

中国語と日本語の同形異義語の 意味の違いの可視化

お茶の水女子大学

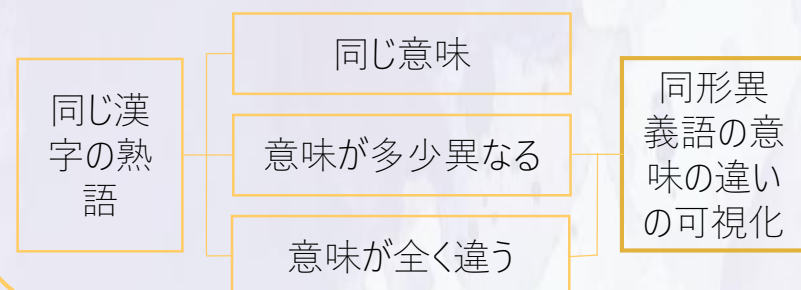
HAN HE 伊藤 貴之

研究背景

中国語も日本語も漢字が多く、全く同じ漢字を使った言葉もたくさんある。例えば、“検討”、“質問”、“連想”のような言葉。

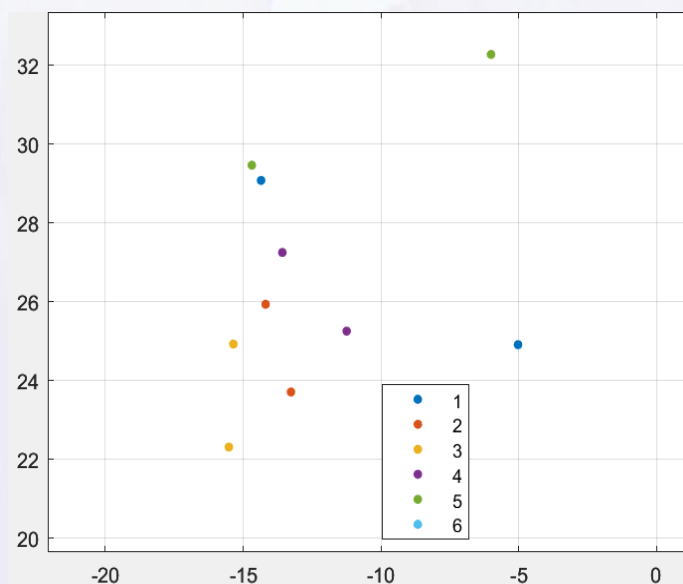
しかし、本当に同じ意味を伝えているだろうか？品詞の分類も同じだろうか？

単語	中国語の意味	日本語の意味
大丈夫	一人前の男	問題ない
得体	言動がその場にふさわしいこと	本当の性質や姿勢
汽車	エンジンによって路上を走る車	蒸気動力によって動く鉄道車両
前年	昨年の前年	ある年の前年
入口	口の中に葉入れること	ものごとの最初の段階



研究方法

1. 中国語でも日本語でも使われる熟語を探す
2. 中国語の意味と日本語の意味を両方英語で記述する
3. SimCSE対比学習法を用いて「文埋め込み」を算出する
4. 3.で計算した2つのベクトルの差分を計算する
5. 1.-4.をいろんな単語について反復し、何種類のパターンがあるかを列挙する



5つの例に対するの可視化結果

関連研究

・中国語と日本語の比較の研究

中国語と日本語の違いについては、漢字の発音、漢字の意味、数詞の違いなど、多くの研究がなされている。

・自然言語処理と「文埋め込み」などの研究

「文埋め込み」方法開発の歴史はわずか6年間で非常に多様な方法が生み出された。例えば、単語ベクトル加重平均法、BERT法、SimCSE対比学習法。

・中国語と日本語の同形異義語の意味の違いの可視化に関する研究はまだ見つかっていない。

今後の課題

・同形異義語の自動検索

- 同形異義語よりも同じ意味の熟語のほうが多い。
- 同形異義語を自動抽出する方法を検討する。

・翻訳

- 中国語と日本語で意味を表すのなら、できるだけ翻訳の表現を揃えたり統一したい。
- 中国語と日本語の両方を英語に翻訳して、両言語を公平に扱うことを検討する。

・「文埋め込み」の組み合わせ

- 多くの単語は複数の意味を有するのに対して、本研究ではその中の1つの意味しか用いていない。