

K-POP の韓日翻訳支援を目的とした対訳辞書自動作成

樋口 桜

今日、日本で起こっている韓流ブーム・K-POP ブームによって韓国語で歌われている K-POP 音楽に触れる機会が増えており、その人気に伴い、韓国語の K-POP 曲を日本語に翻訳する需要が高まっている。現在 K-POP の日本語訳を掲載している Web サイトは存在しているがその数は多いとは言えない。また、日本語訳されている曲は人気や知名度のあるアーティスト・曲が中心であるため、自分の知りたい曲が日本語訳されているとは限らない。しかし、韓国語を学習したことのない人にとって口語的な表現や新語など特有の表現が含まれている歌詞を自分で翻訳をすることは容易ではない。また、既存の辞書には歌詞特有の表現のような歌詞中の出現頻度の低い語が載っている割合は低く、新語などを翻訳することは困難である。

そこで本研究では、既存の手法では抽出困難であった歌詞中の出現頻度の低い語を抽出するため、既存手法より高い精度で低頻度語の対訳表現を抽出する手法を提案する。本研究では K-POP1,280 曲の韓国語歌詞と日本語訳歌詞の対訳文 49,367 ペアをデータとして用いた。それら対訳文を形態素解析ソフト MeCab を用いて形態素解析し、その結果をもとに韓国語、日本語の各単語の出現回数及び同じ対訳文内に同時に出現する回数を数える。さらに得られた結果から韓国語単語と日本語単語間の類似度を計算し、類似度の高いペアから順に対訳表現として抽出を行った。既存手法の尺度では、韓国語と日本語の共起頻度が 1 の場合に類似度がすべて 0 になってしまうという問題点があったが、本研究ではそれを改善し共起頻度 1 の場合でも類似度が算出できるようにした。また、提案手法の有効性を検証するために、既存の尺度を用いた 3 手法を再現し、適合率、低頻度語の訳語対抽出順位の比較を行った。

結果、共起頻度 1 の訳語対抽出においては既存手法を上回る結果となり、共起頻度 1 の訳語については提案手法の有効性を示すことができた。さらに高い精度で対訳表現の抽出を行うために、既存の手法を組み合わせるなどして、高頻度語から低頻度語まで網羅的に抽出できる有用なものにしていくことが今後の課題である。

(指導教員 辻慶太)