合性を調べている。これは、意味属性の概念階層上の類似度計算による選択制限で実現している。類似度計算式を以下に示す。

$$sim(x, y) = \begin{cases} \frac{2 \times L_{xy}}{l_x + l_y} & \text{if } x \notin y \\ 1 & \text{if } x \in y \end{cases}$$

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、閾値以上であれば、候補とする。以下に記述する類似度計算はこの式を指す。

関連質問群の特性を利用する手法

関連質問群の特性を利用する手法では、先行詞候補に優先順位をつけず、直前の質問文の格要素全てを候補としている。また、関連質問群では、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ので、最初の質問文の格要素全てを先行詞候補リストに加えている。この最初の質問文の格要素全てを主題句と呼ぶことにする。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概念階層上の類似度が低いときには、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。また、補完された質問文候補群の生成は、照応表現の意味属性と先行詞候補との意味属性の概念階層上の類似度計算による選択制限のみで行なう。

最低限の選択制限のみを行なう手法

最低限の選択制限のみを行なう手法では、先行文脈全ての格要素を先行詞候補にしている。この手法における先行詞候補リストも優先順位をもたない。補完された質問文候補群の生成は、照応表現の意味属性と先行詞候補との意味属性の概念階層上の類似度計算による選択制限のみで行なう。

文の補完の整合度

以下に述べる知識源を利用するいずれの手法とも、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、質問文中の照応表現(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平均値とした。

⁴判定詞「だ」については「Aが(は)Bだ」においてAとBが同格の関係にあることを表すので、A,Bともに主語(主題)として扱う。またBを修飾する名詞句はAとの関連がないためリストに入れない。

3.4 知識源による補完された質問文の決定

照応表現の補完された質問文候補群について、各々、知識源との結束の度合を測り、いずれの候補が最も適切であるか判断する。ここでは、与えられた質問文に対する知識源に対する結束性の度合がその質問に対して求まった解の「良さ」と相関があると仮定し、質問文に対して一問一答型質問応答システムが返す解候補のスコアならびにその分布を用いる。解の「良さ」を測る尺度としては、解のスコアの最大値を用いる手法と、解候補群のスコア分布を考慮する手法[6]を提案する。

解のスコアの最大値を用いる手法

これは、解のスコア自身がその解の良さを表すと仮定する手法である。まず、照応表現の補完された各質問文候補を、一問一答型質問応答システムに入力し、それぞれ最上位の解のスコアを得る。そして、最も高いスコアを持つ質問文候補が知識源との結束性が高いと考え、適切に補完された質問文として決定する。

解候補群のスコア分布を考慮する手法

石下らは、優先順位型質問応答の解のスコア分布により、入力された質問の解答が適切に見つけられているかが判断可能であると述べている[6]。その手法は、まず、解候補集合のスコア分布が正解集合に対するスコア分布と、不正解集合に対するスコア分布の混合分布であると仮定し、EMアルゴリズムにより二つのスコア分布に分離する。この二つのスコア分布のパラメタを比較することで、解答が適切に見つけられているか否かを判断している。

この手法で、解答が適切に見つけられたと判断された質問は、知識源との整合性がとれた質問であると考え、照応表現の補完された質問文候補群を優先順位型質問応答[4]に入力し、正解集合と、不正解集合のスコア分布のパラメタを比較する。ここで、比較するパラメタは二つのスコア分布の平均値の差とし、その差が最大の質問文候補を適切に補完された質問文と決定する。

4 評価実験ならびに考察

NTCIR-5 QAC3[1]のテストセット(50シリーズ、計360問)を用いて評価実験を行なった。ベースラインとして、知識源との結束性を利用せず、補完の整合度の最も高い補完質問文を用いるシステムを用意した。また、精度の上限を知るために、手作業で照応解析した質問文による質問応答処理を行なった。評価指標はQAC3におけるMMF1値[1]を用いた。各手法におけるMMF1値を表2に示す。

表 2: MMF1 値の比較

先行詞の制約	スコア分布	スコア最大値	整合度のみ
言語学的知見	0.174	0.166	0.164
関連質問群特性	0.147	0.147	0.136
選択制限のみ	0.169	0.157	0.133
手作業		0.242	

言語学的知見を用いる手法とスコア分布を考慮する手法との組合せの精度が最も高いことが分かる。また、スコア分布を考慮する手法は他の先行詞の制約手法を用いても、安定して一番良い結果がでており、うまく補完された質問文の決定が行なえていると思われる。

次に、表3に言語学的知見、選択制限のみを用いた場合の失敗解析の詳細を示す。先行詞抽出失敗は、それ以前の質問のいずれかで取得に失敗した解を先行詞としていることが主な原因である。この解決には、先行詞抽出精度

表 3: 失敗段階の解析

手法	成功		失敗				
	照応成功 解答成功	照応失敗 解答成功	先行詞 抽出失敗	補完候補群 生成失敗	補完文 決定失敗	質問応答 解答不可	その他
言語学的知見+ スコア分布を考慮	11.0% (34)	13.5% (42)	26.5% (82)	21.0% (65)	9.7% (30)	18.1% (56)	0.3% (1)
言語学的知見+ スコア最大値	9.4% (29)	13.2% (41)	29.4% (91)	20.3% (63)	11.0% (34)	16.5% (51)	0.3% (1)
言語学的知見+ 整合度のみ	9.4% (29)	11.9% (37)	29.4% (91)	21.0% (65)	10.6% (33)	17.4% (54)	0.3% (1)
選択制限のみ+ スコア分布を考慮	11.3% (35)	11.3% (35)	17.4% (54)	14.5% (45)	29.4% (91)	15.2% (47)	1.0% (3)
選択制限のみ+ スコア最大値	10.6% (33)	8.7% (27)	18.1% (56)	13.5% (42)	37.1% (115)	11.0% (34)	1.0% (3)
選択制限のみ+ 整合度のみ	6.8% (21)	9.0% (28)	19.4% (60)	15.2% (47)	35.2% (109)	13.5% (42)	1.0% (3)

を向上させることはもちろんのこと、それ以外の部の性能向上が必要である。

選択制限のみの場合、文脈上の制限が緩いので、補完文決定失敗の割合は、知識源を照応解析に用いることの評価を表す直接の指標である。これを見ると、スコア分布を考慮する手法も、スコアの最大値を利用する手法も、整合度のみの場合よりは良いものの、テストセット全体の30%前後が補完文決定に失敗しており、あまりうまくいっていないように思われる。しかし、どちらの手法もそれが正しく機能するためには、システムが自らの知識源から正解を導き出せる質問であることが必要である。

知識源との結束性に基づく手法が質問文の曖昧性を解消する能力を見積もるために、a) 質問文候補が複数生成され、かつ、そのうちの1つ以上が適切な補完になっている、b) 適切に補完すれば一問一答型質問応答システムが正解を導き出せること、この二つの条件を満たす質問に限定してMMF1値を調べた。表4に示す。これより、

表 4: 知識源が曖昧性を解消する能力 (MMF1 値)

手法	言語学的知見	選択制限のみ
スコア分布	0.478 (49問)	0.436 (82問)
スコア最大値	0.437 (44問)	0.374 (83問)

表2と比較して高い精度であることが分かり、知識源を利用することの有効性が示されている。

次に、表5に関連質問群の型の違いによるMMF1値の比較を示す。表5から分かるように、スコアの最大値

表 5: 関連質問群の型の違い

手法	収集型	ブラウジング型
言語学的知見+ スコア分布を考慮	0.177	0.166
言語学的知見+ スコア最大値	0.178	0.137
選択制限のみ+ スコア分布を考慮	0.180	0.146
選択制限のみ+ スコア最大値	0.166	0.135

を見る手法は、スコア分布を考慮する手法と比較すると、同一事物を話題にしつづける収集型に対し、話題が変化するブラウジング型は性能が大きく下回る。これは、基礎としている質問応答のスコアの性質により、多くの文節を持つものほど、つまり先行文脈を多く含む質問文ほど、知識源との照合部が多くなり、解スコアが高くなることによる。このため先行文脈に影響をうけやすく、話題の変化に弱いと考えられる。対して、スコア分布を考慮する手法は、基礎としている質問応答のスコアの性質の影響は小さい。この手法と言語学的知見の組み合わせは、どちらの型でも安定して良い結果を出せる、文脈の変化にもより頑健な手法であると言える。

5 おわりに

本稿では、知識源を用いて関連質問群に回答する手法を提案するとともに、その際に、文脈でどの程度制約をつけるべきか考察した。候補の正解集合と不正解集合のスコア分布の平均値の差を用いる手法は、文脈に過度に依存することなく、知識源と照応表現が補完された質問との整合性を適切に表現しているといえる。

より解抽出精度の高いシステムを目指すには、一問一答型質問応答の精度向上が必要である。それが、相乗して適切に補完された質問を選択する性能を上げるだろう。

参考文献

- [1] T.Kato et al. An Overview of NTCIR-5 QAC3. In *Proceedings of the Fifth NTCIR Workshop Meeting*, pages 361-372, 2005.
- [2] Y.Murata et al. Question Answering Experiments at NTCIR-5: Acquisition of Answer Evaluation Patterns and Context Processing using Passage Retrieval. In *Proceedings of the Fifth NTCIR Workshop Meeting*, pages 361-372, 2005.
- [3] M.Matsuda et al. Answering Questions of IAD Task using Reference Resolution of Follow-up Questions. In *Proceedings of the Fifth NTCIR Workshop Meeting*, pages 361-372, 2005.
- [4] Tatsunori Mori. Japanese Question-answering System Using A* Search and Its Improvement. *ACM Transactions on Asian Language Information Processing*, to appear.
- [5] Shigeko Nariyama. Grammar for Ellipsis Resolution in Japanese. In *Proceedings of the 9th International Conference on TMI*, pages 135-145, 2002.
- [6] 石下円香, 森辰則. 優先順位型質問応答の解スコア分布に基づくリスト型質問応答. 情報処理学会自然言語処理研究会報告, 2005-NL-169, 2005.
- [7] 磯江健史, 竹井光子, 相沢輝明. ゼロ代名詞連体格の自動検出システム. 言語処理学会第10回年次大会, pages 765-768, 2004.
- [8] 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.