

# 関心抽出に基づく話題構築支援に関する研究

## Topicalization Support Based on Interest Extraction

吉村宏樹<sup>\*1</sup>  
Hiroki Yoshimura

大澤幸生<sup>\*1</sup>  
Yukio Osawa

<sup>\*1</sup> 筑波大学大学院ビジネス科学研究科  
Graduate School of Business Science, University of Tsukuba

This paper describes a Topicalization Support from a set of news programs, i.e., extracting the interests of a newscaster and television watchers. We focus onto news programs and analyze the relation of keywords that the newscaster used in the TV program and the audience-rate rating using the quantification theory. Based on the results, we propose the technique to presume television-watchers' interests and to harness it for distilling the future interests of castered comments.

### 1. はじめに

考えや意見を述べる時、話し手は自分自身の関心に基づいて発言していると考えられる。対話には当然、聞き手が存在するため、話し手は自身の関心を把握した上で、聞き手の関心を考慮して発言することによって、さらに話題は盛り上がりやすくなる。認知科学においても Grice は人間同士の会話における協調性を重要な要素として捉えている[Grice 1978]。特に不特定多数が視聴するテレビ番組では、聞き手である視聴者の関心を掴み、視聴率を維持または向上できれば、それはテレビ局、出演者やCMを提供する企業にとってビッグチャンスとなる。視聴率は、マーケティングの調査データとしても利用され、世間の関心のある事柄を数値化したデータである[ビデオリサーチ]。そのためできる限り番組内容に、番組制作者側の主張を盛り込んだ上で、視聴者の関心を反映させることが重要である[吉村 03]。

そこで本論文では、コメントを述べる機会の多いニュース番組の発言内容に着目する。その上でキャスタ自身の関心と視聴者の関心を把握して、キャスタが主張したいコメントを考えるための話題構築支援について提案する。本提案では、まずキャスタの関心事である発言内容からキーワードを抽出し、キャスタの関心があるキーワードを抽出する。次にそのキーワードと視聴率の関係から、視聴者に関心があるキーワードを抽出するために、数量化 I 類を使って解析する。そして、キャスタにとって関心は低いが見聴者には関心の高いキーワードを抽出することで、次回以降の番組のコメント内容に活かす手法を提案する。素材として、TBS 系列で放送されている News23(月・金、22時 54分-)のコーナーの一つである多事争論を取り上げた[News23]。多事争論では、メインキャスターの筑紫哲也氏が時事問題などについて、筑紫氏独自の観点からコメントをするコーナーである。本論文では、Web 上に公開されている内容から 2002 年 10 月 3 日から 2002 年 12 月 16 日までを取り上げた。

以下、2 節で数量化 I 類を利用した解析手順の詳細を述べ、3 節で解析結果、4 節で視聴者の関心のあるキーワードを抽出し、話題構築支援のための方法について考察を述べる。最後に 5 節で、まとめと今後の展望について述べる。

### 2. 解析手順

本論文の解析手順について、実際に多事争論のデータを用

連絡先: 吉村宏樹, 筑波大学大学院ビジネス科学研究科, 〒112-0012 東京都文京区大塚 3-29-1, Tel:0465-80-2232, yosimura@gssm.otsuka.tsukuba.ac.jp

いて紹介する。解析手順は大きく 2 ステップ、筑紫氏の関心を推定する解析前処理ステップと視聴者の関心を推定する数量化 I 類解析ステップに分けられる。

#### 2.1 解析前処理ステップ

前処理ステップでは、数量化 I 類で解析するデータを準備するほか、頻繁に出現するキーワードからキャスタの関心を推定することを行う。以下に解析前処理ステップの 3 プロセスについて説明する。

##### (1) 形態素解析によるキーワード抽出

まず筑紫哲也氏の発言内容を茶釜[茶釜]によって、形態素解析し、単語列だけを抽出する。本処理では、名詞、形容詞、副詞、動詞、未知語を残し、助詞や助動詞は、頻出単語であるが、通常キーワードになりえないのでストップワードにする。

##### (2) 出現頻度の高い単語のリストアップ

次に(1)のキーワード抽出から出現頻度の高い単語順にリストアップを行う。このことにより、筑紫氏にとって関心の高い単語がわかる。

#### 2.2 数量化 I 類による解析ステップ

数量化 I 類は、複数のカテゴリ変数から一つの外的基準への影響と効果を数量化して解析する手法である[岩坪 87]。ただし、キーワードの有無といったカテゴリ変数を扱う場合は、ダミー変数を導入する。

本分析では、カテゴリ変数を前節で抽出したキーワードが多事争論の各オンエアで発言されなかった場合を「0」、発言された場合を「1」としたダミー変数を導入した。外的基準は視聴率データを用いた。視聴率データは、ビデオリサーチ社が発行するデータ集をもとに、News23 の平均視聴率を利用した[ビデオリサーチ 02]。以上のデータを用いて数量化 I 類の解析を行う。本論文では数量化 I 類の解析結果である各単語が視聴率に与える影響(カテゴリスコア)を用いる。このことにより、視聴者にとって関心の高い単語がわかる。

### 3. 解析結果

表 1 は、前節 2.1 で求められた筑紫氏が発言した単語中で出現頻度の高い上位 10 位である。筑紫氏にとって関心の高い単語は、「国」問題「アメリカ」の順で上位にランクされている。全体的に国内政治や国際関係に関する単語が目立つ。前処

理ステップから頻繁に出現するキーワードは彼の関心ともいえるキーワードであるといえる。

表 2 は、前節 2.2 で求められた数量化 I 類で解析した結果であるカテゴリスコア表の上位 10 位である。視聴者にとって関心の高い単語は、「北朝鮮」「小泉」「一つ」の順で上位にランクされている。これは、視聴率にプラスの要因となる単語は視聴者にとって関心の高いキーワードであるといえる。

表 1. 出現頻度の高い単語

出現順位	キーワード
1	国
2	問題
3	アメリカ
4	日本
5	戦争
6	議論
7	自分
8	一つ
9	世界
10	関係
10	言葉
10	形

表 2. カテゴリスコアの高い単語

アイテム	カテゴリスコア
北朝鮮	2.58322
小泉	1.17335
一つ	0.69505
政府	0.57371
言葉	0.20029
主義	0.14406
自分	0.14343
世界	0.04047
形	0.00531
拉致	-0.01301

#### 4. 考察

本節では、筑紫哲也氏カテゴリスコアによる視聴率予測と内容について考察する。また Keygraph[大澤 99]を用いた話題構築支援についても述べる。

表 1 と表 2 を比較すると、上位 10 単語中で両者が一致している単語は、「一つ」、「言葉」、「世界」、「形」であった。これらの単語から、筑紫哲也氏と視聴者は抽象度が高く概念的な単語に対して関心が一致している。一致していない単語を見ると、筑紫氏は、「国」や「日本」の「問題」や「アメリカ」の「戦争」といったことに関心が高いことがわかる。それに対して、視聴者は、「北朝鮮」、「小泉」氏や「政府」について関心が高いといえる。このことから両者の関心の高いキーワードには若干のズレがあるといえる。

また Keygraph と数量化 I 類による解析結果から得た視聴者の関心の高いキーワードを用いて、将来の話題の例を考えてみることにする。Keygraph は、出現頻度の高い単語群の共起関係から文書の主張を取り出すことができる情報視覚化ツールである[大澤 99]。図 1 は、Keygraph を利用して、視覚的に筑紫哲也氏が多事総論で発言した 3 か月分の主張を表している。視聴者にとって関心の高いキーワードである「北朝鮮」と「小泉」氏や「一つ」と「政府」を中心に、Keygraph の図を見てもわかる。視聴者の関心の高い「北朝鮮」に対して関心が高いと考えられる。また「問題」や「国」といったキーワードは筑紫氏にとって関心の高いキーワードと結びついていることがわかる。さらに「小泉」は、筑紫氏にとって、話題の中心ではないこともわかる。

これらを踏まえると、今後の話題の案として、「国の一つの問題」という話題が、両者にとって関心の高い話題だと考えられる。具体的な話題の案として「小泉首相や政府の問題の一つは、北朝鮮の核問題である」といった考えから独自の論を展開することが、筑紫氏と視聴者の関心が一致すると考える。

本手法は、例えば放送後に更新される Keygraph を見ながら、数量化 I 類の解析結果による視聴者の関心の高いキーワードを考慮することで、キャストにとって、新しい話題の発見と示唆を与えることになる。

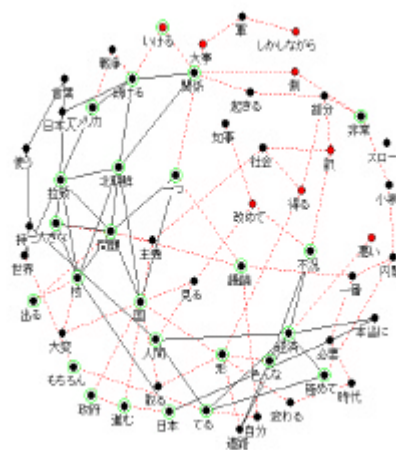


図 1. Keygraph による 3 か月分の発言結果

#### 5. まとめ

本論文では、ニュース番組に着目し、キャストと視聴者の関心を抽出して、それぞれの関心の違いを気づかせることによって、次回のコメントに生かす提案手法について述べた。本手法によって、話し手と聞き手の関心の違いを抽出することができた。話し手は聞き手との関心の違いを認識することで、聞き手の関心にも配慮したコメントを作成することができる。

その分析対象として筑紫氏のコメント内容と視聴率との関係から、筑紫氏と視聴者の関心の高いキーワードに違いがあることがわかった。具体的には筑紫氏の関心の高いキーワードを出現頻度から抽出し、この出現頻度の高い単語と番組平均視聴率から数量化 I 類を用いて、視聴者の関心の高いキーワードを抽出した。このことにより、客観的なデータの分析結果から筑紫氏と視聴者の両者が関心の高いキーワードを発見し、主観的な方法で新しい話題を発見していく手段の可能性を示した。

今後は、数量化 I 類の解析結果であるカテゴリスコアの結果の妥当性や視聴率と相関の高いキーワード抽出なども課題として追記しておく。そして視聴率と相関の高いキーワードを抽出することによって、視聴者である世間の関心の抽出へとつなげていきたい。

#### 参考文献

- [Grice 78] Grice, H.P.: Logic and conversation, Syntax and Semantics Vol.3, Academic Press, 1978.
- [ビデオリサーチ] 株式会社ビデオリサーチ: マーケティングデータ, <http://www.videor.co.jp/service/media/index.htm>
- [吉村 03] 吉村宏樹: ニュース番組におけるニュースキャストの関心抽出とその応用, 第 51 回人工知能学会 SIG-FAI 研究会, 2003.
- [News23] TBS News23 <http://www.tbs.co.jp/news23/>
- [茶筌] 茶筌 <http://chasen.aist-nara.ac.jp/>
- [岩坪 87] 岩坪秀一: 数量化法の基礎, 朝倉書房, 1987.
- [ビデオリサーチ 02] ビデオリサーチ: テレビ視聴率速報 関東地区, Vol.48-51, 2002.
- [大澤 99] 大澤幸生, Nels E. Benson, 谷内田正彦: KeyGraph: 単語共起グラフの分割 統合によるキーワード抽出, 電子情報通信学会論文誌 Vol. J82-D1, No. 2, pp. 391-400, 1999.