

文末の述語表現を利用した質問テキストの解析

柳瀬 隆史¹ 丸元 聡子² 難波 功³ 落谷 亮⁴

† 富士通研究所

‡ 計量計画研究所

¹tyana@jp.fujitsu.com

²smarumoto@ibs.or.jp

³namba@jp.fujitsu.com

⁴ochitani@jp.fujitsu.com

1 はじめに

製品やサービスに関する顧客からの質問への対応を行うために、多くの企業がコールセンターと呼ばれる顧客対応窓口を設けている。従来の対応手段は電話や FAX が主であったが、近年では電子メールや Web フォームで質問を受け付け、回答をメールで返信するところが増えている。この方法は、回答する側では質問とそれに対する回答をそのままログとして蓄積することができるため、過去の類似質問を検索して回答文を再利用したり、よく尋ねられる質問 (FAQ) の分析に利用することができ、回答作成業務の効率化が期待できる。

質問ログに対して FAQ の分析や類似質問検索を高精度に行うためには、ログテキストを解析して質問内容の中心部分をあらかじめ抽出し、さらに質問や状況説明などの質問文中での役割別に分類することが有効である。我々はこれまでに、抽出された質問内容の中心部分のみを検索に用いた類似質問検索において、質問ログの単なる全文検索と比較して検索精度が大幅に向上することを確認している [1]。

従来我々が行ってきた質問テキストの解析では、文末の特徴表現に基づいたパターンマッチングによって、「質問表現」「行為表現」などの談話要素に分類して質問内容の中心部分を抽出する。しかし、パターンによっては談話要素の判別が曖昧になってしまうこともあり、特に「行為表現」と「状況表現」の判別が難しいパターンが多い。

そこで本稿では、「行為表現」と「状況表現」の判別に焦点を当て、文末の特徴表現の前に来る動詞やサ変名詞などの述語表現も組み合わせるパターンを作成することにより、より正確な談話解析を行う手法を提案する。また、質問ログからこれらの述語表現を機械的に収集して分類する方法についても述べる。

2 質問テキストの談話解析

2.1 談話要素

本研究で対象とする質問ログテキストは、パソコンやインターネットに関する質問を受け付けるサポー

ト窓口に送られてきたものである。これらの質問テキストを文単位で詳細に分析すると、障害内容を述べている文や実際の質問が書かれている文など、いくつかの型に分類することができる。他にも、挨拶文やお礼など質問内容とは関係が薄いと思われる記述も少なくない。そこで、本研究では質問内容の中心部分として以下に示す 4 種類の談話要素を定義する。また、本稿でいう談話解析とは、質問テキスト中の各文 (複文の場合は単文または節に分割する) に対して、これらの談話要素に相当するかどうか、相当するならば 4 種類のうちのどの談話要素に分類されるかを判別する処理を示すものとする。

- 行為表現 (action)
質問者が実際に行った行為や操作を示す表現。
- 状況表現 (statement)
質問時の状況を示す表現。障害状況を客観的に述べた表現など。
- 質問表現 (question)
実際の質問を示す表現。
- 意図表現 (wish)
質問者が実現したいことを示す表現。

実際の質問テキストに対する談話解析の例を図 1 に示す。ここでは、談話解析の結果、いずれかの談話要素に相当すると判定された箇所に対応するタグが付与されている。

2.2 文末の特徴表現に基づく談話解析

質問テキスト中の文がどの談話要素に相当するかについては、文末に来る特徴的な表現に着目すれば判断できることが多い。例えば、文末が「～できません。」となっている文は、状況表現であると判断することができる。そこで談話解析は、文末の特徴表現のパターンマッチングによって行っている。照合するパターンは、表 1 のように談話要素の種類別にまとめられており、例えば「設定の方法を教えてください。」という文は「質問表現」というように、照合したパターンによって談話要素が決定される。

こんにちは。お世話になります。
 <action>いまネットワークの設定をしています。</action>
 <wish> ネットワークでNETBEUIプロトコルを追加したいのですが</wish>、
 <statement> 一覧に表示されません。</statement>
 <question> どうやって追加すればいいのか教えてください。</question>
 宜しくお願い致します。

図 1: 質問テキストに対する談話解析例

表 1: 文末の特徴表現の例

談話要素	特徴表現
行為表現	～してしまいました ～してみました
状況表現	～できません ～になりました
質問表現	～でしょうか ～教えてください
意図表現	～したいのですが ～しようと思っています

3 動詞を利用した談話解析

3.1 動詞を組み合わせた照合パターン

談話解析に用いる文末の特徴表現パターンには、「思う」「教える」「できる」など質問文で頻繁に用いられるものを除いて、原則として動詞は含まれていない。これは、1つのパターンでなるべく多くの文を照合させるようにするためである。しかしパターンによっては、どの談話要素に分類すればよいか、特に「行為表現」と「状況表現」の判別が困難なものがある。

判別が難しい特徴表現としては表 2 に示すものが挙げられる。例えば、「～しています」というパターンには「CD-ROM を使用しています。」「エラーが発生しています。」の両方の文が照合するが、実際は前者は行為表現、後者は状況表現である。このパターンだけでは両者が同じ談話要素に判別されるため、正確な判別ができない。

そこで、それらのパターンの直前にある動詞に着目する。例えば「使用する」という動詞は、パソコンに関する質問・障害報告という範囲で考えれば、質問者の行為を指す可能性が高く行為表現として使われることが多いと考えられる。また「発生する」という動詞はエラーや不具合などの状況を指し、状況表現となる可能性が高い。従って、これらの動詞を特徴表現パターンと組み合わせることによって、さらに正確な判別ができると考えられる。

なお、行為表現の文の主語は質問者自身であり、状況表現の文の主語はパソコンなどの物や状況そのもの

など「人」以外になる場合が多く、文の主語を特定することによっても判別が可能と考えられる。しかし、実際の質問文では主語が明示されない場合も多く、明示されたとしてもパターンマッチングによる談話解析では、主語の判別は難しい。今回は、動詞収集のための談話要素判別に主語の特定を用いることにする。

表 2: 談話要素判別の難しい特徴表現

	特徴表現の例
サ変動詞を含むもの	「～する」「～します」 「～した」「～しました」など
継続を表すアスペクトを含むもの	「～(し)ている」 「～(し)ています」 「～(し)ております」など

3.2 動詞の収集

文末の特徴表現と組み合わせる動詞は、行為表現に属するものと状況表現に属するものに分けて質問テキストログから収集する必要がある。ここでは、動詞を機械的に収集し、状況表現と行為表現に分類する方法について述べる。

3.2.1 状況表現動詞の収集

先にも述べたように、行為表現と状況表現の判別は、主語が「質問者自身」か、「人以外」になるかで判断することができるが、今回談話解析に使用したパソコンに関する質問テキスト中の文には、以下に示すような特徴がある。

- 文中で主語が明示されている場合はほとんど「人以外」であり¹、状況表現と判別できる。
- 主語が明示されていない場合は、主語は「質問者自身」「人以外」どちらの可能性もあるため、談話要素の判別はできない。

¹質問者以外の人为主語になっていて明示されている場合もあるが、ごくまれであるので除外して考える

そこで、まず次に示す手順で、状況表現になる動詞を収集する。

1. 述語表現収集用質問テキストから、文末が「動詞+特徴表現」の形になっている文を選別する。特徴表現は表2に示すものとする。
2. 1で選別した文について、主語が明示されているかどうか判定する。次に示す条件をいずれも満たす文を主語の明示されている文とする。
 - (a) 「名詞+格助詞『が』」が含まれていること
 - (b) (a)と文末の「動詞+特徴表現」の間に、用言(動詞・形容詞・形容動詞)が含まれていないこと
3. 主語が明示されている文は、状況表現であるとみなして、特徴表現の前にある動詞を取り出す。

3.2.2 行為表現動詞の収集

以下に示す特徴表現については、そのままのパターン照合だけで容易に行為表現と判断することができると考えられる。つまり、質問ログテキストに対してこれらのパターンで照合を行ってマッチした文の談話要素は行為表現となる。

- 試行を表す動詞を含む
～(し)てみる、～(し)てみた、～(し)てみます、～(し)てみました
- サ変名詞+「を行う」
～を行う、～を行います、～を行いました、～を行った

そこで、行為表現になる動詞は以下の手順で取り出す。

1. 状況表現動詞の収集で選別された、主語の明示されていない文から特徴表現の前にある動詞を取り出す。さらにその中で、状況表現になる動詞と重複しているものあれば取り除く。
2. 1の動詞と、上に挙げた特徴表現を組み合わせでパターンを作成する。
3. 2で作成したパターンで質問ログテキストに対して照合を行い、パターンにマッチする文があれば、その動詞を行為表現になる動詞として取り出す。

4 実験と考察

今回対象とする質問ログテキストとして、パソコンの操作やインターネット接続に関する質問(19836件、105735文)を用意した。これらを述語表現収集用質問テキスト(13224件、70822文)と評価用質問テキスト(6612件、34913文)に分け、動詞の収集実験と評価実験を行った。

4.1 動詞の収集実験

まず、述語表現収集用質問テキスト70822文を用いて、3.2.1節に示した方法で状況表現となる動詞を収集した。文末が「動詞+表2に示す特徴表現」になっていた文は9315文で、うち主語が明示されている文は1423文であった。このうち主語が「人」であったものが9文あり、これらを除いた1414文から156語の動詞を取り出した。

次に、主語が明示されていない7892文に対して、頻度2回以上で出現する動詞を取り出したところ367語あった。このうち、状況表現になる動詞として先に取り出したものを除いた270語について、3.2.2節に示す方法で行為表現となる動詞を収集した。その結果、うち37語の動詞について特徴表現と組み合わせたパターンによる照合でマッチした文があったので、行為表現になる動詞とした。

さらに、残った233語の動詞については、質問テキストからその動詞を含む文を抜き出して分析し、人手で行為表現か状況表現かの判別を行った。内訳は、「思う」や「願う」など他の談話要素の判別にすでに用いている動詞などを除き、行為表現138語、状況表現67語であった。表3に収集された動詞の例を挙げる。

表3: 収集された動詞の例(サ変動詞は「する」を省略)

行為表現 (自動で収集) 37語	使う、外す、調べる、 使用、クリック、インストール 確認、実行、設定、など
行為表現 (人手で判別) 138語	押す、聞く、落とす、 購入、選択、増設、 変更、バージョンアップ、など
状況表現 (自動で収集) 156語	入る、消える、起きる 起動、終了、失敗、 フリーズ、ハングアップ、など
状況表現 (人手で判別) 67語	固まる、変わる、鳴る 機能、続行、再現、 復帰、対応、回復、など

4.2 評価実験および考察

評価実験では、評価用質問テキスト34913文から文末が「動詞+表2の特徴表現」になっている文4567文を取り出して、手作業で談話要素の判別を行った。行為・状況以外の談話要素に判別された文や判別が困難であった文などを除き、いずれかの談話要素に判別された3745文(行為1385文、状況2360文)を正解セットとした。

次に、正解セットの各文に対して、パターン照合による談話要素の判別を行った。照合に用いたパターンは次の3種類である。

(a) 表2の特徴表現パターンのみ、動詞は組み合わせない

例:「～しました」→行為表現
「～ています」→状況表現

(b) 収集された動詞と特徴表現を組み合わせたパターンのみ

例:「～使っています」→行為表現
「～発生しました」→状況表現

(c) (a)と(b)の併用

ここで(b)のパターンでは、収集されていない動詞を含む文については照合されないで、判別漏れ(談話要素の判別が出来ない)が生じる。(c)の併用パターンでは、(b)のパターンによる照合で判別漏れとなった文に対して(a)のパターンで再度照合と判別を行うので、判別漏れは生じない。

正解率による評価結果を表4に示す。特徴表現のみのパターンと比較して、動詞を組み合わせたパターンを併用することにより、正解率が大幅に向上していることが確認できる。動詞の収集方法による比較では、人手判別により収集した動詞も加えることによる正解率の伸びはそれほど大きくなく、自動収集による動詞だけでもある程度の正解率の向上を得ている。

一方、動詞を組み合わせたパターンのみによる照合では、正解率の向上は見られない。これは、先に述べた判別漏れがあるためである。その詳細を表5に示す。この表から分かるように、行為表現では、人手判別により収集した動詞を追加することにより判別漏れが大幅に減っており、自動収集だけでは判別に十分な動詞が収集できていないことが分かる。

また、判別誤りについて原因を調べたところ、同じ動詞で行為表現と状況表現の両方になっているものがあることが分かった。例えば、「起動する」という動詞の場合、「質問者自身」と「システム側」の両方を主語としてとり、「試しにPCを起動しました。」という文では行為表現、「PCが自動的に起動しました。」という文では状況表現になる。この場合談話要素を正確に判別するには、主語の有無や文中の他の語句の情報などをさらに考慮する必要がある。しかし、このような動詞は他に「開始する」「終了する」などでそれほど多くはなく、多くの動詞は行為表現と状況表現に分類することが可能であることが分かった。

5 おわりに

本稿では、パターン照合による談話解析をより正確に行うために、質問テキストから収集した動詞を文末の特徴表現に組み合わせた照合パターンを用いる手法を提案した。評価実験の結果、動詞と特徴表現を組み合わせたパターンを、従来の特徴表現のみのパターンと併用することにより、談話要素の判別

表4: 評価結果(正解率)

	照合パターン	動詞収集方法	
		自動収集のみ	自動収集+人手判別
行為	特徴表現のみ	59.4%	59.4%
	動詞組み合わせのみ	50.6%	75.3%
	併用パターン	78.1%	84.5%
状況	特徴表現のみ	70.4%	70.4%
	動詞組み合わせのみ	70.2%	74.0%
	併用パターン	92.4%	93.1%

表5: 動詞組み合わせパターンの評価詳細

	動詞収集方法	正解	判別誤り	判別漏れ
行為	自動収集のみ	50.6%	6.4%	43.0%
	自動+人手	75.3%	6.8%	17.9%
状況	自動収集のみ	70.2%	1.6%	28.2%
	自動+人手	74.0%	3.7%	22.3%

精度が大幅に向上することが確認できた。今後はさらに細分化した談話要素による談話解析を検討しており、談話要素の詳細な判別が必要になるところで今回提案した手法や実験で得られた知見を応用していきたいと考えている。

参考文献

- [1] 柳瀬隆史, 難波功, 木田敦子, 落谷亮, 談話構造を利用した質問応答事例の検索, 言語処理学会第7回年次大会, pp.30-33, 2001.
- [2] 森田良行, 松木正恵, 日本語表現文型, アルク, 1989.
- [3] 山口秋穂, 秋本守英, 日本語文法大辞典, 明治書院, 2001.
- [4] 黒橋禎夫, 日笠亘, 藤井綱貴, 入力質問と知識表現との柔軟なマッチングに基づく対話的ヘルプシステム, 2001年情報学シンポジウム, pp.11-18, 2001.
- [5] 佐藤円, 佐藤理史, ネットニュース記事群の自動パッケージ化, 情報処理学会論文誌, Vol.38, No.6, pp.1225-1234, 1997.