

大規模言語モデルを用いた リフレーミング表現の自動生成とその評価

河野 誠也^{1,a)} 湯口 彰重¹ 吉野 幸一郎¹

概要：リフレーミングとは、ある枠組みで捉えられている物事の枠組みを外して違う枠組みで見ることを指し、その結果として言い換えが生じる。例えば、ネガティブな内容を異なった側面から捉えてポジティブにリフレーミングし言及することで、対話において話し手と聞き手の良好な関係の維持や自己肯定感の向上を期待することができる。このようなネガティブな発言をポジティブにリフレーミングするような言語的配慮能力は、社会性を備えた対話システムを実現する上で重要である。そこで本研究では、ネガティブな表現をポジティブに言い換えるリフレーミングの事例を、ウェブや書籍など複数の情報源から人手で収集し、対話応答生成タスクで事前学習された大規模言語モデルを用いてネガティブな表現をポジティブに言い換えるリフレーミングの生成モデルの学習とその評価を実施した。

キーワード：リフレーミング, 条件付き言語モデル, 言語生成, 言い換え

1. はじめに

リフレーミングとは、短期療法や家族療法、ナラティブセラピー等で用いられる技法であり、ある枠組みで捉えられている物事の枠組みを外して、違う枠組みで見ることを指す [1], [2], [3]。発話によるネガティブな感情の発露は、話し手と聞き手の双方にストレスを与える可能性がある。このため、ネガティブな発言をポジティブにリフレーミングすることは対話継続性向上の観点から重要で、リフレーミングにより話し手の自己肯定感の向上や話し手と聞き手の良好な関係の維持を期待することができる。このように、ネガティブな発言をポジティブにリフレーミングするような言語的配慮能力は、社会性を備えた対話システムを実現する上で重要である。

そこで、本研究では、ネガティブな表現をポジティブにリフレーミングする言語生成モデルを検討する。一般にこうした言い換えの研究では、言い換え元と言い換え先のペアからなる大規模なパラレルデータを必要とする [4]。しかし、あらゆる目的のためにこうした大規模パラレルデータを準備することは現実的ではなく、文アライメントに基づく教師なしの手法などが提案されてきた [5]。しかしながら、完全に教師なしでリフレーミング表現の生成モデルを構築することは困難を伴う。そのため、本研究ではまず

最初のステップとして、ネガティブな表現をポジティブにリフレーミングした事例を、ウェブや書籍など複数の情報源から人手で収集し、リフレーミング表現生成モデルを学習するためのパラレルコーパスを構築した。次に、対話応答生成タスクで事前学習された大規模言語モデルを、構築したコーパスを用いてファインチューニングすることで、ネガティブな表現をポジティブに言い換えるリフレーミング表現生成モデルを学習した。しかしながら、事前学習モデルが扱う対話応答生成タスクとリフレーミングでは、ややタスクに隔たりがあるためファインチューニングがうまく機能しない場合がある。そこでこのギャップを埋めるために、本研究では追加の学習データとしてリフレーミングと関連度が高い対話応答ペアを自動収集し、リフレーミング表現の生成タスクに加えて、対話応答生成タスクを同時にするようにモデルを拡張した。

評価実験では、構築したリフレーミング表現の自動生成モデルの性能と限界について検証した。具体的には、事前学習モデルの種類、対話用例を用いたデータ拡張の有無、モデルが扱うタスクの種類に着目した複数の設定に基づいたリフレーミング表現の自動生成モデルを比較した。また、リフレーミング表現生成モデルの評価結果を踏まえて、今後の研究の展開に向けた課題を議論した。

2. 関連研究

これまでに、発話文の感情極性やスタイルの変換に着目

¹ 理化学研究所ガーディアンロボットプロジェクト、
京都府相楽郡精華町光台 2-2-2 国際電気通信基礎技術研究所 3 階
^{a)} seiya.kawano@riken.jp

した研究は数多く行われてきた [3], [6], [7]。感情極性の変換では、単純には入力文（ないし発話、単語）と意味が真逆な文を生成する [8]。例えば、料理のレビューの極性変換を考える場合、「本当にあたりさわりのない中華料理でした。」というレビューはネガティブに捉えられる。しかし、「あたりさわりのない」という表現は「難点が特に思い当たらずおいしい」と解釈することができる。そこで、「あたりさわりのない」を「おいしい」と変換することで、よりポジティブな「本当においしい中華料理でした。」に変換することができる。また、対話応答生成のタスクでは、目標応答の感情ラベルを陽にモデルに生成条件として与えることで、与えた感情ラベルを反映しつつ対話履歴に対しても適切な応答を生成しようとする制御可能なニューラル会話モデルが提案されている [9], [10]。このような変換タスクでは元々の入力文の構造や属性に依存しないトピックは変換後の文に保持されるが、変換前の文と変換後の文の真理条件的意味が等しくなる保障はない。

一方で、スタイル変換のタスクでは、入力文の意味的内容を保持しつつ意味以外の情報を制御する言い換えを生成することに焦点を当てる。例えば、元々の文の難解な表現を平易に言い換えたり、文の丁寧さのレベルを制御したりする研究が行われている [4], [11]。本研究が対象とする、ポジティブな表現をネガティブにリフレーミングする言い換えタスクは、スタイル変換のタスクと類似する。例えば、「私は自己中心的です」という文をリフレーミングすることを考える場合「私は自分を持っている」等と言い換えることができる。このような変換は、変換後の文の真理条件的意味は完全には等価ではないにせよ、フレームを変えて見た場合、一面では等価であるともみなすことができる（少なくとも、その個人や物事に対して共起しやすいパターンであり矛盾しない）。このような婉曲的な言い換えは、「皮肉」の生成とも関連する [12], [13]。例えば、相手の自己中心的な行動を非難する意図があり、「とても自分を持っているんだね」等と言うような場合、リフレーミングとほぼ同様の婉曲的な言い換えを生成することができる。しかしながら、皮肉が他者に対する悪意に紐づいたものである一方で、リフレーミングでは、より広範な概念を扱う。特に、本研究では、リフレーミングの中でも、社会的幸福や認知能力の向上を意図して行われるネガティブな表現をポジティブに言い換えるリフレーミングに着目した言い換え生成モデルの検討を実施する。

3. リフレーミング表現の自動生成方法

スタイル変換を含む言い換えの生成タスクは、同一言語内の機械翻訳の問題と考えることができるため系列変換モデルと相性が良い。通常、系列変換モデルは、数十万文対を超える大規模なパラレルコーパスを用いて学習されるが、リフレーミングタスクに着目した日本語のスタイル

変換コーパスには限りがある。このような問題に対して、大規模な言語資源を用いて事前学習された大規模言語モデルを、目標タスクの少量のデータでファインチューニングするようなアプローチが有望であることが知られている [14], [15]。そこで、本研究では、ネガティブな表現をポジティブにリフレーミングした事例を、ウェブや書籍など複数の情報源から人手で収集し、リフレーミング表現生成モデルを学習するための小規模のパラレルコーパスを構築する。次に、対話応答生成タスクで事前学習された大規模言語モデルを構築したコーパスを用いてファインチューニングすることで、ネガティブな表現をポジティブに言い換えるリフレーミング表現生成モデルを学習する。しかしながら、事前学習モデルが扱う言語生成タスクと目標タスクの隔たりが大きい場合ファインチューニングがうまく機能しない可能性がある。そこで本研究では、このギャップを埋めるために追加の学習データとして、リフレーミングと関連度が高い対話応答ペアを自動収集し、リフレーミング表現の生成タスクに加えて、対話応答生成タスクの学習を同時にするようにモデルを拡張する。人手で小規模なデータしか用意できないような場合に、それらをシードデータとして類似するデータを Web などから収集し利用する手法は、古くから言語モデル [16] や対話モデル [17], [18] の構築で用いられてきた。本研究では、こうした手法で収集されたデータをモデルの追加学習に用いることで、より自然なリフレーミングを実現しようとする。

3.1 条件付き言語生成モデリング

本研究が扱う、リフレーミング表現生成タスクは、ネガティブな表現 $N = [x_1, \dots, x_T]$ が与えられたとき、それをポジティブに言い換えたリフレーミング表現 $P = [y_1, \dots, y_{T'}]$ を生成することである（以下、N/P 変換）。また、同時に、ポジティブな表現 P をネガティブな表現 N にリフレーミングするような生成タスクも扱う（以下、P/N 変換）。ここで、 x_* 、 y_* は単語、 T 、 T' はそれぞれ N と P の単語数である。このような言い換え生成タスクは、リフレーミング前の文とその極性、リフレーミング後の文の極性、の三つを条件として、リフレーミング前の文に対するリフレーミング表現を生成する条件付き言語生成（系列変換）タスクとして定式化できる。本研究では、ポジティブ/ネガティブの相互変換を実現するために、変換前の表現の先頭にその文の極性、末尾に変換後の文の極性を示す単語を付加する。また、以下の目的関数を最大化するようにして条件付き言語モデルを学習する*1。

*1 より具体的には、ターゲットのリフレーミング表現の予測結果と正解表現の間の交差エントロピー誤差を学習に用いる。

$$J(\theta) = \frac{1}{|D_{N \rightarrow P}|} \sum_{(P_i, N_i) \in D_{N \rightarrow P}}^{M_{N \rightarrow P}} \log[p(P_i | N_i, \theta)] + \frac{1}{|D_{P \rightarrow N}|} \sum_{(P_j, N_j) \in D_{P \rightarrow N}}^{M_{P \rightarrow N}} \log[p(N_j | P_j, \theta)] \quad (1)$$

ここで、 D_* は、それぞれ N/P、P/N 変換についての学習データ集合、 θ は条件付き言語モデルのパラメータ、 p は目標リフレーミング表現の生成確率である。ただし、このような変換モデルを学習するための利用可能なデータセットのサイズには制限があるため、本研究では、条件付き言語モデルのパラメータ初期値 θ として、対話応答生成タスクで事前学習された 1.6B のパラメータを持つ Transformer Encoder-Decoder モデル*2 を利用する [19], [20]。しかしながら、本研究で利用する事前学習モデルは、主に SNS 上での雑談対話を学習データとしており、事前学習モデルが扱う言語生成生成タスクと目標タスクの隔たりが大きい場合ファインチューニングがうまく機能しない可能性がある。そこで、リフレーミングと関連度が高い対話応答ペアを自動収集し学習データに追加する。より具体的には、N/P 変換と、P/N 変換に加えて、通常の対話応答生成に関する事例を学習データに混在させる。

3.2 学習用データセットの構築

3.2.1 リフレーミング事例の収集

本研究では、リフレーミングの中でも、社会的幸福や認知能力の向上を意図して行われるポジティブな表現をネガティブに言い換えるリフレーミングを対象として言い換えの事例を Web や書籍から人手で収集した。より具体的には、「リフレーミング」、「ネガポジ辞典」、「ネガティブ ポジティブ 言い換え」などの検索ワードを手掛かりとした調査を実施し、主に教育心理学、カウンセリング、自己啓発等を扱う媒体から 1,323 件のリフレーミング事例を抽出し、単言語パラレルコーパスとして整理した。例えば、「あきらめが早い」というネガティブな表現については、「潔い」、「決断が早い」、「気持ちの切り替えが早い、上手」などの複数の言い換え表現がレファレンスとして割り当てられる。ただし、構築したコーパスは単語単位、フレーズ単位、文単位等、変換の粒度については特に統制されていないことに注意されたい。

3.2.2 リフレーミング表現を含む対話応答ペアの自動収集

本研究では、収集したリフレーミング事例について可能なリフレーミングの変換ペアの組み合わせに対応する対話応答ペア（対話履歴とそれに対する応答）を一般的な雑談対話コーパス*3 からパターンマッチングに基づいて抽出し、条件付き言語モデルの追加の学習データとした。より具体的には、それぞれのリフレーミング事例について、リフレー

ミング前の表現を対話履歴の発話のいずれかに含み、かつ、応答発話に変換後のリフレーミング表現が含まれる対話応答ペアを抽出した。ただし、構築したデータセットには、リフレーミング表現の表記ゆれや、リフレーミング事例の変換の粒度が統制されていない等の問題があるため、単純なパターンマッチングで得られる対話応答ペアの事例数は少なくなるといった問題がある。そこで、実際には、雑談対話コーパスの発話における単語 n-gram のいずれかと、与えた表現の類似度が 0.9 以上であれば、その発話は与えた表現を包含するものとみなした。類似度は、事前学習済みの BERT (Bidirectional Encoder Representations from Transformers) モデルによって得られる分散表現ベクトル間の cos 類似度 [21] によって計算した。

4. 評価実験

評価実験では、構築したリフレーミング表現の自動生成モデルの性能と限界について検証する。具体的には、事前学習モデルの種類、対話用例を用いたデータ拡張の有無、モデルが扱うタスクの種類に着目した複数の設定に基づいたリフレーミング表現の自動生成モデルを比較する。

4.1 データセット

本研究で用いるデータセットは、1,323 件のリフレーミングの事例から構成される。このうち本研究では、1,190 件の事例を学習用、259 件の事例を評価用データとして使用した。ただし、収集したリフレーミングの事例は、入力表現に対して複数のレファレンス表現があるため、実際には、与えた表現に対する変換後のリフレーミング表現の可能な組み合わせ（学習用データ: 2,083 件、評価用データ 679 件）をモデルの学習と評価に用いた。

4.2 モデル設定

事前学習モデルの種類、モデルが扱うタスクの種類、データ拡張の有無に着目して、複数の設定に基づいたリフレーミング表現の自動生成モデルを学習、比較する。まず、事前学習モデルについては、以下の三つのモデルを用いる場合について比較する。

- **Twitter:** 日本語 Twitter コーパスで事前学習したモデル。
- **Empdial:** Twitter データで学習された事前学習モデルを、感情な状況について発話する「話し手」と、それに応答する「聞き手」の 2 話者による共感的な対話によってファインチューニングしたモデル。
- **Persona:** Twitter データで学習された事前学習モデルを、話者の個性に一貫した対話によってファインチューニングしたモデル。

さらに、モデルが扱うタスクの種類、データ拡張の有無の組み合わせに基づいたリフレーミング表現生成モデルを

*2 <http://github.com/nttclab/japanese-dialog-transformers>

*3 本研究では、Twitter データを用いた。

比較する。具体的には、以下の設定に基づいたモデルを比較する。

- **N/P**: ネガティブな表現をポジティブに言い換える事例でモデルを学習する場合 (N/P 変換)。
- **P/N**: ポジティブな表現をネガティブに言い換える事例でモデルを学習する場合 (P/N 変換)。
- **N/P+P/N**: N/P 変換と P/N 変換の両方の事例でモデルを学習する場合。
- **N/P+P/N+Aug.**: N/P 変換と P/N 変換の事例に加えて、リフレーミング事例を含む対話応答ペアでモデルを学習する場合。

学習したモデルの評価は、言語生成モデルにおける標準的な評価指標である Perplexity (PPL) を用いた自動評価により実施する。

5. 評価結果

表1と表2に、学習したリフレーミング表現生成モデルの評価セットに対する Perplexity (PPL) を示す。ここで、表1は、N/P 変換事例に対して PPL を計算した場合、表2は、P/N 変換事例に対して PPL を計算した場合に対応する。

表1 N/P 変換事例に対する PPL

モデル設定/事前学習モデル	Twitter	Persona	Empdial
N/P	24.17	25.30	24.33
N/P+P/N	23.84	24.96	24.05
N/P+P/N+Aug.	23.62	24.51	23.87

表2 P/N 変換事例に対する PPL

モデル設定/事前学習モデル	Twitter	Persona	Empdial
P/N	25.13	25.97	24.95
N/P+P/N	24.38	25.44	24.28
N/P+P/N+Aug.	24.21	25.13	24.12

表1より、N/P 変換のタスクにおいては、事前学習モデルとして **Twitter** を用いた場合で、最も小さい PPL を示した。他方、**Persona** を用いる場合で、最も大きい PPL を示した。また、表2より、P/N 変換のタスクにおいては、事前学習モデルとして **Empdial** を用いる場合で、最も小さい PPL を示した。他方、N/P 変換のタスクと同様に、**Persona** を用いる場合に、最も大きい PPL を示した。対話応答生成のタスクにおいては、**Twitter** を用いたモデルが、他のモデルと比較して、様々な雑談対話データに対して高い PPL を示す傾向があるという調査結果 [20] があるが、リフレーミング表現の生成タスクにおいてはその知見と反する結果が得られた。この原因として、**Twitter** では、特定の場面や状況に特化した対話応答生成には課題がある一方で、**Twitter** は、様々な話題や様々な背景を持

つ話者による膨大な対話データを背景としているため、その言語モデルは様々な文を生成する可能性がある。また、**Empdial** は、共感的な対話を学習データとしており、本研究が扱うようなリフレーミングと類似する現象が学習データに含まれていた可能性がある。一方で、**Persona** の学習に用いられた対話データは、対話における話者の個人性の一貫性に着目したものであり、様々な話題や対話状況をカバーするようなものではない。そのため、N/P 変換と P/N 変換のタスクの両方で最も大きい PPL を示したのだと考えられる。

さらに、表1、表2より、リフレーミング表現生成モデルを学習する場合、N/P 変換や P/N 変換といった個別のタスクのみ扱うように学習されたモデル (**N/P**, **P/N**) と比較して、N/P 変換と P/N 変換の両方のタスクを単一のモデルで学習することで (**N/P+P/N**)、N/P 変換と P/N 変換の両方のタスクで PPL が向上することが確認できた。本研究でモデルの学習に用いた P/N 変換事例は、もともとは、収集した N/P 変換事例のソースとターゲットを単純に反転することで作成されたものである。したがって、N/P 変換事例と P/N 変換事例の両方を学習データとして変換モデルを学習するという事は、ネガティブ文とポジティブ文の間の相互情報量の最大化とも関連する。相互情報量の最大化による学習アプローチは、対話応答生成モデルの学習にもしばしば用いられる [22], [23]。具体的には、通常のターゲット文に対する交差エントロピー誤差に加えて、ソース文とターゲット文の間の相互情報量に基づいた誤差を追加してモデルを学習することで、生成応答の一貫性が向上することが報告されている。また、リフレーミング表現生成モデルを学習するための追加データとして、リフレーミング事例が含まれる対話応答ペアを利用する場合 (+Aug.)、P/N 変換、N/P 変換タスクの両方で、PPL が向上することが確認できた。この結果は、事前学習モデルが扱う対話応答生成タスクとリフレーミングのタスクのギャップを埋めることで、言語生成モデルのファインチューニングの効率を向上できる可能性を示唆している。

表3、表4に実際に、**N/P+P/N+Aug.** モデル (**Twitter** を事前学習モデルとして使用) を用いたリフレーミング事例を示す。生成例は、モデル出力のソフトマックス分布からランダムに単語をサンプリングした場合の結果である。これらの結果からわかるように、提案モデルにおける、一部の生成例では、明らかにリフレーミングととれないような事例が生成されるなど、依然として改善の余地がある。例えば、「頑固」というネガティブ表現に対して、「こだわらない」という、意味的に反対の表現 (ポジティブな表現にもなっていない) が生成されている。また、「お金がたまると」というポジティブ表現に対して、「無駄遣い」という、元の文に対して矛盾するような事例も散見された。

表 3 N/P+P/N+Aug. モデル (Twitter) を用いた N/P 変換事例

入力ネガティブ表現	出力ポジティブ表現の候補 (5件)
うるさい	元気がいい, 活発な, 元気がある, 元気が良い, 明るい
頑固	意志が強い, 信念がある, こだわらない, 信念が強い, 頑固な
暗い	落ち着いている, 芯が強い, 落ち着いた, 情熱的な, 冷静
人をうらやむ	正義感が強い, 思いやりがある, 人を大切にしている, 向上心がある, 思いやりの強い
狡猾	頭脳明晰, 物おじしない, 判断力がある, 計算高い, 狡猾な

表 4 N/P+P/N+Aug. モデル (Twitter) を用いた P/N 変換事例

入力ポジティブ表現	出力ネガティブ表現の候補 (5件)
お金がたまる	無駄遣い, 無駄遣いが多い, 無駄遣いをする, 散財, 使いすぎ
礼儀を重んじる	かたくなるしい, 口が悪い, 調子に乗らない, 口下手, 口うるさい
斬新な発想が期待できる	ふざける, 考えが浅い, 話が聞けない, ふざけた, つまらない
感受性が豊かな	涙もろい, 泣き虫, 感情的な 傷つきやすい, 短気な
周りをよく観察している	気が散りやすい, 人付き合いが下手, 目立ちたがる, 人に流されやすい, 調子にのらない

6. おわりに

本研究では、ネガティブな表現をポジティブにリフレーミングする言語生成モデルを検討した。具体的には、まずネガティブな表現をポジティブにリフレーミングした事例を、ウェブや書籍など複数の情報源から人手で収集し、リフレーミング表現生成モデルを学習するためのパラレルコーパスを構築した。次に、収集したパラレルデータとリフレーミング事例を含む対話応答データを用いて、事前学習済みの大規模言語モデルをファインチューニングすることで、ネガティブな表現をポジティブに言い換えるリフレーミング表現生成モデルを学習した。評価実験では、データ拡張に基づいたモデルが最も高いリフレーミング表現の生成性能を持つことを自動評価により確認した。

今後の課題としては、生成されたりフレーミング表現の質についてより厳密に評価するために、Perplexity による自動評価だけではなく、人手による評価を行う必要がある。また、本研究で構築したパラレルコーパスのサイズは依然として小さいため、さらなる拡張が必要である。これには、ブーストラップアルゴリズムを用いた半教師あり学習によるリフレーミング事例の自動収集や、クラウドソーシング

などを活用した人手収集などを検討する。また、本研究で収集したりフレーミングの事例は、単語単位、フレーズ単位、文単位、とその変換の単位は統制がなされていない。また、リフレーミングという行為は、その動機や目的に応じて、複数のタイプに分類することができるが、本研究ではそれらを明示的に区別していない [3], [24]。したがって、対話システムを始めとした各種アプリケーションでの応用を考えると、リフレーミング表現生成モデルがどのような単位の入力を受け付け、またどのような単位の出力を生成するか、そしてどのようなタイプのリフレーミングをするかを、モデルが制御できるような枠組みを検討する必要がある。

謝辞

本研究は、JSPS 科研費 22K17958 の助成を受けた。

参考文献

- [1] Robson Jr, J. P. and Troutman-Jordan, M.: A concept analysis of cognitive reframing., *Journal of Theory Construction & Testing*, Vol. 18, No. 2 (2014).
- [2] 竹田葉留美: コラム: 出来事の視点を変えてポジティブに考える ~ リフレーミングを活用したストレスマネジメント ~, *情報の科学と技術*, Vol. 67, No. 3, pp. 121–122 (2017).
- [3] Ziems, C., Li, M., Zhang, A. and Yang, D.: Inducing Positive Perspectives with Text Reframing, *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 3682–3700 (2022).
- [4] Kajiwar, T. and Komachi, M.: Building a monolingual parallel corpus for text simplification using sentence similarity based on alignment between word embeddings, *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pp. 1147–1158 (2016).
- [5] 梶原智之, 小町守: 平易なコーパスを用いないテキスト平易化, *自然言語処理*, Vol. 25, No. 2, pp. 223–249 (2018).
- [6] Jin, D., Jin, Z., Hu, Z., Vechtomova, O. and Mihalcea, R.: Deep learning for text style transfer: A survey, *Computational Linguistics*, Vol. 48, No. 1, pp. 155–205 (2022).
- [7] Tran, Q. H., Zukerman, I. and Haffari, G.: A hierarchical neural model for learning sequences of dialogue acts, *Proc. of ACL*, Vol. 1, pp. 428–437 (2017).
- [8] Samanta, B., Agarwal, M. and Ganguly, N.: Fine-grained Sentiment Controlled Text Generation, *arXiv preprint arXiv:2006.09891* (2020).
- [9] Zhou, H., Huang, M., Zhang, T., Zhu, X. and Liu, B.: Emotional chatting machine: Emotional conversation generation with internal and external memory, *Proc. of AAAI* (2018).
- [10] Huang, C., Zaiane, O., Trabelsi, A. and Dziri, N.: Automatic dialogue generation with expressed emotions, *Proc. of NAACL-HLT*, Vol. 2, pp. 49–54 (2018).
- [11] Fu, L., Fussell, S. and Danescu-Niculescu-Mizil, C.: Facilitating the Communication of Politeness through Fine-Grained Paraphrasing, *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Pro-*

- cessing (*EMNLP*), pp. 5127–5140 (2020).
- [12] Zhu, M., Yu, Z. and Wan, X.: A Neural Approach to Irony Generation, *arXiv preprint arXiv:1909.06200* (2019).
 - [13] Joshi, A., Bhattacharyya, P. and Carman, M. J.: Sarcasm generation, *Investigations in Computational Sarcasm*, Springer, pp. 119–127 (2018).
 - [14] Kenton, J. D. M.-W. C. and Toutanova, L. K.: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, *Proceedings of NAACL-HLT*, pp. 4171–4186 (2019).
 - [15] Ruder, S., Peters, M. E., Swayamdipta, S. and Wolf, T.: Transfer learning in natural language processing, *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: Tutorials*, pp. 15–18 (2019).
 - [16] Yoshino, K., Mori, S. and Kawahara, T.: Incorporating semantic information to selection of web texts for language model of spoken dialogue system, *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, IEEE, pp. 8252–8256 (2013).
 - [17] Csáky, R., Purgai, P. and Recski, G.: Improving Neural Conversational Models with Entropy-Based Data Filtering, *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pp. 5650–5669 (2019).
 - [18] Akama, R., Yokoi, S., Suzuki, J. and Inui, K.: Filtering Noisy Dialogue Corpora by Connectivity and Content Relatedness, *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 941–958 (2020).
 - [19] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L. and Polosukhin, I.: Attention is all you need, *Advances in neural information processing systems*, Vol. 30 (2017).
 - [20] Sugiyama, H., Mizukami, M., Arimoto, T., Narimatsu, H., Chiba, Y., Nakajima, H. and Meguro, T.: Empirical Analysis of Training Strategies of Transformer-based Japanese Chat Systems, *arXiv preprint arXiv:2109.05217* (2021).
 - [21] Yoshino, K., Ikeuchi, K., Sudoh, K. and Nakamura, S.: Improving Spoken Language Understanding by Wisdom of Crowds, *Proceedings of the 28th International Conference on Computational Linguistics*, pp. 2606–2612 (2020).
 - [22] Li, J., Monroe, W., Ritter, A., Jurafsky, D., Galley, M. and Gao, J.: Deep Reinforcement Learning for Dialogue Generation, *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pp. 1192–1202 (2016).
 - [23] Li, J., Galley, M., Brockett, C., Gao, J. and Dolan, B.: A Diversity-Promoting Objective Function for Neural Conversation Models, *Proceedings of NAACL-HLT*, pp. 110–119 (2016).
 - [24] 兼折友美子, 畦地博子ほか: 困難事例に対応する看護師のリフレーミングを促す技術, *高知女子大学看護学会誌*, Vol. 39, No. 1, pp. 43–50 (2013).