

平成24年度 情報工学コース卒業研究報告要旨

石川 研究室	氏 名	田 中 麻 祐 子
卒業研究題目	推敲支援のための日本語文の語順整序	

日本語は語順が比較的自由であるため、語順を強く意識しなくても、意味の通じる文を書くことができる。しかし、語順に関する制約がないわけではないため、文法的には間違っていないものの読みにくい文が生成されることがある。例えば、「彼が私が気になっていたPCの不具合を直してくれた。」という文は、そのままの語順では読みにくい。長い修飾句を持つ文節は前方に位置する傾向が強いことを考慮し、「私が気になっていたPCの不具合を彼が直してくれた。」と語順を変更できれば自然に読むことができるようになる。

そこで本論文では、日本語文に対して、より読みやすくなるように文節を並び替える手法を提案する。これまでも語順整序を行う手法が幾つか提案されている。これらの多くは、前処理として係り受け解析を施し、その結果を用いて語順整序を行うことを想定しており、入力文が読みにくい語順である場合に係り受け解析の精度が低下することを考慮していない。一方、本手法では、係り受け解析と語順整序を同時に行う。係り受けと語順の適切さを同時に考慮することにより、読みやすい語順を同定することができる。具体的には、入力文に対して考えられる語順と係り受け構造のあらゆるパターンの中から、最適なパターンを確率モデルを用いて決定する。その際、考えられるパターンの数は膨大であるため、効率的な解の探索が必要となるが、本手法では、以下の3つの前提から、従来の係り受け解析で利用されてきた動的計画法を改良し、近似解を効率的に探索する。

1. 入力文の係り受け構造は日本語の構文的制約(後方修飾性, 非交差性)を満たす
2. 語順整序前と後の係り受け構造は同一である
3. 語順整序後の語順でも日本語の構文的制約を満たす係り受け構造となる

京都大学テキストコーパスに収録されている新聞記事を用いて評価実験を行った。本実験では、コーパス中の係り受け情報を利用し、文意は取れるものの読みにくい文を擬似的に作成し(図1参照)、それをテストデータとして用いた。比較のために、係り受け解析を行った後に語順整序を行う従来手法を用意した。実験の結果、本手法の語順一致率は、2文節単位の一致率(2文節ずつ取り上げた時の文節の順序関係が元の文のそれと一致しているものの割合)が76.1%(29,541/38,838)、文単位の一致率が22.5%(195/865)であった。図2は、本手法による語順整序の結果、文全体の語順が元の文と完全に一致した例である。本手法は従来手法を上回る語順一致率を達成しており、本手法の有効性を確認した。

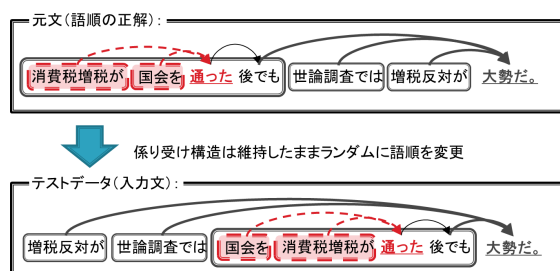


図1: テストデータの作成例

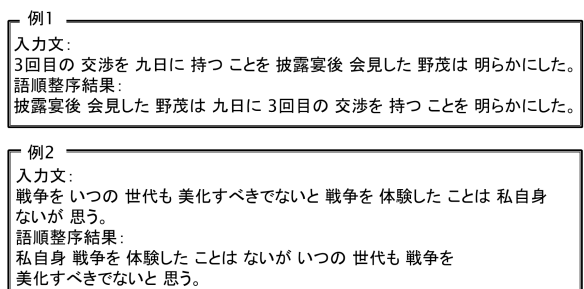


図2: 本手法による語順整序結果の例