

# メッセージの背後に潜む「問い」の抽出

## Extraction of Questions Behind Messages

松村 真宏

Naohiro Matsumura

大阪大学大学院経済学研究科

Graduate School of Economics, Osaka University

matumura@econ.osaka-u.ac.jp, <http://www2.econ.osaka-u.ac.jp/~matumura/>

河原 大輔

Daisuke Kawahara

情報通信研究機構

National Institute of Information and Communication Technology (NICT)

dk@nict.go.jp

岡本 雅史

Masashi Okamoto

東京工科大学片柳研究所

Katayanagi Advanced Research Laboratories, Tokyo University of Technology

okamotoma@media.teu.ac.jp

黒橋 禎夫

Sadao Kurohashi

京都大学大学院情報学研究科

Graduate School of Informatics, Kyoto University

kuro@i.kyoto-u.ac.jp, <http://nlp.kuee.kyoto-u.ac.jp/~kuro/>

西田 豊明

Toyoaki Nishida

(同上)

nishida@i.kyoto-u.ac.jp, <http://www.ii.ist.i.kyoto-u.ac.jp/~nishida/>

**keywords:** question extraction, case frame

### Summary

To overcome the limitation of conventional text-mining approaches in which frequent patterns of word occurrences are to be extracted to understand obvious user needs, this paper proposes an approach to extracting questions behind messages to understand potential user needs. We first extract characteristic case frames by comparing the case frames constructed from target messages with the ones from 25M sentences in the Web and 20M sentences in newspaper articles of 20 years. Then we extract questions behind messages by transforming the characteristic case frames into interrogative sentences based on new information and old information, i.e., replacing new information with WH-question words. The proposed approach is, in other words, a kind of classification of word occurrence pattern. Qualitative evaluations of our preliminary experiments suggest that extracted questions show problem consciousness and alternative solutions – all of which help to understand potential user needs.

### 1. はじめに

アンケートの自由記述欄やコールセンター、電子掲示板に寄せられるメッセージには、商品やサービスへの要望や評判や不満など人々の生の声が溢れている。そのような人々の声を新商品開発やサービス向上に活かすために、メッセージから注目すべきパターンを抽出するテキストマイニングに近年注目が集まっている [那須川 01]。ここで述べた「注目すべきパターン」は主にメッセージに現れる文字列の頻出パターンのことであり、顕在化しているニーズを汲み取るために有用である。

しかし、当然のことながら顕在化しているニーズは同業他社も既に気づいていることが多いため、そのようなニーズに基づいたマーケティング戦略は価格競争や他社製品・サービスとの差別化などの点から得策ではない [コトラー 04]。したがって、有能なマーケターほど人々の潜在的なニーズに注目し、人々に新たな価値観を気づかせることにより市場を開拓していく。例えば、

- 「コーヒー 1 杯で 500 円は高い」という不満の背後にある「1 杯 180 円で立ち飲みもできる店」といった要望 [Associé 03]
- 「和食レストランを利用する」という目的の背後にある「家族連れでレジャーを楽しむ」といったライフ・スタイル [IT Pro 03]

などが潜在的なニーズになり、新たな価値観を人々に提供する。このようなニーズはメッセージに表出していないので、これまではテキストマイニングで抽出すべき対象にならなかった。そこで本論文では、あらゆるメッセージはその背後に潜む「問い」に向けて発せられているという仮定 [中村 02] に基づき、人々の潜在的な「ニーズ」をこの問いに注目して取り出すことを試みる。

マーケットリサーチの分野では、フォーカスグループ (調査対象に関心を持つ人たち) へのグループインタビューによって潜在的なニーズを発掘することが試みられている [ヴォーン 99]。しかし、グループインタビューは手間と費用がかかるので、多くの人から意見を集めることや

定期的に行うことは難しい。また、膨大なデータから主観的な意見や評価文を含む文を取り出すことを狙ったアプローチ [立石 01, Morinaga 02, 村野 03, Kushal 03, Yu 03] は、顕在化されたニーズにのみ着目しており、潜在的なニーズやメッセージの背後にある問いには注目していない。アンケートから要求意図を含む文を判定する試み [大塚 04] は本研究の問題意識に近いが、我々はもっと広い枠組みとして問いを扱っている。

本論文はまず 2 章で新情報と旧情報の観点からメッセージの背後に潜む問いを定義する。続く 3 章と 4 章で特徴的な格フレームとメッセージの背後に潜む問いを抽出する手法について述べる。5 章では提案手法によって実際に抽出された特徴的な格フレームと問いについて考察し、6 章で従来手法との比較を行なう。7 章で残された課題について述べ、最後に 8 章で本論文を締めくくる。

## 2. 問いと新情報と旧情報

本研究で対象とするアンケートの自由記述欄や電子掲示板には「ファイバーウィッグへのクチコミ投稿」や「あなたが水の節約のためにしていることを教えてください」などのテーマが与えられている。参加者（アンケートの回答者・電子掲示板のユーザ）はこの与えられたテーマから想起される事柄のうち、相手（アンケートの実施者・電子掲示板の読者）に伝えたい内容についてのメッセージを回答もしくは投稿する。

ここで、参加者がメッセージに込めた内容を明らかにするために、新情報と旧情報の概念を導入する。旧情報とはある状況において知っていることが前提となっている知識であり、新情報とは新たに付け加えられる知識である [福地 85]。上述の場合だと、与えられたテーマは旧情報であり、相手に伝えたい内容が新情報となる。ここで、新情報と旧情報はペアになっていることに着目し、本論文では問いを以下のように定義する。

【定義 1】新情報を引き出す疑問文をメッセージの背後に潜む「問い」とする。

つまり、テーマが与えられたときに連想する事柄が問いであり、その問いに向けて発せられるメッセージを介して相手に新情報が伝わると仮定するのである。

なお、新情報と旧情報の区別は世界知識や共有知識などメッセージの外にある状況に依存するので、客観的に判断することは難しい。しかし、旧情報は意志疎通のために共有されていることが前提となっている情報であることから、新情報より頻繁に用いられていると考えられる。そこで本論文では、メッセージ中の語のうち、相対的に頻度の高い語を旧情報、頻度の低い語を新情報と見なすことにする。また、新情報と旧情報は機能文法における概念であるので、混乱を避けるために、問いによって引き出される新情報を焦点と呼ぶことにする。

表 1 格フレームの例

用言	走る
ヲ格 (直前格)	道路, 路線, 道, コース, ...
ガ格	車, バス, 自転車, 車両, ...
用言	走る
ガ格 (直前格)	衝撃, ショック, 危機感, ...
二格	業界, 市場, 社内, 金融界, ...
ノ格	コメ, 事件, テロ, 最大級, ...
用言	食べる
ヲ格 (直前格)	料理, 弁当, ご飯, 朝食, ...
ガ格	子供, 全員, 人, 園児, ...

## 3. 特徴的な格フレームの抽出

### 3.1 用言の格フレーム

テキストマイニングでは語順や語の共起関係が利用されることが多いが、これは細かいところは無視して文の骨子を軸にテキストを分析したいという考えからである。しかし、日本語は語順が比較的自由なので、語順や語の共起関係からだけでは文の意味を掴むことは難しい。本研究では最終的には文の背後に潜む問いを取り出すことを狙っているので、文の意味や構造まで捉える必要がある。そのための一つのアプローチとして、本研究では、用言（動詞、形容詞、名詞+判定詞）を基準として、取り得る格要素の関係を表した用言の格フレーム（以下では格フレームと略す）を利用する。用言には多義性があるが、用言の意味は用言の直前の格要素（直前格）によってほぼ一意に決定されるので、用言と直前格の組を単位として用例を収集し、格要素の意味属性の類似度に基づいてクラスタリングを行うことによって用言の多義性の問題は解決できる [河原 02]。本研究で用いる格フレームは全て河原らの手法 [河原 05] により自動構築したものであり、3.2 節で用いるコーパス ( $D_{news}$  と  $D_{www}$ ) はそれぞれ 20 年分の新聞記事データ (約 2000 万文) から自動構築した 317924 個の用言についての格フレームと、Web ページ (約 2500 万文) から自動構築した 485350 個の用言についての格フレームからなる。

実際に新聞記事から自動構築された格フレームの一部を表 1 に示す。「走る」の格フレームを見ると、直前格に「道路」や「路線」がくるときはガ格に「車」や「バス」がくることがわかる。また、直前格に「衝撃」や「ショック」がくるときはノ格に「コメ」や「事件」、二格に「業界」や「市場」がくることがわかる。なお、ノ格だけ例外的に直前格にかかる格要素を表している。

### 3.2 格フレームの特徴度

膨大なデータから構築された格フレームコーパス ( $D_{news}$  や  $D_{www}$ ) は、用言と格要素の一般的な関係を表している。一方、解析対象となる膨大なテキストデータ (対象データ) から構築された格フレームコーパス ( $D_{target}$ )

は、対象データにおいてよく使われている用言と格要素の関係を表している。したがって、 $D_{target}$  を  $D_{www}$  や  $D_{news}$  と比較すれば、対象データではよく使用されるが一般にはあまり使用されない特徴的な格フレームが抽出できる。そのような特徴的な格フレームを抽出するために、 $D_{target}$  中の格フレーム  $c$ <sup>\*1</sup>の特徴度（格フレーム特徴度） $u(c)$  を以下の式 (1) で定義する。

$$u(c) = \frac{f_t(c)}{\log(f_w(c) + f_n(c) + e)} \quad (1)$$

ここで、 $f_t(c)$  は  $D_{target}$  における  $c$  の頻度、 $f_w(c)$  は  $D_{www}$  における  $c$  の頻度、 $f_n(c)$  は  $D_{news}$  における  $c$  の頻度であり、 $e$  は自然対数である。つまり、格フレーム  $c$  が  $D_{www}$  にも  $D_{news}$  にも出現していないときは  $u(c) = f_t(c)$  となり、 $D_{www}$  や  $D_{news}$  に多く出現しているほど  $u(c)$  の値は低くなる。

例えば、「まぶたにつく」という格フレーム  $c_1$  は 5・1 節で後述する  $D_{target}$  に 42 回出現しているが、 $D_{news}$  と  $D_{www}$  には一度も出現していないので、

$$\begin{aligned} u(c_1) &= \frac{42}{\log(0+0+e)} \\ &= 42 \end{aligned}$$

となる。また、「拍車をかける」という格フレーム  $c_2$  は、同じく 5・1 節の  $D_{target}$  で 1 回、 $D_{news}$  で 6844 回、 $D_{www}$  で 627 回出現しているので、

$$\begin{aligned} u(c_2) &= \frac{1}{\log(6844+627+e)} \\ &= 0.11 \end{aligned}$$

となる。したがって、「まぶたにつく」の格フレームの方が「拍車をかける」よりも特徴的であると判断できる。

格フレームは、用言の前に格要素を配置するように展開すれば、そのまま意味の通じる文になる。したがって、提案手法で得られた特徴的な格フレームを展開すれば、対象データから特徴的な文が抽出できる。

### 3・3 表記揺れ

対象データから格フレームを抽出しても、表記揺れの問題が残る。例えば、

まぶたにつく  
まぶたに付く  
瞼に付く

は全く同じ意味の格フレームである。一般に読みが同じで漢字が異なる場合も多く、例えば「つく」と読む漢字にも「付く、着く、突く、憑く」などいろいろある。そこで本研究では、格フレームの読みが全て一致していれば

ば同じ意味の文だと見なし、最も平仮名が用いられていない格フレームに統一した。上記の場合だと「まぶたにつく」と「まぶたに付く」はそれぞれの特徴度を「瞼に付く」の特徴度に加えてから削除する。語の読みは形態素解析システム JUMAN [黒橋 99] を用いて調べた。ただし、「洗濯」と「選択」のような同音異義語の判別はできないので、本論文ではそこだけ人手で判断した。また、

まぶたにつく  
まぶたにくつつく

も同じ意味であるが、読みが異なる語同士の類似度の判定については今後の課題とする。

## 4. メッセージの背後に潜む問いの抽出

2章での定義から、新情報を引き出す疑問文が問いである。本研究では、用言と格要素からなる格フレームに着目し、新情報が用言か格要素かによって問いの構造を大きく2タイプに分類する。

用言が新情報のときは問いの焦点は用言になるので、用言を「どうなる/どうする」に置換することによって問いを抽出する。一方、格要素が新情報のときは問いの焦点は格要素になるので、その格要素を「何」に置換することによって問いを抽出する。格要素が2つ以上ある場合も同様であるが、格要素の1つがノ格の場合は「格要素+の+格要素」で一つの対象を表しているのでまとめて扱う。なお、新情報と旧情報の判断は、用言と格要素の対象データにおける出現頻度を比較し、格フレームごとに最も出現頻度の低い語を新情報、それ以外の語を旧情報と見なすことによって行う。

これらをまとめると、格フレームから問いを作成するルールは以下ようになる。

ルール1：用言が新情報の場合 直前格がガ格のときは用言を「どうなる」、それ以外のときは用言を「どうする」に置き換え、文末に「？」をつける。

ルール2：格要素が新情報の場合 格要素を「何」に置き換え、文末に「？」をつける。

ルール3：格要素にノ格が含まれる場合 ノ格の格要素は必ず直前格に係っている [河原 05] ので、「ノ格の格要素+の+直前格」を一つの格要素として扱う。

例えば「瞼に付く」という格フレームの場合、「瞼」と「付く」のどちらが新情報になるかによって、異なる問いが生成される。

- 瞼（新情報）に付く（旧情報）  
→ 何に付く？
- 瞼（旧情報）に付く（新情報）  
→ 瞼に どうする？

また、「目の周りに付く」のように格要素にノ格が含まれる場合は、以下のように「目の周り」をまとめて扱う。

- 目の周り（新情報）に付く（旧情報）

\*1 1つの格フレームに複数の用例がクラスタリングされているときは1つの用言に対して同じ格に複数の要素が存在するので、そのときは用言と格要素のそれぞれの組み合わせを格フレームと呼ぶことにする。

→ 何に付く？

- 目の周り (旧情報) に付く (新情報)

→ 目の周りに どうする？

上記の例において「何」「どうなる/どうする」に置き換えられた語が問いの焦点となる。

なお、特徴的な格フレームから作られる問いは新規性の高い要望や問題意識への視点を含んでいるので重要である。したがって、同じ問いに変換される格フレーム  $c$  の格フレーム特徴度  $u(c)$  の和をその問いの特徴度 (問い特徴度) とすることにより、膨大なテキストデータから特徴的な問いを抽出できる。

## 5. 実 験

### 5.1 アットコスメ からの格フレームおよび問いの抽出

3 章, 4 章で提案した手法の有効性を検討するために、アットコスメ\*2 に寄せられたメッセージを解析した。アットコスメは 1999 年 12 月に開設された化粧品情報の Web サイトであり、ユーザによる商品のクチコミ投稿が 100 万件以上寄せられているサイトである。

「ファイバーウィッグへのクチコミ投稿」に投稿された 4161 投稿 (約 1.6 MB, 約 84 万字) を河原らの手法 [河原 05] を用いて解析すると 4150 個の用言についての格フレームが得られ、さらに格フレームの表記揺れを吸収すると 3877 個の用言についての格フレームが残った。その格フレーム特徴度の上位 20 を表 2, 下位 20 を表 3 に示す。「ファイバーウィッグ」はマスカラの商品名であり、表 2 の格フレームを見ると「瞼に付く」「まつげに慣れる」「繊維が落ちる」といったユーザの声を知ることができる。一方、表 3 の格フレームには「拍車をかける」「人が多い」といったファイバーウィッグとは関係のない格フレームが並んでいることがわかる。格フレームは、対象データ内では「まぶたに付くどころか」「まぶたについちゃうかな...」「まぶたについた時に」「まぶたに付くし」といった多様な表現によって記述されていたものであるが、格フレームを用いることによって表現の多様性が吸収でき、用言と格要素という本質的な関係だけを取り出すことに成功している。また、新聞記事データや Web ページなど世の中の一般的な話題を網羅的に集めたテキストデータから作成した格フレームコーパスと比べることによって、アットコスメに特徴的な文が的確に得られていることがわかる。

また、得られた格フレームから抽出した問いのうち、問い特徴度 (格フレームの特徴度の和) の上位の問いとその焦点を表 4 に示す。表 4 を見ると、「ファイバーウィッグ」に対するユーザの要望や問題意識が問いと焦点の対応によってはっきりと見てとれる。例えば、「何を 使う?」という問いの焦点には「MF のウォーターブルーフ」「ランコムのエターニシル」など様々なマスカラの商品名が

挙げられており、「ブラシが どうなる?」という問いの焦点には「細い」や「フィット」といったユーザの意見が見られる。このように相対的な旧情報である「使う」や「ブラシ」を軸として新情報との関係を整理することにより、別々に述べられているユーザの意見を定型的に理解することができる。

### 5.2 アンケートからの格フレームおよび問いの抽出

次に、日経エコロジー Web がアンケートで集めた「主婦の中の声\*3」のうち、「設問 1 あなたが水の節約のためにしていることを教えてください」に対して寄せられた 531 件 (約 40 KB, 約 2 万字) の自由回答文から格フレームコーパスを構築し、特徴的な格フレームの抽出および問いの抽出を行なった。格フレームは 206 個、表記揺れを吸収すると最終的に 191 個得られ、その格フレーム特徴度の上位 20 を表 5, 下位 20 を表 6 に示す。格フレーム特徴度の上位の格フレームには「風呂の洗濯に使う」「風呂の水を使う」といった水の再利用に関する意見から「こまめに水道を止める」「タンクにペットボトルを入れる」といった水の節約に関する意見まで幅広く見られる。一方、格フレーム特徴度の下位の格フレームには「水位を下げる」「汚れをふき取る」を除いて水の節約との関係がはっきりしない格フレームが多い。ここでも、新聞記事や Web ページから構築した格フレームと比較することによってアンケートに特徴的な格フレームが得られていることがわかる。

また、表 7 に示す問い特徴度の高い問いに着目すると、水の用途、水の再利用、洗濯時の工夫など水の節約についてのアンケート回答者の視点が整理されていることがわかる。また、5.1 節で用いたアットコスメのデータに比べると約 40 分の 1 の小規模なデータなので得られる格フレームも少なくなるが、それでも特徴的な格フレームが取り出せていることがわかる。

### 5.3 考 察

5.1 節, 5.2 節での分析から、特徴度の高い格フレームには対象データに特徴的な意見が多く含まれており、また格フレームから得られた問いからはユーザの非明示的な要望や問題意識が浮かび上がってくるのがわかった。ここでさらに問いと焦点の関係を分析すると、問いが対象焦点型、行為焦点型、評価型、言い換え型の 4 タイプに分類できることがわかった。

タイプ 1: 対象焦点型 問いの焦点に商品名や用途など具体的な対象が列挙される問いが対象焦点型である。例えば、表 4 中の「何を 使う?」という問いの焦点には「MF のウォーターブルーフ」「ランコムのエターニシル」といったマスカラの商品名が挙げられ、表 7 中の「何に 使う?」の問いの焦点には「掃除」「植

\*2 <http://www.cosme.net/>

\*3 <http://web.archive.org/web/20030803122514/eco.nikkeibp.co.jp/eco/ecommu/index.html>

表 2 アットコスメにおける格フレーム特徴度の高い格フレーム

順位	格フレーム	u	f <sub>t</sub>	f <sub>w</sub>	f <sub>n</sub>
1.	瞼に 付く	80.00	80	0	0
2.	まつげに 慣れる	40.00	40	0	0
3.	繊維が 落ちる	37.00	37	0	0
4.	カールが 落ちる	36.00	36	0	0
5.	まつ毛が 伸びる	35.00	35	0	0
6.	長いが 出る	35.00	35	0	0
7.	ブラシが 大きい	34.00	34	0	0
8.	まつ毛が 長くなる	33.00	33	0	0
9.	マスカラを 使う	26.32	57	6	0
10.	ボリュームが 出る	23.35	65	25	0
11.	マスカラを 塗る	23.00	23	0	0
12.	クレンジングで 落ちる	20.00	20	0	0
13.	マスカラを 付ける	19.71	35	13	0
14.	ブラックを 買う	16.00	16	0	0
15.	マスカラを 試す	16.00	16	0	0
16.	ボリュームが 足りない	14.00	14	0	0
17.	毛に なれる	14.00	14	0	0
18.	下が 黒くなる	14.00	14	0	0
19.	ビューラーを 使う	14.00	14	0	0
20.	ボリュームが 欲しい	13.91	24	9	0

表 3 アットコスメにおける格フレーム特徴度の低い格フレーム

順位	格フレーム	u	f <sub>t</sub>	f <sub>w</sub>	f <sub>n</sub>
1.	拍車を かける	0.11	1	627	6844
2.	人が 多い	0.11	1	6565	1
3.	とと 思う	0.11	1	6050	28
4.	気を使 用	0.12	1	2148	1559
5.	印象を 受ける	0.12	1	2281	1175
6.	力を 発揮	0.12	1	1266	2096
7.	力をつ ける	0.12	1	1168	2067
8.	条件を 満たす	0.12	1	1738	1388
9.	効果が出 る	0.13	1	428	2230
10.	事が 多い	0.13	1	2406	39
11.	手をつ ける	0.13	1	1021	1378
12.	度が 高い	0.13	1	933	1330
13.	声を かける	0.13	1	1288	809
14.	声を出 す	0.13	1	2010	1
15.	金がか かる	0.13	1	1174	620
16.	群を 抜く	0.13	1	551	1182
17.	感を与 える	0.13	1	670	1041
18.	意見が 多い	0.14	1	343	1288
19.	効果を 狙う	0.14	1	178	1390
20.	問題が ない	0.14	1	540	988

表 4 アットコスメにおける問い特徴度の高い問いとその焦点

順位	問いと焦点	Sum of u
1.	何を 使う? (MFのウォータープルーフ、ランコムのエターニシル、エレガンスの下地、サンプル、テストイ... (他 142 個))	269.82
2.	ブラシが どうなる? (細い、フィット、良い、太い、ダメだ、デカイ、あらい、好きだ、汚れる、大きめ、ソフトだ、恐ろし... (他 17 個))	165.02
3.	まつ毛が どうなる? (当たる、長くなる、下がる、硬くなる、絡まる、カール、ボリュームアップ、抜ける、すだれ、パリ... (他 35 個))	152.08
4.	まつ毛に どうする? (絡まる、見とれる、のる、もってくる、つける、なれる、合う、いい、重い、悪い、なじむ、憧れる、... (他 32 個))	124.42
5.	マスカラを どうする? (上塗り、探す、渡り歩く、重ねる、合わせる、落とす、にじむ、使い分ける、リピート、支持、試す、... (他 21 個))	116.81
6.	何に 付く? (まぶた、目じりのほう、瞼、トコ、内側、きわ、まぶたの目尻、目の回り、目のきわ、まぶたの周辺... (他 29 個))	115.20
7.	ボリュームが どうなる? (ダメダメ、物足りない、得る、足りない、欲しくなる、凄い、印象、全然、少ない、いまいちだ、違... (他 18 個))	85.25
8.	何が 落ちる? (まつげのカール、繊維、粉、ディグニータ、にももの、毛の方、色、カス、カール、元のカール、毛... (他 26 個))	81.48
9.	繊維が どうなる? (頼りない、多い、細い、ポロポロ、フィルム、絡む、付く、絡みつく、汚くなる、飛び散る、固まる、... (他 26 個))	66.46
10.	何で 落ちる? (マイクレ、パウダー、クレンジング、リムーバー、汗、オイル、あくびの涙、ぬるま湯、洗顔、ピッ... (他 22 個))	63.22
11.	私に どうする? (強い、イマイチ、十分だ、ナチュラルだ、NG、よい、合う、駄目だ、最高だ、きつい、ピッタリ、救... (他 33 個))	61.08
12.	何に 使う? (仕上げ、泊りの日、マスカラのあと、下地のあと、スポーツの際、薄化粧の際、薄化粧の際、最後... (他 39 個))	58.74
13.	マスカラが どうなる? (多い、溶ける、にじむ、見つかる、なくなる、のる、苦手だ、落ちる、好評、取れる、抜ける、欲しい... (他 17 個))	57.01
14.	何を 買う? (ブラウン、ヴィセのを、落とし、ナチュラルブラウン、ファイバーウィッグ、マスカラコート、ブ... (他 22 個))	56.76
15.	何が 良い? (伸び、バサバサ、アナスイの方、クチコミの評判、発色、デジャブのマスカラ、皆さんのクチコミ... (他 26 個))	54.75
16.	何を 付ける? (カール、ボリュームマスカラ、フルアンドソフト、カバーガールのマラソン、アンブリシル、ワン... (他 43 個))	52.42
17.	何を 使用? (クレンジング、ヴァティジニアス、ヘレナのヴァティジニアス、コンタクト、ドラマティカルア... (他 31 個))	51.22
18.	まつ毛を どうする? (実現、伸ばす、演出、カール、コーティング、コーミング、痛める、太くする、長くする、作る、強調... (他 18 個))	51.01
19.	何を 購入? (エレガンスの下地、メイペリンのフルアンドソフト、ナチュラルブラウン、ファイバーウィッグ... (他 14 個))	46.78
20.	何が 出る? (艶、力、ミニブラシ、効果、茶色、ボリューム、差、涙、インパクト、ハリ、ツヤ、雰囲気、まつげのポリ... (他 17 個))	46.48

表 5 アンケートにおける格フレーム特徴度の高い格フレーム

順位	格フレーム	$u$	$f_t$	$f_w$	$f_n$
1.	風呂の洗濯に使う	21.00	40	4	0
2.	湯を使う	15.89	64	72	39
3.	洗濯に使う	15.65	52	14	11
4.	風呂の水を使う	12.00	12	0	0
5.	洗濯に使用	9.00	9	0	0
6.	水を使う	7.31	38	840	904
7.	風呂のやりに使用	6.09	8	1	0
8.	洗濯に利用	6.00	6	0	0
9.	湯で洗濯	6.00	6	0	0
10.	風呂の湯で洗濯	6.00	6	0	0
11.	水をせんとくに利用	5.00	5	0	0
12.	洗濯に水を使う	5.00	5	0	0
13.	こまめに水道を止める	5.00	5	0	0
14.	掃除に使う	4.52	8	0	7
15.	湯を使用	4.05	13	14	8
16.	やりに使う	4.00	4	0	0
17.	風呂の水を使用	4.00	4	0	0
18.	タンクにペットボトルを入れる	4.00	4	0	0
19.	風呂の湯をつかう	3.00	3	0	0
20.	花にあげる	3.00	3	0	0

表 6 アンケートにおける格フレーム特徴度の低い格フレーム

順位	格フレーム	$u$	$f_t$	$f_w$	$f_n$
1.	気を配る	0.13	1	903	1268
2.	汚れを落とす	0.15	1	448	195
3.	気にかける	0.16	1	347	184
4.	人数が少ない	0.18	1	185	49
5.	油を使う	0.18	1	117	103
6.	水につける	0.19	1	206	0
7.	剤を入れる	0.20	1	102	53
8.	機を購入	0.20	1	66	85
9.	水をくみ上げる	0.21	1	15	102
10.	蛇口をひねる	0.22	1	68	27
11.	利用を控える	0.22	1	44	50
12.	量を考える	0.23	1	59	21
13.	身体を洗う	0.23	1	78	0
14.	事を聞く	0.23	1	70	0
15.	皿を洗う	0.24	1	51	14
16.	間を空ける	0.24	1	65	0
17.	米をとぐ	0.24	1	37	22
18.	水位を下げる	0.25	1	18	33
19.	量を加減	0.26	1	36	10
20.	汚れをふき取る	0.27	1	14	26

表 7 アンケートにおける問い特徴度の高い問いとその焦点

順位	問いと焦点	Sum of $u$
1.	何に使う? (そうじ, 植物のやり, 庭の植物, 植木鉢, 洗濯, 植木, 花のやり, 植木のやり, ふろの洗濯, 植物, ... (他 7 個))	59.46
2.	何を使う? (風呂, 桶, 洗剤, 風呂の湯, 米, 汁, ふろの水, 水, シャワー, 風呂の水, 機, 米の汁, 湯, コップ, フ... (他 2 個))	38.77
3.	洗濯にどうする? (つかう, 使用, 回す, 利用)	18.00
4.	何に使用? (植木の撒き, 庭の撒き, 掃除, 風呂のやり, やり, まき, 撒き)	13.09
5.	水をどうする? (くみ上げる, つかう, 替える, とめる, 変える, 止める, 出す, もらう, 使用, 流す, 利用)	11.25
6.	湯をどうする? (張る, 使いきる, はる, つかう, 使用, 少ない, 利用, 活用)	10.91
7.	何を使用? (シャワー, 機, 風呂の水, コマ)	6.77
8.	湯でどうする? (洗う, 洗濯)	6.58
9.	何にあげる? (草木, 花, 植物)	6.00
10.	何に水を使う? (やり, 洗濯)	6.00
11.	何で洗濯? (風呂の湯)	6.00
12.	何を利用? (風呂の残り, 残り, 風呂の湯, 風呂の水)	5.49
13.	何を少なくする? (水の量, 水の出, 容量, 出, 洗剤の量, 量)	5.48
14.	何に利用? (洗濯, 掃除, せんとく, 洗いの時, 時)	5.41
15.	こまめに何を止める? (水道)	5.00
16.	水を何に利用? (せんとく)	5.00
17.	水を何に使う? (掃除, 植木鉢, 洗濯, 植物)	5.00
18.	タンクに何をを入れる? (おもり, ペットボトル)	5.00
19.	何をを入れる? (おもり, ボトル, ペットボトル, 剤, コマ)	4.91
20.	水でどうする? (洗う, やる, すすぐ, 洗濯, すます)	4.51

物のやり（植物の水やりの解析間違い）」などの水の使い道についてのアイデアが挙げられている。直前格が新情報になる問い（「何を～?」「何に～?」「何が～?」など）がこのタイプになる。

タイプ2：行為焦点型 問いの焦点に行為が列挙される問いが行為焦点型であり、例えば表7中の「水をどうする」の問いの焦点には「くみ上げる」「つかう」「替える」などの用途が挙げられる。用言が新情報からなる問い（「～がどうなる?」「～にどうする?」「～でどうする?」など）にこのタイプが含まれる。

タイプ3：評価型 問いの焦点に評価に関する意見が列挙される問いが評価型である。例えば、表4中の「ブラシがどうなる?」の問いの焦点には「細い」「太い」「ダメだ」「好きだ」といった評価に関する意見が挙げられる。用言が新情報になる問いにこのタイプが含まれる。

タイプ4：言い換え型 電子掲示板やアンケートなどから得られるユーザの生の声を分析する場合、ユーザごとの表記や語彙の揺れが起こる。例えば、表7中の「洗濯にどうする?」の問いの焦点には「つかう」「使用」「回す」「利用」など同じ意味で使われている語が並んでいる。このような言い換えに相当する意見が得られる問いが言い換え型である。用言が新情報になる問いにこのタイプが含まれる。

タイプ1の直前格が新情報になる問いについては、直前格の種類（ガ格、ヲ格、ニ格など）によって焦点を主体、手段、目的といった下位区分に分類できる。また、用言が新情報になるタイプ2、タイプ3、タイプの問いは、用言の種類（動詞、形容詞、名詞+判定詞）と類似度を見ることによって自動分類が可能である。つまり、焦点に形容詞がくる場合はタイプ3の評価型、焦点に動詞がくる場合は、列挙される動詞の意味的類似度が高いときにはタイプ4の言い換え型、そうでないときはタイプ2の行為焦点型であると判断できる。動詞の意味的類似度は分類語彙表\*4などのシソーラスを用いて求めることができる[河原02]。したがって、提案手法は1章で紹介した先行研究よりも広い枠組みから潜在的なニーズの発掘を試みていることが分かるだろう。

これら4タイプの具体的な応用として例えばマーケティングの分野を対象とすると、対象焦点型は競合商品の理解、行為焦点型は行動パターンの理解、評価型は価値判断の多様性の理解、言い換え型は語彙の多様性の理解につながる知見が得られると考えられる。あらかじめ対象が絞られていれば、品詞や特定の語ごとに頻度に重み付けを与えることによって、目的に応じたタイプの問いを選択的に取得することも可能である。このように、ユーザの生の声を問いと焦点という視点から分類することによって、ユーザの多様な視点や価値観を把握することが容易になる。様々な観点からメッセージの背後に潜む問いを

意識することが新しい発想の着眼点となり、新商品開発や販売戦略立案の糸口となるであろう。

また、特徴度の高い格フレームを眺めると、新聞記事やWebページにはほとんど出現せず対象データにだけ出現する格フレームと、新聞記事やWebページにも出現するがそれ以上に対象データに出現する格フレームの性質の異なる2種類が得られていることがわかる。例えば、「臉に付く」や「洗濯に使用」などは $D_{www}$ や $D_{news}$ には全く現れない格フレームであるのに対し、「ボリュームが出る」や「湯を使う」などは $D_{www}$ や $D_{news}$ にも現れる格フレームである。前者は新規性の高い意見、後者は強い意見を表していると考えられるが、両者を区別した問いの抽出と評価は今後の課題である。

## 6. 重要文抽出法との比較

提案手法が対象としている問いの精度を客観的に評価することは難しい。そこで本章では、キーワード抽出法を利用して取り出した重要文と提案手法により得られる特徴的な格フレーム（表2と表5に示した格フレーム）を比較し、提案手法の有用性について考察する。重要文の抽出はTFIDF法[Salton 83]により語に重みを付与し、その重みの和の大きい文を取り出すことにより行う。文書 $d$ における語 $w$ のTFIDFによる重み $w(d, w)$ は以下の式(2)より求められる。

$$w(d, w) = \frac{tf(d, w)}{N \log \frac{df(w)}{df(w)} + 1} \quad (2)$$

ここで $tf(d, w)$ は文書 $d$ における語 $w$ の出現頻度、 $df(w)$ は語 $w$ を含む文書数、 $N$ は文書総数であり、 $w(d, w)$ は少数の文書に偏って用いられる語の重みを高く算出する。式(1)の $u(c)$ も特定の対象に偏って用いられる格フレームの特徴度を高く算出しており、TFIDF法と同様のアイデアに基づいている。

ここで、 $D_{news}$ と $D_{www}$ に含まれる語の出現頻度を近似的に文書総数だと見なし、「ファイバーウィッグへのクチコミ投稿」に投稿された4161投稿、および「設問1あなたが水の節約のためにしていることを教えてください」に寄せられた531件の自由回答文をそれぞれ1文書にまとめて $d_{cosme}$ と $d_{survey}$ にすると、それぞれの文書の語の重みはTFIDF法と同様のアイデアで求めることができる。出現頻度の近似値に基づいた語 $w$ の重み $w'(d, w)$ は以下の式(3)で表される。

$$w'(d, w) = \frac{tf(d, w)}{N \log \frac{df'_w(w) + df'_n(w)}{df'_w(w)} + 1} \quad (3)$$

ここで、 $tf(d, w)$ を文書 $d$ における語 $w$ の出現頻度、 $df'_n(w)$ を $D_{news}$ から推定した語 $w$ を含む文書数、 $df'_w(w)$ を $D_{www}$ から推定した語 $w$ を含む文書数である。表8に $w'(d_{cosme}, w)$ の上位20語、表9に $w'(d_{survey}, w)$

\*4 <http://www.kokken.go.jp/katsudo/kanko/goihyo/>

表 8 アットコスメにおける語の重みの上位 20 語

順位	語	$w'$	$tf$	$df_w$	$df_n$
1.	にくい	527.06	527	0	0
2.	じみだ	346.04	346	0	0
3.	ない	144.82	1147	1375	212
4.	長い	123.77	268	3	1
5.	ロング	92.14	493	104	10
6.	やすい	89.01	89	0	0
7.	カラ	81.99	397	62	22
8.	効果	73.29	1018	539107	38174
9.	カール	66.02	332	75	2
10.	自然だ	58.01	58	0	0
11.	マス	54.39	397	738	69
12.	ボリューム	53.88	419	1191	54
13.	まつげ	51.48	286	128	2
14.	綺麗だ	51.01	51	0	0
15.	確かだ	49.01	49	0	1
16.	太い	42.00	42	0	0
17.	きれいだ	35.00	35	0	0
18.	ブラシ	34.26	251	758	109
19.	づらい	29.00	29	0	0
20.	楽だ	29.00	29	0	0

表 9 アンケートにおける語の重みの上位 20 語

順位	語	$w'$	$tf$	$df_w$	$df_n$
1.	洗濯	35.50	259	736	111
2.	ない	33.33	265	1375	212
3.	風呂	32.41	330	13216	199
4.	水	31.09	361	55190	14236
5.	湯	20.24	205	12494	1093
6.	食器	11.77	99	2245	248
7.	歯磨き	8.84	48	113	20
8.	シャワー	8.29	75	4254	208
9.	めだ	8.00	8	0	0
10.	使用	7.09	77	26048	5300
11.	トイレ	6.96	67	7612	1286
12.	節水	6.24	29	51	289
13.	掃除	5.44	46	2361	215
14.	利用	4.99	58	55525	17698
15.	水道	4.56	34	859	380
16.	米	4.49	43	7172	8337
17.	物	4.27	50	60611	10164
18.	植木	4.21	24	149	40
19.	洗面	4.16	13	10	0
20.	洗顔	4.07	19	52	0

表 10 アットコスメにおける文重要度の上位 10 文

順位	文	Sum of $w'$
1.	確かに落ちにくいシダマにならなくていいんだけどブラシが大きくてなんだか塗りにくい	1329.00
2.	マスカラは、にじみにくい、落ちにくいに限ります	1190.49
3.	それとブラシが大きいので、皆さん使いにくいといいたので多少けれど私の場合そんな言う程使いにくくは無かったです	1145.26
4.	にじみにくいのはいいのですが、ブラシが大きくて塗りにくいです	1135.18
5.	にじみにくい(コンタクトをつけているので目薬をしばしばさすのですが、下瞼に転写しにくくなります)	1069.62
6.	でも、乾いてから綿棒で落とせば、すぐ落ちるし、使いにくさも、落ちにくさでカバーして	1058.80
7.	手持ちの落ちやすいマスカラがこのマスカラのおかげで落ちにくいマスカラに変化したのでその点では良かったかも	1055.99
8.	円だし、にしては重ねづけしてもダメになりにくいしにじみにくいので、まあまあかなあという印象です	1055.20
9.	ボリューム・にじみにくさはが、全くカールをキープしてくれないので、私はヘレナの上にマスカラコートとして使ってます	960.58
10.	ブラシが大きすぎて塗りにくいし、「塗るつけまつげ」っていう程ロング&ボリュームが出てるようには感じなかったです	928.25

表 11 アンケートにおける文重要度の上位 10 文

順位	文	Sum of $w'$
1.	湯船には給湯器で湯を張り、湯は浴槽の半分以下で止める風呂の残り水を洗濯に使用シャワーヘッドを節水用に変えた溜め洗い、溜めすぎを心がけるトイレのタンクに節水コマを取り付け、必要以上に流さない荒いものをすすぐときもできるだけ少ない水が流れるように調節する	246.70
2.	洗濯を風呂の残り湯で行う歯磨きの時など水を出しっぱなしにしない食器洗いの時、不要な水を出さないようにする	237.60
3.	洗濯にお風呂の残り湯を使う、洗い物の時水を流しっぱなしにしない、トイレの水を少なく流す	194.89
4.	風呂水をトイレや洗濯、庭のみずやりに使ったり、お風呂に入る時には、水をいれたペットボトルをいれて、水量を増やしたりしています	173.62
5.	お茶碗を洗う時に水の量を少なくお風呂の水で洗濯植物にあげる水はお米のとぎ汁などで	172.09
6.	蛇口はしっかり閉める風呂の残り水は選択やそうじ、植木の水遣りなどに使う水を勢いよく出さない	169.48
7.	洗濯に使わなかった風呂の残り湯をトイレの水を流すのに使う	159.53
8.	風呂水を洗濯に使用し、食器洗い時や歯磨き時等こまめに水を止める	158.89
9.	食器を洗うとき水を出したままにしない、食器の洗剤を流すときは洗い桶に水が落ちるようにして流し、その水でシンクの下の洗いする	154.55
10.	洗濯に風呂の残り湯を使う・水を出しっぱなしにしない、こまめに止める	152.57

の上位 20 語を示す。なお、語の重みは名詞、形容詞、カタカナ語、サ変名詞に対してのみ算出し、それ以外の語の重みは 0 とする。

このようにして付与された語の重みの 1 文ごとの和(文重要度)が大きいほどその文が重要文であると見なされる。アットコスメにおける文重要度の上位 10 文を表 10、アンケートにおける文重要度の上位 10 文を表 11 に示す。重要文抽出では一文がそのまま抜き出されるので文脈の理解が容易であるが、取り出される重要文は全体の一部なのでユーザの意見の分布を知ることは難しい。一方、格フレームは複数の文に共通するものほど格フレーム特徴度が高くなるので一種の自動要約になっており、ユーザの意見の分布を大きく捉えることができる。また、格フレームを問いへ変換してしまえばユーザ全体の意見を整理して概観することができる。重要文抽出においても自動要約[奥村 05]の技術を用いれば全体のユーザ全体の意見を要約することはできるが、5.3 節で述べたように問いへの変換は対象焦点型、行為焦点型、評価型、言い換え型からの視点が得られ、これらは従来の自動要約の範囲外である。以上の考察より、ユーザの意見を大局的に整理できるという点で提案手法は重要文抽出および自動要約よりも優れていると言うことができよう。

## 7. 残された課題

今回用いたコーパスは新聞記事と Web ページから作成したものであり、アットコスメ や環境問題対策とは分野が異なる。これはすなわち、新聞や Web ページは読むが化粧品や環境問題対策にはあまり詳しくない人にとって特徴的な格フレームが得られていることになり、普段から化粧品や環境問題対策に関心のある専門家から見れば重要な格フレームを取りこぼし面白くない格フレームを抽出している可能性がある。適切な格フレームを得るために対象に応じたコーパスを用意する必要がある。

また、格フレームは 1 つの用言と格要素の組み合わせしか表せないのが、元の文が重文や複文の場合は意味を捉えきれない場合がある。また、含意表現に潜む意味をくみ取ることも難しい。例えば「マスカラが目落ちる」という文は「マスカラが目落ちる(から嫌だ)」という意味かもしれないし、「マスカラが目落ちる(のは気にならない)」という意味かもしれない。また、格フレームでは否定文と肯定文の区別がないので、「まぶたにつく」と「まぶたにつかない」が同じ格フレームになる。

問いを作成する際には出現頻度を利用して新情報と旧情報を判断しているが、これは大局的な視点から見たときの大胆な仮定であり、全ての語が新情報になる可能性をもっている。新情報の選び方によって幾通りもの問いを作成することができるが、そのようにして得られる問いの評価については本論文では論じていない。

さらに、新情報が人名や地名や時相名詞のときは「誰」

や「どこ」に変換したほうが自然な問いが得られるが、現在は考慮していない。JUMAN の出力として得られる品詞細分類情報を利用すれば対処可能である。

表記揺れへの対策として 3.3 節で簡単な処理を行っているが、本質的な解決策ではないので、「まぶたにつく」と「まぶたにくつつく」のような同じ意味だが読みが違う文の判断ができない。本研究で用いている JUMAN の最新版(ver.5.0以降)では同じ語の表記バリエーションを扱うために「代表表記」を出力できるようになっており、今後はこちらを利用する予定である。

以上が現状で明らかになっている問題点であり、今後の課題として検討していく予定である。

## 8. ま と め

本論文では、メッセージの中に込められたユーザの要望や問題意識を抽出するための方法として、特徴的な格フレームの抽出法およびメッセージの背後に潜む「問い」とその「焦点」の抽出法を提案した。このような問いに注目したテキストマイニングは全く新しい試みであるが、本論文で基礎的な検討はできたと考えている。最近ではウェブログから人々の評判を検索するサービス\*5や検索キーワードのランキングサービスが次々と登場し、インターネット上のデータを利用して社会の動向を探ることが本格的に行なわれるようになってきたが、提案手法はさらに踏み込んで世の中の潜在的なニーズの発掘に応用できる可能性を秘めている。今後は、実際にマーケティングや商品開発の現場で使ってもらいながら、提案手法の評価と改良を進めていく予定である。

## 謝 辞

本研究の着想は、科学技術振興機構 PRESTO「自然現象・社会動向の予兆発見と利用」において大澤幸生氏(当時筑波大学助教授、現東京大学助教授)と中村洋氏(当時 JST リサーチスタッフ)との議論によって得られました。また、本研究の一部は学術創成研究「人間同士の自然なコミュニケーションを支援する知能メディア技術」の支援を受けて行われました。また、匿名の査読者の方からは示唆に富む貴重なコメントを多数頂きました。記して感謝致します。

## ◇ 参 考 文 献 ◇

- [Associé 03] Associé, 日経 BP 出版センター, 第 3 号, pp. 92, 2003.
- [福地 85] 福地肇: 談話の構造, 大修館書店, 1985.
- [IT Pro 03] IT Pro ウェブページ, モノを買うにはワケがある, 2003. (<http://web.archive.org/web/20031207141556/http://itpro.nikkeibp.co.jp/free/ITPro/OPINION/20030723/1/>)

\*5 例えば blogWatcher (<http://blogwatcher.pi.titech.ac.jp/>) やテクノラティ (<http://www.technorati.jp/>) など。

- [河原 02] 河原大輔, 黒橋禎夫: 用言と直前の格要素の組を単位とする格フレームの自動構築, 自然言語処理, Vol.9, No.1, pp. 3-19, 2002.
- [河原 05] 河原大輔, 黒橋禎夫: 格フレーム辞書の漸次的自動構築, 自然言語処理, Vol.12, No.2, pp. 109-131, 2005.
- [コトラー 04] フィリップ・コトラー (著), フェルナンド・トリアス・デ・ベス (著), 恩蔵 直人 (翻訳), 大川 修二 (翻訳): コトラーのマーケティング思考法, 東洋経済新報社, 2004.
- [黒橋 99] 黒橋禎夫, 長尾真: 日本語形態素解析システム JUMAN version 3.62, 1999.
- [Kushal 03] Kushal Dave, Steve Lawrence, and David M. Pennock: Mining the Peanut Gallery: Opinion Extraction and Semantic Classification of Product Reviews, The Twelfth International World Wide Web Conference (WWW2003), 2003.
- [Morinaga 02] S. Morinaga, K. Yamanishi, K. Tateishi, and T. Fukushima: Mining Product Reputations on the Web, Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD2002), ACM Press, 2002.
- [村野 03] 村野誠治, 佐藤理史: 文型パターンを用いた主観的評価文の自動抽出, 言語処理学会 第 9 回年次大会, pp.67-70, 2003.
- [中村 02] 中村洋: 「XはYだ。」と「XがYだ。」の意味の違いについて, 人工知能基礎論研究会 SIG-FAI-A103-10, pp. 55-60, 2002.
- [那須川 01] 那須川哲哉: コールセンターにおけるテキストマイニング, 人工知能学会誌 Vol. 16, No. 2, pp. 219-225, 2001.
- [奥村 05] 奥村学, 難波英嗣: テキスト自動要約, オーム社, 2005.
- [大塚 04] 大塚裕子, 内山将夫, 井佐原均: 自由回答アンケートにおける要求意図判断基準, 自然言語処理 Vol. 11, No. 2, pp. 21-66, 2004.
- [Salton 83] G. Salton and M.J. McGill: Introduction to modern information retrieval, NY, McGraw-Hill, 1983.
- [立石 01] 立石健二, 石黒義英, 福島俊一: インターネットからの評判情報検索, 情報処理学会研究報告 NL-144-11, pp. 75-82, 2001.
- [Yu 03] Hong Yu and Vasileios Hatzivassiloglou: Towards Answering Opinion Questions: Separating Facts from Opinions and Identifying the Polarity of Opinion Sentences, Empirical Methods in Natural Language Processing (EMNLP-2003), pp. 129-136, 2003.
- [ヴォーン 99] S. ヴォーン (著), J. シナグブ (著), J.S. シューム (著), 井下 理 (翻訳), 柴原 宜幸 (翻訳), 田部井 潤 (翻訳): グループ・インタビューの技法, 慶応義塾大学出版会, 1999.

〔担当委員: 阿部 明典〕

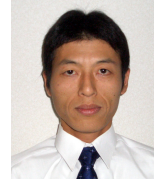
2006 年 5 月 8 日 受理

## 著者紹介



松村 真宏 (正会員)

1998 年大阪大学基礎工学部システム工学科卒業。2000 年同大学院修士課程修了。2003 年東京大学大学院工学系研究科博士課程修了。博士(工学)。東京大学大学院情報理工学系研究科学術研究支援員を経て, 2004 年 3 月大阪大学大学院経済学研究科専任講師。ひらめきを生むコミュニケーション環境のデザインの研究, 誰でも参加して楽しめるフィールドマイニングの研究に従事。



河原 大輔

1997 年京都大学工学部電気工学第二学科卒業。1999 年同大学院修士課程修了。2002 年同大学院博士課程単位取得認定退学。東京大学大学院情報理工学系研究科学術研究支援員を経て, 2006 年独立行政法人情報通信研究機構研究員, 現在に至る。構文解析, 省略解析, 知識獲得の研究に従事。



岡本 雅史

1994 年早稲田大学政治経済学部卒業。1996 年京都大学大学院人間・環境学研究科修士課程修了。1999 年同大学院博士後期課程単位取得認定退学。1996 年日本学術振興会特別研究員, 2002 年東京大学大学院情報理工学系研究科学術研究支援員を経て, 2006 年 4 月東京工科大学片柳研究所研究所客員助教授。博士(人間・環境学)。認知言語学, 語用論, コミュニケーション論の研究に従事。電子情報通信学会 HCG 第 3 種研究会ヴァーバル・ノンヴァーバル・コミュニケーション研究会委員長。



黒橋 禎夫 (正会員)

1994 年京都大学大学院工学研究科電気工学第二専攻修士課程修了。博士(工学)。2006 年 4 月より京都大学大学院情報学研究科教授。自然言語処理, 知識情報処理の研究に従事。情報処理学会平成 5 年度研究賞, 同平成 14 年度山下記念研究賞, 言語処理学会 10 周年記念論文賞, 同平成 17 年論文賞等を受賞。



西田 豊暁 (正会員)

1977 年京都大学工学部卒業。1979 年同大学院修士課程修了。1993 年奈良先端科学技術大学院大学教授, 1999 年東京大学大学院工学系研究科教授, 2001 年東京大学大学院情報理工学系研究科教授を経て, 2004 年 4 月京都大学大学院情報学研究科教授。会話情報学, 社会知のデザインの研究に従事。日本学術会議連携会員。本学会理事・編集委員長。