

[国語古文]

### 3 機械翻訳と表層による類似度計算を用いた古文問題解答



横野 光 ((株) 富士通研究所)

センター試験国語の古文問題は、本文として古典作品からの一部が提示され、それに関する文法問題や傍線部現代語訳問題、内容理解に関する問題が出題される。本稿では内容理解問題解答について述べる。

本文として取り上げられる作品の多くは物語であり、内容理解問題に関しては現代文の小説問題と同様に、登場人物の心情などを問う問題が出題される。古文問題は古文が読めるかどうかが主に問われるため、現代文問題のように深い内容の理解はそれほど求められず、解答に該当する個所を現代語に訳すことができれば、表層的な手がかりを元に解答することができると思われる。

内容理解問題解答ではまず本文を統計的機械翻訳器 Moses<sup>☆1</sup>によって現代語に翻訳する。翻訳モデルの学習に用いたコーパスは『日本古典文学全集』(小学館)の電子化データである。翻訳例を図-1に示す。統計的機械翻訳の精度は学習コーパスのサイズに依存する。ソルバーで用いている対訳コーパスは日英翻訳などに用いられるコーパスに比べて規模が小さく、内容語に関する誤訳や訳しきれなかった語が多かった。その一方で、助動詞のような機能語に関しては比較的正しく翻訳できている。

この翻訳した本文と問題中の各選択肢とを比較して類似度を計算する。類似度の計算では正解の選択肢に含まれる情報は問題の傍線部の周辺に存在し、その選択肢に対応する本文の個所は互いに近い位置にあると仮定している。選択肢の各節に対して本文

夜もいたく更けたるに、初秋風いと涼しく吹き出でて、宵月夜のころなれば、月は西の山の方にうち傾きながら、影のさやけくさしわたるに、しばし立ちやすらひて、

夜もたいそう更けているのに、そうしたら風がまことに涼しい吹き出でて、宵の月のころなので、月が西の山の方に、首をかしげながら、影がきらめいさしてきたのに、しばらく立ってかたでさえて、

図-1 翻訳結果(上:本文,下:現代語訳)

中の節との形態素単位でのコサイン類似度を計算し、最も類似度が高い節を選択肢の節に対応する本文の個所とする。選択肢の各節に対応する本文中の個所間の距離を求め、その逆数を選択肢の本文に対する類似度とし、最も高い類似度の選択肢を解として出力する。

2016年度模試タスクにおける結果は、内容理解問題4問中2問正解であり、本文の現代語訳を人手で修正したものを使用した場合の正答数は3問であった。ソルバーでは形態素の表層一致で類似度を計算しているが、翻訳によっては類義語で訳される場合があるため単語間の意味的な類似性を考慮する必要がある。また、問題解答という面からは、本文との内容の近さを測るだけでなく“本文にない記述を含むかどうか”といった不正解の選択肢を見つけてというアプローチも必要である。

(2017年3月30日受付)

■横野 光 (正会員) yokono.hikaru@jp.fujitsu.com

東京工業大学研究員、国立情報学研究所特任助教を経て現在(株)富士通研究所所属。専門は計算言語学、自然言語処理。博士(工学)。

☆1 <http://www.statmt.org/moses/>