

Q&A サイトにおける質問と回答の分析 (2)

— 文の構造と属性を中心に —

栗山和子¹⁾ 神門典子¹²⁾

本研究では、Q&A サイトにおける質問と回答を分析し、質問者が複数の回答群からベストアンサーを選択するとき、質問と回答の文章としての構造と属性が影響を与えているかどうかを調べるため、質問と回答の構造について考察する。

まず、Q&A サイトに投稿された質問と回答を分析することにより、文章の属性として文章の難易度と文字種の割合を計算した。次に、文書の構造を明らかにするため、質問と回答に含まれる構成要素を抽出し、質問と回答の表現のパターンについて考察した。

結果として、質問と回答の構造と属性には、カテゴリごとにいくつかの特徴的なパターンが見られることがわかった。

Analysis of Questions and Answers in Q&A Site (2) — Based on Document Structures and Attributes —

KAZUKO KURIYAMA¹⁾ and NORIKO KANDO¹²⁾

In this paper, we analyzed questions and answers in Q&A site to investigate whether document structures and attributes of questions and answers affected each questioner's selection of the best answer for his/her own question.

First, we computed readability of questions and answers and ratios of a variety of characters in them as attributes. Then, we showed a list of elements included in the document structures, and labeled the questions and the answers with the elements. Finally, we found that there were some patterns of the document structures and attributes of questions and answers by categories.

1. はじめに

近年、Q&A サイトと呼ばれる質問回答サイトの利用者が増加している^{12),17)}。Q&A サイトの仕組みは、利用者が匿名 (ユーザー ID) で質問を投稿し、その利用者とは別の利用者が回答を投稿するというものである。一般的に、質問あるいは回答をすることで得られる金銭的な報酬はなく、回答に対して質問者や他の利用者が直接あるいは投票などで付与する簡単な評価 (ポイントやコメント等) 得られるだけである。利用者が質問・回答を行う理由は様々であるが、その行動は利用者の自発的なものであり、記述される情報・知識の内容や表現形式は多種多様であるため、質問に対して質問者の目的に沿った回答が得られるとは限らない。

三浦ら¹⁰⁾によれば、コミュニティ Q&A サイトの利用者が質問・回答を行う理由としては、質問者においては、全体的に外発的動機 (回答が得られることによる実質的な利益獲得) と内発的動機 (回答が得られることによる内的な充足) が多く、回答者においては、援助的動機 (質問者を助けること) が最も多い。また、質問者の 6 割以上は、質問する前に他の手段を用いて質問の内容についての調査を行っているが、回答者の半数以上は、回答する前に質問について調査を行っていない。つまり、回答者は、必ずしも、正しい回答を知っているから回答を行うわけではないと考えられる。

本研究では、利用者が内容以外にどのような観点で、回答する質問やベストアンサーを選んでいるのか検討するため、まず、質問文の表現が回答の内容や表現にどのように影響を与えているかを調べる。本稿では、「Yahoo!知恵袋」の研究機関提供用データ^{17),18)}を使用し、質問と回答を分析することにより、文章から構造と属性を抽出し、質問文・回答文の構造と属性が回答者による回答する質問の選択や質問者によるベストアンサーの選択に関連しているかどうか考察する。

2 節では、文章の構造や難易度を情報アクセス技術として研究している関連文献について簡単に述べる。3 節では、文章としての質問と回答の構造と属性について分析し、いくつかの表現のパターンを示す。4 節では、本稿のまとめを述べ、本研究の今後の予定と課題を示す

¹⁾ 白百合女子大学

Shirayuri College

¹²⁾ 国立情報学研究所

2. 関連研究

個別の著者の文体の特徴や、あるジャンルに属する文章の構造・表現形式などについて、文章の構成要素（文字、単語、文節、文、段落等）を用いて統計的な分析や比較を行う研究は、19世紀から計量文体学、計量言語学などの分野で伝統的に行われてきた^{6),11)}。本節では、その長い歴史をひもとくことはせず、文章の構造あるいは難易度を情報アクセス技術として研究している関連文献についてだけ簡単に述べる。

三池ら⁹⁾は、文書構造に基づく検索と抄録生成を行う全文検索システムの評価のため、文の意味的役割として「話題」、「背景」、「従来問題」、「目的」、「特徴」、「結果」、「結論」、「課題」の8種類を定義し、人手で科学技術論文に付与することにより、システムの意味的役割抽出規則の評価実験を行っている。

神門^{1),3),4)}は、日本語の科学技術分野の論文の構造を分析し、全文データベース検索への応用を検討した。また、日本語と英語の新聞記事に対しても、テキストの構造を機能的側面から、「主記」、「解説」、「背景」、「見通し」、「意見」という構成要素を定義し、複数の被験者を用いて認識の一致度についての検証を行っている²⁾。この実験では、被験者間の一致度は8割以上であり、新聞記事の構成要素としての妥当性が確かめられている。

Teufelら⁵⁾は、科学技術論文を対象として、「AIM」、「TEXTUAL」、「OWN」、「BACKGROUND」、「CONTRAST」、「BASIS」、「OTHER」という文タイプ(annotation scheme)を定義し、機械学習を用いて文タイプを同定することによって、テキスト自動要約に利用している。

関ら¹⁵⁾は、新聞記事からの主観的情報の自動抽出の精度の検証のため、神門²⁾の構成要素を拡張した、「主記」、「解説」、「背景」、「著者の意見」、「見通し」、「識者の意見」という6つの文タイプを定義している。

日本語の難易度については、佐藤ら¹⁴⁾が、教科書コーパスに基づく難易度推定システム『帯』を構築し、一般に公開している。また、教材以外のテキストに対する難易度も測定可能にするため、国立国語研究所が編纂している「現代日本語き言葉コーパス」を利用して、標準的なレベルの難易度に関する予備調査と設定を行っている⁷⁾。この調査において、佐藤らは、中学3年に対応するレベル（『帯』の13段階の9）を標準的な難易度（7段階の4）のとして難易度の変換式を設定している。

柴崎ら¹⁶⁾も、教科書コーパスを用いて、日本語のリーダーティの測定式を提案し、文の長さ、文字種（平仮名、漢字、カタカナ、ローマ字）の割合、文法の複雑さ、語彙の難易、

文中の情報量などを用いてリーダーティ公式決定の実験を行っている。

本研究では、回答者が回答する質問を選択するとき、あるいは、質問者が自分の質問に付けられた回答からベストアンサーを選択するとき、質問文・回答文の難易度や読みやすさが、その選択に何らかの影響を与えているのではないかと考えている。その仮定を検証するための予備的な調査として難易度の判定を行うが、難易度の評価手法を提案することが目的ではないので、難易度の判定ツールとしては『帯』(obi-2)^{7),13),14)}を用いる。質問・回答の属性としては、文の長さ（文字数）、文字種、難易度を用い、質問、ベストアンサー、その他の回答でどのような傾向があるかについて簡単に考察する。

3. 質問と回答の分析

3.1 使用データの概要

本研究では、Q&Aサイトのデータとして「Yahoo!知恵袋 研究機関提供用データ 国立情報学研究所 (NII) 提供版 ver.1」¹⁸⁾（以下、知恵袋データ）を使用する。知恵袋データは、2004年4月1日から2005年10月31日に「Yahoo!知恵袋」ベータ版に投稿された質問と回答から抽出されたもので、解決済みの質問3,116,009件、質問者が選んだベストアンサー3,116,008件、その他の回答10,361,777件から成る。1つの質問には、1つのベストアンサーと0個以上のその他の回答が存在する。質問者ID、回答者IDは、本来のYahoo!JAPAN IDからユニークな番号に変換されているが、同じIDは同じ番号に変換されているため、番号により利用者を識別することは可能である。

「Yahoo!知恵袋」では、質問はトピックのカテゴリごとに分けられ、質問者が質問投稿時にカテゴリを選択するようになっている。現在のカテゴリは、カテゴリ内の質問と回答を含んだまま、ベータ版以降何度が再構成されて数も増えているため、知恵袋データのカテゴリと現在の「Yahoo!知恵袋」のカテゴリとは名称や階層構造が異なっている。また、知恵袋データでは、ベータ版当時のカテゴリ構造のデータは提供されていないため、本研究の分析では、カテゴリ間の階層関係はないものとして扱う。

本研究では、以前の論文⁸⁾で質問タイプの分類に使用した5つのカテゴリ、「数学、サイエンス」、「恋愛、人間関係の悩み」、「パソコン、周辺機器」、「料理、グルメ、レシピ」、「健康、病気、ダイエット」のうち、「情報検索型」（サーチエンジンや質問応答システムによる情報検索や図書館のレファレンス・サービスなどを利用すれば適切で客観的な回答が得られるタイプ）の質問が8割以上であった「パソコン、周辺機器」、「社会調査型」（適切な回答の基準がなくQ&Aサイトに参加している人に個人的な意見・嗜好・経験などを尋ねること

とよって質問者が主観的に適切な回答や結論を決定するアンケート調査のようなタイプ)の質問が9割以上であった「恋愛、人間関係の悩み」、「情報検索型」と「社会調査型」の質問がほぼ半々であった「料理、グルメ、レシピ」の3つのカテゴリーの質問と回答のデータを使用する。手作業で文章の構造の分析には、以前の論文⁸⁾で質問タイプの分類するため、上記の3カテゴリーの全データから無作為に抽出した300件(1カテゴリー100件ずつ)の質問とその回答のセットを使用する。

カテゴリーごとの質問件数、1質問あたりの平均文字数、各質問(Q)に付随する回答の件数の平均、ベストアンサー(以下、BA)の平均文字数、その他の回答(以下、NA)の平均文字数を表1に示す。

表1 カテゴリーごとの質問件数・平均文字数・平均回答数

カテゴリー	質問件数	Q 平均文字数	平均回答数	BA 平均文字数	NA 平均文字数
恋愛、人間関係の悩み	210,105	130.8	6.7	133.5	92.2
パソコン、周辺機器	171,848	109.9	2.8	140.2	90.3
料理、グルメ、レシピ	37,714	70.8	4.3	115.7	72.4

1質問あたりの平均回答数は、「恋愛、人間関係の悩み」(6.7)が最も多く、「パソコン、周辺機器」(2.7)が最も少ない。表には記載していないが、全286カテゴリーの1質問あたりの平均文字数と平均回答数はそれぞれ95.4と4.3である。

3.2 文中の文字の種類

3つのカテゴリーの質問(Q)、BA、NAに含まれる文字の種類を、記号、英数字、カタカナ、ひらがな、漢字の5種類に分け、その割合を調べた。それらの平均を表2に示す。文字は半角と全角の区別をせず、どちらをも1文字として数えている。全体的な傾向が見られるように、文字種の割合ごとの件数のヒストグラムを図1、図2、図3に示す。縦軸は文書数(件数)、横軸は文章全体に占める各文字種の割合で、区間は[0,1]、階級幅は0.05である。

表2と図1~3からわかるように、「恋愛、人間関係の悩み」では、カタカナの割合が少なく、ひらがなが多い傾向にある。「パソコン、周辺機器」は、他の2カテゴリーに比べて、英数字とカタカナの割合が多く、漢字の使用が少ない。これは、パソコン関連の用語に英数字やカタカナが多いためと考えられる。また、情報源として本文の中でURLを示している場合、URLに含まれる英数字もそのまま1文字ずつ数えているので、英数字の割合が高くなっていることがある。いずれのカテゴリーでも、質問・BAと比べて、NAは記号の割合が多く、

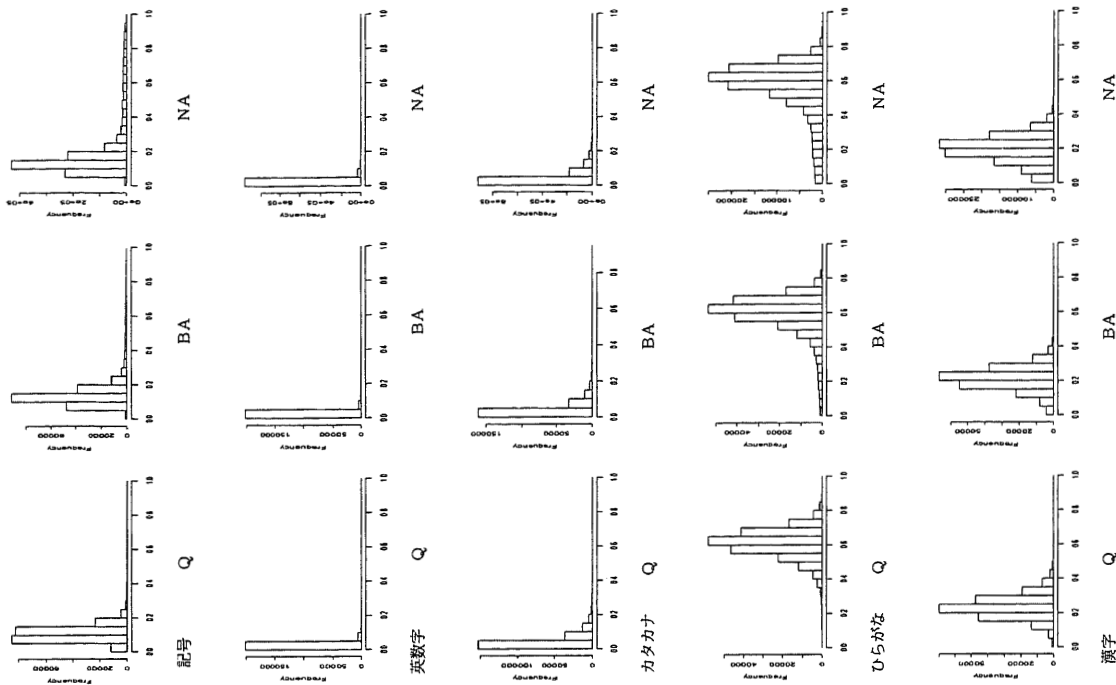


図1 「恋愛、人間関係の悩み」: Q, BA, NAの文字種の割合(%)のヒストグラム

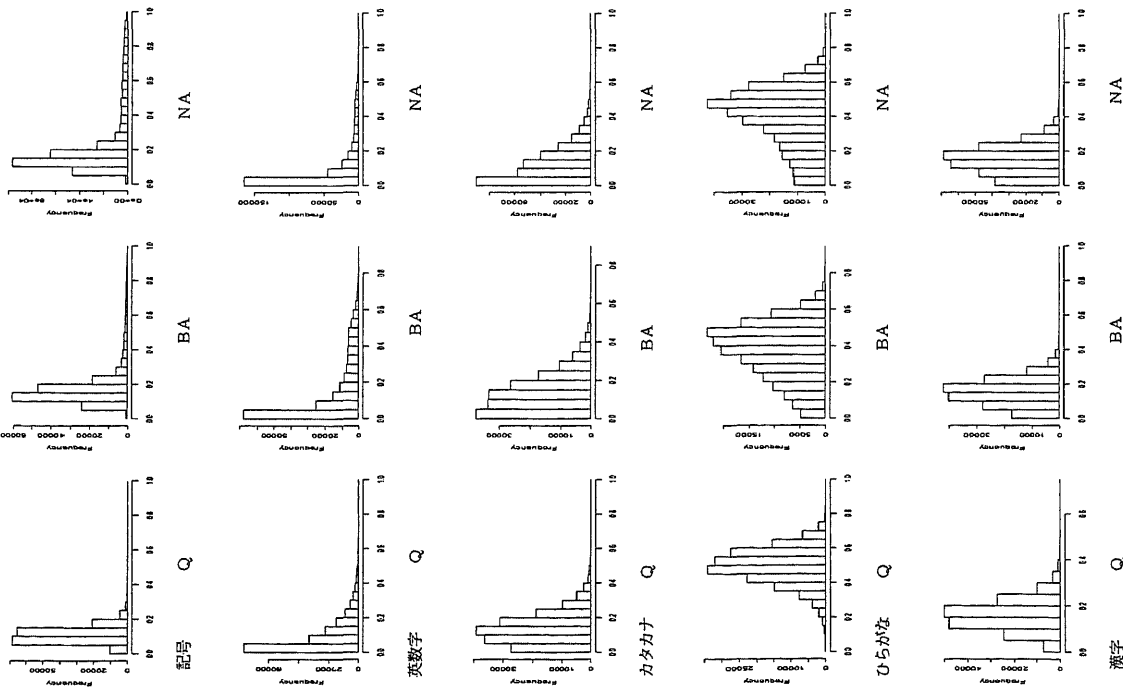


図 2 「パソコン、周辺機器」: Q, BA, NA の文字種の割合 (%) のヒストグラム

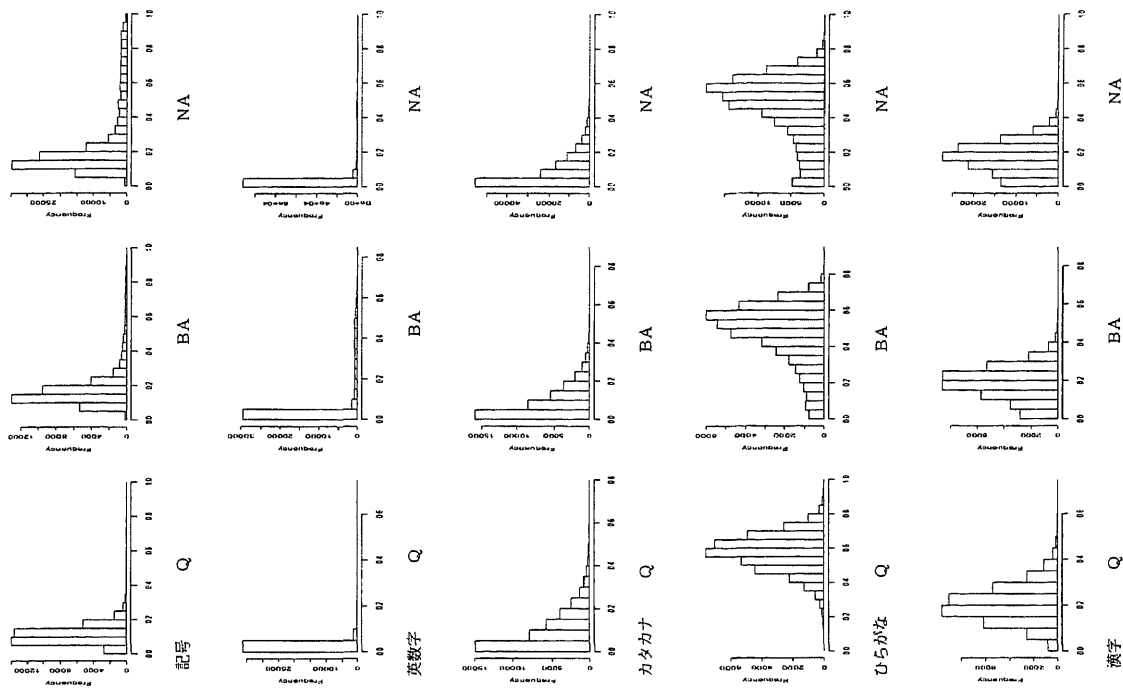


図 3 「健康、精気、ダイエット」: Q, BA, NA の文字種の割合 (%) のヒストグラム

表 2 Q, BA, NA の文字種の割合 (%) の平均

カテゴリ (Q/BA/NA)	記号	英数字			カタカナ			ひらがな			漢字
		Q	BA	NA	Q	BA	NA	Q	BA	NA	
恋愛、人間関係の悩み	Q	11.3	1.0	3.7	60.8	23.4					
	BA	16.2	0.9	3.1	58.7	21.0					
	NA	21.1	1.0	3.1	55.4	19.4					
パソコン、周辺機器	Q	11.1	9.7	13.9	49.4	15.9					
	BA	18.4	15.8	13.8	36.7	15.3					
	NA	23.1	10.2	12.2	39.3	15.2					
料理、グルメ、レシビ	Q	11.5	0.8	9.7	57.7	20.5					
	BA	18.8	7.1	8.8	46.7	18.7					
	NA	25.6	2.9	8.8	45.7	17.1					

漢字の割合が少ない。

カテゴリ間で文字種の比較を行うと、全体的には上記のような傾向があるが、質問と BA の文字種の割合の相関係数を計算したところ、相関は見られないが、カテゴリ内では文字種の文字種が直接 BA の文字種に影響を与えているわけではないが、カテゴリ内では文字種の割合に一定の傾向があると考えられる。質問・回答を作成するときに、利用者が文字種を意識して文章を作成したり、回答を選択したりすることは無いと思われ、そのカテゴリ内でよく見られる質問・回答の表現に合わせて、表現や文章の読みやすさを調整している可能性が考えられる。

3.3 文章の難易度

質問と回答の難易度に差があるかどうか調べるため、以前の論文⁸⁾で使用した 300 件の質問、その BA, NA の文章の難易度を『帯』(obi-2)^{7),13),14)}を用いて計算した。『帯』の難易度は 1~13 の 13 段階で、1~6: 小学 1~6 年、7~9: 学 1~3 年、10~12: 高校 1~3 年、13: 大学、である。

カテゴリごとの質問、BA, NA の難易度の平均を表 3 に示す。(括弧内は分散)。難易度の分布を箱ひげ図として図 4 に示す。

表 3 Q, BA, NA の難易度の平均 (各カテゴリ 100 件)

カテゴリ	Q	BA	NA
恋愛、人間関係の悩み (100 件)	6.6 (1.9)	6.9 (3.2)	6.7 (5.0)
パソコン、周辺機器 (100 件)	6.8 (2.8)	8.3 (3.5)	7.5 (6.4)
料理、グルメ、レシビ (100 件)	5.7 (3.6)	6.5 (3.3)	6.3 (6.5)

表 3, 図 4 からわかるように、質問よりも回答の方が平均では難易度が高く、範囲に広が

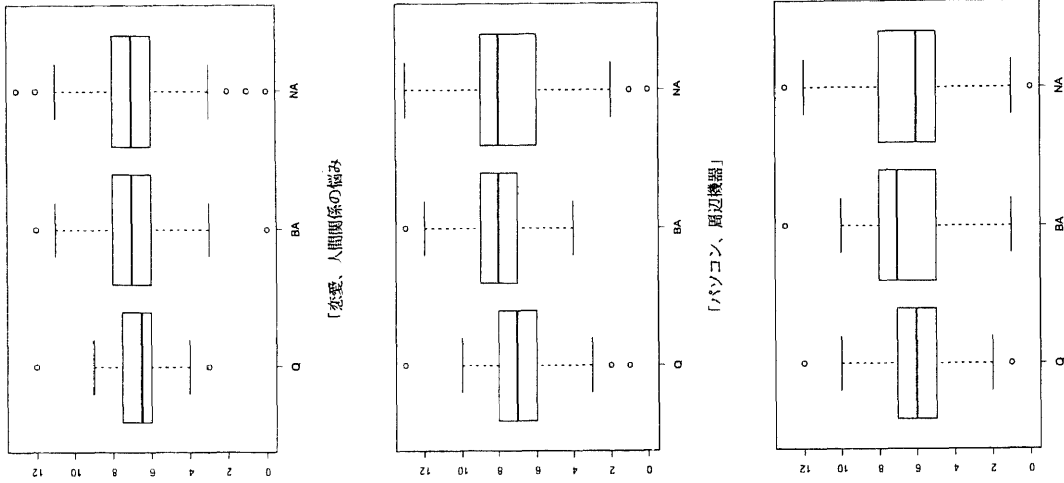


図 4 Q, BA, NA の難易度の平均 (各カテゴリ 100 件)

りがある。BAの難易度が質問の難易度以上である割合は、「恋愛、人間関係の悩み」では69%、「パソコン、周辺機器」では87%、「料理、グルメ、レシピ」では75%である。複数の回答を持つ質問（NAを持つ質問）の件数は、「恋愛、人間関係の悩み」では94件、「パソコン、周辺機器」では70件、「料理、グルメ、レシピ」では85件であるが、そのうち、BAの方がNAよりも難易度が高かったのは、それぞれ、15件（6.3%）、19件（27.1%）、16件（18.8%）であった。

以上のことから、回答者は回答する質問文よりも日本語として難易度が高い回答文を作成する傾向があることがわかる。また、質問者は自分の質問文よりも難易度が高い回答をBAとして選択することが多いが、必ずしも、難易度が一番高いものがBAに選ばれるわけではないと考えられる。ただし、これらは、分析に使用した3カテゴリにおける傾向であり、カテゴリが異なれば、異なる結果になる可能性もある。

3.4 質問と回答の構造

カテゴリごとの質問（Q）、BA、NAの文の数の平均を表4に示す。括弧内は分散である。

表4 Q、BA、NAの文の数

カテゴリ	Q	BA	NA
恋愛、人間関係の悩み	5.3 (16.5)	5.9 (25.1)	4.5 (79.6)
パソコン、周辺機器	4.5 (18.3)	5.6 (30.0)	3.9 (26.2)
料理、グルメ、レシピ	3.0 (5.0)	5.1 (21.8)	3.7 (30.7)

表4からわかるように、BAに比べてNAの文の数の平均は少ない。ただし、分散からわかるように、文の数は質問・回答ごとのばらつきが大きい。「料理、グルメ、レシピ」の質問では、分散が小さく、ばらつきは比較的小さいが、これは質問の内容におすすめのレシピや食材の説明や調理法を尋ねるものなど、文章が比較的単純であるものが多いことが原因であると考えられる。

3.1項の表1に示したように、知恵袋データの質問と回答は、長さ（文字数）が100文字前後、文の個数が平均3~6件であり、学術論文や新聞記事と比較して文章全体が短く、構造が明確ではない。一文ずつを排他的に構成要素として分けてしまうと、一文の中に複数の構成要素（役割）が含まれている場合、構成要素がもれてしまうので、本研究では、一文ずつに構成要素のラベルを付与することはせず、各要素が質問文あるいは回答文の中に含まれているかどうかとその組み合わせのパターンを調べることにする。構成要素は、文献^{(2),(15)}を参考にし、質問と回答とは異なる構成要素を使用する。構成要素のラベルを付与する質

問とBAのデータは、前項で使用した300組（1カテゴリ100組ずつ）である。

質問の構成要素としては、以下の5つを使用する。

- (Q1) 挨拶：回答者に対する挨拶・呼びかけ
 - (Q2) 背景：質問の背景・状況説明
 - (Q3) 本文：質問本文
 - (Q4) 予想：回答の予想
 - (Q5) 意見・感想：質問に関する自分の意見・感想
- 回答の構成要素としては、以下の4つを使用する。
- (A1) 挨拶：質問者に対する挨拶・呼びかけ
 - (A2) 本文：回答本文
 - (A3) 意見・感想：質問・自分の回答に関する意見・感想
 - (A4) 他回答：同じ質問への回答への言及（参照、意見、感想）

構成要素のラベルの付与は、筆者の一人が行った。付与結果として、表5、表6に各構成要素を含む質問・BAの件数を、表7に質問とBAの構成要素の組合せのパターンとその件数を示す。ただし、組合せは構成要素の出現順序を考慮せずに数えているので、構成要素は表7に示した順序で質問とBAに出現するとは限らない。つまり、以下で組合せについて言及するとき、質問とBAの中での構成要素の出現順序は表記の順序とは限らない。

表5 Qの構成要素

カテゴリ	Q1	Q2	Q3	Q4	Q5
恋愛、人間関係の悩み	9	75	100	4	21
パソコン、周辺機器	29	74	100	7	5
料理、グルメ、レシピ	12	65	100	10	5

表6 BAの構成要素

カテゴリ	A1	A2	A3	A4
恋愛、人間関係の悩み	9	99	12	1
パソコン、周辺機器	4	99	17	1
料理、グルメ、レシピ	4	100	41	1

表 7 構成要素の組合せ

Q1	Q2	Q3	Q4	Q5	A1	A2	A3	A4	LH	PE	FG
○	○	○	○			○	○		0	1	0
○	○	○	○	○	○	○			1	2	1
○	○	○			○				1	0	0
○	○	○			○				2	2	0
○	○	○				○			0	3	5
○	○	○							4	20	4
○	○	○			○	○			0	0	1
○	○	○							1	0	1
○	○	○							0	1	0
○	○	○							1	0	2
○	○	○							0	4	4
○	○	○							1	0	0
○	○	○							2	2	2
○	○	○							14	3	2
○	○	○							0	0	1
○	○	○			○				4	1	1
○	○	○			○		○		0	1	0
○	○	○							6	6	13
○	○	○							39	29	29
○	○	○							0	0	2
○	○	○							2	0	1
○	○	○							3	0	0
○	○	○							0	0	1
○	○	○							0	1	1
○	○	○							1	0	1
○	○	○					○		15	21	10
○	○	○							1	0	1
○	○	○							1	2	12
○	○	○							1	1	0
○	○	○							0	0	5

LH: 恋愛、人間関係の悩み, PE: 料理、グルメ、レシビ, FG: 構成要素を含む

表 5~7 から、質問と BA の構成要素の組合せは、どのカテゴリでも、「(Q2) 背景-(Q3) 質問本体-(A2) 回答本体」というパターンが最も多いことがわかる。また、カテゴリごとにみると、「恋愛、人間関係の悩み」では「(Q2) 背景-(Q3) 質問本文-(Q5) 意見-(A2) 回答本文」、「パソコン、周辺機器」では「(Q1) 挨拶-(Q2) 背景-(Q3) 質問本文-(A2) 回答本文」、「料理、グルメ、レシビ」では「(Q2) 背景-(Q3) 質問本文-(A2) 回答本文-(A3) 意見」と「(Q3) 質問本文-(A2) 回答本文-(A3) 意見」というパターンの件数が他の 2 カテゴリに比べて多くなっていることがわかる。

文章の内容からその原因を検討すると、「恋愛、人間関係の悩み」では、まず、質問の最初や質問本文の中で、人間関係の背景や現在の状況を説明し、質問本文の途中や質問の最後で質問内容についての自分の意見や感想を述べることが多いので、「(Q2) 背景」と「(Q5) 意見・感想」を含むパターンが多くなると考えられる。「パソコン、周辺機器」では、質問者が質問の最後に「よろしくお願ひします」や「どなたかわかる方、教えてください」などという文を付けることが多いので、「(Q1) 挨拶」を含むパターンが多くなる。「料理、グルメ、レシビ」では、質問に対する直接的な回答以外に、自分の体験による意見や感想を付け加えることが多いので、「(A3) 意見・感想」を含むパターンが多くなる。

以上のことから、質問の構成要素と BA の構成要素の組合せには、どのカテゴリの質問と BA にも当てはまる一般的な組合せ以外に、特定のカテゴリだけでよく見られるコミュニケーションの形式を反映したパターンが存在することがわかる。本稿では、質問と回答の構造を明らかにするために、含まれている構成要素のみを数えたが、構成要素の出現順序を考慮に入れて詳細な分析を行えば、カテゴリごとに特徴的な構成要素の組合せのパターンを抽出できる可能性も考えられる。

4. おわりに

本研究では、利用者の目的や意図によって質問・回答文の表現が異なるかどうか、また、質問文の表現が回答の内容や表現にどのように影響を与えているかを調べるため、質問と回答の文章の構造（構成要素）と文章としての属性（文字種と日本語としての難易度）について分析を行った。その結果、カテゴリごとに見たとき、質問と回答に含まれる文字種の割合には一定の傾向があることがわかった。また、回答は質問に比べて難易度が高いことが多いが、必ずしも、一番難易度が高い回答がベストアンサーに選ばれているわけではないことがわかった。さらに、質問と回答の構造には、カテゴリごとの構成要素の組合せのパターンがあることがわかった。これらの原因としては、カテゴリごとによく見られる表現形式との関連

が考えられるが、どれくらいの影響があるのかについてはさらに詳細な分析と検証が必要である。

今後の課題として、以下のようなことを行う予定である。

- データ数を増やして、複数人で構成要素の付与を行い、一致度を統計的に検証する。
- 質問・回答の難易度や構造が回答者の質問への回答行動・質問者の回答の選択行動に影響を与えているかどうか分析・検討する。
- 質問と回答の構造について、構成要素の出現順序を考慮に入れた分析を行う。
- 以前の論文⁸⁾で提案した質問タイプの分類と質問・回答の構成要素との関連について分析し、自動分類に使用できるような特徴を抽出できかどうか検討する。

謝辞 本研究の実施にあたっては、ヤフー株式会社が国立情報学研究所に提供した「Yahoo!知恵袋 研究機関提供用データ 国立情報学研究所 (NII) 提供版 ver.1」¹⁸⁾を利用した。

参 考 文 献

- 1) 神門典子：文献の機能構造を用いた全文データベース検索の試み、学術情報センター紀要, Vol.7, pp.29-47 (1995).
- 2) 神門典子：認識特性に基づくテキスト構造の分析：日英新聞記を例として、学術情報センター紀要, Vol.8, pp.107-129 (1996).
- 3) Kando, N.: Text-level structure of research articles and its implication for text-based information processing systems, *Proc. of the 19th British Computer Society Annual Colloquium on Information Retrieval Research*, Aberdeen, Scotland, UK, pp.68-81 (1997).
- 4) Kando, N.: Text structure analysis as a tool to make retrieved documents usable, *the 4th International Workshop on Information Retrieval with Asian Languages*, Taipei, Taiwan, pp.126-135 (1999).
- 5) Teufel, S. and Moens, M.: Summarising scientific articles - experiments with relevance and rhetorical status, *Computational Linguistics*, Vol.28, No.4, pp.409-445 (2002).
- 6) 金明哲：テキストデータの統計科学入門, 岩波書店, 東京 (2009).
- 7) 小島健輔, 佐藤理史：現代日本語書き言葉均衡コーパスに対する難易度付与, 電子情報通信学会技術研究報告, Vol.109, No.84, TL2009-3, pp.13-18 (2009).
- 8) 栗山和子, 神門典子：Q&A サイトにおける質問と回答の分析, 情報処理学会研究報告. 情報学基礎研究会報告, Vol.2009-FI-95, pp.1-8 (2009).
- 9) 三池誠司, 住田一男：文の意味役割解析に基づく全文検索, 情報処理学会研究報告. 情報学基礎研究会報告, Vol.FI-94-76, pp.17-24 (1994).

- 10) 三浦麻子, 川浦康至：人はなぜ知識共有コミュニティに参加するのか：質問行動と回答行動の分析, 社会心理学研究, Vol.23, No.3, pp.233-245 (2008).
- 11) 村上征勝：真贋の科学：計量文献学入門, 朝倉書店, 東京 (1994).
- 12) OKWave: 教えて!goo. <http://oshiete.goo.ne.jp/> (参照 2009-10-15).
- 13) 佐藤理史：帯2：日本語テキストの難易度推定.
<http://kotoba.nuee.nagoya-u.ac.jp/sc/obi2/obi.html> (参照 2009-10-15).

- 14) Sato, S., Matsuyoshi, S. and Kondoh, Y.: Automatic Assessment of Japanese Text Readability Based on a Textbook Corpus, *Proc. of LREC'08* (Calzolari, N., Choukri, K., Maegaard, B., Mariani, J., Odijk, J., Piperidis, S. and Tapias, D., eds.), Marrakech, Morocco, European Language Resources Association (ELRA) (2008).
- 15) 関洋平, 神門典子, 青野雅樹：日本語と英語の文タイプの自動付与とその特徴素, 電子情報通信学会技術研究報告, Vol.NLC2005-28, PRMU2005-55, pp.23-28 (2005).
- 16) 柴崎秀子, 沢井康孝：国語教科書コーパスを応用した日本語リーダビリティ構築のための基礎研究, 電子情報通信学会技術研究報告, Vol.107, No.246, TL2007-32, pp.19-24 (2007).

- 17) Yahoo!JAPAN: Yahoo!知恵袋. <http://chiebukuro.yahoo.co.jp/> (参照 2009-10-15).
- 18) Yahoo!JAPAN: 「Yahoo!知恵袋」データの提供について.

<http://research.nii.ac.jp/tde/chiebukuro.html> (研究機関用データは国立情報学研究所 IDR 事務局から配布されている) (参照 2009-10-15).