

# 読者の特性に応じた効果的な表題の作成方法

千田 恭子, 篠原靖志 ((財) 電力中央研究所),

senda@criepi.denken.or.jp, sinohara@criepi.denken.or.jp

## 1 はじめに

近年、要約生成に類似した研究の一つとして、文章の本文から表題を自動生成する研究が行われている [3][2][1]。これらのほとんどは、従来の要約生成の手法に似ており、文書中の用語の頻度分布、文書構造、特定の文型等に基づいて、著者が最も述べたい内容が書かれた箇所を推定して抽出し、表題に加工している。つまり、それらは著者の考えを重視した表題生成といえる。

しかし、そのような視点で抽出した内容が、必ずしも、想定読者の理解や関心を得るのに優れているとは限らない。たとえば、著者が重要と判断して言及している事柄が、必ずしも読者の関心をひくとは限らない。また、たとえば、著者が好んで使う特殊な用語が定義文なしで抽出されてしまうと、読者には理解できない可能性もある。そこで本研究では、読者を重視した表題生成の枠組みを検討する。

読者を重視した表題を生成するためには、文書で説明したい内容ごとに、「表題にどんな要素を含むと、想定読者の関心を得やすいのか (what to say)」と「それはどんな表現で述べた時か (how to say)」を明らかにすることが重要と、本研究では考える。そこで、専門的な技術成果を専門外の読者にアピールする表題のつけ方を明らかにするため、これまでに、同種の技術成果を説明する文書で、専門外の読者の関心をひくことに長けた筆者がつけた表題と、それほど長けていない筆者がつけた表題、もしくはそのようなことを意識せずにつけられた表題を比較し、前者の表題では、後者の表題とは違ってどんな要素が含まれるかと、含まれる同じ要素については、その表現に違いがあるかを調べた。具体的には、同種の技術成果に関する、新聞記事の見出しと論文・報告書の表題とを比較した。

その結果、表題・見出しの表現を、6つの要素 (開発技術の目的、方法、長所、対象、動作、形態を表す要素) に大別し、出現頻度から前者3つを任意要素、後者3つを必須要素に分け、その二つの要素において、どんな要素が表現されるか (what to say) と、それがどのように表現されるか (how to say) という視点から、幾つかの表現パターンを明らかにした [5][6]。

表1は必須要素 (その技術の動作内容を説明する要素) の表現パターンであるが、ここでは「具体的に何をどうする技術なのか」が述べられる場合 (論文・報告書の表題に多い) と、技術の具体的な動作内容は述べずに、単に「何を行うための技術なのか」 (つまり開発目的) が述べられる場合 (新聞の見出しに多い) とに分けられることを明らかにし、それぞれをパターン1、2に分けた。パターン1は更に、専門用語で表現されるか、

それを平易な用語で言い換えるかで、パターン1.1と1.2に分岐する。勿論、前者が論文・報告書に多い表現で、後者が新聞の見出しに多い表現である。

次に表2に示すのは、任意要素のうち、開発技術の特徴を示す要素の表現パターンであるが、ここでは「その技術の実現方法」が述べられる場合 (論文・報告書の表題に多い) と、実現方法は述べずに、「その技術の長所」が述べられる場合 (新聞の見出しに多い) とに分けられることを明らかにし、それぞれをパターン3、4に分けた。パターン3は更に、専門用語で表現されるか、それを平易な用語で言い換えるかで、パターン3.1と3.2に分岐する。勿論、前者が論文・報告書に多い表現で、後者が新聞の見出しに多い表現である。

このように、技術成果を専門外の読者にアピールする場合の、表題における what-to-say と how-to-say の表現パターンを明らかにしたが、これを実際の表題生成の枠組みとして利用するには、これらのパターンが実際にどんな特性の読者に向いているかを検証する必要がある。そのために行った行ったアンケート調査について、本論では報告する。

## 2 アンケートの概要

### 2.1 目的

本アンケート調査の目的は、表1、表2に記載した各表現パターンについて、「それを聞いた表題は、読者にどんな印象を与えるか」と「それは読者の特性 (専門知識の度合い) によって、どう変化するか」を検証することである。このため、3つの専門分野 (送電技術、土木・建築、環境科学) の技術成果を表1、2の表現パターンを組合わせて表現し、その分野への関心と専門知識の度合いの異なる回答者に、その印象を尋ねるアンケート調査を実施した。

### 2.2 方法

本調査は、アンケートへの協力を依頼するアナウンスを見て応募してきた、ある研究機関の職員と、回答者募集のメールに対して返答を返した調査会社のモニターとに対して、主に Web ページを通して、その回答者の各分野に対する専門知識と、幾つかの表題に対する印象を調べるアンケートを行った。

表 1: 必須要素の表題パターン

What to say	パターン 1 (何を行う技術かを表現)		パターン 2 (何を行うための技術かを表現)
	パターン 1.1 (専門用語で)	パターン 1.2 (平易な用語で)	パターン 2.0 (平易な用語で)
対象	放射能半減期を	放射能の寿命を	放射性廃棄物の保管期間を
動作	短縮する	短縮する	短縮する
形態	手法	手法	手法

表 2: 任意要素の表現パターン

What to say	パターン 3 (表現方法を表現)		パターン 4 (技術の長所を表現)
	パターン 3.1 (専門用語で)	パターン 3.2 (平易な用語で)	パターン 4.0 (平易な用語で)
方法	金属燃料 FBR により	燃焼により	
長所			1 万分の 1 に
対象	放射性廃棄物の保管期間を	放射性廃棄物の保管期間を	放射性廃棄物の保管期間を
動作	短縮する	短縮する	短縮する
形態	手法	手法	手法

## 2.3 調査内容

アンケートでは、各回答者の専門知識を調べる質問として、二種類の質問を行った。一つは、その分野への関心の有無と、関心がある場合の情報源（一般の新聞・ニュース番組か、業界紙か、学会・論文誌か）を尋ねる質問である。もう一つは、印象を尋ねる表題に含まれる専門用語について、その用語の意味がわかるかを 5 段階の選択肢（良く分かる、大体分かる、少し分かる、よく分からない、全く分からない）から選ばせる質問である。

前者の質問は、各回答者のその分野への専門知識の度合いを大まかに把握する目安となり、これによって回答者を以下の 4 タイプに分けた。

- 「関心なし」と答えたタイプ（以後、このタイプを Uninterest と呼ぶ。）
- 「関心あり、情報源は一般の新聞やニュース」と答えた回答者のタイプ（以後、このタイプを Interest と呼ぶ。）
- 「関心あり、情報源は業界誌」と答えた回答者のタイプ（以後、このタイプを Engineer と呼ぶ。）
- 「関心あり、情報源は学会誌・論文誌」と答えた回答者のタイプ（以後、このタイプを Researcher と呼ぶ。）

また、後者の質問は、一つ一つの技術に対して、その回答者の知識の度合いを把握する目安となる。

各表題の印象を尋ねるアンケート調査では、表題一つ一つを提示した上で、「その表題はわかりやすいか」「その技術に対して、良い印象を持てるか」「その技術についてもっと知りたいか」を尋ねる質問を行った。（以後、それらの質問に対する答えを、「理解」「好感」「関心」と分類する。）また、前者二つの質問に関して、「わかりにくい」と回答した人、「良い印象を持てる／持てな

い」回答した人には、その理由を選択肢から選ばせた。

### 2.3.1 質問票の内容

ここでは、表題の印象をどのように尋ねたか、具体的な質問票（アンケートページ）の内容を示す。

表題を一題提示した上で、「わかりやすいか（理解）」については、以下の質問を行った。

- この表題はわかりやすいと思いますか。どちらか近い方を選んで下さい。
  - わかりやすい
  - わかりにくい
- わかりにくいと思った場合は、その理由として、あてはまるものを全て選んで下さい。
  - 意味のわからない言葉が使われているから
  - この分野には興味がないから
  - 表題が長すぎて読みづらいから
  - よく知らない分野のことなので、どう評価していいかわからない
  - その他（具体的に説明できる場合は、空欄に記入）

また、「その技術に対して、良い印象を持てるか（好感）」については、以下の質問を行った。

- この表題の技術について好感を持ちましたか。どちらか近い方を選んで下さい。
  - 好感を持った
  - 好感を持たない
- 好感を持った方は、その理由として、あてはまるものを全て選んで下さい。
  - 面白そう
  - 新しそう

- 社会に役立ちそう
- 他の類似技術より優れていそう
- なんとなく良さそう
- その他(具体的に説明できる場合は、空欄に記入)
- 好感を持たれなかった方は、その理由として、あてはまるものを全て選んで下さい。
  - 意味のわからない言葉が使われているから
  - 特に目新しいものではないから
  - よく知らない分野のことなので、どう評価していいかわからない
  - その技術が何の役にたつか分かりにくいから
  - 良い技術とは特に思わないから
  - その他(具体的に説明できる場合は、空欄に記入)

また、「その技術についてもっと知りたいか」については、以下の質問を行った。

- この技術について、もう少し詳しく知りたいと思えますか。あてはまるものを一つ選んでください
  - かなりそう思う
  - ややそう思う
  - あまりそう思わない
  - 全くそう思わない

## 2.4 表題の作成と回答者との組合せ方法

回答者に提示する表題については、一分野につき、3つの新規技術について表題を作成し、一つの技術については、表現パターンの組み合わせを変えた表題を12題用意した(つまり1分野につき36題)。12題の内訳は、必須要素が異なるもの3タイプ(表1の1.1-2.0)×任意要素の異なるもの4タイプ(任意要素のないものと、表2の3.1-4.0)である。

そして回答者には、一分野につき9題の表題の印象を答えてもらう設定にした。そのために、一分野36題の表題を4分割した。分割方法は、必須要素と任意要素が互いに異なる3題の表題を、一つの技術ごとに選択し、3技術×3題で9題とした。また回答者についても、各タイプ別にランダムに4分割して4グループを構成し、各グループの回答者に同じ9題の表題を割り当てた(但し、表題の提示順序は各回答者ごとにランダムにしている)。

## 2.5 回収サンプル数

上記で説明したアンケートを実施した結果、回収したサンプル数の内訳は、表3のようになった。

表3: アンケートの回収数

	1. Uninterest	2. Interest	3. Engineer	4. Researcher
送電	151人	114人	48人	69人
建築	168人	115人	45人	62人
環境	108人	153人	53人	71人

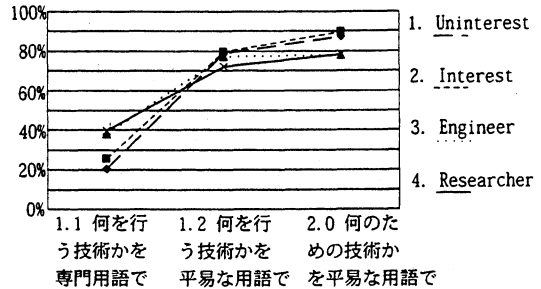


図1: 理解: 分かりやすいと答えた回答者の割合(各タイプ別)

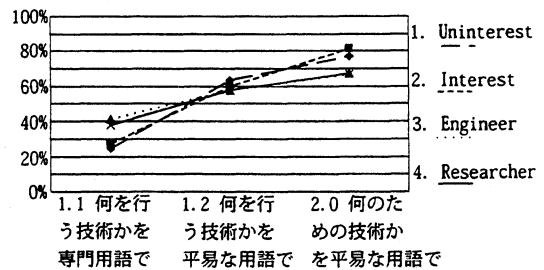


図2: 好感: 良さそう(な技術である)と答えた回答者の割合(各タイプ別)

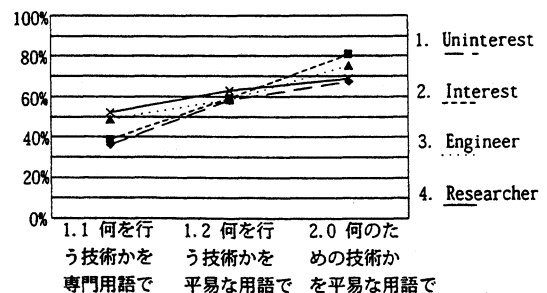


図3: 関心: もっと知りたいと答えた回答者の割合(各タイプ別)

### 3 必須要素の表現パターンによって異なる印象の違い

本論では、必須要素の表現パターンが各タイプの読者にどのような印象を与えるかを、アンケートにより分析した結果について報告する。

図1~図3は、必須要素の表現パターン1.1~2.0を用いた表題に対して、「分かりやすい(理解)」「良さそう(好感)」「もっと知りたい(関心)」という感想を、各タイプの回答者の何%が持ったかを示したものである。(この結果は、調査会社によるアンケートと、筆者の所属先の職員に対するアンケートの結果を合わせたものである。)

「理解」「好感」「関心」のどの図も、全体的にグラフが右上がりであり、各タイプの回答者とも、パターン2.0による表題が、他の要素を用いた表題よりも良い傾向にある。このうち、表3の「関心」のグラフの結果について、各回答者別に $\chi^2$ 乗検定をしたところ、Uninterest、Interest、Engineerタイプ回答者については、有意水準が1%以下であったため、その人数に偏りがあることが確認できた。(Researcherタイプについては、有意水準は5%以下)。また同様に、クラメールの独立係数を算出したところ、Uninterest、Interest、Engineer、Researcherタイプに対して順に、 $r_c = 0.27$ 、 $r_c = 0.35$ 、 $r_c = 0.22$ 、 $r_c = 0.1$ という値となった。(クラメールの独立係数について、 $0.5 > r_c \Rightarrow 0.25$ をやや弱い相関、 $r_c < 0.25$ を非常に弱い相関と判断すると[4]、)専門知識の少ない回答者であるUninterest、Interestタイプは、表現パターンにより関心の度合いが変化する傾向があるが、専門知識の多少あるEngineer、Researcherタイプはそれほど関心の度合いが変化しないと言える。実際、表3の「関心」のグラフ中で、Engineer、Researcherタイプに着目すると、折れ線の起伏がUninterest、Interestタイプより平坦であり、その関心は表現パターンの違いにはあまり左右されにくいと言える。

これより、Uninterest、Interestのような専門知識があまりないタイプの読者には、新聞で多用されるパターン2.0を使った表題が、関心をひく上で有効であることが確認できた。そのため、できるだけ多くの人の「理解」「好感」「関心」を得るには、パターン2.0に基づく表題を生成してつけることが無難と思われる。また、Engineer、Researcherタイプの回答者は、必須要素の表現パターンを変えた表題への反応に、あまり強い相関が見られなかった。そのため、技術成果を報じる文書の中でも、技術速報など、専門家よりの読者を想定した文書については、パターン1.1や1.2を用いた表題をつけることも悪くはないことが示唆されている。

### 4 まとめ

表現パターンの異なる表題について、読者の印象の違いはあるか、それが読者の特性(専門知識の度合い等)によって変わるかを分析した。その結果、必須要素の表現パターンについては、パンフレット等につける万人向けの表題では、開発技術の内容を直接表現するかわりに、開発目的を表現するパターンによる表題が回答者全体から評価される傾向にあった。但し、回答者の専門知識の度合いが増すにつれ、表現パターンによる違いは減少する傾向が見られる。

今後は、任意要素の表現パターンの違いが回答者の印象にどう影響するか、分析を進める予定である。またアンケートでは、表題で用いた専門用語の意味がわかるかを尋ねているが、この回答結果も活用して、知っている用語だけの表題と、知らない用語が入った表題の違いも分析する予定である。

### 5 謝辞

本研究に対してご意見を寄せて下さった東工大精密工学研究所の奥村学助教授、アンケート方法について相談にのって下さった(財)電力中央研究所 経済社会研究所の土屋智子氏に深く感謝致します。

### 参考文献

- [1] A.L. Berger and V.O. Mittal. Ocelot: A system for summarizing web pages. In *Proc. of the 23rd Annual International ACM-SIGIR Conference on Research and Development in Information Retrieval*, pp. 144-151, 2000.
- [2] Rong Jin and Alex G. Hauptmann. Title generation for spoken broadcast news using a training corpus. In *Proc. of the 6th International conference on Spoken Language Processing (ICSLP)*.
- [3] 松本賢治, 伊藤山彦, 谷田泰郎, 柏岡秀紀, 田中英輝. 重要名詞の共起情報を利用した表題生成. 言語処理学会第7回年次大会発表論文集, pp. 375-378, 2001.
- [4] 菅民郎. アンケートデータの分析. 現代数学社, 2000.
- [5] 千田恭子, 篠原靖志. 開発した技術をアピールする表題のつけ方. 情報処理学会自然言語処理研究会報告 111-4, pp. 21-26, 2001.
- [6] 千田恭子, 篠原靖志. 読者の関心をひく表題はどんな特徴を持つか-論文の表題と新聞の見出しを比較した分析-. 言語処理学会第7回年次大会発表論文集, pp. 421-424, 2001.