

# 会話中の単語に基づいた 話題提供を行う雑談支援システムの提案

木村 有里<sup>1</sup> 丸山 一貴<sup>1</sup>

**概要:** 対面型コミュニケーションの必要性が問われる現代において、無言状態の気まづさが原因で友人以外との雑談を苦手と感じる若者がいる。これまでも雑談を支援する研究やシステムは存在するが、会話の転換そのものをサポートするシステムは少ない。また、それまでの話の流れを考慮せず関連のない話題を提示することで、転換の際に相手に違和感を感じさせる場合がある。そこで本論文では、話題が途切れた際に気まづさを感じさせずに話題転換のできるシステムを提案する。雑談に大切な話の流れを考慮し、会話中の話題に沿った話題候補を提示する。候補にはプロトタイプとしてニュース記事の見出しを用いた。本稿では会話内容を音声認識でテキスト化できたと想定して、口語体のテキストから現在の話題を単語抽出し、評価を行った。

## Providing Topic Candidates Based on Picking Up Terms and Retrieving Topics

YURI KIMURA<sup>1</sup> KAZUTAKA MARUYAMA<sup>1</sup>

**Abstract:** In recent years, face-to-face communication skills become increasingly required. Silence in chatting with others, not close friends, may cause us awkward. There are many research and systems that support chat to avoid the silence, but few systems that support changing topics in spite of the importance of the choice of topics related to one just talked. In this paper, we propose a system that provides topics related to the contents of the conversation. A prototype system we implemented extracts some keywords from the conversation, fetches some news articles from a web service and shows their headlines to the users. The user experiments are also included.

### 1. はじめに

近年、SNS やマイクロブログなどの様々なコミュニケーションツールの登場により、私たちはいつでもどこでも手軽にコミュニケーションを取ることができるようになった。しかし齋藤 [1] によれば、このような現代日本の状況では、自然なコミュニケーション力を支えてきた基盤が崩れだしているという。世間話ももはや自然な技ではなく、対面型コミュニケーションを技として捉え直す必要があると述べている。

このように対面型コミュニケーションが求められている現代において、それを苦手と感じる若者が多いことも現状

である。実際に新卒採用を行う企業を対象としたアンケートでは、選考にあたり重視するものは、13 年連続で「コミュニケーション能力」が第 1 位である [2]。一方、大学生のコミュニケーションに対する意識調査では、「初対面や性格の合わない人と 1 対 1 で話している時」を苦手とする結果が出ている [3]。齋藤 [1] も、友達としか話せない若者などの会話相手の幅が狭い人はコミュニケーション力が十分でないとして述べているため、社会のニーズに現代の若者は応えられていないと言える。

我々は対面型コミュニケーションを苦手と感じる要因を調べるため、大学生を対象に先行調査アンケートを行った。その結果、15 名中 11 名が「会話が途切れた際に気まづい」と感じていることが明らかとなった。

本論文ではこの会話中の気まづさに着目し、新たな話題

<sup>1</sup> 明星大学 情報学部  
School of Information Science, Meisei University

候補を提示することで、気まずさを感じさせず話題の転換をできるシステムを提案する。

また、システムの利用対象者を選ぶにあたり、知り合いではあるが親密度の低い間柄を選択した。仲の良い間柄であれば沈黙はそこまで気にならず、初対面なら相手のことを知らないので質問することは多くあると考えたからである。

齋藤は話題の転換に対する自身の考えとして以下のようにも述べている。「話と話の間に脈絡がないのは文脈力（著者の造語で、文脈を的確につかまえる力のこと）の欠如である」「脳の中は編み目のようになっていて、話の中で出てきた言葉が脳の中で別の言葉やイメージにつながる」「話題を転換する場合は、少しずつ話題をずらして、知らぬ間に自分の話したい本題に入っている、というのが望ましい。」[1]

我々も齋藤の考えには同意しており、コミュニケーション力を培うには相手の発した単語や現在の話題から発想を膨らませ、話の流れに沿った話題転換を行うべきだと考える。

しかしこれまでの会話支援の研究には、ユーザプロフィールの利用 [4] や、話題提供アプリ [6] などがあるが、これら 2 点の研究は直前の話題とは脈絡のない話題を提示するため、相手は転換に違和感を感じてしまう。また、会話中の単語を画像検索し提示する研究 [5] もあるが、そこから思いつく話題の内容まではシステムとしてサポートしていない。

そこで本論文では、雑談の自然な転換をサポートすることを目的とし、会話中の話題に関連した話題候補を気まずさを感じさせずに提示するシステムを作成する。

## 2. 関連研究

### 2.1 雑談支援

笹嶋ら [4] は Twitter からユーザのプロフィール図を作成し、所属や好きなもの、趣味などを可視化して話題のきっかけとなるものを提示している。共通の話題に入りやすくなることが期待されるが、話題の転換に関しては言及していない。

清水ら [5] は、発話内容を音声認識して利用するウェアラブルインタフェース「しゃべりカス」を提案している。発話内容を形態素解析し、最近発言された名詞で画像検索を行いリアルタイムで表示している。表示された名詞から脈絡のある話題転換が期待される点で本論文と共通するが、そこから思いつく話題をシステムとして提示する、という点で異なる。

株式会社日産は車内での話題提供を目的としたアプリ「ドライブトーク」[6] を提供している。パラメータとして相手との関係性や親密度を設定することで、その場に最適な話題を提示してくれる。話題の転換に注目したシステム



図 1 Memorium[10]

Fig. 1 Memorium[10].

だが、「自分でやるのが好きなスポーツについて語り合おう」や「次の休みの予定について語り合おう」など直前とは関連のない話題となるため、転換に違和感や不安を抱かせる可能性がある。

### 2.2 単語抽出

単語抽出の分野は幅広く研究されており、特に白松ら [7] の計算手法は会話の文脈を考慮した注目度の高い語の抽出が可能である。注目されている語ほど継続されて参照されやすいという仮定に基づいて設計された参照確率は、細かい推移を追うのに適していて発話ごとに更新される。会話に特化した単語抽出手法という点で関連深いですが、本論文では違うアプローチを取る。詳細は第 3.1 章で述べる。

高橋ら [8] は、対話型システムのための短い発話文を対象とした重要語の抽出を行っている。Web を用いた TF-IDF 法による値を採用しているため Web 上で頻出する語はスコアが低くなり、会話中に話題となっても抽出されない可能性がある。そのため本論文では、会話の文章だけを用いる TF-IDF 法を採用する。

### 2.3 提示方法

北井ら [9] は、テーマパークを歩く 2 人を対象にした、場所に紐づいた情報を音声で提供するシステムを提案している。しかし音声での情報提供は 2 人の会話を妨げる可能性があるため、気まずさを感じさせない情報提供を目的とする本論文とは相性が悪い。

渡邊ら [10] は、偶発的な新しい情報と出会う機会を提供する、眺めるインタフェース「Memorium」を提案している。ユーザが記入したキーワードの書かれたカードが水槽の中をふわふわと動き、衝突したカード同士から新しい情報のカードが生まれる。ユーザは気に留めた時だけこのシステムを眺めるため、ユーザへの負担をかけることの少ない環境型アプリケーションとなっている。このユーザへの負担の少なさに注目し、本論文では Memorium のデザインを提示方法として採用する。

### 3. 提案手法

#### 3.1 話題抽出

白松ら [7] は、TF-IDF による単語の頻度を利用した検索では、発話を細粒化した際にはどの単語の頻度も 0~1 回になるため、文脈を考慮した語の抽出には不向きであると述べている。一方 TF-IDF により重要語の抽出を試みた高橋ら [8] は、ある程度まとまった会話文を対象として抽出を行うことでこの問題は解決できると述べている。

そこで本論文では、会話文を時系列にある程度の長さでまとめて 1 ドキュメントとみなし、TF-IDF 法を利用して会話中の名詞を話題語として抽出する手法をとる。話題を提供する間隔は平均的な 1 話題の長さを考慮して暫定的に 2 分間おきとした。しかし今後の課題として、間隔を変えて検証実験を行い最適なタイミングを選択する、もしくはユーザが使用中に自由に間隔を調整できるような画面設計を行うなどが挙げられる。

また、ドキュメント数は会話時間によって増加するため、最初の 1 ドキュメントおよびデータが少ない初期の出力にはいい結果が期待できないコールドスタート問題が発生する。この問題はユーザプロファイルの作成 [4] などを行うことで将来的にはサポートできると考えているが、本論文では検討の対象外とする。

#### 3.2 話題候補の提示

笹嶋ら [4] や清水ら [5] のように、話題のきっかけとなるような単語や画像を提示する研究は多くあるが、そこから話を膨らませることは対話者のスキルに依存する。本論文では対話が苦手な人を主なシステムの対象としているため、ドライブトーク [6] のような、そのまま提示するだけで会話として成立するような具体的な候補を提示する必要がある。

そこで本論文では、抽出した単語から最新のニュース記事見出しを検索し話題候補として提示する。記事見出しは短文かつ興味を引くため、複数並んでいても気になる記事をすぐ見つけることができる。また最新のニュース記事には話題性があり、お互いが知らなかった新しい情報の発見や、片方が知らなかった情報を共有するような行動が期待できる。

#### 3.3 提示方法

先行研究の提示方法には北井ら [9] の音声手法などがある。一方、支援システムを利用することで会話に集中できなくなり、ユーザに負担をかけてしまうケースもある。会話は話の内容に集中し、話題が一段落するか気が向いた時にシステムを利用できるような提示方法が、会話支援システムには相応しいだろう。そのため我々は、会話をしながら眺めることに適したデザインである Memorium[10] を

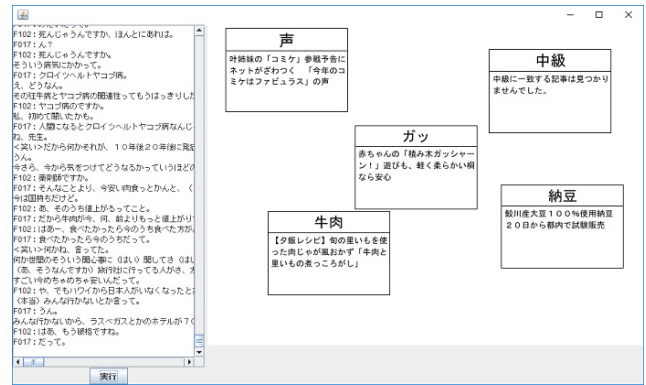


図 2 システム実行例

Fig. 2 Topics provided by our pilot system.

参考にシステムを作成する。

### 4. 実装

提案システムは Java 言語で作成し、GUI 部分には Java の GUI ツールキットである Swing を使用した。話題となるキーワードは、形態素解析ツール Mecab\*1 を用いて抽出した名詞を用いる。話題候補のニュース記事取得には、抽出した名詞の Yahoo!ニュース \*2 検索結果の URL を取得し用いる。

以下にシステムの流れを示す。

- (1) 会話内容を音声認識でリアルタイムにテキスト化 (本実験では未実装)
- (2) テキストを一定文字数で区切って形態素解析を行い全ての名詞を抽出 (Mecab)
- (3) 抽出メソッドによりスコア値を求め、1 位のものを現在の話題として 1 つ提示
- (4) 話題の単語から、ニュース記事検索結果を取得し記事の見出しタイトルを提示

話題を提供する間隔として 2 分間を設定したため、1 分間に話す速度の目安と言われている 300 文字×2 の 600 文字でテキストを区切り、1 区切りから 1 つ話題を抽出する。システムの実行例を図 2 に示す。

画面左側のテキストエリアには、会話内容の音声認識結果のテキスト (本実験では文字起こし済みの会話コーパスを使用) を表示する。テキストを 2 分間に話せる文字数を想定して 600 文字ごとに区切り、そのドキュメント内で話題と評価された名詞をひとつ右側の表示エリアに追加する。追加された話題のカードをクリックするとブラウザが立ち上がり、元のニュース記事の Web ページに飛ぶリンクとなっている。表示エリアの話題表示数は最大 5 個までで、新しい話題が追加されると一番古い話題は削除される。

\*1 Mecab: Yet Another Part-of-Speech and Morphological Analyzer, <http://taku910.github.io/mecab/>

\*2 Yahoo!ニュース, <http://news.yahoo.co.jp/>

#### 4.1 Mecab システム辞書の使用

通常の Mecab の辞書のみでは、最近現れた新語や地名などの固有名詞、専門用語などに対応していないため、システム辞書である mecab-ipadic-NEologd<sup>\*3</sup> を併用した。

mecab-ipadic-NEologd は web 上から得た新語を週 2 回の更新で追加することで、通常の辞書より形態素解析の精度を高めた辞書である。本論文の形態素解析において重要なことは口語体の性質を踏まえ単語を抽出することである。最近話題になった語や番組名、人名などは会話中では話題の中心となる可能性が高いため抽出する必要がある。そこで通常の辞書とシステム辞書を併用した状態で解析に差が出るかの比較を行った。例として会話コーパス内にある「水素水って世界一受けたい授業から話題になったよね」という内容の解析結果を挙げる。通常辞書を使用して品詞ごとに分割した解析結果は「水素/水/って/世界一/受け/たい/授業/から/話題/に/なっ/た/よ/ね」となり、システム辞書を併用した結果は「水素水/って/世界一受けたい授業/から/話題/に/なっ/た/よ/ね」となった。水素水という単語は以前から存在していたものの、メディアで取り上げられ話題となったことから 2016 年に新たに辞書へと追加された名詞である。また、「世界一受けたい授業」といったテレビ番組タイトルが抽出できるようになったことも明らかである。この結果から、本論文では口語体ではシステム辞書を使用することが適していると仮定し、使用を決定した。システム辞書の使用有無はプログラム上で変更可能とし、検証結果を見て切り替えることも考慮する。

### 5. 実験

本章ではテキストから話題を抽出する手法の有用性を検証する。

#### 5.1 実験方法

実験では、1 対 1 の雑談を想定した会話コーパスを複数用意する。全 3 文書中、1 文書は著者と友人の会話を文字起こししたもの、2 文書は日本語自然会話書き起こしコーパス [12] を使用した。

会話コーパスの先頭から 600 文字ごとに分割したものを 1 ドキュメントとみなし、そこから TF-IDF 法を用いて話題を抽出し模範解答と比較を行う。ドキュメントからは以下の基準に従って話題を手動で抽出し、模範解答とした。

- ドキュメント中に最低 1 度は出現した名詞のうち、ドキュメント内の話題として最も相応しい語
- 違う話題が複数あると判断した場合、順位をつけ 2 つまで選択する

会話例として示す図 3 は、解析を行った会話コーパスの一部である。また、このコーパスの一部が含まれるドキュ

<笑い>(え)息がくさくないかなあ。 ごま油を落として一、(はい)ご飯にかけて食べる。  
大丈夫かな。 F102: あ、もうじ、全部手作りって感じですか。  
F102: ニンニクですか。 F102: ん？  
F102: うん。 F102: 全部自分で作って。  
今日朝、なっ、キムチ納豆のご飯。 F102: そうそうそうそう。  
F102: ああ、くさい。 F102: え、キムチと納豆は作りませんよ。  
絶対食べれない。 <笑い>それは買ってきませうけど。  
F102: おいしいんだって、これが。 F102: 何か巻いたら笑わせる、納豆作って。  
かなり。 F102: からいですよね。  
F102: からくてー、キムチと納豆を混ぜてー (コーパスから一部抜粋)

図 3 会話コーパス [12] の例

Fig. 3 Conversation corpus[12] example.

表 1 コーパスの解析結果

Table 1 Analysis results of the conversation corpus.

単語	重要度 (TF-IDF)
納豆	0.3598
キムチ	0.2998
逆	0.1199
ご飯	0.1199
味	0.0710

表 2 抽出の実験結果

Table 2 Experimental results of term extraction.

	コーパス 1	コーパス 2	コーパス 3
ドキュメント数	13	21	17
(A) 解答 1 位と結果 1 位が一致	9	8	7
(B) 解答 2 位と結果の 1 位, 2 位が一致	2	7	5
(C) 一致せず	2	5	5

メントからは表 1 のような解析結果が得られた。表 1 の単語は重要度が高いほど会話内で頻出していた単語であり、一番重要度の高い「納豆」が会話中の話題であると推定できる。

#### 5.2 実験結果

第 5.1 章の手順で実験を行い、模範解答との正答数を表 2 に示す。例えば図 1 の解析結果の模範解答が 1 位「キムチ」、2 位「納豆」だったとすると、模範解答 2 位の「納豆」と結果 1 位が一致するため、表の「解答 2 位と結果の 1 位, 2 位が一致」とみなされる。

この表 2 の結果より、少なくとも解答の一部と一致したドキュメントは全体の 7 割以上となった。

また、解答と一致せず不正解となった解析結果 (表 2 内の C) の一例を表 3 で示す。このドキュメントでの模範解答は「08 小隊 (アニメ作品のタイトル)」だったが、解析結果は表 3 のようになり、上位と一致することはなかった。ドキュメント内では確かにこのアニメの内容である「迷子になる主人公」について話されており、表にある通り 08 小隊という単語も抽出はできていた。しかし単語の出現頻度を重視した結果、タイトル名よりも頻出した別の単語が選ばれてしまった。

<sup>\*3</sup> mecab-ipadic-NEologd, <https://github.com/neologd/mecab-ipadic-neologd>

表 3 C における模範解答の TF-IDF 値

Table 3 TF-IDF values of correct terms in case C.

単語	重要度 (TF-IDF)
迷子	0.1232
主人公	0.0478
それ	0.0478
08 小隊	0.0323
分かん	0.0323

## 6. 考察

実験の結果より、7割以上のドキュメントから人間が考える話題の第1位、もしくは第2位の単語を抽出することができたことから、本手法を利用した話題の抽出はある程度妥当ではないかと考えられる。

しかし表3に挙げた不正解例のように、話題そのものの単語は頻出回数が少なく、実際に会話内で出てきた他の単語を重視した結果、話題抽出が上手く行えないドキュメントも多々あった。単語の頻出のみではこれらの検出は難しく、他の手法を合わせて考える必要がある。

また、抽出が失敗する原因として話題抽出のタイミングの問題が挙げられる。一話題にかかる時間は安易に平均化できるものではなく、話者の話すスピードや場の状況によって大きく変化する。そのため常に固定の時間(文字数)を設定するのは話題の区切り方としては相応しくないだろう。改善案としては、一定以上の「間」の検知や、間の直後に現れる「そういえば」などの転換を表す語に注目し、話題の途切れを検知してドキュメントを区切るようなことが考えられる。

## 7. まとめ

本論文では、雑談の自然な話題転換をサポートすることを目的とし、会話中から抽出した話題を用いて話題転換のための候補を提示するシステムを作成した。また、単語抽出メソッドの性能を検証するため、システムが抽出した単語と模範解答の比較実験を行った。実験の結果、7割以上の解答との一致を確認し、抽出メソッドとして一定の性能を証明した。

今後は話題候補としてシステムが提示したニュース記事の有用性を確認する被験者実験を行う予定だ。

## 参考文献

- [1] 齋藤 孝: コミュニケーション力, 岩波新書 (2004).
- [2] 経団連: 2016年度新卒採用に関するアンケート調査結果 (2016-11-15): 入手先 (<http://www.keidanren.or.jp/policy/2016/108.html>) (2016.12.15).
- [3] 飯塚 一裕: 大学生のコミュニケーション意識について—テキストマイニングによる分析—, 愛知教育大学研究報告. 教育科学編, Vol.59, pp.49-53(2010).

- [4] Ryota Sasajima, Kohei Otake, Makoto Oka, Akito Sakurai: Proposal of a Visualization Method to Support Informal Communication Using Twitter Attributes, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), No.9172, pp.107-114(2015).
- [5] 清水 大悟, 安村 通晃: しゃべりカス: 発話の視覚化を用いたウェアラブルインタフェース, 情報処理学会研究報告, No.28, pp.1-8(2009).
- [6] こんな時に! ドライブトーク講座 - 日産ドライブナビ: 入手先 (<http://drive.nissan.co.jp/DRIVETALK/>) (2016.12.15).
- [7] 白松 俊, 駒谷 和範, 尾形 哲也, 奥乃 博: 会話文脈に応じた関連情報提示タスクのための文脈類似度計算手法の開発, 人工知能学会言語・音声理解と対話処理研究会資料, Vol.50, pp.57-62(2007).
- [8] 高橋 勲, 田添 丈博, 椎野 努: 対話型システムのための短い発話文の意図推定, 言語処理学会年次大会発表論文集, Vol.15, pp.458-461(2009).
- [9] 北井 美穂, 丸山 一貴: テーマパークにおける場所に紐付いた話題提供システム, インタラクシオン 2016 論文集, pp.207-210(2016).
- [10] 渡邊 恵太, 安村 通晃: ユビキタス環境における眺めるインタフェースの提案と実現, 情報処理学会論文誌, Vol.49, No.6, pp.1984-1992(2008).
- [11] MeCab: Yet Another Part-of-Speech and Morphological Analyzer: 入手先 (<http://taku910.github.io/mecab/>) (2016.5.25).
- [12] 日本語自然会話書き起こしコーパス: 入手先 (<https://nknet.ninjal.ac.jp/nuc/templates/nuc.html>) (2016.8.21).