

定義を問う質問応答を目的とした表現パターンの有効性の分析

岩田 匡史† 上西 康広† 石田 健二‡ 榊井 文人‡ 河合 敦夫‡ 井須 尚紀‡
†三重大学工学部 ‡三重大学大学院工学研究科
E-mail:{iwata,uenishi,ishida,masui,kawai,isu}@ai.info.mie-u.ac.jp

1. はじめに

質問応答システムは TREC[7]や NTCIR[2]などの評価型ワークショップを中心に研究されてきた。

従来は、主に名称や値を回答とする factoid 型質問応答への対応が関心の対象となっていた[3][4]。ところが、応用場面を考えた場合、factoid 型質問応答は必ずしも有効ではない。たとえば、新語彙の獲得を目的としたユーザーが質問応答システムを利用する場面では、factoid 型質問文よりもむしろ用語の定義を問う質問文に重きを置いた質問応答システムが求められている[1]。

新聞や辞書において、用語の定義について書かれた文章には「(用語)とは」や「(用語)は～である」のような特定の表現パターンが多用されることが報告されており[5][8]、このような特定の表現パターンを利用して用語の定義について書かれた文書を抽出する手法が提案されている[9]。しかしながら、いかなる表現パターンを用いた場合が、回答の網羅性や精度の上で有利となるかという点に着目した調査・議論はなされていない。

そこで本研究では、物事の定義について書かれた文章から人手で表現パターンを取り出し、これらの表現パターンを用いて得られた回答の精度、網羅性をパターン毎に比較した。その結果を分析し、回答抽出に有効なパターンについて考察する。

以下 2 章では、表現パターンの取得方法について説明する。3 章では、表現パターンを用いた回答抽出結果について述べる。4 章では、結果について考察を行う。

2. 表現パターンの収集

本章では、物事の定義について書かれた文章の規則性を見出し、表現パターンとして抽象化を収集する方法について説明する。

まず、定義について書かれた文章を新聞記事から抽出する。用語の定義を問う質問文 15 問を用意

し、それらの回答を含む記事を毎日新聞 2000 年と 2001 年の 2 年分から取り出した。その結果、用語の定義についての記述を含む新聞記事、225 記事が得られた。

次に、回答を含む記事から「(用語)とは～」などといった表現パターンを人手で抽出する。上で述べた 225 記事中で 2 回以上出現したパターンを人手で収集した結果、10 個の表現パターンが得られた。本実験では、上記の 10 個のパターンに、我々が有効であると考えた 13 個のパターンを加えた合計 23 パターンを用いて回答抽出を行う。

表 1. 収集したパターン

識別子	パターン
A	「キーワード」とは～タイプ
B	「キーワード」とは～
C	「キーワード」は～タイプ
D	キーワードとは～タイプ
E	キーワードとは～
F	キーワードは～タイプ
G	キーワードは～
H	～タイプ「キーワード」～
I	～タイプ、キーワード～
J	～「キーワード」というタイプ～
K	～「キーワード」の存在～
L	～キーワードというタイプ～
M	～キーワードの存在～
N	～キーワード～同タイプは～
O	～キーワード～に関するタイプ～
P	キーワードについて～説明する～
Q	キーワードは～するもの
R	キーワードの内容は～
S	～なキーワード
T	～(キーワード)
U	～た「キーワード」
V	～たキーワード
W	(段落の始まり)キーワードは～

取得した表現パターンは大きく以下の2つに分かれる。1つはキーワード（定義を問う対象）とタイプ（キーワードの下位概念）を用いるパターンである。例えば、定義を問う対象がボブスレーの場合、キーワードが“ボブスレー”，タイプが“競技”となる。表現パターンの例としては「ボブスレーとは～競技」などが挙げられる。2つ目は、キーワードのみを用いるパターンである。例えば「ボブスレーというのは」などが挙げられる。

3. 表現パターンの評価

本研究で収集した表現パターンの有効性を評価するためには、パターンに対する入力としてキーワードとそれに対応したタイプのセットを用意する必要がある。例えば、キーワード“ボブスレー”に対応したタイプ“競技”のセットを用意する。

本実験では、評価用にキーワードとタイプのセットを35セット用意した。それらを2章で述べた表現パターンに当てはめて、回答を新聞記事中から抽出した。回答を抽出する知識源として、毎日新聞2000年と2001年の2年分を使用した。

このようにして抽出された回答の適合性を人手で判断して、「正解」と「誤り」の2つに分類した。パターン毎に抽出した回答数、正解数、回答の精度（式1）、R値（式2）を求めた。1個以上の回答を出力したパターンの結果を表2に示す。

$$\text{Precision} = \frac{\text{回答数}}{\text{正解数}} \dots \text{式 (1)}$$

$$\text{R値} = \frac{\text{正解が得られた質問数}}{\text{質問総数}} \dots \text{式 (2)}$$

4. 考察

本章では、実験結果に対して考察する。4章1節では、タイプを伴うパターンについて考察する。2節では、タイプを伴わないパターンについて考察する。

4.1. タイプを伴うパターン

タイプを伴うか否かでパターンを見ると、タイプを伴うパターンによって得られた回答は、理論的にはかなりの精度が期待される。ところが、タイプを伴うパターンの多くでは回答が得られなかった。このことは、主に以下の3つの原因が考えられる。

1つ目の原因として、タイプの適切性に問題があったと考えられる。今回はキーワードに対して1つのタイプを用意したが、本来は複数のタイプを用いて回答抽出を行うことが望ましい。例えば、キーワード「ボブスレー」に対するタイプとして、「競技」のほかにも、「種目」などのタイプが考えられる。このように、キーワードに対し、複数のタイプを用意し、再度実験を行うことでより多くの回答が得られる。

2つ目の原因として、質問文の試行数が少ないことが考えられる。この問題は、質問の試行数を増

表2. パターン毎の結果

識別子	パターン	回答数	正解数	正解が得られた質問数	R値	精度
v	～たキーワード	716	52	18	0.514	0.064
t	～(キーワード)	590	21	4	0.114	0.036
w	(段落の始まり)キーワードは～	84	24	10	0.286	0.202
g	キーワードは～	75	23	10	0.286	0.253
u	～た「キーワード」	16	10	4	0.114	0.625
e	キーワードとは～	11	0	0	0.000	0.000
l	～キーワードというタイプ～	5	2	2	0.057	0.400
c	「キーワード」は～タイプ	1	1	1	0.029	1.000
d	キーワードとは～タイプ	1	0	0	0.000	0.000

やすことで解決される。

3つ目の原因として、検索対象記事が少ないことが考えられる。今回は、検索対象として新聞記事2年分を用いたが、さらに大規模なコーパス (World Wide Web など) を用いて再度実験を行うことでより多くの回答が得られると考えられる。

4.2. タイプを伴わないパターン

パターン u “～た「キーワード」” とパターン v “～たキーワード” の違いはキーワードに鍵括弧が付くか否かである。表 2 を見ると鍵括弧を伴うパターンの方が精度が高く、鍵括弧付きのキーワードを含む文章は、キーワードの定義文である傾向が強いと言える。この原因として、鍵括弧の付いたキーワードを含む文章は、そのキーワードが最初に現れる文書だと考えられる。また、キーワードの定義を説明する場合、キーワードが最初に出現した文章付近であることが多いと推測される。これらから、鍵括弧の付いたキーワードを含む場合の方が精度が高いと考察される。

パターン t “～ (キーワード) ～” を用いて得られる回答は、括弧内に出現するキーワードが単なる省略表現を含む場合や、正解とするには情報が不足している曖昧な回答が多い。この問題点は、“～た (キーワード) ～” など、より条件の厳しいパターンを用いることで解消可能と考える。

パターン g と w の差異は“キーワードは～” というパターンが段落の最初に現れるか、文中 (または段落の最初に) 現れるかである。これらのパターン間の精度を比較すると、差異はみられない。このことから、キーワードが段落の始めに現れるか否かは、定義文を抽出するための手掛かりとしては重要でない傾向があると言える。

5. おわりに

本論文では、定義について書かれた文書から表現パターンを取得した。取得した各パターンを用いて新聞記事から定義文を抽出し、その結果を人手で調査した。さらに、各パターンの有効性を精

度・網羅性の観点から調査した。その結果、定義文を抽出するために有効であるパターンが把握できた。

今後は、調査する質問文やパターンを増やし、追調査を行う予定である。

参考文献

- [1] Ellen M. Voorhees “Overview of the TREC2001 Question Answering Track” In E.M Voorhees and D.K.Harman, editors, Proceeding of the Tenth Text Retrieval Conference (TREC2001), pages 42-51, 2002
- [2] NTCIR (NII-NACSIS Test Collection for IR Systems) Project URL <http://www.research.nii.ac.jp/ntcir/>
- [3] Junichi Fukumoto, Tsuneaki Kato, Fumito Masui. Question Answering Challenge for Five ranked answers and List answers Overview of NTCIR4 QAC2 Subtask 1 and 2, pp.283-290
- [4] Jun'ichi FUKUMOTO “Question Answering Challenge (QAC-1) An Evaluation of Question Answering Task at NTCIR Workshop 3”, Proceedings of the Third NTCIR Workshop, 2002
- [5] 黒橋禎夫, 長尾真, 佐藤理史, 村上雅彦. “専門用語辞典の自動的ハイパーテキスト化の方法” 人工知能学会誌, Vol.7, No.2, pp.336-345, 1992.
- [6] H Saggion, R Gaizauskas “Proceedings of FLAIRS 2004” dcs.shef.ac.jp
- [7] Text REtrival Conference (TREC) Home Page URL, www.trec.state.tx.us/
- [8] 西野文人, 橋本三奈子, 落谷亮. “テキストからの用語とその定義文の抽出” 言語処理学会第5回年次大会発表論文集, pp.124-127, 1999
- [9] 藤井敦 “World Wide Web を用いた事典知識情報の抽出と組織化” 電子情報通信学会論文誌 Vol.J85-D-II No.2 2002.