

質問応答におけるルールに基づく回答抽出

野中清孝* 宮武明義* 鱒目正志*

Rule Based Answer Extraction in Question Answering System

Kiyotaka NONAKA Akiyoshi MIYATAKE Masashi SAWARAME

Abstract

Generally in Question Answering System, user asks about something to computer and computer answers its question. However, the development of the system in which computer asks and user replies will be expected in the future. These systems can be used to front desk system of company, questionnaire on web, and so on. In this research we propose the method in which the answer word for a question sentence is extracted by Extraction Rules based on semantics. Extraction Rule is constructed with the procedure which extracts the answer and a semantics pair for a question sentence and an answer sentence. In this paper, we describe the overview of this system, formalization for semantics of Japanese sentences, Extraction Rules to extract an answer word and the execution results for several examples.

Key Words : Answer Extraction, Natural Language Processing, Question Answering System

1. はじめに

近年、質問応答システムの研究では、ユーザの自然言語による質問に対してコンピュータが保持するデータベースやインターネット上に蓄積されている情報資源の中から回答としての語句や文を切り出し、適切な回答を生成して応答するCRUQ (Computer Reply for User's Question) 型を対象とする研究が殆どである。しかし、これとは逆にURCQ(User Reply for Computer's Question) 型の質問応答システム、すなわちコンピュータが質問を行い、その質問に対してユーザが自然言語で任意の応答を行う質問応答システムの開発も期待されている。URCQ型の質問応答システムの応用としては、企業や病院などの受付システムやウェブ上でアンケートを実施するなど様々な応用システムが考えられる。URCQ型の質問応答システムを実現する上での技術課題は、自然言語によるユーザの応答を解析し、ユーザがいかなる回答をしているかを語句として適切に抽出することにある。ここでの問題点は、自然言語による応答表現の多様性に対する適応性と一般のユーザが応答することから質問応答という文脈からはずれた応答をする場合についての対処を考える必要がある。

本研究では、URCQ型の質問応答システムを実現する上での上記の問題を解決するために、任意の日本語単文の意味を形式的に定めるとともに、任意の日本語単文を解析し、ここで提案する意味形式に変換する意味解析モジュールを開発した。また、質問文と応答文の意味から回答としての語句を抽出するための回答抽出ルール群を構築し、このルール群を質問文と応答文に適用するモジュールを作成した。そして、これらの2つのモジュールを結合し、日本語の質問文とそれに対する応答文を引数として渡すと、回答語句、抽出の成功率、適応したルールを、また、抽出に失敗した場合には、抽出失敗の情報とユーザの応答文の意味を結果として返す汎用的なプログラムコンポーネントを実現し、今後開発される様々なURCQ型質問応答システムにおいて利用できるようにした。

本稿では、開発した回答抽出コンポーネントの概要、日本語単文の意味の形式化、回答抽出ルールについて述べるとともに、具体的な質問応答に対するこのコンポーネントモジュールの実行結果を示す。なお、日本語単文の意味解析については、意味の形式的定義については詳述するが、実現方法については、簡単に説明することに留め、詳細については改めて報告する。

* 香川高等専門学校詫間キャンパス 情報工学科

2. 回答抽出コンポーネントの概要

回答抽出コンポーネントは質問文とそれに対する応答文を入力とし、基本的に応答文の意味と応答文に含まれる答えとなる語句(回答)を切り出して出力するモジュールである。

回答抽出における基本的な処理の流れは、まず自然言語による質問文(QB)と応答文(RB)の意味解析を行い、それぞれの文に対する意味を次節で述べる意味形式に変換する。次に、その形式化された質問文の意味と応答文の意味を回答抽出のためのプロダクションルールとして記述された回答抽出ルール郡の条件部にマッチするか順次照合し、マッチした場合それぞれのルールの手続きにおいて、応答文の形式的に記述された意味から回答語句の抽出やその抽出の正しさに係る抽出信

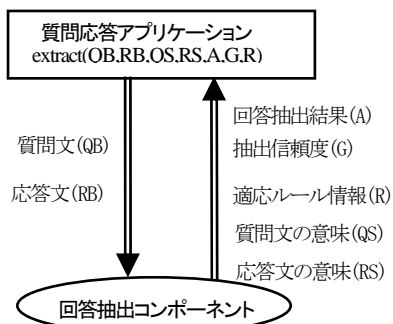


図1 入出力インターフェース

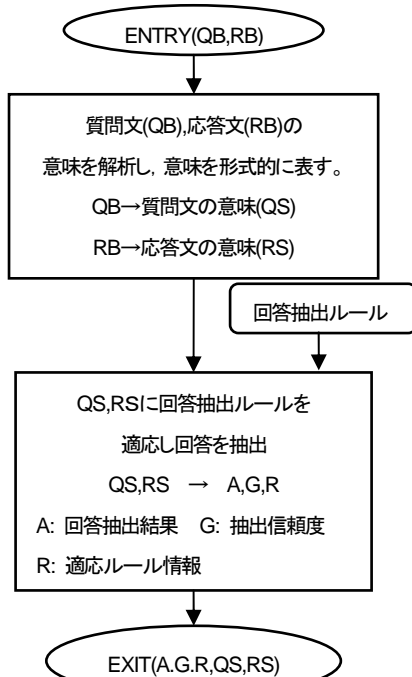


図2 処理の流れ

頼度、また適応したルールの情報などを求めている。図1に回答抽出コンポーネントの入出力インターフェースを、図2に基本的な処理の流れを示す。

3. 日本語単文の形式的意味表現

回答抽出コンポーネントでは意味解析モジュールを用いて質問文と応答文の意味解析を行い、その意味を形式的記述に変換する。この節では、まず日本語単文がどのような意味形式に変換されるかについて詳述し、次に、意味解析モジュールについて述べる。回答抽出コンポーネントの意味解析部では、意味の形式的記述を、以下に示す述語論理形式(sem/3)で表す。

sem (K, P, PW)

- K : 文の意味分類とモダリティ
- P : 文の基本的内容 (格解析情報)
- PW : Pに対する関係補助命題

上で示すように、質問文及び応答文の形式的意味の定義は述語名をsemとし、発話文の意味分類とモダリティ(K)、発話文の基本的内容(P)、Pに対する補助命題(PW)の3つの引数で構成している。以下各引数の形式について述べる。

3.1 意味分類とモダリティ(K)

発話文の意味分類とモダリティ(K)は文の役割の意味(M1)、内容分類(M2)、断言度(C)、話者の表現上の気持ちを分類した内省表現(I)の4組で表す。

$$K = [M1, M2, I, C]$$

各引数は後で説明するとして、意味分類とモダリティについてイメージを簡単につかんでもらうために、例文で具体的に説明する。例えば、以下の3つの文の表現に関する分類Kは発話文から次のように解析される。

- ① 「私は東京に行きたいと思うであろう。」
K=[報告, 希望, 推量, 思考]
- ② 「あなたは男ですか。」
K=[真偽質問, 関係, 断定, '']
- ③ 「はい、そうかもしれません」
K=[慣応, 肯定, 曖昧, '']

上の例で①の文に対する意味分類とモダリティは、希望について思考することを推量的に報告していることを表している。また、②はある関係について断定的に真偽質問(YesNo 質問)をしていることを表し、さらに、③は、慣用的に使われる応答で曖昧に肯定的

(Yes)応答していることを表す。

(1) 文の役割の意味(M1)と内容分類(M2)

意味解析によって、M1には文の役割の意味が代入される。役割の意味として、ここでは文を依頼、報告、格質問(5W1H質問)、真偽質問(Yes No質問)、格応答、慣用応答、慣用表現に分類している。また、M2はM1それぞれに対し文の意味的内容をさらに細かく分類したものである。例えば、報告文では、文の内容として希望、存在、状況、関係、可能、状態、許可、禁止、義務、運動、協力行為に分類される。表1にM1とM2に関する分類を示す。

表1 文の意味分類

役割の意味(M1)	内容の分類(M2)
依頼	依頼、命令
報告	希望、存在、状況、関係、可能、状態、許可、禁止、義務、運動、協力行為
格質問	
真偽質問	
格応答	未設定
慣用応答	肯定、否定、曖昧、了解、拒絶、不明
慣用表現	激励、感謝、賞賛、慰め、非難、不満、同情、陳謝、嘆き、驚き、呼びかけ、挨拶(X)、時候(X)、天気(X)、感情文(X)、感覚文(X)

表1について補足すると、格応答として分類される文は、例えば、「名詞+助詞」だけからなる文等で、その意味は文脈からしか判断できないものを表している。また、慣用的に使われる文の分類としては、応答として使われる慣用応答とそれ以外の慣用表現に分類し、それらの意味の詳細をM2に設定している。なお、慣用表現でXなる引数はさらに詳細な意味情報が与えられることを示している。

(2) 断言度 (C)

断言度は文の内容について、話者がどの程度断言をしているかの情報である。ここでは、断言、推量、曖昧の3つに分類している。

$C \in \{ \text{断言, 推量, 曖昧} \}$

(3) 内省表現 (I)

内省表現の形式を以下に示す。

$I = [IK, ID]$

IK: 内省表現 ID: 内省動詞の原形

IK ∈ { 思考, 推測, 祈り, 願望, 体感, 目感, 耳感, 記憶 }

内省表現とは、文末に表れる話者の内省的な表現を意味的に分類したものである。例えば「試験を受けたと思います」のように、文の前半で表現したいことを述べ、文末に「思う」、「考える」、「希望する」、「祈る」、「感じます」等の語句を加えて、話者の気

持ちが表現されている場合である。また、内省動詞とはその表現のために使用されている動詞の原形である。内省表現(IK)の分類を上で示したが、ここで、体感、目感、耳感、記憶は、順に「～と感ずる」、「～と見ます」、「～と聞いています」、「～と記憶しています」なる表現に対応させている。

3.2 文の基本的内容 (P)

文が表す基本的内容は構文解析及び文の表層に現れる格構成の解析結果を基に、主要動詞の原型(V)、肯定/否定(N)、時制(T)、格リスト(CL)、疑問詞情報(CW)の5引数を持つ述語pとして表している。形式は以下の通りである。

$p(V, N, T, CL, CW)$

V : 主要動詞 $V \in \{ \text{動詞原形} \}$

N : 肯否 $N \in \{ \text{肯定, 否定} \}$

T : 時制 $T \in \{ \text{過去, 現在} \}$

CL : 格リスト $CL = [C_1, C_2, C_3, \dots]$

$C_i = \text{name}(\text{obj}, \text{pre})$

name: 格の名称, obj: 格対象, pre: 格判定語句

CW : 疑問詞情報 $CW = [W, \text{Case}, CW]$

格リストは文を構成する格(Ci)の集合であり、その出現順序は表層文に現れる順序と一致する形式をとる。それぞれの格は、name(obj, pre)なる形式をもち、nameは格の名前、objは格対象、prepは表層文において使われている格助詞などの格判定語句である。表2に意味解析で解析される格の種類を示す。

表2 格の種類

格の名称	格記号	備考
主格	csub	意味上の主語
対象格	cobj	動作の対象
源泉格	csou	動作の出発点
目標格	cgoa	動作の着地点
理由格	crea	動作の理由
道具格	cins	動作の道具・手段
場所格	cloc	動作の場所
時間格	ctim	動作の時間
共同格	cadv	動作を共にする対象
副詞格	cwit	動作を補足する句
数量格	cqua	数量

疑問詞情報(CW)は文中に疑問詞を含む場合、その疑問詞の分類(W)と可能性のある格記号(Case)、疑問詞補助語句(A)から構成している。疑問詞の分類としてwhat, who, when, where, why, which1, which2, howmany, howの9種類に分類しているが、この分類と日本語の疑問詞との対応を表3に示す。

表3 格の分類

疑問詞記号	疑問詞(ひらがなのみ記載)
what	なに, なん
who	だれ, どなた
when	いつ, いつごろ, いつぐらい
where	どこ, いづこ, いずこ
why	なぜ, どうして, なにゆえ
which1	どの
which2	どれ, どちら
howmany	いかほど, いくら, いくつ, いくつくらい, いくつぐらい, いくつほど, いくついで, どれくらい, どれぐらい, どれほど, どのく
how	どんな, どのような, どんなふうな, いかなる, どのような, いかがな, いか, どのような, どんなふう

また、格については疑問詞の問う意味素性と疑問詞の後に続く格助詞等から判定している。さらに、疑問詞補助語句は疑問詞の後に続く係り受けの関係にある名詞がある場合にその語句を設定している。この語句は一般に単位に相当する語句が多いと考える。具体的な例文で説明すると、以下の3つの文の表現に関する疑問詞情報は次のように解析される。

- ① 「あなたはどこに住んでいますか」 [where, cloc, ""]
- ② 「あなたの血液型は何型ですか」 [what, cobj, "型"]
- ③ 「体重はいくらですか」 [howmany, pred, ""]
- ④ 「なぜ東京に行くのですか」 [why, crea, ""]

3.3 関係補助命題(PW)

一般に質問応答においては、ある事物と事物の関係について質問したり、応答したりすることが非常に多い。この場合、「Pです(か)」、「AのBはPです(か)」、「AはPです(か)」、「AはBのCです(か)」等、文末に「です」、「である」等の語句を用い、端的に事物と事物の関係を表す構文が多く使われる。ここでは、これらの構文を持つ文を関係文と呼ぶ。これらの事物と事物の関係を以下の形式で表したものが関係文に対する補助命題である。

PW=関係(R,T1,T2,N,T,C)

R : 事物の関係 N : 肯否
T1 : 事物1 T : 時制
T2 : 事物2 C : 断言度

上記の関係補助命題は、例外もあるが基本的には事物T1と事物T2が関係R(T2のRがT1)であることを示している。またN, T, Cはそれぞれ肯否, 時制, 断言度であり、発話文の基本的内容等で述べたものと同一の情報である。なお、関係補助命題は、質問応答におい

て応答文が関係文である場合に、この形式からユーザ側の回答を容易に抽出するために使用している。

3.4 意味解析モジュール

意味解析モジュールは、格解析部、用言部解析部、意味記述への変換部から構成される。前節で述べた意味形式は日本語単文の意味解析モジュールで解析された文の構成的意味、内容的意味を基礎に形式化して再構成したものである。格解析部は格助詞及び格判定語句を基礎に文の格構成を解析するが、その詳細については参考文献1を参照されたい。

用言部解析部では、文の用言部を詳細に解析し、疑問詞解析、時制解析、肯定/否定解析、文のモダリティ(表現内容)解析、内省解析、推量及び断言度解析、動作/行為の進行解析を行っている。疑問詞解析では用言部末尾や格解析での疑問詞情報から質問、同意、断定の区別を解析する。質問のときは、真偽と5W質問の区別といかなる格に関する質問かについて解析する。また、モダリティ解析では、文の表現内容が、命令、依頼、状態、許可/禁止、可能、意志、義務、関係、存在、運動であるかについて解析している。

4. 回答抽出ルール

回答抽出ルールは回答を抽出するためのルールである。回答抽出ルールの記述は、ルールを適応するかを判定する条件部、条件が成立するとき応答文の意味から回答を取り出す回答抽出部、抽出結果出力部で構成される。条件部はそのルールが適応可能であるための質問文の意味パターンと応答文の意味パターンを記述している。そして、実際の質問文と応答文の意味をこの意味パターンとマッチすることにより、ルール適応の条件判定としている。

回答抽出処理は、以下に示す手続きaruleを呼び出すことに実行される。手続きの入力引数は質問文の意味(QS)と応答文の意味(RS)であり、出力引数としては回答抽出結果(A)、抽出信頼度(G)、適応ルール情報(R)から構成される。

arule (QS, RS, A, G, R)

QS : 質問文の意味形式
RS : 応答文の意味形式
A : 回答抽出結果
G : 抽出信頼度
R : 適応ルール情報

回答抽出ルールを構築するに当たり、条件部の質問文と応答文の意味パターンを以下に示すように分類

し、質問文と応答文の意味パターンの全ての組み合わせに対し構築し、この分類を適応ルールの主番号に対応させている。さらに、この分類(主番号)の下位の分類として、時制、文の断言度等の条件パターンに対応してルールの枝番号を割り当てて構築している。

〈質問文の意味パターンの分類〉

- ・一般文による5W質問
- ・一般文による真偽質問
- ・関係文による5W質問
- ・関係文による真偽質問

〈応答文の意味パターンの分類〉

- ・一般文による応答
- ・関係文による応答
- ・慣用応答文による応答
- ・上記以外の文による応答

なお、一般文とは、いわゆる一般動詞が使われている文であり、関係文とは「AのBはCです。」などの事物と事物の関係を表す文のことをさす。また、慣用応答文とは「わかりません。」、「そうです」など応答において慣用的に使われる文を意味する。以下、出力引数について説明する。

(1) 回答抽出結果(A)

回答抽出結果(A)は抽出された回答語句(K)、質問に対する応答文の断言度(C)、回答抽出ルール信頼度(RC)の3つ組([K, C, RC])で構成される。

ここで、回答(K)は、回答語句の文字列、“拒絶”、“不明”のいずれかを取る。また、断言度(C)は発話文の表現に関する分類の節で述べた断言度(C)と同一のものである。また、回答抽出ルール信頼度(RC)は抽出ルール自体の確実性を現すもので、“確実推論”、“曖昧推論”、“non”のいずれかの値が設定される。

(2) 抽出信頼度(G)

抽出信頼度(G)は、抽出された語句の信頼度(G1)と、G1が曖昧や失敗の場合のその理由が示せる場合のコメント(G2)の2つの組から構成される。抽出された語句の信頼度は、応答文の断言度と。成功度は解答抽出結

果(A)の要素であるユーザの応答文の断言度(C)と回答抽出ルール信頼度(RC)より決定する。値としては成功、曖昧、超曖昧、失敗の4つをとるが、表4にC及びRCとG1の関係を示す。

(3) 適応ルール情報(R)

適応したルールの主番号、枝番号およびルール分類名の2つの要素から構成され、以下の形式をとる。

R=[[適応ルール主番号, 枝番号], 適応ルール分類名]

ルール分類名は、5W関係・報告一般などの値をとる。これは関係文で5Wの質問が行われ、一般文で報告の応答がなされた場合のルールであることを示す。

5. 回答抽出コンポーネントの実行例

以下に病院での受付における質問応答を例にとり、入力としての質問文(QB)、応答文(RB)と出力としてのそれぞれの意味形式(QS, RS)及び回答抽出結果(A)、抽出信頼度(G)、適応ルール情報(R)を示す。

実行例1

〈質問文とその意味〉

QB= あなたの名前は何ですか
 K = [5W質問(what), 関係, non, 断定]
 P = p(である, 肯定, 現在, [csub(私の名前, は)], [what, pred, non])
 PW= 関係(名前, xvar(何), 私, 肯定, 現在, 断定)

〈応答文とその意味〉

RB= 私の名前は詫間太郎です
 K = [報告, 関係, non, 断定]
 P = p(詫間太郎である, 肯定, 現在, [csub(私の名前, は)], non)
 PW= 関係(名前, 詫間太郎, 私, 肯定, 現在, 断定)

〈回答抽出結果の出力〉

A = [詫間太郎, 断定, 確実推論]
 G = [成功, non]
 R = [[100, 7], 5W関係・報告関係]

この例は、関係文で5W質問を行い、関係文で応答した場合の例である。質問文の中での語句「あなた」が意味記述では「私」に変換されていることに注意された。対話においては私・あなたの意味が対話者によって反対の意味になるので、意味解析では、応答文の発話者自身を指し示す語句は「私」になるようにしている。

実行例2

〈質問文とその意味〉

QB= 咳は出ますか
 K = [真偽質問(質問), 運動, non, 断定]
 P = p(出る, 肯定, 現在, [csub(咳, は)],

表4 断言度と抽出成功度の関係

断言度	ルール信頼度	抽出信頼度
断言	確実推論	成功
断言	曖昧推論	曖昧
推量	確実推論	曖昧
推量	曖昧推論	曖昧
曖昧	確実推論 曖昧推論	超曖昧
それ以外	—	失敗

<p>[yesno, 質問, non])</p> <p>PW= non</p> <p>〈応答文とその意味〉</p> <p>RB= いいえ, 咳は出てないと思います</p> <p>K = [報告, 運動, 思考, 推量]</p> <p>P = p(出る, 否定, 現在, [csub(咳, は)], non)</p> <p>PW= non</p> <p>〈回答抽出結果の出力〉</p> <p>A = [NO, 推量, 確実推論]</p> <p>G = [曖昧, non]</p> <p>R = [[200, 1], 5W一般・報告一般]</p>

この例は一般文で5W質問を行い、一般文で応答した場合の例である。応答に「～と思います」なる語句があるため、抽出信頼度は曖昧と解析されている。

実行例3

<p>〈質問文とその意味〉</p> <p>QB= 今日の体温は何度ですか</p> <p>K = [5W質問(what), 関係, non, 断定]</p> <p>P = p(である, 肯定, 現在, [csub(今日の体温, は)], [what, pred, 度])</p> <p>PW= 関係(体温, xvar(何度), 今日, 肯定, 現在, 断定)</p> <p>〈応答文とその意味〉</p> <p>RB=朝の7時には38度2分でした</p> <p>K = [報告, 関係, non, 断定]</p> <p>P = p(38度2分である, 肯定, 過去, [ctim(朝の7時, には)], non)</p> <p>PW= 関係(is, 38度2分, non, 肯定, 過去, 断定)</p> <p>〈回答抽出結果の出力〉</p> <p>A = [38度2分, 断定, 曖昧推論]</p> <p>G = [曖昧, 警告(時制)]</p> <p>R = [[100, 3], 5W関係・報告関係]</p>

この例は関係文で5W質問を行い、一般文で応答した場合の例である。質問の時制と応答の時制が異なるため、抽出信頼度は曖昧と解析され、コメントとして時制に関する警告がでている。

実行例4

<p>〈質問文とその意味〉</p> <p>QB= あなたはどこに住んでいますか</p> <p>K = [5W質問(when), 行為, non, 断定]</p> <p>P = p(住む, 肯定, 現在, [csub(私, は), xvar(cloc(どこ,)), [where, cloc, non]])</p> <p>PW= non</p> <p>〈応答文とその意味〉</p> <p>RB= 僕は詫間町香田に住んでいます</p> <p>K = [報告, 行為, non, 断定]</p>

<p>P = p(住む, 肯定, 現在, [csub(私, は), cloc(詫間町香田,)), non)</p> <p>PW= non</p> <p>〈回答抽出結果の出力〉</p> <p>A = [詫間町香田, 断定, 確実推論]</p> <p>G = [成功, non]</p> <p>R = [[200, 12], 5W一般・報告一般]</p>
--

この例は一般文で5W質問を行い、一般文で応答した場合の例である。抽出信頼度は成功と解析されている。

6. おわりに

本稿では、日本語による質問文と応答文から質問に対する回答を抽出する一つの手法を提案した。この方法の特徴は文の意味を形式的に定め、質問文と応答文の意味をこの形式に変換していること、また、文の意味パターンを基礎にして回答抽出ルールを記述し、質問に対する回答を応答文の形式的意味記述から取得しているところにある。この方式の評価については今後の課題であるが、評価それ自体難しい面がある。それは、適切に回答を抽出する確率だけが問題ではなく、ユーザの応答が質問と無関係な応答をした場合にもその応答の意味を正確に解析しているかを評価しなければならないからである。この点から言えば、この研究の評価は提案している意味の形式化の妥当性、さらには、開発した日本語文の意味解析モジュールの精度が問題となるべきである。自然言語処理の研究分野において意味の問題は最も重要で難解なテーマであるが、ここで提案する意味の形式化がこの分野で何等かの参考になればと考える。

参考文献

- 野中清孝：格判定語句による日本語文の構造的意味解析、詫間電波工業高等専門学校研究紀要、第31号、pp. 233-258, 2003
- 市村由美, 齋藤佳美, 酒井哲也, 國分智晴, 小山誠：固有表現抽出と解答タイプ体系が質問応答システムの性能に与える影響、電子情報通信学会論文誌、D-II, 情報・システム, II-パターン処理, pp.1067-1079, 2005
- 加藤恒昭, 福本淳一, 榊井文人, 神門典子：質問応答技術は情報アクセス対話を実現できるか、情報処理学会研究報告、2004-NL-162, pp. 145-150, 2004
- 藤畑勝之, 志賀正裕, 森辰則：係り受けの制約と優先規則に基づく数量表現抽出、情報処理学会研究報告、IPSJ SIG Notes, pp. 119-125, 2001