

# Why 型質問に対する回答抽出

森本 格行

立命館大学大学院理工学研究科  
n\_morimoto@nlp.cs.ritsumei.ac.jp

福本 淳一

立命館大学理工学部情報学科  
fukumoto@cs.ritsumei.ac.jp

## 1 はじめに

現在、日本において日本語の質問応答システムのワークショップとして Question Answering Challenge(QAC)[1]がある。この評価ワークショップでは、基本的に、質問文が、人名、地名、組織名や人工物名などの名詞、日付や時間、数量表現となる「fact」を問う質問を対象としている。質問文にはいわゆる「5W1H」の疑問詞のうち「Why」型の質問文は設定されていない。

Why 型質問とは「なぜ…ですか」という形の質問文のことである。例えば「なぜ電車は遅れたのですか」といった質問に「信号が故障したから」と答えるものである。この例で質問文は「結果」を回答は「原因」を表しており、質問と回答の間に因果関係がある。Why 型質問に答えるためには、因果関係ような何らかの意味的な関係で質問とむすばれている部分が回答となる可能性がある。

文間の意味的な関係を定義しているものとして Rhetorical Structure Theory(RST)[2][3]がある。RSTは、意味的なまとまりをもつスパン間に成り立つ関係を記述したものである。スパンには「Nucleus」「Satellite」があり、関係として Evidence, Concession, Elaboration, Background など 25 種類が定義されている。

本研究では、RSTの例文を用いた分析から Why 型質問に利用できる関係を探し出し、入力された質問文とそれらの関係が成り立つ部分を回答として返す手法を提案する。まず、Why 型質問に利用できる関係をみつけるために RST の例文から Why 型の質問と回答を生成することで、質問と回答が成り立つのかを検証する。その結果得られた関係が一般的な文章においても適応できるのかを調べるために新聞記事を用い検証する。これにより、関係が一般的な文章においても適応できることがわかる。さらに新聞記事内から関係を抜き出す際に利用できる言語表現も得ることができる。

以下、二章では RST の例文についての分析、三章では新聞記事を用いた分析、四章で回答抽出の方法、五章で実験と評価について述べる。

## 2 RST の例文を用いた Why 型 QA の生成

25 種類の RST の例文から Why 型 QA を生成し、質問と回答が成り立つのかを分析した。その際、「Nucleus」「Satellite」のそれぞれに対して、一方に「なぜ」と「のですか」を付加することで質問を生成し、一方をそのまま回答とした。その結果、「Evidence, Motivation, nVC, Purpose, Means」の関係を持つものが Why 型 QA を生成できる可能性があることがわかった。以下に例を示す。N は「Nucleus」、S は「Satellite」を表す。(なお例文の訳は筆者によるものである。)

### Purpose

[例] スチールを作るために,s

鉄鉱石、石灰などの材料を採鉱する必要があった。N

質問:なぜ鉄鉱石、石灰などの材料を採鉱する必要があったのですか?

回答:スチールを作るために

「なぜ採鉱する必要があるのか」という理由を問う質問に対して「スチールを作るために」というように理由を回答している。これより N を質問文、S を回答として Why 型 QA が生成できる。

### Evidence, Motivation, nVC

これらの関係についても Purpose と同様 N を質問文、S を回答として Why 型 QA が生成できる。

### Means

[例] 海底からの跳ね返り音を探知機で記録することでs  
海底の構造をイメージすることができる。N

質問:なぜ海底からの跳ね返り音を探知機で記録する必要があるのですか  
回答:海底の構造をイメージすることができるから

「なぜ跳ね返り音を記録するのか」という理由を問う質問に対して「海底の構造をイメージすることができるから」というように理由を回答している。これより  $S$  を質問文、 $N$  を回答として Why 型 QA が生成できる。また、下のように「どのように」を  $N$  の文頭につけ質問文し  $S$  をそのまま回答とすることで手段を問う質問も生成できる。

質問:どのようにすれば海底の構造をイメージすることができますか?  
回答:海底からの跳ね返り音を探知機で記録することで

「どのようにすれば海底の構造をイメージすることができるのか」という手段を問う質問に対して「海底からの跳ね返り音を探知機で記録することで」という手段を回答する QA を生成できる。

### 3 新聞記事を用いた分析

前章で示した方法が一般に適用できるのかを確かめるため、新聞記事内にある Why 型質問に利用できる関係を持つ文を抜きだし分析した。その際、毎日新聞 2000 年 4 月の 25 記事を使用した。その結果、質問と回答を生成することができたものを以下の例を示す。四角で囲われた表現は関係抜き出す際利用する手がかり語である。

#### Evidence, Motivation, nVC, Purpose

- a1. 信号が故障したため  $S$  電車が遅れたのです。 $N$   
質問:なぜ電車が遅れたのですか  
回答:信号が故障したため
- a2. のどの渇きを癒したいから  $S$  水を飲みます。 $N$   
質問:なぜ水を飲むのですか  
回答:のどの渇きを癒したいから
- a3. 水を綺麗にしたことで  $S$  湖に魚が増えた。 $N$   
質問:なぜ湖に魚が増えたのですか  
回答:水を綺麗にしたことで
- a4. 手当がでないので  $S$  残業する気になれない。 $N$   
質問:なぜ残業する気になれないのですか  
回答:手当がでないので

a1 において「なぜ電車が遅れたのですか」という質問に対して「信号が故障したため」という原因を回答している。この例と同様 a2, a3, a4 の場合も Why 型 QA として原因、理由を回答している。

- a5. 単一通貨・ユーロの導入などで統合が深化する欧州で、地方分権の充実を避けて通れない。 $N$  なぜなら、統合は中央政府の権限の縮小を意味し、欧州に直結する地域政治の重要性が増すから  $S$  だ。 $S$

質問:単一通貨・ユーロの導入などで統合が深化する欧州で、地方分権の充実を避けて通れないのはなぜですか

回答:なぜなら、統合は中央政府の権限の縮小を意味し、欧州に直結する地域政治の重要性が増すから

a5 では、2 文に跨って関係を持つときも「避けて通れないのはなぜですか」という質問に対して「なぜなら、統合は中央政府の権限の縮小を意味し、欧州に直結する地域政治の重要性が増すから」という理由を回答している。

#### Means

- b1. 消費者に価格決定の主導権を与えることによって  $S$  不動産を変えたい。 $N$

質問:なぜ消費者に価格決定の主導権を与えるのですか

回答:不動産を変えたいから

- b2. 映画スターを差しこむことで  $S$  注目を集めたい意図のようだ。 $N$

質問:なぜ映画スターを差しこむのですか

回答:注目を集めたい意図のようだから

- b3. 法案は、大量生産・消費・廃棄という社会から脱却し廃棄物の効率的な利用を進めることにより  $S$  環境への影響が少ない「循環型社会」の実現を目的としている。 $N$

質問:なぜ法案は、大量生産・消費・廃棄という社会から脱却し廃棄物の効率的な利用を進めるとしているのですか

回答:環境への影響が少ない「循環型社会」の実現を目的としているから

b1 で「なぜ消費者に価格決定の主導権を与えるのですか」という理由を問う質問に対して「不動産を変えたいから」というように理由を回答している。b1 と同様に b2, b3 もの場合も Why 型 QA として理由を回答している。

## 4 回答抽出方法

ここでは、回答抽出方法を説明する。まず、入力された質問から新聞記事検索を行い上位10記事を使用する。次に得られた新聞記事から設定した関係をもつ部分を抜き出し「Nucleus」「Satellite」を判定する。抜き出した部分と質問とを照合し意味的に結ばれている部分（「Nucleus」もしくは「Satellite」）を回答として出力する。以上の検索や照合には、名詞、動詞、形容詞を利用する。

Evidence, Motivation, nVC, Purposeについては手がかり語「ため」「から」「ので」「タ形(過去形)+こと」「タ形(過去形)+ことにより」「タ形(過去形)+ことによって」を含む一文を取り出し、手がかり語の前を「Satellite」その後ろを「Nucleus」とした。また、以上の手がかり語の後ろの品詞が助詞、助動詞の場合、2文を抽出し手がかり語を含む文が「Satellite」一文前が「Nucleus」とし取り出す。また Means については手がかり語「ル形(現在形)+こと」「ル形(現在形)+ことにより」「ル形(現在形)+ことによって」を含む一文を取り出し、手がかり語の前を「Satellite」その後ろを「Nucleus」とした。

## 5 実験・評価

以上の手法をもとに Why 型質問に回答するシステムを構築した。このシステムの精度をみるために、人手によって毎日新聞記事から考えられた Why 型質問 23 問を実験に使用した。各質問に対する解答は新聞記事内から読み取れるものである。また、解答は各質問に一つ用意している。

このシステムでは、回答は文で出力される。そのため、用意した解答と同じ内容を表している文を出力に含まれている場合を正解とし、正解が含まれていない場合、もしくは回答が出力されていない場合は不正解とした。評価に用いた質問文とそれぞれの質問の正解不正解を表1に示す。表1内の「No.」は質問番号を表し、「正不」の「○」は正解、「×」は不正解を表している。

実験から、正解が9問、不正解14問という結果を得た。またこの際平均出力数は1.5であった。例えば表1のNo.2の質問に対して「昨年9月30日午前10時35分、茨城県東海村の核燃料加工会社ジェー・シー・オー（JCO）東海事業所の転換試験棟でウラン溶液の混合作業中、臨界を防ぐ形状になっていない沈殿槽に制限量を超える

ウランをバケツで大量投入したことで」と正しく回答した。このときシステムは「ことで」に注目し新聞記事から図1の文を抽出し、質問が作られたNと意味的な関係で結ばれているSを回答として出力している。

昨年9月30日午前10時35分、茨城県東海村の核燃料加工会社ジェー・シー・オー（JCO）東海事業所の転換試験棟でウラン溶液の混合作業中、臨界を防ぐ形状になっていない沈殿槽に制限量を超えるウランをバケツで大量投入したことで、S 核分裂反応が連続して起こる国内初の臨界事故が発生した。N
---

図1: 新聞記事から抽出された文

## 6 考察

実験で正解を出力できた質問は、前章で示した例のように言語表現を用いて新聞記事から抽出した範囲にその質問が作られた部分があり、意味的な関係で結ばれている部分を回答として抽出できたものであった。

一方、正解を出力できなかった質問については原因として以下のものがあつた。まず、用意された正解が名詞だけで文になっておらず、回答を抽出できなかった表1のNo.17の解答は「日中国交回復の記念」で、本システムが返した回答は「市民を元気づける機会になればと期待しているから」であった。

表1のNo.3の解答は「キャンデーやチョコなどお菓子をもらうため」であり、この質問と解答は図2に示す記事から生成されたものである。この解答は、文章の内容を理解した結果得られるものであり本システムでは正解できなかった。

動物の衣装を着たり、お姫様にふんしたり。 「トリック・オア・トリート」（ごちそうをくれないといたずらするぞ）とかわい声で脅される。 「はいはい」とキャンデーやチョコをあげる。
---

図2: 新聞記事 001030263(一部省略)

表1のNo.8の解答は図3に示す内容からつくられたものである。これを得るためには図3に示す下線部「ことから」に注目しなければならないが、この表現を登録していなかったため抽出できなかった。

表 1: 評価に用いた質問文と質問の正解不正解

No.	正不	質問文
1	○	なぜ火星に生物がいる可能性があると言えるのですか
2	○	茨城県東海村での臨界事故はなぜ起こったのですか
3	×	なぜキリバス共和国は世界で最も早く日の出を迎えるのですか
4	×	なぜハロウィンに子供たちは仮装するのですか
5	×	なぜ昭和天皇結婚記念切手は、発売されなかったのですか
6	×	なぜ小淵首相は亡くなったのですか
7	○	なぜ二酸化マグネシウムという金属化合物は注目を集めているのですか
8	×	なぜ琵琶湖に生息するピワマスという魚はアメノウオと呼ばれているのですか
9	○	なぜ幼児のうちから正しく箸をもてるように習慣づけるのがいいのですか
10	×	なぜマラソン選手の高橋尚子はキューちゃんと呼ばれているのですか
11	×	なぜ妊娠中に眠くなるのですか
12	×	ワールドカップのバレーボールにラリーポイント制を導入したのはなぜですか
13	○	なぜ予防接種が必要なのですか
14	×	エビやカニをゆでると赤くなるのはなぜですか
15	○	ポケットモンスターを見ていた子供たちが、けいれん、めまい、吐き気といった発作をおこしたのはなぜですか
16	×	なぜカーリングは氷上のチェスと呼ばれているのですか
17	×	なぜ中国からパンダが来ることになったのでしょうか
18	×	なぜ赤穂浪士は討ち入りをしたのですか
19	×	なぜギリシャはオリンピックの開会式一番に入場行進するのですか
20	○	なぜエルニーニョ現象は発生するのですか
21	×	なぜ関西では鰻丼のことをまむしと呼ぶのですか
22	○	顔見世にでる歌舞伎役者の名前を書いたまねきに名前を書くとき墨と酒を混ぜるのはなぜですか
23	○	なぜ構造改革のビジョンを明確に示す必要があるのですか

湖の固有種で、雨が降って増水すると川を上ることから「アメノウオ」とも呼ばれ、燕村に「瀬田降って志賀の夕日や江鮭」の句がある。

図 3: 新聞記事 000821253(一部省略)

Theory and Text Analysis”. Discourse Description: Diverse linguistic analyses of a fundraising text. ed. by W.C. Mann and S.A. Thompson, pp.39-78, 1992.

## 7 おわりに

本手法に基づいたシステムを実装し、23 問の質問文を使い実験を行った。結果、23 問中 9 問正解で約 40% の精度であった。

今後の課題として、前章の考察で上げた不正解になった原因を解決し、システムの精度向上を目指す予定である。

## 参考文献

- [1] J.Fukumoto, T.Kato and F.Masui. “Question Answering Challenge (QAC-1) Question answering evaluation at NTCIR Workshop 3”. Working notes of the Third NTCIR Workshop Meeting Part IV: QAC, National Institute of Informatics, pp.1-10, 2002.
- [2] W.C.Mann and S.A.Thompson. “Rhetorical Structure Theory:A Theory of Text Organization”. USC ISI Technical Report ISI/RS-87-190, pp.15-82, 1987.
- [3] W.C.Mann, Christian M.I.M. Matthiessen and S.A.Thompson. “Rhetorical Structure