

選択性を考慮した語彙エントレインメント尺度

守屋彰二¹ 佐藤志貴¹ 徳久良子¹ 赤間怜奈^{1,2} 横井祥^{1,2} 乾健太郎^{3,1,2}

¹ 東北大学 ² 理化学研究所 ³ MBZUAI

shoji.moriya.q7@dc.tohoku.ac.jp

{shiki.sato.d1,tokuhisa,akama,yokoi,kentaro.inui}@tohoku.ac.jp

概要

語彙エントレインメントとは、対話の参加者の語彙選択が徐々に同調してくる現象である。先行研究では、語彙項目の一致頻度に基づいてこの現象が定量化されてきたが、このやり方では、単に他に適切な表現が無いから相手と同じ表現を使用しているのか、それとも相手に同調して「あえて」その表現を使用しているのかを区別できない。そこで本研究では、選択性、つまり「あえて」の度合いを明に考慮した尺度を提案する。まず、他に自然な選択肢が存在するかどうかを言語モデルと言語資源を用いて計算し、次に、当該の表現を選択する必然性が小さい場合に大きなスコアを与える。また実際の対話データを用いた実験により、提案尺度によってはじめて捕捉できるエントレインメント現象を明らかにし、対話システムへの応用可能性について議論する。

1 はじめに

人同士の対話において、対話の進行に伴い話し方や振る舞いが徐々に同調してくることが知られている。この現象はエントレインメントと呼ばれ、語彙 [1, 2]、統語構造 [3, 4]、文体 [5]、韻律 [6, 4, 7, 8]、ターンテイキング [9]、対話行為 [2] など、対話の様々な要素において観測される。先行研究では、エントレインメント度合いの高い対話では、ユーザの対話意欲や対話を介したタスクの成功率も高くなりやすいことが報告されている [3, 10]。これを踏まえ、エントレインメントの分析を通じて対話システムの性能や対話の質を測る試み [10, 11] や、エントレインメントを対話戦略に組み込む試み [12, 13] が行われている。これらの研究では、対話におけるエントレインメントを測る指標として、語彙エントレインメントが頻繁に用いられてきた [10, 2]。語彙エントレインメントとは、対話に参加する話者の語彙選択が徐々に同調してくる現象である [14, 1]。

先行研究では、対話相手と使用する語彙項目が一致する頻度に基づいて、語彙エントレインメントを定量化する手法が提案されてきた [10, 15]。しかし、対話相手と同じ表現を用いている場合においても、語彙エントレインメントの度合いに差があることが考えられる。例えば、他に適切な表現が無いから対話相手と同じ表現を使用している場合より、対話相手に同調してあえてその表現を使用している場合の方が、エントレインメントの度合いが高いといえる。

そこで、我々は選択性を明に考慮した語彙エントレインメント尺度を提案する。具体的には、他に自然な選択肢が多く存在するかどうかを言語モデルと言語資源を用いて計算し、当該の表現を選択する必然性が小さい場合に高い語彙エントレインメントスコアを与える。実際の対話データを用いた実験により、「意図的に」表現を選択している度合いを適切に考慮できていることを定性的に確認した。

2 関連研究

2.1 語彙エントレインメント

話者がある概念に対して特定の表現を用いて言語化を行ったとき、その後の対話において対話相手と同じ表現を用いて言語化することがある（**語彙エントレインメント**） [14]。一度対話相手と同じ表現が使用されると、Grice の量の格率 [16] を始めとする他の様々な語彙選択の要因に反していたとしても、同じ表現を使用し続ける場合があることが知られている。Brennan ら [1] はこの現象を、一方の話者がある言語化を提案したのに対し、他者がその提案を受け入れることでその表現が定着すると考えた¹⁾。

1) Brennan らはこの合意を “conceptual pact” と呼んだ。

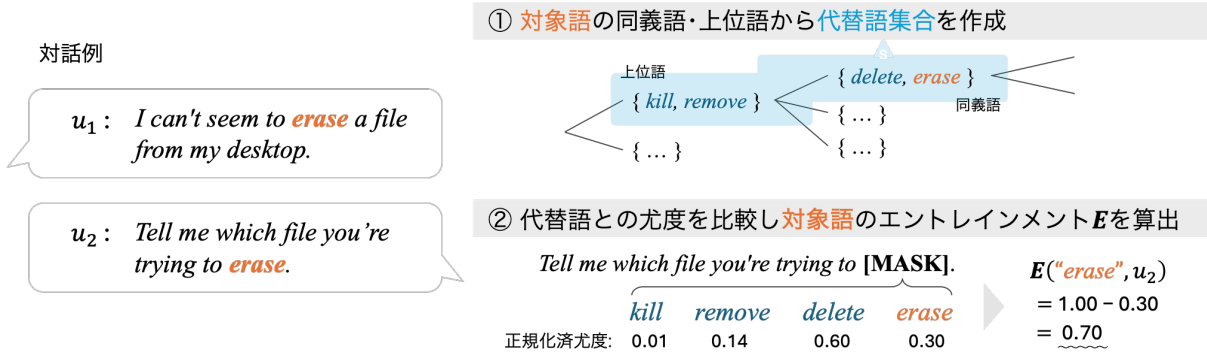


図1 提案手法の概要

2.2 既存の語彙エンタレインメント尺度

先行研究では、対話相手と使用する語彙項目が一致する頻度に基づいて、語彙エンタレインメントを定量化する手法が提案されている。

Nenkova ら [10] は、対話コーパス内の高頻度語に着目し、それぞれの話者が使用した単語の頻度差に基づいて語彙エンタレインメントを定量化した。具体的には、語彙エンタレインメント $E(V)$ を、特定の語彙群 V 内の語彙項目 w を2人の話者 S_1, S_2 が使用した回数をもとに、以下のように表現した。

$$E(V) = - \sum_{w \in V} \left| \frac{\text{count}_{S_1}(w)}{\text{ALL}_{S_1}} - \frac{\text{count}_{S_2}(w)}{\text{ALL}_{S_2}} \right| \quad (1)$$

ここで、 ALL_{S_i} は各話者が対話中に使用した語彙項目の総数であり、 $\text{count}_{S_i}(w)$ は各話者が対話中に語彙項目 w を使用した回数である。語彙群 V には、コーパス全体において使用頻度が高い25単語が用いられている。この尺度は、自然な対話かどうかを判断する要因になることが示されており、後続の様々な研究において広く使われた [17, 18]。

一方で、Shi ら [15] は、高頻度語に基づく尺度が語彙エンタレインメントを広く解釈し過ぎていることを指摘し、参照表現に着目した新たな尺度を提案した。具体的には、ある話者が使用した名詞句に対し、対話相手のその後の発話において、修飾語や活用形を無視した上で同じ名詞句を用いている頻度をもとに語彙エンタレインメントを定量化した。本稿では、Shi らの手法によって算出されるスコアを**頻度スコア** S_{count} と呼称する。

このように、先行研究では対話相手と使用する語彙項目が一致する頻度に基づいて語彙エンタレインメントを定量化している。3節では、既存手法の抱える問題点について述べ、これを解消する方法を提案する。

3 提案手法

2.2節で述べたように、既存手法では同じ表現を使用している頻度に基づいて語彙エンタレインメントを定量化している。しかし、頻度に基づく手法では、他に適切な表現が無いから対話相手と同じ表現を使用しているのか、それとも対話相手に同調してあえてその表現を使用しているのかを区別することが難しい。そこで本研究では、どの程度「**意図的に**」選択しているかを考慮した語彙エンタレインメント尺度を提案する。

提案手法の概要図を図1に示す。まず、①対話相手が以前使用した語彙項目に対し、その同義語や上位語から代替語集合を作成する。これらは、話者がある概念を言語化する際の語彙選択の候補を表現している。次に、②集合内の各語彙項目に対し、代替できる可能性を言語モデルの尤度をもとに算出する。これにより、選択した語彙項目の他にどれだけ語彙選択の余地があるかを定量化している。

続けて、具体的な手法を形式的に述べる。まず、対話中のあるターン j の発話 u_j において、対話相手が以前使用した語彙項目 w と同じ表現を用いている事例に注目する。次に、対象の語彙項目の同義語や上位語を取得し、それらを代替語集合 V_w とする。その後、対象の語彙項目が他の代替語と比較してどの程度意図的に選択されているかを、言語モデルの尤度をもとに以下の式で算出する。

$$E(w|u') = 1 - p(w|u') \quad (2)$$

ここで、 u' は1ターンの発話 u_j から該当の語彙項目 w を MASK トークンに置換した文である。また、 $p(w|u')$ は、代替語集合 V_w における尤度の総和が1になるように正規化した値である。対象語彙項目 w の他に文脈上自然な代替語が存在する場合、 $p(w|u')$ はより低い値を取ることから、 $E(w|u')$ が

表1 「選択性」を考慮してエンタテインメント度合いを定量化出来ている事例

話者	発話	S_{count}	S_{choice}
(a) B	You all have much trouble up there, with insects		
A	Uh, well, not a whole lot. I think our cabbages, and broccolis and that we have more trouble with, that type of thing, with the	1	0.593
(b) B	Well, what kind o grass do you grow up there?		
A	Do you grow, uh, Bermuda or, Well, our grass , like I said is just grass . We really didn't plant anything.	2	0.032
(c) A	Do you like to cook?		
B	Yes, I love to cook.		
A	Yeah, I love to cook, too,	0	0.816

高くなることが想定される。この値を、1ターンの発話に含まれる代替語が存在する全ての語彙項目²⁾に対し算出し、その最大値を取ることで、そのターンにおける**選択性スコア** S_{choice} を算出した。

4 実験

各対話におけるターン毎の頻度スコアおよび選択性スコアを算出し、本手法がより直感に沿った語彙エンタテインメントの定量化を行えているかどうかについて分析する。

4.1 実験設定

データセット 本研究では、書き起こしの雑談対話コーパスである Switchboard corpus [19] を採用した。Switchboard corpus [19] は、2名の話者の電話による雑談を収録したコーパスであり、対話分析研究において広く使用されている。過去のエンタテインメント分析においても同コーパスが頻繁に使用されている [10, 2]。本研究では、同コーパスに含まれる1126対話を対象とした。参照する対話相手の先行発話数は、局所的な(固定窓幅における対話文脈の)言語的エンタテインメントを測定するフレームワークである Local Interpersonal Distance (LID) [11] のアイデアに基づき、2ターンとした。

提案法で用いる資源 尤度を算出するための**言語モデル**として、RoBERTa-base³⁾を使用した。また、同義語および上位語を取得するための**言語資源**として、WordNetを使用した⁴⁾。対象とする語彙項目からはNLTK⁵⁾ライブラリの Stopwords に含まれる語

2) 語彙選択は名詞だけでなく、動詞や形容詞などにおいても行われることが考えられる。そのため、対象を代替語が存在する全ての語彙項目とした。

3) <https://huggingface.co/roberta-base>

4) 下位語については、文脈から判断することが困難であるため、代替語集合に含めていない。

5) <http://www.nltk.org>

彙項目や記号を除外した。

4.2 実験結果

実験結果を図2に示す。横軸は正規化されたターンを表しており、左端が対話開始時、右端が対話終了時を示している。縦軸は各スコアの値を表しており、上方向(青)のグラフが我々の尺度で算出される**選択性スコア**、下方向(赤)のグラフがShiらの尺度で算出される**頻度スコア**を示している。

本手法で対象とする語彙項目は、Shiらが対象としている語彙項目の大半を網羅できていることがわかった。コーパス中の全対話において、頻度スコアが高い箇所のうち89.7%では**選択性スコア**もある程度の値を示していた。これは、Shiらの尺度が対象とする名詞は**選択性**をもつ語彙項目が多く、ほとんどが我々の尺度が対象とする語彙集合に含まれるためである。

また、表1より、本手法を使用することで対話相手と同じ表現を「意図的に」選択している度合いを適切に考慮できていることを定性的に確認できた。図2を見ると、頻度スコアが高い値を示している箇所では、**選択性スコア**が高い値を示す箇所(a)もあれば、低い値を示す箇所(b)もあることがわかる。これらの箇所の対話例および算出された各スコアを表1に示す。(a)では“trouble”の代わりに同義語の“difficulty”を使って表現することもできるが、対話相手と同じ表現が用いられている。一方で、(b)では、“grass”を代替できる可能性のある同義語や上位語がほとんど無い。ここから、(b)の事例に比べ(a)の事例の方がエンタテインメント度合いが高いことが想定され、実際に**選択性スコア**もそれに合った値となっている。ここから、**選択性**を考慮したスコアを算出することが出来ていることを定性的に確認した。

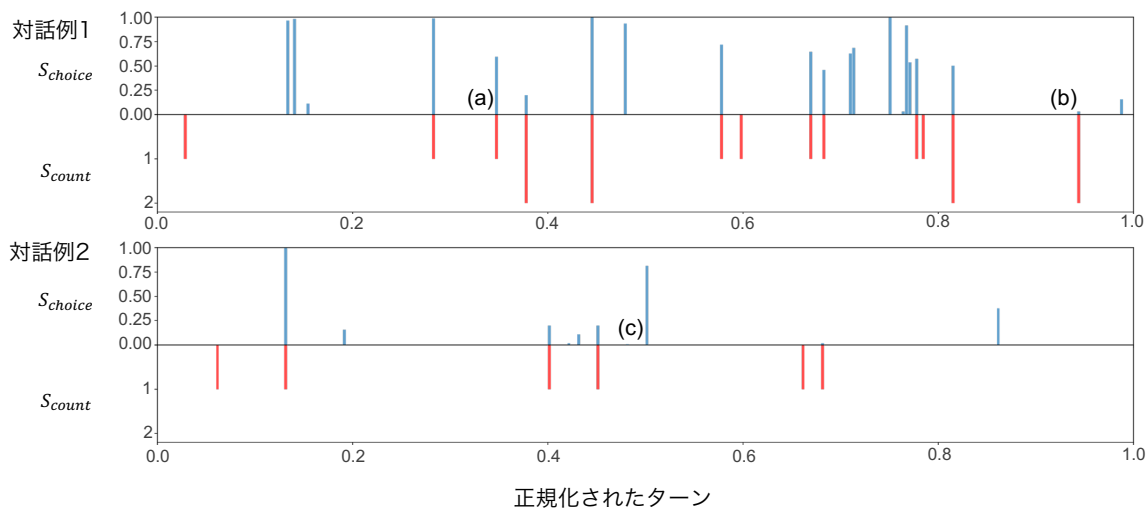


図2 実験結果. 上方向(青)が我々の尺度で算出される選択性スコア, 下方向(赤)がShiらの尺度で算出される頻度スコア.

また、動詞や形容詞など、本手法のみが対象としている語彙項目においても直感に沿ったスコアを算出できていることが確認できた。図2より、選択性スコアが一定の値を示す箇所において、頻度スコアが0になっている箇所が見受けられた。これらを定性的に確認すると、表1(c)のように、“love”に比べ選択されやすいと考えられる上位語の“like”により高い尤度が割り当てられており、選択性スコアが高くなっていることが確認できた。ここから、語彙エンタインメントスコアを算出する上で、名詞以外の語彙項目についても対象にできることが示唆される。

5 議論

5.1 提案手法の問題点

提案手法では、1ターンに含まれるエンタインメント表現の回数の情報が失われてしまっている。選択性スコアは、1ターンに代替語がある語彙項目が複数ある場合、それらのエンタインメント度合いの最大値を取ることで、そのターンのエンタインメント度合いを定量化している。そのため、発話中にエンタインメント度合いの高い表現が複数ある場合でも、その中で最もエンタインメント度合いが高い表現のみが考慮されてしまう。この問題に対し、既存手法での頻度に基づいた尺度と組み合わせることで、1ターンあたりのエンタインメント度合いをより適切に定量化ができることが考えられる。

5.2 提案手法の応用可能性

本提案手法の応用として、エンタインメントを考慮した対話システムの応答選択への活用が考えられる。本手法では、エンタインメント度合いの算出に言語モデルの尤度を活用している。そこで、ユーザが過去に使用した語彙項目や対話システムの応答生成時に得られる各トークンの尤度を活用したデコーディングアルゴリズムやランキングシステムを構築することで、語彙エンタインメントを考慮した応答を生成できることが期待される。

6 おわりに

本研究では、選択性を考慮した語彙エンタインメント尺度を提案した。実際の対話データに対し、本手法を用いて語彙エンタインメントスコアを算出することで、対話相手と同じ表現を意図的に選択している度合いを考慮できていることを定性的に確認した。今後の研究では、本手法を既存手法と組み合わせることでさらに精緻に語彙エンタインメントの度合いを定量化することを試みるとともに、対話システムへの応用手法について模索していきたい。

謝辞

本研究は JSPS 科研費 JP21J22383, JP21K21343, JP22K17943, JP22H03654, JST CREST JPMJCR20D2 の助成を受けたものです。本研究を進めるにあたり, TohokuNLP の坂口慶祐氏, 塩野大輝氏, 谷口雅弥氏, 松崎孝介氏に有益なご助言をいただきました。ここに深く感謝申し上げます。

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 Raquel Fernandez, Wolfgang Minker, Giuseppe Carenini, Ryuichiro Higashinaka, Ron Artstein, and Alesia Gainer, editors, **Proceedings of the 17th Annual Meeting of the Special Interest Group on Discourse and Dialogue**, pp. 310–318, Los Angeles, September 2016. Association for Computational Linguistics.
- [3] David Reitter and Johanna D. Moore. Predicting success in dialogue. In Annie Zaenen and Antal van den Bosch, editors, **Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics**, pp. 808–815, Prague, Czech Republic, June 2007. Association for Computational Linguistics.
- [4] Arthur Ward and Diane Litman. Measuring convergence and priming in tutorial dialog. **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] 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.
- [7] Rivka Levitan, Agustín Gravano, Laura Willson, Štefan Beňuš, Julia Hirschberg, and Ani Nenkova. Acoustic-prosodic entrainment and social behavior. In **Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human language technologies**, pp. 11–19, 2012.
- [8] 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 **2015 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA)**, pp. 392–395. IEEE, 2015.
- [9] Nick Campbell and Stefan Scherer. Comparing measures of synchrony and alignment in dialogue speech timing with respect to turn-taking activity. In **Eleventh Annual Conference of the International Speech Communication Association**, 2010.
- [10] Julia Bell Hirschberg, Ani Nenkova, and Agustin Gravano. High frequency word entrainment in spoken dialogue. 2008.
- [11] Md Nasir, Sandeep Nallan Chakravarthula, Brian Baucum, David C. Atkins, Panayiotis Georgiou, and Shrikanth Narayanan. Modeling interpersonal linguistic coordination in conversations using word mover’s distance, 2019.
- [12] Seiya Kawano, Masahiro Mizukami, Koichiro Yoshino, and Satoshi Nakamura. Entrainable neural conversation model based on reinforcement learning. **IEEE Access**, Vol. 8, pp. 178283–178294, 2020.
- [13] 金崎翔大, 河野誠也, 湯口彰重, 桂井麻里衣, 吉野幸一郎. エンtrainメントスコアを用いた応答ランキングとその自動評価. 言語処理学会第 29 回年次大会 発表論文集, pp. 1963–1968, 2023.
- [14] Simon Garrod and Anthony Anderson. Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. **Cognition**, Vol. 27, No. 2, pp. 181–218, 1987.
- [15] Zhengxiang Shi, Procheta Sen, and Aldo Lipani. Lexical entrainment for conversational systems. In Houda Bouamor, Juan Pino, and Kalika Bali, editors, **Findings of the Association for Computational Linguistics: EMNLP 2023**, pp. 278–293, Singapore, December 2023. Association for Computational Linguistics.
- [16] Herbert P Grice. Logic and conversation. In **Speech acts**, pp. 41–58. Brill, 1975.
- [17] Guillaume Dubuisson Duplessis, Caroline Langlet, Chloé Clavel, and Frédéric Landragin. Towards alignment strategies in human-agent interactions based on measures of lexical repetitions. **Language Resources and Evaluation**, Vol. 55, pp. 353–388, 2021.
- [18] Andreas Weise, Vered Silber-Varod, Anat Lerner, Julia Hirschberg, and Rivka Levitan. “talk to me with left, right, and angles” : Lexical entrainment in spoken hebrew dialogue. In **Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics**, 2021.
- [19] John J Godfrey, Edward C Holliman, and Jane McDaniel. Switchboard: Telephone speech corpus for research and development. In **Acoustics, speech, and signal processing, ieee international conference on**, Vol. 1, pp. 517–520. IEEE Computer Society, 1992.