

古文は日本語といえるか？

塩原健太郎 23B00611

東京工業大学理学院

はじめに

古文は日本語と呼べるのかについて、すなわち古文と現代の日本語がどれだけ乖離しているのかを調べる。

方法

各時代ごとの文献に対してテキストマイニングを行い、登場する語彙を比較する。全語彙についての比較や、頻出する語彙についての比較などから、時代によって言語が移り変わる様子を数値に落とし込む。

結果

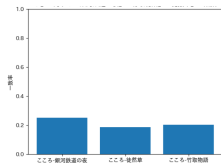


図 1: 全語彙の一致率比較

確かに『銀河鉄道の夜』が最も『こころ』に一致しているが、相関としては弱そうである。

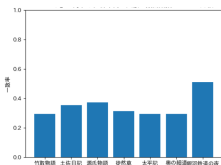


図 2: 頻出語彙 TOP50 の一致率比較

助動詞や助詞なども含めたうえで、頻出語彙に絞って比較したところ、『こころ』に対して年代の比較的近い『銀河鉄道の夜』が他の文献に比べて 1.7 倍ほどの一致率を記録した。この頻出語彙は、助詞や助動詞や、基礎的な動詞などが多くを占めており、文法などの運用の近さが表れていそうである。

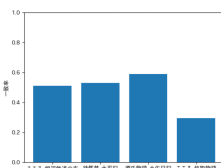


図 3: 年代の近い文献同士の頻出語彙比較

『こころ』と『銀河鉄道の夜』の例だけでなく、『徒然草』と『太平記』、『源氏物語』と『土佐日記』など、年代が近い文献同士の一貫率は高くなる傾向にあることがわかる。

考察

現代から遡って文献の頻出語彙の比較をしていて意外だったのは、時代が離れかたに比例して一致度合いも下がっていくだろうと予想していたのに、同時代の文献だけ一致率が跳ね上がり、他の時代についてはある程度の値で停滞していたことだ。

鎌倉 室町に流行った表現

こそぞ也候

特に太平記や徒然草などについては、平安の貴族言葉に上記のような武家らしい言葉使いなどが入り込んだことによって、むしろ一致率が一時的に下がったのではないかと考えた。

全文を通して常に良く使われた語句

てにをはのと

上記のような、文法を為すうえでの根幹となるような語句については時代を問わず一貫するものがあつたこともわかった。一見した際の印象は古文と現代文で全く違うように思うが、根本的には大きな変化は起きていないのではないかと考えた。

おわりに

古文は日本語と呼べるのかについて、頻出語彙の一致度を測ることによって確かめた。最終的には、時代によって流行り廃りはあるが、日本語としての根幹の部分は平安から現代までを通して連続性を保っている、すなわち古文も日本語と呼んで全く問題がないというように結論付けた。

参考文献

高橋雄一. "現代日本語の機能語のリスト作成について: 現古文対照辞書の作成に関連して." 専修大学人文科学研究月報 285 (2017): 1-17.