

音声対話におけるユーザ発話からの口癖検出

大野 誠寛†

松原 茂樹‡

†名古屋大学大学院国際開発研究科 ‡名古屋大学情報基盤センター/JST-CREST

1 はじめに

音声・言語技術の進展を背景に、これまでに様々な音声対話システムが開発されてきた。その多くは、タスクを正確かつ効率的に達成することが重視され、システムの振る舞いはユーザに依存することなく画一的であった。しかし、ユーザとシステムとの間で親しみのあるコミュニケーションを実現するためには、システムはユーザの個性を理解し、言い方や言い回しなど、応答様式をユーザに応じて使い分けることが望まれる。例えば、コミュニケーションの場において、人は、言い回しが自分と似ている応答に親しみを覚えることが報告されており [1]、ユーザの特徴的な言い方を捉えて、それを適度に真似た応答をすることにより、ユーザにシステムへの親近感を与えることができる。

そこで、本論文では、ユーザの個性に応じて柔軟に応答する対話システムの実現を目指し、その要素技術として、ユーザの口癖を検出する手法を提案する。本手法では、複数ユーザの各発話における口癖の分布に着目して、口癖らしさを示すスコアを定め、あらゆる形態素 n-gram に対して口癖スコアを求める。CIAIR 車内音声対話コーパス [2] のドライバ発話を用いて口癖検出実験を行い、本手法の有効性を確認した。

2 口癖の特徴

口癖とは、発話するとき、よく使う言葉遣いのことであり、話者の個性を示す現象の一つである。複数の話者がそれぞれ、トピックの異なる複数の対話を遂行したとき、口癖となるフレーズは、トピックワードやストップワードと比較して、分布傾向が異なると考えられる。

図 1 に、ストップワード、口癖フレーズ、トピックワードの分布傾向の模式図を示す。この図は、ある話者による、あるトピックの対話で、各フレーズがどの程度発話されたか（頻度割合）を示すグラフである。トピックワードは、話者に関係なく、ある特定のトピックに関する発話のみに出現し、ストップワード（助詞や助動詞といった機能語など）は、話者やトピックに関係なく、どのような発話でも出現する。一方、口癖フレーズは、ある特定の話者の発話のみに出現し、トピックに関係な

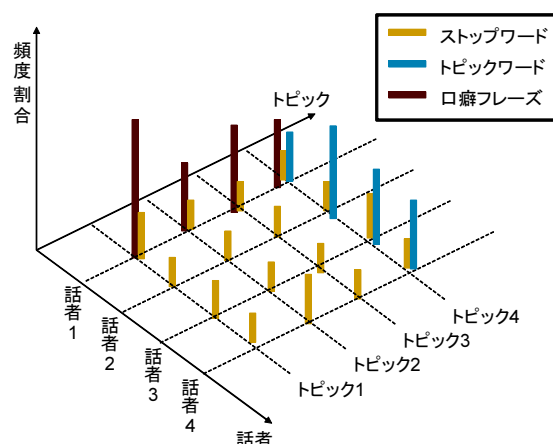


図 1: ストップワード、トピックワード、口癖フレーズの分布

く発話される傾向が強いと考えられる。すなわち、口癖とは、以下のような特徴を備えている。

特徴 1 その口癖を発する話者はそれを頻繁に発する

特徴 2 その口癖を発する話者は偏る

特徴 3 トピックに関係なく発せられる

実際に、CIAIR 車内音声対話コーパス [2] を用いて、ストップワード、トピックワード、口癖フレーズ、それぞれの代表的なフレーズに対して分布を調査した。CIAIR 車内音声対話コーパスには、ドライバとナビゲータとの間で遂行された対話の音声データの書き起こしが収録されている。各対話には、トピックとして店検索や道案内などが設定されており、発話ごとにトピックのラベルが付与されている。また、すべての発話に対して、ChaSen[3] を用いて形態素解析を施している。辞書には、店名などの固有名詞を独自に追加した Unidic[4] を用いた。分析には、ドライバ 497 人分の発話データを用いた。分析データの基礎統計を表 1 に示す。なお、発話単位とは、発話を 200ms 以上のポーズで区切った単位である。

図 2 に、口癖フレーズ「そいじゃ」、トピックワード「和食」、ストップワード「が」の分布を示す。図 1 と同様に、あるドライバによる、あるトピックの対話で、各フレーズがどの程度発話されたか（頻度割合）を示す

表 1: 分析データ

ドライバ数	497
トピックの種類数	63
発話単位数	20,736
形態素数	141,183

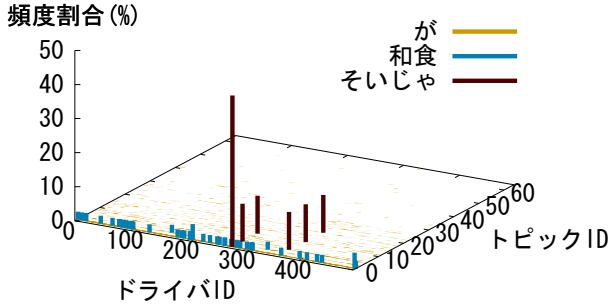


図 2: ストップワード「が」、トピックワード「和食」、口癖フレーズ「そいじゃ」の分布

グラフである。「和食」は、レストラン検索に関するトピック（図 2 上のトピック ID は 1）のトピックワードである。これらの分布は、図 1 と同様の傾向を示している。なお、全体的に若い ID をもつトピックに分布が集中しているのは、各ドライバがすべてのトピックについて対話しているわけではないため、各トピックの ID を、そのトピックについて対話しているドライバの数が多い順に付与したためである。

以上に加えて、口癖は、ひとまとまりとして発話されるものであるため、

特徴 4 単位性（ひとかたまりのフレーズとして見なせる度合い）[5] が強い
を備えている。

3 口癖検出手法

本手法では、話者 1 人分の発話から形態素 n-gram を抽出し、抽出した全ての形態素 n-gram に対して、各 n-gram の口癖らしさを示すスコア（以下、口癖スコア）を求める。口癖スコアが閾値以上の形態素 n-gram を、その話者の口癖であるとして出力する。本手法の流れを図 3 に示す。以下で、各処理について説明する。

1. 形態素 n-gram の抽出

話者 1 人分の発話から、形態素 n-gram を抽出する。口癖フレーズは発話単位をまたがないとして、発話単位ごとに、その中の形態素列から、あらゆる形態素 n-gram を獲得する。獲得した n-gram のうち、話者 1 人分の発話において、1 回しか発話さ

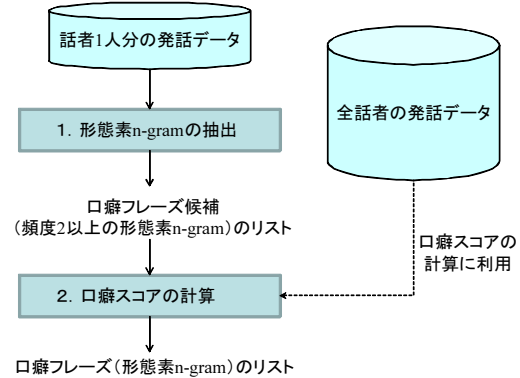


図 3: 口癖検出の流れ

れていない n-gram は口癖にはなりえないと考えて排除し、その他の n-gram を口癖フレーズの候補として次の処理に渡す。

2. 口癖スコアの計算

上記で抽出された全 n-gram に対して、口癖スコアを計算し、閾値 m 以上の口癖スコアを持つ n-gram を、この話者の口癖であるとして出力する。口癖スコアの計算式については次節で詳述する。

3.1 口癖スコア

本研究では、2 節で挙げた口癖の各特徴を持った形態素 n-gram を口癖として検出する。特徴 1 から 4 の各特徴を捉えるべく、各特徴に対して以下のようなスコア $FP_i (1 \leq i \leq 4)$ をそれぞれ定めた。どのスコアも、数値が高ければ高いほど、その特徴の度合いが強まる計算式になっており、これらの積を口癖スコア $FP = FP_1 \times FP_2 \times FP_3 \times FP_4$ とする。

なお、以下では、ある形態素列（形態素 n-gram）を $w_\alpha (1 \leq \alpha \leq N_w)$ 、話者を $s_\beta (1 \leq \beta \leq N_s)$ 、トピックを $t_\gamma (1 \leq \gamma \leq N_t)$ と記す。 N_w, N_s, N_t はそれぞれ、口癖フレーズ候補の総数、話者の総数、トピックの総種類数を示す。話者 $s_{\beta'}$ で発話された、ある形態素列 $w_{\alpha'}$ に対する各スコアの計算式 $FP_i(w_{\alpha'})$ を以下に示す。

● 特徴 1

全話者により発話された頻度に対して話者 $s_{\beta'}$ により発話された割合が高く、かつ、話者 $s_{\beta'}$ により発話された頻度そのものも多いとして、以下のスコアを定めた。 $tf(w_{\alpha'}, s_{\beta'})$ は、話者 $s_{\beta'}$ が形態素列 $w_{\alpha'}$ を発話した頻度を示す。なお、割合だけで算出すると、1 人の話者だけが発話している場合、その頻度に関係なく、一定値（= 1）となるため、頻度との積をとっている。

$$FP_1(w_{\alpha'}) = tf(w_{\alpha'}, s_{\beta'}) \times \frac{tf(w_{\alpha'}, s_{\beta'})}{\sum_{\beta=1}^{N_s} tf(w_{\alpha'}, s_{\beta})} \quad (1)$$

● 特徴 2

形態素列 $w_{\alpha'}$ を発話する話者が偏ると、「形態素列 $w_{\alpha'}$ が発話されたとき、その話者が誰であるか」のエントロピーが小さくなる。そこで、このエントロピーの逆数を取り、特徴 2 を示すスコアとした。なお、このスコアに idf を用いることも考えられるが、口癖となるフレーズは、トピックワードと比べて、それを口癖としない話者によっても低頻度であれば発話される可能性が高く、頻度を考慮しない idf では話者に偏りはないと算出される恐れがあるため、本研究ではエントロピーを用いた。

$$FP_2(w_{\alpha'}) = \frac{1}{H(S|w_{\alpha'})+1} = \frac{1}{-\sum_{\beta=1}^{N_s} P(s_{\beta}|w_{\alpha'}) \log P(s_{\beta}|w_{\alpha'}) + 1} \quad (2)$$

● 特徴 3

形態素列 $w_{\alpha'}$ はトピックに関係なく発話されるのであれば、「形態素列 $w_{\alpha'}$ が発話されたとき、その発話されたトピックが何であるか」のエントロピーが大きくなる。そこで、このエントロピーを特徴 3 を示すスコアとした。

$$FP_3(w_{\alpha'}) = H(T|w_{\alpha'}) + 1 = -\sum_{\gamma=1}^{N_t} P(t_{\gamma}|w_{\alpha'}) \log P(t_{\gamma}|w_{\alpha'}) + 1 \quad (3)$$

● 特徴 4

単位性を調べるためのスコア関数 $LScore$ や $RScore$ として、用語抽出に関する従来研究 [5, 6] では、左右に接続する形態素の異なり語数や頻度、C-value など様々な関数が利用されているが、本研究では、エントロピーを利用する。単語 $w_{\alpha'}$ の左右に接続する単語のエントロピーが小さいほど、特定の形態素が接続しやすく単位性が低いことを意味し、逆に、大きいほど、様々な形態素が接続しやすいことになり、高い単位性を持つと考えられる。ここで、 L を形態素列 $w_{\alpha'}$ の左側に接続する形態素、 R を形態素列 $w_{\alpha'}$ の右側に接続する形態素とし、以下を単位性を示すスコアとした。

$$FP_4(w_{\alpha'}) = \sqrt{H(L|w_{\alpha'}) \times H(R|w_{\alpha'})} = \sqrt{-\sum_x P(l_x|w_{\alpha'}) \log P(l_x|w_{\alpha'}) \times -\sum_y P(r_y|w_{\alpha'}) \log P(r_y|w_{\alpha'})} \quad (4)$$

以上の式 (1) から (4) の積を形態素列 $w_{\alpha'}$ の口癖スコア $FP(w_{\alpha'})$ とし、 $FP(w_{\alpha'}) \geq \theta$ である形態素列 $w_{\alpha'}$ を口癖フレーズとして出力する。

4 評価実験

本手法の有効性を評価するため、実際の対話データを用いた口癖検出実験を実施した。

表 2: テストデータ

ドライバ数	50
トピックの種類数	36
発話単位数	2,259
形態素数	14,310
正解口癖フレーズ数	44
正解口癖フレーズ未付与のドライバ数	25

4.1 実験概要

表 1 のデータの中からランダムに選択したドライバ 50 人分の発話をテストデータとして、これに対して口癖検出実験を実施した。なお、口癖スコアの閾値 θ は実験的に $\theta = 2.5$ とし、口癖スコアの計算時には、2 節の分析時に用いた表 1 のデータ全体を使用した。評価は、テストデータに対して、ドライバごとの正解口癖リストを作成し、それを正解データとして、本手法による検出結果と比較することにより行った。

正解データは、人により口癖だと感じるフレーズが異なることを考慮し、複数人が口癖だと感じるフレーズとなるように作成した。具体的には、作業員 2 名にドライバごとの口癖リストをそれぞれ作成させ、各作業員による 2 種類の口癖リストの積集合を正解データとした。なお、作業員には、各ドライバの対話音声聞き、かつ、その書き起こしテキストを見て、口癖と思われるフレーズを判断するように指示し、その結果、口癖がないと判断することも許した。また、積集合を取る際は、作業員間で完全一致した口癖フレーズだけでなく、他方の作業員による口癖フレーズと部分一致する、もしくは、それを包含する口癖フレーズも含めた。表 2 に、正解データを作成したテストデータの基礎統計を示す。なお、各作業員がドライバごとに列挙した口癖フレーズ数の合計は、それぞれ、110 個と 78 個であった。

評価指標として、以下の指標を用いた。

● 口癖フレーズに関する再現率と適合率

$$\text{再現率} = \frac{\text{一致した口癖フレーズ数}}{\text{正解データの口癖フレーズ数}}$$

$$\text{適合率} = \frac{\text{一致した口癖フレーズ数}}{\text{検出結果の口癖フレーズ数}}$$

● 口癖がないドライバに関する再現率と適合率

$$\text{再現率} = \frac{\text{口癖未付与ドライバの一致数}}{\text{正解データで口癖未付与のドライバ数}}$$

$$\text{適合率} = \frac{\text{口癖未付与ドライバの一致数}}{\text{検出結果で口癖未付与のドライバ数}}$$

なお、口癖フレーズが一致しているか否かは、完全一致だけでなく、部分一致、包含も許して判断しており、再現率と適合率の計算式における分子の数は必ずしも一致しない。

表 3: 口癖フレーズに関する再現率, 適合率

再現率	適合率	F 値
29.55% (13/44)	51.52% (17/33)	37.56%

表 4: 口癖がないドライバに関する再現率, 適合率

再現率	適合率	F 値
72.00% (18/25)	56.25% (18/32)	63.16%

4.2 実験結果

表 3 に, 口癖フレーズに関する再現率と適合率, および, これらの調和平均である F 値を示す. なお, 完全一致した口癖フレーズの数 7 個であった. 本手法の F 値は 37.56% であり必ずしも高い値ではないが, 正解データ作成時の作業員間で再現率, 適合率, F 値をとると, それぞれ, 33.64%(37/110), 52.56%(41/78), 41.02% であり, これを考慮すると, 本手法がある程度の検出性能を有していることがわかる.

次に, 表 4 に, 口癖がないドライバに関する再現率と適合率, F 値を示す. F 値において 63.16% と比較的高い値を達成している.

4.3 考察

表 5 に, 本手法により検出された口癖フレーズの例を, 正解データ中の口癖フレーズとともに示す. ドライバ ID は CIAIR 車内音声対話コーパスに付与されているもので, f と m はそれぞれ女性, 男性を意味している. また, 各口癖フレーズは, 縦棒「|」により区切って表示しており, 空欄になっている箇所は口癖がないと判断していることを意味する.

まず, 話者 f2024 や m2209, m3126 に対しては, 正解データとほぼ同じ口癖フレーズを検出できている様子がわかる. 一方, 話者 f3026 や m3269 に対しては, 固有名詞や数詞を含む形態素 n-gram が口癖として検出された. これは, これらの形態素 n-gram が他の話者で発話されていなかったため, 高い口癖スコアが与えられたためだと思われる. 固有名詞や数詞など, 偶然その話者だけが発話する可能性のある形態素への対応が必要となる. また, 話者 m3254 に対しては, 正解データ中の口癖「じゃあ」を検出することができなかった. これは, 「じゃあ」が他の話者でも頻繁に発話されており, 口癖スコアが低下したためである. 実際, m3254 で発話された「じゃあ」に対する口癖スコアは 0.004 であり, 非常に低い値であった.

表 5: 検出した口癖フレーズの例

ドライバ ID	検出結果	正解データ
f2024	んえーと	えーと
m2209	そいじゃ	そいじゃ
m3126	よろしく って感じ	よろしく
f3026	琵琶湖	だけど んだけど
m3269	三泊四日	
m3254		じゃあ

5 おわりに

本論文では, ユーザ発話から各ユーザの口癖を検出する手法を提案した. 本手法では, 複数ユーザの各発話における口癖の分布に着目して, 口癖らしさを示すスコアを定め, 閾値以上の口癖スコアをもつ形態素 n-gram を口癖フレーズとして出力する. 対話コーパスの話者発話を用いて口癖検出実験を行った結果, 口癖フレーズの検出は F 値 37.56%, 口癖がないドライバの同定は F 値 63.16% であった.

今後は, 口癖に関して詳細に分析しスコア関数をさらに精緻化することなどにより, 口癖検出の性能向上を図る予定である.

謝辞 本研究は一部, 科学技術振興機構 CREST により実施した.

参考文献

- [1] 西原, 松村, 谷内田. QA サイトにおける質問に適した回答の判定. NLP 若手の会 第 2 回シンポジウム予稿, 発表 8, 2009. http://yans.anlp.jp/modules/menu/main.php?page_id=14&op=change_page.
- [2] N. Kawaguchi, S. Matsubara, K. Takeda, and F. Itakura. CIAIR in-car speech corpus –influence of driving status–. *IEICE Trans. Inf. Sys.*, Vol. E88-D, No. 3, pp. 578–582, 2005.
- [3] 松本, 高岡, 浅原. 形態素解析システム『茶筌』 version 2.4.3 使用説明書, 2008.
- [4] 伝, 小木曾, 小椋, 山田, 峯松, 内元, 小磯. コーパス日本語学のための言語資源: 形態素解析用電子化辞書の開発とその応用. *日本語科学*, No. 22, pp. 101–122, 2007.
- [5] 池野, 濱口, 山本, 井佐原. Web 文書集合からの専門用語獲得. *情報処理学会論文誌*, Vol. 47, No. 6, pp. 1717–1727, 2006.
- [6] 中川, 湯本, 森. 出現頻度と接続頻度に基づく専門用語抽出. *自然言語処理*, Vol. 10, No. 1, pp. 27–45, 2003.