

# 文の構造を可視化した翻訳後編集インターフェース

岸本 裕大<sup>†</sup> 中澤 敏明<sup>‡</sup> 河原 大輔<sup>†</sup> 黒橋 禎夫<sup>†</sup>

<sup>†</sup>京都大学 大学院情報学研究科      <sup>‡</sup>科学技術振興機構

## 1 はじめに

グローバル化が進んだ社会において、言語の壁を乗り越えるために翻訳がますます重要になってきている。従来、高品質な翻訳を必要とする際には人手による翻訳が行われてきたが、近年の機械翻訳の精度向上に伴い、機械翻訳結果を下訳として利用し、これを人間が後編集するという方法が注目されている。実際、機械翻訳の精度がある程度高いヨーロッパ言語間では後編集が良く使われており、後編集を行うためのインターフェースの開発や後編集を行う専門家を育成する教育が行われている [1][2]。

しかしながら、日英や日中といった語順が大きく異なる言語間では機械翻訳精度が低いため、誤った機械翻訳結果を読む時間や翻訳断片の取舍選択に時間を費やさかえって作業効率下がってしまい、現時点では後編集は現実的ではない。我々は構文解析に基づいた機械翻訳を利用し、「入力文解析結果」、「入力文語順の翻訳結果」、「翻訳結果」の3段階で文を可視化することで機械翻訳が「何をおこなったか」を一目で理解することができる後編集インターフェースを提案する。

## 2 関連研究

ヨーロッパ言語に対応した後編集インターフェースは Aziz ら [1] や Schwartz ら [2] などが開発している。Aziz らのインターフェースは特定の機械翻訳システムに依存しないように設計されている。Schwartz らは、目的言語のみしか話せない人でも後編集を適切に行うことができるインターフェースを開発している。しかしながら日英のような語順が大きく入れ替わる言語ペアを考慮したシステムは無い。

文構造や文章全体の構造の可視化に関する関連研究としては Hengbin ら [3] の研究がある。文構造が一目

### Post-Editing User Interface Using Visualization of a Sentence Structure

Yudai KISHIMOTO<sup>†</sup>, Toshiaki NAKAZAWA<sup>‡</sup>,  
Daisuke KAWAHARA<sup>†</sup> and Sadao KUROHASHI<sup>†</sup>

<sup>†</sup>Graduate School of Informatics, Kyoto University  
Kyoto 606-8501

<sup>‡</sup>Japan Science and Technology Agency  
Kawaguchi-shi, Saitama 332-0012

<sup>†</sup>kishimoto@nlp.ist.i.kyoto-u.ac.jp {dk,kuro}@i.kyoto-u.ac.jp

<sup>‡</sup>nakazawa@pa.jst.jp

でわかるように構文木を用いて可視化を行っているほか、各単語の品詞情報や文法的役割も表示している。本研究は、構文木の表示はせず文を装飾することで文構造を可視化している点で Hengbin らと異なっている。

## 3 文構造の可視化を用いた後編集の効率化

後編集を行う際に語順が大きく異なると、機械翻訳によって「何がおこったのか」を把握することが難しいという問題が発生する。この問題を解決するために、本論文では構文解析に基づく機械翻訳を用いた上で、入力文の文構造を可視化した「入力文解析結果」、入力文の語順を維持した状態で機械翻訳を行った「入力文語順の翻訳結果」、そして最終的な機械翻訳結果である「翻訳結果」の3段階で文の表示を行う。これによって翻訳作業者が効率的に翻訳の質を把握することができ、更に翻訳結果を以下のようにさまざまなレベルで柔軟に利用できるようになるので、翻訳作業時間の短縮につながると期待される。

- 1) 「入力文解析結果」の時点で誤りを発見した場合や「翻訳結果」の内容が明らかに破綻していると判断できる場合、機械翻訳を利用することを諦めて一から翻訳作業者が翻訳を行う。
- 2) 語順が大きく異なっても翻訳結果が部分的に利用可能である場合、「入力文語順の翻訳結果」を下訳として利用し、語順を入れ替える。
- 3) 語順の入れ替えも含めて翻訳結果が妥当な場合、「翻訳結果」を下訳として若干の修正を行う。

また翻訳作業者にとって木構造の表示は馴染みが少ないことが多く、読む時間がかかってしまい作業効率下がるとい問題がある。そこで木構造を使わず、直感的に文構造が分かるように、ボックスやフォントサイズ、文字色を利用することによって、翻訳の質をより効率良く把握できるように文を可視化する。

## 4 翻訳後編集インターフェースの実装

本論文で提案するインターフェースの入力は、入力文と機械翻訳結果の品詞・構文解析結果、および入力文と

## 入力文解析結果



## 入力文語順の翻訳結果



## 翻訳結果



i

next, the present state of production and sale of agriculture and forestry are clarified and the promotion of local industry, and development of a specialty is described.

編集結果を送信

図 1: 翻訳後編集インターフェース

機械翻訳結果の単語の対応が記されたリスト(翻訳マッピング)が記載されているテキストファイルであり、インターフェースが処理した結果は JavaScript ファイルで出力される。この JavaScript ファイルを Web ブラウザ上で図 1 のように表示できる。

本インターフェースでは、文全体を「節構造を示すボックス」、「補文や連体修飾等の埋め込み文を示すボックス」、「文節や句を示すボックス」の3つのボックス単位に分割して表示することで文構造を可視化している。文全体の head や節構造・埋め込み文の head といった文中において重要な語についてはフォントサイズや文字色を変更している。これらのシンプルな方法によって文構造が一目で理解できると同時に、「入力文解析結果」内での誤りを容易に発見することも可能だと考える。なおこの際、品詞と係り受け関係を利用して文構造の可視化を行っている。したがって、本インターフェースは品詞と係り受け関係を付与できる言語であればいかなる言語にも適用することが可能である。

「入力文解析結果」「入力文語順の翻訳結果」「翻訳結果」の3段階で文を可視化する際、図 1 では“翻訳結果”の文を2つの形式で表示している。まず、上の文については「文節や句を示すボックス」等を利用した形式で表示している。このボックスは翻訳者の手で語順を入れ替えることを考慮し、内容を編集することが可能であると同時に、ドラッグアンドドロップで各ボックスの順序を入れ替えることができる。下の文については、ボックスを利用した形式では文の細部修正が難しいことを考慮して、テキスト形式で表示している。

また、「入力文解析結果」と「入力文語順の翻訳結果」の2つを比較する際、単語・句レベルでの翻訳結果が正しいかどうかを判断することが最も重要である。そこで本インターフェースでは「入力文解析結果」と「入力文語順の翻訳結果」のボックスサイズを調整することで、上下に並んだ時に対応している単語同士が同じ

位置にくるように調整している。またマウスポインタを句ボックスに合わせると対応している単語をハイライトするようにも設定している。

## 5 おわりに

本論文では、語順が大きく入れ替わる言語間における文の構造を可視化した翻訳後編集インターフェースを提案した。本インターフェースを利用することによって効率的に翻訳結果の質を把握し、柔軟に利用できるようになり、翻訳作業時間の短縮につながると期待される。今後の課題としては、本インターフェースの有効性について調査することと、品詞や係り受け関係を付与できない言語への拡張方法の検討が挙げられる。

## 参考文献

- [1] Wilker Aziz, Sheila C. M. De Sousa, and Lucia Specia. Pet: a tool for post-editing and assessing machine translation. In *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, pp. 3982–3987, 2012.
- [2] Lane Schwartz, Timothy Anderson, Jeremy Gwinup, and Katherine Young. Machine translation and monolingual postediting: The AFRL WMT-14 system. In *Proceedings of the Ninth Workshop on Statistical Machine Translation*, pp. 186–194, 2014.
- [3] Hengbin Yan and Jonathan Webster. Collaborative annotation and visualization of functional and discourse structures. In *Proceedings of the 24th Conference on Computational Linguistics and Speech Processing (ROCLING 2012)*, pp. 366–374, 2012.