

人間は、日常的に様々な機会情報を文章化し表現している。日常生活においては、他人と一緒に料理や掃除をするときは自分が依頼したい作業内容を文章化し相手に伝えている。仕事をする上で、複雑な手順を伴う作業を複数人に伝達する必要があるときは、作業手順ごとに行う動作を文章化することで、多くの人に複雑な作業を正確に伝達することができる。人間の持つ文生成の能力は、自分以外の相手と円滑に生活したり、仕事を進めていくうえで必要不可欠な能力といえる。

機械に人間同様の文生成の能力を実装することができれば、我々の生活や仕事を大幅に効率化できると考えられる。例えば、日常生活における料理や掃除の作業について、普段の作業内容を自動で文章化し記録しておくことができれば相手への情報伝達の時間を短縮できる。また、仕事で複雑な手順を伴う作業について、事前に作業内容を撮影し、これを自動で文章化できればマニュアル作成の手間が省けると共に、無意識に行っている作業内容も自動記録されることでより正確に作業を伝達することができる。さらには、目の不自由な人が周囲の環境を把握するために、機械が周囲の環境について説明する文を自動生成することができれば、目の不自由な人がより安心して生活を送れるようになる。人間の伝えたい内容を自動的に文章化する機能は、我々人間の生活や仕事を効率化することができたり、人間の能力拡張に有効活用できると考えられる。

日常の生活や仕事の効率化や、人間の能力拡張に向けて、これまで数多くの文生成に関するタスクが設定され、文生成手法の研究が為されてきた。例えば、画像キャプションタスクは画像中の情景や活動を説明する文を生成するタスクである。画像キャプションタスクのデータセット MSCOCO は、同じ画像に対して 5 つの正解文が付与されている。正解文から画像中の同じ人物に対して異なる単語や語句を使って説明していることが分かる。例えば、同じ画像中の“二人の女性”に対して“a woman and another woman”や“two women”という表現で説明する。同じ画像中の“人々”に対して“dozens of people”や“a bunch of people”という表現で説明する。また、それぞれの単語は適切な順番で並び、人間が理解可能な文を構築している。例えば、正解文“People are walking across the street at an intersection.”は、文頭から“主語 (People)”→“述語 (are walking)”→“副詞句 (across the street at an intersection)”という順番で並び、人間が定めた文法構造規則に従った文を構築している。

また、文頭から“動作主格 (People)”→“動詞 (are walking)”→“対象格 (across the street)”→“場所格 (at an intersection)”という順番で並び、人間が定めた意味構造的規則に従った文を構築している。上述の通り、画像に対して、正しくかつ柔軟に単語や語句を選択しながら、人間が定めた構造規則に従って単語や語句を並べることで文を生成している。また、画像の視覚的情報を説明するタスクの他にも、長い文書を要約する文書要約タスク、質問文に対して適切な回答文を生成する質問応答タスク、自然な会話文を生成する対話生成タスクなど様々な文生成に関するタスクがある。文書要約タスクでは、文書中の重要な情報を説明する単語を正確にかつ柔軟に選択し、人間が定めた規則に従って選択した単語を適切な語順に並べ、正しい構造を持つ文を生成する必要がある。質問応答タスクや対話生成タ

スクも同様に、回答文や会話文中で使用され得る単語を正確かつ柔軟に選択し、人間が定めた規則に従って選択した単語を適切な語順に並べ、正しい構造を持つ文を生成する必要がある。上述の通り、人間同様の文生成の能力を機械に実装するために、タスクや状況に応じて正確かつ柔軟に単語や語句を選択し、人間の定めた文法構造的規則や意味構造的規則に基づいた正しい構造の文を生成する必要がある。

しかしながら正確かつ柔軟な単語選択を実現しつつ、人間の定めた規則に基づいた正しい構造の文を生成することは難しい。例えば、テンプレートベースの手法は、予め定めたテンプレート（構造）に従いながら単語選択を行うため正しい構造の文を生成可能だが、柔軟な単語や語句の選択はできない。大量の文書データから尤もらしい単語順を学習し文を生成する **ngram** ベースの手法は、連続する単語順のみを考慮しており、離れた単語同士の関係を考慮できないことから適切な構造をもつ文の生成が困難である。近年急速な発展を遂げているニューラルネットワークベースの手法は、連続する単語順のみならず離れた単語同士の関係性を学習する **Transformer** と呼ばれる手法が提案されている。離れた単語同士の関係性を学習することで、尤もらしい文の構造規則を暗に学習することができ、適切な構造をもつ文の生成が可能である。しかしながら大量の文書データを学習したニューラルネットワークは、データ内で頻出する単語を使った文の確率を高くする傾向がある。頻出単語が偏って選択され、柔軟な単語や語句の選択は困難である。上述の通り、正確かつ柔軟に単語や語句を選択しながら、正しい構造をもつ文の生成を実現することは困難である。

本論文では、人間同様の文生成の能力を機械に実装するために、正確かつ柔軟に単語や語句を選択し、人間が定めた構造的規則に基づき正しい構造を持つ文の生成を実現する方法を検討する。具体的には、文の構造的規則として、文法構造的規則と意味構造的規則の二種類の規則の活用方法についてそれぞれ検討し、各規則の文生成への有効性を検証する。文法構造的規則として文脈自由文法を採用し、文脈自由文法を適用して構築される構文木を生成する。適切な構文木の探索アルゴリズムとして広大な探索範囲を効率的に探索可能なモンテカルロ木探索アルゴリズムを使用する手法を提案する。タスクの設定としては、文の主要な要素となる重要単語のセットを入力として、それらの単語を使用しながら、正確かつ柔軟な単語や語句の選択と、正しい文法構造を備えた文の生成を目指す。次に、意味構造的規則として格構造ラベルを採用し、文生成時に格構造ラベルを条件として与え、与えられた格構造ラベルに対応する単語を選択しつつ、適切な格構造ラベルの順番を同時に推定する手法を提案する。タスクの設定としては、画像を入力として、画像に対応する正しくかつ柔軟な単語や語句を選択しつつ、画像を説明する文として正しい意味構造をもつ文の生成を目指す。画像の特徴抽出と、適切な格構造ラベルの順番推定と、格構造ラベルに対応する正しくかつ柔軟な単語推定とを同時に行う **end-to-end** 構造の **Neural Network** を提案する。さらに、**Neural Network** が適切な文の意味構造を学習するために、離れた単語同士の関係性や格構造ラベル同士の関係性を学習可能な **Transformer** ベースの画像キャプション手法をベースラインとして使用する。各手法の実験では、文法構造的規則や意味構造的規則の文生成への有効性を検証する。