

# エントレインメント分析に基づく用例選択モデルの提案

## Response Selection on EBDM-Dialogue System based on Entrainment Analysis

水上 雅博\*      吉野 幸一郎      Graham Neubig†      中村 哲  
Masahiro Mizukami      Koichiro Yoshino      Graham Neubig      Satoshi Nakamura

<sup>1</sup> 奈良先端科学技術大学院大学 情報科学研究科

<sup>1</sup> Nara Institute Science and Technology, Information Science

**Abstract:** Entrainment is a factor in dialogue that affects not only human-human dialogue but also human-machine interaction. While the entrainment on the lexical level is well documented, less is known about how entrainment affects dialogue on a more abstract, and structural level. In previous works, we investigated the effect of entrainment on lexical choice given dialogue acts. From these results, we build the example selection on the EBDM dialogue system that would like to entrain with users in a similar manner.

## 1 はじめに

エントレインメント (Entrainment) は、同調傾向やシンクロニー (Synchrony) とも呼ばれ、対話中の話者間において、話し方や声の調子などの振る舞いが同調、類似する現象を指す。この現象は、語彙 [1, 2], 統語構造 [3, 4], 文体 [5, 6], 韻律 [7, 8, 4, 9], 発音 [10], ターンテイク [11, 12], 対話行為 [2] など、対話中の多様な要素で観測されることが知られている [13]。

また、エントレインメントが対話のタスク成功率や自然性、対話意欲 (Engagement) と相関することが報告されており [14], エントレインメントの分析を通して対話システムの性能や対話の質を測る試みが行われている。一方で、エントレインメントを対話中に考慮し、対話行為選択や応答選択に利用する取り組みは未だなされていない。

そこで本研究では、エントレインメントを考慮した用例ベース対話システムの応答選択手法に取り組む。この手法では、対話行為によって語彙のエントレインメントの度合いが変化する現象に着目し、ユーザの現在の対話行為とそれに対する応答におけるシステムの対話行為から次の発話のエントレインメントの度合いを推定し、適切なエントレインメントが生じている応答を選択する。ユーザの対話行為に合わせてエントレインメントの度合いを変化させることで、ユーザと対話システム間でより高い自然性と対話意欲を持った対話

の実現が期待される。

## 2 エントレインメントと関連研究

対話のエントレインメントを分析する研究として、聞き役対話における聞き役側の相槌 (Backchannel) のエントレインメント傾向の分析がある。具体的には、ピッチ (F0) や声量 (Power) の相関関係を示す研究 [9] や、発話のタイミングのみでなく、対話のターン管理や主導権についてエントレインメントが起きることを示す研究 [15] がある。さらに、対話において重要な要素である対話行為に着目した分析 [2] も行われている。この研究では、対話行為の選択傾向においてもエントレインメントが生じ、また、語彙のエントレインメントの度合いには発話の対話行為が非常に強く影響していることが分かっている。

また、人対システムの対話においては、エントレインメントを対話のモデルや戦略に組み込む研究もなされている。例えば、エントレインメントを考慮することで発話タイミングを高精度に予測する研究 [11] や、エントレインメントを誘発することで、より音声認識の成功率を高める対話戦略を持った情報提示対話システムを構築する研究 [16] がある。さらに、これらを統合し、ユーザのエントレインメントを制御・検出することで音声認識とターン管理の性能を向上する研究 [13] もある。

\*連絡先：奈良先端科学技術大学院大学情報科学研究科  
〒630-0101 奈良県生駒市高山町 8916-5  
情報科学研究科 知能コミュニケーション研究室  
E-mail: masahiro-mi@is.naist.jp

†現在、カーネギーメロン大学所属。

### 3 語彙と対話行為のエントレインメント関係

#### 3.1 エントレインメントの分析手法

語彙のエントレインメント分析 [1] では、語彙におけるエントレインメントの度合いを示す数値としてエントレインメント指標 (Entrainment Score) を提案している。エントレインメント指標は、それぞれの話者が発した単語頻度の差であり、その具体的な値  $En(V)$  は、特定の語彙群  $V$  と両話者の 1-gram 言語モデル確率  $P_{S_1}(w)$  および  $P_{S_2}(w)$  を用いて、以下のように表現される。

$$En(V) = - \sum_{w \in V} |P_{S_1}(w) - P_{S_2}(w)| \quad (1)$$

ここで、語彙群  $V$  はコーパス全体において最も出現頻度の高い 25 単語 (25MFC; 25 Most Frequent Words in Corpus) を用いる。先行研究において、この語彙群 25MFC を用いて計算したエントレインメント指標は、タスクの成功率と高い相関を示している。

我々は、対話行為と語彙のエントレインメント分析のために、先述のエントレインメント指標を拡張した。発話の対話行為を考慮した条件付 1-gram 言語モデル確率を用いて、以下のように表現される。

$$En(V|d) = - \sum_{w \in V} |P_{S_1}(w|d) - P_{S_2}(w|d)| \quad (2)$$

この拡張されたエントレインメント指標では、ある対話行為  $d$  が与えられた際に生じる語彙のエントレインメントの度合いを示している。

さらに、エントレインメント指標の拡張として、対話相手以外の話者とのエントレインメント指標との勝率の利用を提案する (エントレインメント指標比率; Entrainment Score Ratio)。これにより、対話のターン数や語彙数の影響で直接比較することが困難なエントレインメント指標が比較可能となる。エントレインメント指標比率は次の手順で計算される。まず、特定の話者において、その話者の条件付き言語モデル確率  $P_{S_{\text{target}}}(w|d)$  と、対話している相手話者との条件付き言語モデル確率  $P_{S_{\text{partner}}}(w|d)$  から、対話相手とのエントレインメント指標  $En_{\text{partner}}(V|d)$  を計算する。

$$En_{\text{partner}}(V|d) = - \sum_{w \in V} |P_{S_{\text{target}}}(w|d) - P_{S_{\text{partner}}}(w|d)| \quad (3)$$

同様に、特定の話者の条件付き言語モデル確率  $P_{S_{\text{target}}}(w|d)$  と、対話と無関係な話者群の条件付き言語モデル確率  $P_{\text{non-partner}_i}(w|d)$  から、対話外の相手とのエントレ

インメント指標  $En_{\text{non-partner}_i}(V|d)$  を計算する。

$$En_{\text{non-partner}_i}(V|d) = - \sum_{w \in V} |P_{\text{target}}(w|d) - P_{\text{non-partner}_i}(w|d)| \quad (4)$$

この対話相手とのエントレインメント指標と対話外の相手とのエントレインメント指標を比較し、対話相手とのエントレインメント指標が対話外の相手とのエントレインメント指標より高くなる確率を計算することで、エントレインメント指標比率  $R(V|d)$  が得られる。

$$R(V|d) = \frac{1}{N} \sum_{i \in N} \begin{cases} 1 & (En_{\text{partner}}(V|d) > En_{\text{non-partner}_i}(V|d)) \\ 0.5 & (En_{\text{partner}}(V|d) = En_{\text{non-partner}_i}(V|d)) \\ 0 & (En_{\text{partner}}(V|d) < En_{\text{non-partner}_i}(V|d)) \end{cases} \quad (5)$$

このエントレインメント指標比率は、エントレインメント指標の勝率であるため、1 に近いほど対話中の話者間に強くエントレインメントが生じていることを示す。逆に、エントレインメントが生じていない場合は 0.5 に近くなり、対話中の話者間にエントレインメントと逆の現象が生じている場合 0 に近くなる。

#### 3.2 エントレインメントの分析結果

対話行為と語彙のエントレインメント分析 [2] の結果、語彙のエントレインメントは対話全体で一様な強さを持つのではなく、特定の対話行為において強く生じていることが明らかになった。

具体的には、挨拶 (Conventional closing, Conventional opening)、相槌 (Acknowledge)、表明 (Statement non opinion, Statement opinion) などの対話行為においてエントレインメントは強く生じ、逆に謝罪 (Apology)、指示 (Action directive)、ネガティブな否定でない回答 (Negative non-no answer) などの対話行為においてはエントレインメントは生じなかった<sup>1</sup>。これらのことから、人間は対話において、対話行為を考慮してエントレインメントを行うかどうかを決定していることが分かる。

### 4 エントレインメントを考慮した応答選択モデル

対話行為と語彙のエントレインメント分析の結果から、自然なエントレインメントを行う対話システムを

<sup>1</sup>Switchboard dialogue act corpus を対象にした分析結果。詳細は [2] を参照。

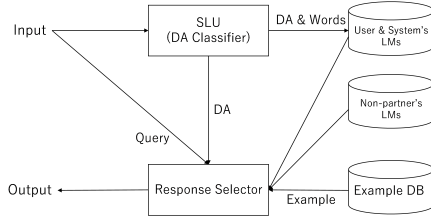


図 1: 提案法の全体図

構築する．まず，対話システムの全体像を図 1 に示す．この対話システムでは，用例ベース対話システム [17] の枠組みに沿って用例選択を行うが，用例選択の際に入力文との類似度のみでなく，エントレインメントの度合いを考慮して用例を選択する．そのため，応答選択はユーザから与えられたユーザ発話  $q'$  に対して，適当なシステム応答  $r_j$  を持つ用例  $\langle q_j, r_j \rangle$  を用例データベース  $e$  から探索する処理を，ユーザ発話  $q'$  とクエリ発話  $q_j$  との類似度  $\text{Sim}(q', q)$  と，その用例がユーザに対するエントレインメント度合いの妥当性  $\text{Entr}(r_j)$  の両者の最大化を行うことで決定する．

$$\hat{r} = \underset{\langle q_j, r_j \rangle \in e}{\text{argmax}} (\lambda \text{Sim}(q', q_j) + (1 - \lambda) \text{Entr}(r_j)). \quad (6)$$

このエントレインメント度合いの妥当性  $\text{Entr}(r_j)$  について，3章で述べたエントレインメント指標では，相手と同じ発話を対話システムが返したとき，言語モデル確率の差は 0 になり，最もエントレインメントが強く生じていることになる．しかしながら，エントレインメントは対話行為ごとに適切な度合いがあり，常にオウム返しが無効な戦略ではない．つまり，対話行為に応じて適切なエントレインメントの度合いを推定し，それに合わせた応答文を選択しなければならない．そこで，我々はエントレインメント指標比率を用いて，エントレインメント分析の結果と同程度のエントレインメント指標比率を持つ応答をそれぞれの発話文に対して推定する手法を提案する．具体的には，ユーザの対話行為を与えられた条件付き言語モデル確率  $P_{user}(w|d)$  と対話システムの対話行為に対応する条件付き言語モデル確率  $P_{system}(w|d)$  を用いる．これらを事前に用意された対話と無関係な話者の発話から学習された  $N$  個の条件付き言語モデル確率群  $P_{non-partner_i}(w|d)$  ( $i \in N$ ) と比較してエントレインメント指標比率を計算し，応答選択に利用する．応答選択を行う際，応答候補群の発話  $r_j$  を対話システムの言語モデルにそれぞれ追加し，応答候補ごとの条件付き言語モデル確率  $P_{system+r_j}(w|d)$  を計算する．これを用いてエントレインメント指標比率  $R(V|d, P_{user}, P_{system+r_j})$  を求める．

$$\text{En}_{\text{participants}}(V|d) = - \sum_{w \in V} |P_{user}(w|d) - P_{system+r_j}(w|d)| \quad (7)$$

$$\text{En}_{\text{non-partner}_i}(V|d) = - \sum_{w \in V} |P_{non-partner_i}(w|d) - P_{system+r_j}(w|d)| \quad (8)$$

$$R(V|d, P_{user}, P_{system+r_j}) = \frac{1}{N} \sum_{i \in N} \begin{cases} 1 & (\text{En}_{\text{participants}}(V|d) > \text{En}_{\text{non-partner}_i}(V|d)) \\ 0.5 & (\text{En}_{\text{participants}}(V|d) = \text{En}_{\text{non-partner}_i}(V|d)) \\ 0 & (\text{En}_{\text{participants}}(V|d) < \text{En}_{\text{non-partner}_i}(V|d)) \end{cases} \quad (9)$$

この応答候補ごとの条件付き言語モデル確率から計算されたエントレインメント指標比率  $R(V|d, P_{user}, P_{system+r_j})$  が，事前に計算された理想とするエントレインメント指標比率  $R_{ideal}(V|d)$  との差から，エントレインメントの妥当性を示す  $\text{Entr}(r_j)$  を計算する．

$$\text{Entr}(r_j) = 1 - |R(V|d, P_{user}, P_{system+r_j}) - R_{ideal}(V|d)| \quad (10)$$

これにより，理想的なエントレインメントの度合いに近い応答文を選択することが可能となり，より自然なエントレインメントの再現が期待される．

## 5 まとめ

本稿では，対話行為と語彙のエントレインメント分析について報告した．また，その結果に基づいて，エントレインメントの度合いを考慮した用例選択を行う用例ベース対話システムの応答の選択手法を提案した．この手法により，対話行為ごとに理想的なエントレインメント度合いに近い応答文を選択し，出力することが可能になる．

## 謝辞

本研究は，JST, CREST の支援を受けたものである．

## 参考文献

- [1] Susan E Brennan and Herbert H Clark. Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, Vol. 22, No. 6, p. 1482, 1996.

- [2] Masahiro Mizukami, Koichiro Yoshino, Graham Neubig, David Traum, and Satoshi Nakamura. Analyzing the effect of entrainment on dialogue acts. In *Proc. SIGDIAL*, 2016.
- [3] David Reitter and Johanna D Moore. Predicting success in dialogue. In *Proc. ACL*, 2007.
- [4] Arthur Ward and Diane Litman. Measuring convergence and priming in tutorial dialog. In *University of Pittsburgh*, 2007.
- [5] Kate G Niederhoffer and James W Pennebaker. Linguistic style matching in social interaction. *Journal of Language and Social Psychology*, Vol. 21, No. 4, pp. 337–360, 2002.
- [6] Cristian Danescu-Niculescu-Mizil, Michael Gamon, and Susan Dumais. Mark my words!: linguistic style accommodation in social media. In *Proceedings of the 20th international conference on World wide web*, pp. 745–754. ACM, 2011.
- [7] Michael Natale. Convergence of mean vocal intensity in dyadic communication as a function of social desirability. *Journal of Personality and Social Psychology*, Vol. 32, No. 5, p. 790, 1975.
- [8] Rachel Coulston, Sharon Oviatt, and Courtney Darves. Amplitude convergence in children’s conversational speech with animated personas. In *Proc. IC-SLP*, Vol. 4, pp. 2689–2692, 2002.
- [9] Tatsuya Kawahara, Takashi Yamaguchi, Miki Uesato, Koichiro Yoshino, and Katsuya Takanashi. Synchrony in prosodic and linguistic features between backchannels and preceding utterances in attentive listening. In *APSIPA*, pp. 392–395. IEEE, 2015.
- [10] Jennifer S Pardo. On phonetic convergence during conversational interaction. *The Journal of the Acoustical Society of America*, Vol. 119, No. 4, pp. 2382–2393, 2006.
- [11] Nick Campbell and Stefan Scherer. Comparing measures of synchrony and alignment in dialogue speech timing with respect to turn-taking activity. In *INTERSPEECH*, pp. 2546–2549, 2010.
- [12] Štefan Beňuš, Agustín Gravano, Rivka Levitan, Sarah Ita Levitan, Laura Willson, and Julia Hirschberg. Entrainment, dominance and alliance in supreme court hearings. *Knowledge-Based Systems*, Vol. 71, pp. 3–14, 2014.
- [13] Rivka Levitan. Entrainment in spoken dialogue systems: Adopting, predicting and influencing user behavior. In *HLT-NAACL*, pp. 84–90, 2013.
- [14] Ani Nenkova, Agustín Gravano, and Julia Hirschberg. High frequency word entrainment in spoken dialogue. In *Proc. ACL*, pp. 169–172. Association for Computational Linguistics, 2008.
- [15] Rivka Levitan, Stefan Benus, Agustín Gravano, and Julia Hirschberg. Entrainment and turn-taking in human-human dialogue. In *AAAI Spring Symposium on Turn-Taking and Coordination in Human-Machine Interaction*, 2015.
- [16] Andrew Fandrianto and Maxine Eskenazi. Prosodic entrainment in an information-driven dialog system. In *INTERSPEECH*, pp. 342–345, 2012.
- [17] 水上雅博, Lasguido Nio, 木付英士, 野村敏男, Graham Neubig, 吉野幸一郎, Sakriani Sakti, 戸田智基, 中村哲. 快適度推定に基づく用例ベース対話システム. 人工知能学会論文誌, 2016.