

# 照応関係を利用した日本語質問文の生成

高木 想一郎      飯田 龍      徳永健伸  
 東京工業大学 大学院情報理工学研究所  
 {takagi, ryu-i, take}@cl.cs.titech.ac.jp

## 1 はじめに

雑談のような非目的達成型の対話も含めた知的な対話システムを実現するには、相槌のような受動的な応答だけでなく、能動的に話題を提示し、円滑な対話を行う必要がある。このようなシステムを実現するために、本研究ではシステムがユーザにある話題について質問する機能を、以下の2つの部分問題に分けて処理をすることを考える。

1. これまでの対話履歴から質問の内容とすべき話題を選択する
  2. 選択された話題に対し、適切な質問文を生成する
- このうち1つ目の問題は、これまで入力されたクエリからユーザが必要とするページを推測するといった情報検索と関連すると考えられる。2つ目の問題は Web などの文章から解答を問う表現を生成する問題と考えられる。

本研究では特にこの文章からの質問生成に焦点を当てる。文章からの質問生成は一見自然言語での質問応答の逆問題のように見えるが、基本的な質問応答の処理では質問文から解答とすべき文章の範囲(パッセージ)を抽出し、その範囲内に出現する表現を解答として出力する。これに対し、質問生成の処理では文章中に散在する解答に関連する表現を選択し、選択した表現から質問文を生成する。このため、質問応答で利用されるパッセージのような限られた範囲だけではなく、文章全体から質問に必要な内容を探索する必要がある。

質問生成の課題は一般的な文章生成と同様に以下の2つの処理に分割して考えることができる [1]。

- 内容生成：文章から質問に関連する表現を収集する
- 表層生成：内容生成の結果に基づいて文法的に正しい文、もしくは文章を質問文として生成する

このうち、本稿では特に1つ目の内容生成に必要な表現の収集について議論する。既存の質問生成手法 [2] [3] では、解答を含む文にある表現のみから質問を生成している。しかし 2.2.1 で説明するように、我々が作成した評価用データの中で約7割が解答が出現していない文を参照して質問を作成する必要があることがわかる。

また質問生成の課題は、文章中の重要箇所を同定する重要文抽出の課題や情報抽出と類似している。ただし、

重要文抽出が文章中で絶対的に重要な箇所の同定を試みるのに対し、質問箇所の同定では文章全体ではなく解答となる表現との相対的な関係を考慮した重要度を考える必要がある。また、情報抽出では文章中に出現する固有表現間の関係を網羅的に捉えることを目的とするが、質問生成の場合は質問文に不必要な表現を抽出してはならないという違いがある。つまり、質問生成の場合は文章中に散在する表現の中から質問文中に出現しても冗長にならない表現を選択するという処理が必要となる。

本研究ではこの処理を文章中の照応関係を手がかりにして実現する。解答となる表現を含む文に出現する照応関係をもとにその先行詞が出現する文を参照し、その文内を探索することで、質問生成に必要な箇所を選択する。

本稿では、2節で評価に利用する評価用データの作成方法について述べ、次に3節で質問生成に利用する内容の同定手法について説明する。4節で2節を用いた評価実験について述べ、考察を行う。最後に5節でまとめる。

## 2 評価用データの作成

本研究では選択された話題に対し自然言語で記述された情報源から質問を生成することを目的とする。具体的には、新聞記事と記事内の固有表現を入力とし、与えられた固有表現が答えとなるような質問文を生成する。以降、与えられた新聞記事と記事内の固有表現の対を課題と呼び、課題で指定された固有表現のことを解答と呼ぶこととする。

### 2.1 課題の作成

評価用データを作成するために、課題を自動的に生成し、それぞれに対して人手で質問文を作成した。課題は NAIST テキストコーパス [4] 内の報道記事 1,000 記事をもとに作成した。NAIST テキストコーパスには 1995 年の毎日新聞の記事を対象に、述語項構造と共参照にタグが付与されている、固有表現と品詞の情報は CaboCha の解析結果を、係り受け関係については京都大学テキストコーパス [5] に付与された関係を参照する。課題を解答の出現位置 3 種類(最初の文、最初の文以外で記事の前

表 1: 質問文の作成に必要と判断された文節の一致

作業員 A \ 作業員 B	必要	必要でない	合計
必要	3,978	6,865	10,843
必要でない	310	29,794	30,104
合計	4,288	36,659	40,947
作業員 C \ 作業員 D	必要	必要でない	合計
必要	2,672	1,141	3,813
必要でない	1,569	34,854	36,423
合計	4,241	35,995	40,236

半, 記事の後半<sup>1)</sup>) と固有表現のクラス (地名, 人名, 組織名, 人工物名, 日付, 金額, 割合) で分類し, 各固有表現クラスに対し, 解答が最初の文に出現する場合には 25 課題, それ以外の出現位置について 50 課題ずつを用意した。

## 2.2 評価用データの作成

2.1 で用意した課題を対象に人手で質問文を作成した。質問文の作成は作業員 A ~ D の 4 人で行い, 作業員 A と B, 作業員 C と D でそれぞれ同じ 500 課題を担当した。課題から質問文を作成する際には, 記事中のどの表現が利用されたかを後で評価に利用するために, 記事中のどの範囲を質問作成に利用したかについてもタグ付与する。

以上の作業の結果, 1,000 課題中 982 課題から質問文を作成できた。残りの 18 課題は, 解答を一意に定める質問文を作成できない, などの理由により作業員が質問を作成しなかった場合であり, これは評価から除外する。

質問文の作成に必要と判断された部分の作業員間の一致数を各作業員がタグ付与した範囲を文節単位で一致しているかどうかで調査した。この結果を表 1 に示す。

作業員によって質問作成に利用した範囲は大きく異なっていることがわかる。これは, 質問に必要な表現の組み合わせに曖昧性があり, その結果作業員が異なる表現の組み合わせを選択したためである。何らかの作業指針を加えることでこの曖昧性を解消できるのかについては今後検討していく必要があるが, 本稿での評価にはこのタグ付与された結果を利用する。

### 2.2.1 必要と判断された範囲の分布

既存の質問生成手法 [2] [3] では, 解答を含む文の表現のみを用いて質問文を生成しているが, 今回作成したデータからもわかるように解答を含む文より前に登場した表現が質問作成に必要とされた課題が 704 課題であり, 全課題の約 7 割は異なる文に出現する表現を参照する必要がある。

質問文に必要と判断された表現の位置とその課題数との関係を図 1 に示す。この図より, 解答を含む文の近傍だけでなく離れた位置に出現する表現も同定する必要が

<sup>1)</sup>ここでは, 記事内の文を文単位で数えた場合に, 全文数の半分以下の箇所に出現するものを記事の前半, それ以外を記事の後半と呼んでいる。

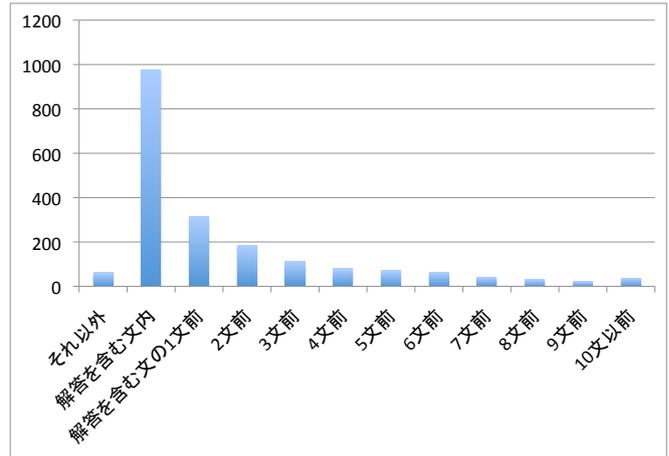


図 1: 質問文の作成に必要な表現の位置と課題数

あることがわかる。

## 3 質問生成に利用する内容の同定

2 節で作成した課題を用い, 文章中の記述のうち質問生成に利用する箇所を同定する。具体的には 2 節のデータのうち, 質問作成に利用したと作業員が判断した箇所を文節単位で選択する問題を扱う。特に, 文章中の (ゼロ) 照応関係をもとに解答の出現する文より前の文脈において質問作成に利用された箇所を同定する。例えば, 評価用データには, 記事 (1) の「クマラトゥンガ大統領」を解答とする質問文の例 (2) を作成し, またその作成の際に記事 (1) の下線の箇所を利用したことがタグ付与されている。この状況で提案手法では, 解答が出現する文の「合意<sub>i</sub>」とその前文に出現する「合意<sub>i</sub>」間の照応関係を参照し, 解答作成のために照応先の情報を参照すべきか, また参照する場合はどの範囲を参照するかをそれぞれ文節単位にその情報を利用するか否かの分類問題として解く。(1) については理想的には「合意<sub>i</sub>」間の照応関係をたどることにより, 前文の文節全体を参照し, その文に出現する「本交渉<sub>j</sub>」からさらに前文の「和平予備交渉<sub>j</sub>」へたどることにより, 最初の文を参照する。この後「スリランカ政府と...トラ」, 「和平予備交渉を再開」の 2 箇所を選択することにより, 文章全体から質問生成に必要な箇所を得る。

- (1) スリランカ政府と、タミル人分離独立軍事組織「タミル・イーラム解放のトラ」は三日、同組織の根拠地ジャフナ半島で 和平予備交渉<sub>j</sub> を再開。  
本交渉<sub>j</sub> の前提となる停戦で、初めて基本的合意<sub>i</sub> に達した。  
合意<sub>i</sub> 内容はクマラトゥンガ大統領 が発表する。
- (2) スリランカ政府と、タミル人分離独立軍事組織「タミル・イーラム解放のトラ」が和平予備交渉を再開し、初めて基本的合意に達したが、その内容を発表するのは誰か？

ここでは (ゼロ) 照応関係はあらかじめ解析されていることを仮定し, 本手法ではどのように質問箇所を取

表 2: 文間質問範囲同定に用いる素性

先行詞の品詞名と基本形
先行詞が記事の最初の文に登場しているか
照応詞の品詞名と基本形
照応関係の種類: ガ格, マ格, ニ格それぞれのゼロ照応, 名詞句共参照のいずれか
先行詞と照応詞が何文離れているか
解答と照応詞の間にある全ての機能語の品詞名と見出し語の基本形

捨選択するかの問題を解く. 具体的には以下の手順に従い, 質問に必要な箇所を選択する.

1. 解答を含む文に出現する照応関係を列挙する.
2. 各照応関係の先行詞が質問作成に必要な否かの分類問題(文間質問範囲同定)を解く.
3. 2の分類の結果, 先行詞が質問作成に必要と判定された文中の全ての文節のうち, 最終的に質問生成に利用する文節か否かの分類問題(文内質問範囲同定)を解く.
4. 2の分類の結果, 先行詞が質問作成に必要と判定された文中に前方文脈と照応関係を列挙する.
5. 4で到達可能な照応関係が列挙されていれば2へ戻る. それ以外の場合は処理を終了する.

文間質問範囲同定問題と文内質問範囲同定問題はそれぞれ機械学習を用いた2値分類問題として扱う.

文間質問範囲同定には表2に示す素性を利用する. 解答と分類対象となる照応詞との位置関係や, 照応の種類(名詞句照応, ガ格ゼロ照応など)が解答との関連性を捉えるための素性に加え, 先行詞の情報を分類に利用する.

また, 文内質問範囲同定には表3に示す素性を利用する. 対象文節の情報に加え, その文節が出現する文がどの程度質問作成に必要なかを文間範囲同定の結果選択された文節数で近似し, その数を素性として利用する.

## 4 評価実験

人間が質問文の作成に必要なと判断した範囲を同定することで, 提案手法の性能を評価した. 2節で作成したデータのうち, 解答を含まない文を用い, 2人の作業者のどちらかが質問作成に必要なと判断した部分を同定すべき正解とみなし, 文節単位での自動同定の評価を行った.

文間と文内の質問範囲同定の学習にはSVM<sup>light</sup>を使用した. カーネルは線形カーネル, パラメタcは1とした. 評価指標には各問題で同定対象となる文節に関して再現率, 精度, F値を用いて評価を行った.

### 4.1 文間質問範囲同定の評価

質問作成に必要なと判断された先行詞を同定することで文間質問範囲同定の性能を評価した. 評価では解答を含む文に出現している2,078の照応詞のみを対象にした.

表 3: 文内質問範囲同定に用いる素性

分類対象の文節が含む見出し語の基本形
分類対象の文節が含む品詞
分類対象の文節に含まれる語句と共参照する語句が以前に登場しているか
分類対象の文節の記事の最初の文に含まれるか
先行詞の品詞と見出し語
分類対象の文節から先行詞に何度の係り受け関係で到達出来るか
分類対象の文節が必要と判断された先行詞を修飾するか
分類対象の文節が必要と判断された先行詞に修飾されるか
分類対象の文節と必要と判断された先行詞の間に機能語が含まれる場合, その品詞と見出し語
並列関係 <sup>2</sup> にある2つのセグメントのうち, 一方に質問作成に必要な先行詞を含み, 分類対象にある文節が他方のセグメントに含まれる必要と判断された先行詞と分類対象の文節で, 一方が括弧の中, 他方が括弧の外にあるか
分類対象の文節を含む文に含まれる, 文間質問範囲同定で必要と判断された先行詞の数

表 4: 文間質問範囲同定の評価結果

	再現率		精度		F 値
P <sub>intra</sub>	0.631	(507/804)	0.624	(507/813)	0.627
BL <sub>intra</sub>	1	(804/804)	0.386	(804/2078)	0.557

P<sub>intra</sub>: 提案手法で文間質問範囲同定を行った場合  
BL<sub>intra</sub>: 全ての照応関係の先行詞を抽出した場合

このうち, 先行詞が質問作成に必要なと判断されている事例は804件であった. この評価データに対し, 5分割交差検定を行った結果を表4に示す. この結果, 提案手法では再現率を低下させるものベースラインに比べ適切に質問箇所を選択しているため, F値で比較した場合数値が向上していることがわかる. ただし, 最終的な評価はこの結果と文内の同定精度の組み合わせとなるため, 同定できなかった約4割を同定するために今後この処理を改善させる必要がある.

### 4.2 文内質問範囲同定の評価

文間質問範囲同定で必要と判断された先行詞を含む文に対して, 同一文中に出現する必要範囲を文節単位で同定することで文内質問範囲同定の性能を評価した.

評価では質問作成に必要なと判断された先行詞を含む文に出現する6,704の文節を用いて行った. このうち文節が質問作成に必要なと判断されている事例は4,437件であった. 提案手法に加えて, 文間質問範囲同定で必要と判断された先行詞を含む文のすべての文節を取り出す方法でも同定を行った. 5分割交差検定の結果を表5に示す. では, 再現率をほとんど低下させることなく, 精度を1割向上させており, 結果的にF値が約2%向上している. 今後は精度を低下させた原因について調査したい.

### 4.3 システム全体の評価

文内質問範囲同定と再帰的な文間質問範囲同定を組み合わせた, 最終的な提案手法を用いて, 質問作成に必要な

表 5: 文内質問範囲同定の評価結果

	再現率	精度	F 値
P <sub>inter</sub>	0.930 (4,127/4,437)	0.726 (4,127/5,683)	0.816
BL <sub>inter</sub>	1 (4,437/4,437)	0.662 (4,437/6,704)	0.797

P<sub>inter</sub>: 文内質問範囲同定を行った場合

BL<sub>inter</sub>: 先行詞を含む文の文節全てを抽出した場合

表 6: システム全体の評価結果

	再現率	精度	F 値
P	0.351 (2,696/7,673)	0.470 (2,696/5,735)	0.402
P <sub>-recur</sub>	0.332 (2,803/7,673)	0.477 (2,553/5,356)	0.392
P <sub>-inter</sub>	0.377 (2,894/7,673)	0.423 (2,894/6,834)	0.399
BL1	0.700 (5,371/7,673)	0.244 (5,371/21,984)	0.362
BL2	0.959 (7,357/7,673)	0.205 (7,357/35,822)	0.338

P: 提案手法

P<sub>-recur</sub>: 提案手法において再帰的な照応関係の抽出を行わない場合

P<sub>-inter</sub>: 提案手法において文内質問範囲同定のかわりに、文に出現する全ての文節を抽出した場合

BL1: 解答を含む文よりも前の文から、解答を含む文と照応関係にある文全てを再帰的に抽出した場合

BL2: 解答を含む文の前方文脈全てを抽出した場合

な文節の同定を行った。同定の対象には 65,363 の文節を用いた。このうち質問作成に必要なと判断されていた事例は 7,673 件であった。提案手法の比較対象として、解答を含む文より前にある文の中から解答を含む文との間に照応関係を持つ文を再帰的に抽出する方法、解答を含む文の前方文脈全てを抽出する方法の 2 種のベースラインを用いて同定を行った。また、提案手法において再帰的に文間質問範囲同定を行ったことの影響を確かめるため、再帰的な同定を行わない場合と比較する。また、文内質問範囲同定の影響を確かめるため、文内質問範囲同定の代わりに先行詞を含む文の文節全てを抽出する手法と比較した。5 分割交差検定の結果を表 6 に示す。

実験の結果、提案手法は 2 種のベースラインを F 値で上回ったが、再現率は約 35 % と、改善の余地は大きい。今回は 2 段階の処理で質問箇所同定を行ったが、今後はこの枠組み以外についても吟味し、また解析に必要な手がかりについても調査したい。しかし、提案手法による質問作成に必要な部分の再現率は、2 種のベースラインの半分以下にとどまった。

また、提案手法が前方文脈全てを抽出する方法と比較して再現率が低かった理由として NAIST テキストコーパスで照応関係のタグ付けのされていない部分が質問作成に必要な表現を含んでいたことがあげられる。照応関係のタグ付けがなされていないが、質問作成に必要なと判断された表現の例に間接照応によって参照される表現がある。間接照応とは記事 (3) の「世界貿易機関<sub>i</sub>」と「初代事務局長<sub>i</sub>」の間にあるような、先行詞が明示的に出現しない照応関係である。

(3) 自由貿易体制を監視し、発展を目指す国際機関「世界貿易機関<sub>i</sub>」が一日スタートする。初代事務局長<sub>i</sub>にはメキシコのサリナス前大統領、イタリアのルジエロ元

貿易相、韓国の金・前商工資源相の三人が立候補。

また、表 6 より、提案手法と文内質問範囲同定を文内の全ての文節を抽出する方法で代用した場合との比較では、性能が大きく変わらなかった。文内質問範囲同定を単体で解く場合 (表 5) に比べて、全ての文節を抽出する方法との差がでなかったのは、文間質問範囲同定の結果が文内質問範囲同定に影響を与えたためと考えられる。文内質問範囲同定には文間質問範囲同定の結果が素性として利用されており、文間質問範囲同定に失敗した場合に文内質問範囲同定の精度が低下すると考えられる。今後、文内質問範囲同定の性能をあげるためには、文内質問範囲同定に利用する素性を見直すとともに、文間質問範囲同定の精度をあげることも必要と考えられる。

表 6 より、再帰的な文間質問範囲同定を行う場合が F 値が高い。この結果は提案手法の段階的に範囲を同定することで、質問に必要な範囲を網羅的に探索することができる可能性を示している。今後はこの範囲同定の枠組みとそれ以外のやり方 (一度に全体の文節を分類する手法) などと比較することで、提案手法の有効性をさらに調査したい。

## 5 おわりに

本稿では文書から質問を生成する際に必要になる表現の同定に焦点を当て、文間質問範囲同定と文内質問範囲同定の 2 つの部分問題を解くことで質問範囲同定を実現する手法を提案した。人間が質問作成に利用した表現をどの程度同定できるかを文節単位で評価し、この結果、提案手法は F 値で単純なベースラインの結果を上回った。しかし、提案手法を用いた場合には再現率は約 36 % しか得ることができないため、今後改善していく必要がある。本研究では照応関係を手がかりに質問に必要な箇所の同定を行ったが、これに加え、bridging reference[6] のような間接照応の関係を考慮して探索範囲を広げる必要があり、これについては今後の課題としたい。

## 参考文献

- [1] R. Dale, E. Hovy, D. Rösner, and O. Stock. *Aspects of automated natural language generation*, Vol. 587. Springer-Verlag, New York, 1992.
- [2] 翠輝久, 河原達也, 正司哲朗, 美濃導彦. 質問応答・情報推薦機能を備えた音声による情報案内システム (音声質問応答; 特集; インタラクシオンの理解とデザイン). *情報処理学会論文誌*, Vol. 48, No. 12, pp. 3602-3611, 2007.
- [3] Donna M. Gates. Generating look-back strategy questions from expository texts. In *Proceedings of the Question Generation Workshop 2008*, 2008.
- [4] 飯田龍, 小町守, 乾健太郎, 松本裕治. NAIST テキストコーパス: 述語項構造と共参照関係のアノテーション. *情報処理学会研究報告 (自然言語処理研究会) NL-177-10*, pp. 71-78, 2007.
- [5] 黒橋禎夫, 長尾眞. 京都大学テキストコーパス・プロジェクト. *人工知能学会全国大会論文集*, Vol. 11, pp. 58-61, 1997/06/24.
- [6] H. H. Clark. Bridging. In *Thinking: Readings in Cognitive Science comprehension*, pp. 411-420. 1977.