

対訳表現候補を手がかりとした用例翻訳手法の提案

大山 鉄郎

近年では、情報技術の発展により、世界中の情報を容易に手に入れることが可能になったが、言語の壁という問題は依然としてある。言語の壁を失くすための技術として、人手を用いずに言語を機械的に変換する、機械翻訳と呼ばれる技術がある。本研究では、機械翻訳の方式の一つである用例翻訳と呼ばれる方式を取り扱う。用例翻訳は、あらかじめ異なる言語で書かれた文のペアを用例として登録しておき、入力文と類似した用例をもとに翻訳を行う形式である。

用例翻訳では、入力文に対して適切な用例を選択して用いることで翻訳が行われる。このとき、翻訳に適している用例を選択するための手がかりとして、入力文と用例の文字列の一致や、文中に含まれる語の概念が近いものかを示す類似度など、様々な尺度が用いられる。

本研究では、用例を選択する際の手がかりとして、対訳表現候補を用いた用例翻訳手法の提案を行った。例として、入力文“出版社について教えてください。”に対する用例を考える。

入力文と用例の文字列の一致で用例を検索すると、“教えてください。”が一致した次の2つの用例候補が得られるとする。

1. ステップを教えてください。 — Teach me the steps, please.
2. ルールを教えてください。 — Please tell me the rules.

このとき、“教えてください。”の対訳表現候補として、“Teach me ~ please”と“Please tell me ~”が得られる。複数の対訳表現候補からの絞り込みには、意味距離と呼ばれる、動詞にかかる目的語の概念の近さが良く用いられる。しかし、例では“出版社”に対して“ステップ”と“ルール”が対応しており、2語とも概念的に遠すぎて比較が難しい。

提案手法では、得られた対訳表現候補で用例を検索し直すことで、“製造社を教えてください。”や“メールアドレスを教えてください。”などの用例を追加で取得する。これらの用例は、対訳表現候補で検索、選択されているため、入力文の表層表現とは一致しないものが取得できる。次に、入力文と用例の意味距離を計算し、対訳表現候補別に意味距離の総和を取る。提案手法によって、意味距離の計算に“製造社”、“メールアドレス”などが追加され、意味距離の差が明確になる。例では、“出版社”と“製造社”の概念が近いため、“製造社を教えてください。”の“教えて”の訳語(“Please tell me ~”)が選択される。仮に、近い概念がなかった場合も、意味距離の総和を取るため、多く出現した対訳表現候補が優先して選択されるようになっている。

提案手法は、高度言語情報融合フォーラム ALAGIN で公開されている日英翻訳エンジン学習・評価用対訳コーパスを用いた実験の結果、既存の意味距離を用いた手法と比較して、翻訳精度 BLEU が平均 6.5% 程度向上することを確認した。

今後の課題として、多義性を持つ対訳表現候補で用例を検索すると、誤った用例を追加してしまうことがあるため、対訳表現候補が用例検索に適切か判定する処理の実装が挙げられる。

(指導教員 関 洋平)