

データサイエンスに関する一考察

～KH Coder を用いた、テキスト型データの分析をめぐって～

1. テーマ設定の理由

数理科の特別授業で、データサイエンスについて知る機会があり、データサイエンスは、これからの社会で求められていることを知った。そこで、テキストマイニングツールの一つである KH Coder を用いてテキスト型データについて考察しようと考えた。

2. 実験について

(1) KH Coder とは

内容分析、テキストマイニングのためのソフト。テキストデータ中にどのような言葉が多く出現していたか、また、どの言葉とどの言葉とがともに出現しているかが一目でわかるのがこのソフトの利点である。

(2) 各テキスト型データの分析と意図

①「山月記」の単語分解

現代文の時間に学習した身近な短編小説であり、場面による言葉の変化が大きいと考えた。

②「大日本帝国憲法」と「日本国憲法」の比較

一般に、この2つの憲法は内容が異なっていると言われている。どこが違うのか、また、似ているところはあるのか、単語を比較して調べようと考えた。

③都道府県魅力度ランキングと広報誌の関係性

近年、テレビやメディアで話題となっているが、集計方法は解らない。そこで、広報誌で使われる語彙から調べようと考えた。

④太宰治作品の分析

前期・中期・後期で作風が全く異なる。それぞれの時期に使われた言葉から、作風の違いを調べようと考えた。

3. 各テキスト型データ種類別の結果と分析

(1)「山月記」の単語分解

総数 4074 文字、李徴の語り 2430 文字（約 60%を占める）

結果及び考察：①「己」が物語の中で最も多く使われた言葉

李徴が自分自身のことを指し、虎としての感情を出す場面に出てくる。

→本文中の虎となった李徴の語り長い事情と一致している。

②比較的多かった「人間」と「虎」

人間から虎になってしまった李徴の苦悩を中心に描いた物語の特徴をよく表していると考えられる。

(2) 「大日本帝国憲法」と「日本国憲法」の比較

結果及び考察：①「大日本帝国憲法」…天皇主権

「日本国憲法」…国民主権

他国との関係、平和を重視

②どちらも議会についての記載が多い

→どちらも国会に重きを置いている。

③「国会」「裁判所」「内閣」「権利」が多く出てきた

→日本国憲法で特徴的なのは、三権分立についての記載が多いこと。

(3) 都道府県魅力度ランキングと広報誌の関係性

結果及び考察

県ごとの大きな違いは見られなかった。広報誌は県民向けであり、生活する上での情報提供を目的としていて、通常それらの情報にはあまり大きな違いがみられないことが明らかになった。

(4) 太宰治作品の分析

前期（1925 年～1938 年、54 作品）

中期（1938 年～1947 年、74 作品）

後期（1974 年～1977 年、50 作品）

結果からの考察

前期：「医者」「顔」「白い」「金」が文章中にまとまって出てくることが多い

→当時の格差社会や世の中への批判ととれる。

中期：「いま」「笑う」「生きる」などがやや多い

→明るい単語が前期に比べ増えており、作風が明るくなったのは、希望や活力が多くなったためとも考えられる。

後期：「自分」「無い」「女」「酒」などの単語が多い

→作品中の単語から太宰の堕落していく人生と重なる部分があるともとれる。

4. 全体の考察

KH Coder をもちいたこの方法では、言葉には複数の解釈があり、すべてを分析的にとらえることは不可能だといえる。しかし、単語に過剰な意味や意義を加えることがなく、分析者の主観が入る余地のないことから、文章の大まかな傾向がとらえられる点はメリットだとも考えられる。

また、小説の分析では、研究者による分析が一般的であったが、文章を構成する要素の数値化により、定量的な分析が可能になったとも言える。

しかし、小説は読み手が人間であることを想定して書かれているものであり、その分析には表現から受ける感覚等からの分析が必要不可欠である。

このことから、人の受ける印象に基づく分析とテキストマイニングによる分析は必ずしも 2 者択一とは限らず、双方の特性を活かした相互補完的な活用を目指すことが理想だと考える。

5. 今後の課題

- ・テキストマイニングの特性をいかすことができるデータを探し、分析していくことが必要である。
- ・得たデータから、社会にとって有益なデータについて考えていきたい。
- ・山月記を場面に分けて分析して、今回の分析や考察をさらに裏付けていくことが必要である。

6. 参考文献

- ・「青空文庫」 <https://www.aozora.gr.jp/cards/000119/card624.html> (参照 2022/09/10)
- ・中島敦 『山月記』 新潮社 (1994/05/15)
- ・「大日本帝国憲法」 <https://www.aozora.gr.jp/cards/001528/card825.html> (参照 2022/09/19)
- ・「日本国憲法」 <https://www.aozora.gr.jp/cards/001528/files/474.html> (参照 2022/09/19)
- ・『最新国語便覧』 浜島書店 2021 pp254 - 257
- ・「都道府県魅力度ランキング」 <https://diamond.jp/articles/-/310887> (参照 2023/01/16)

謝辞

日本工業大学 先進工学部 大宮 望先生にはご指導いただき、大変お世話になりました。

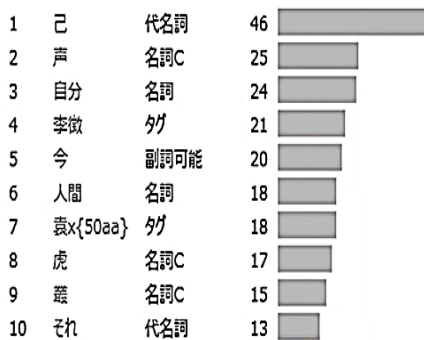


図1 山月記

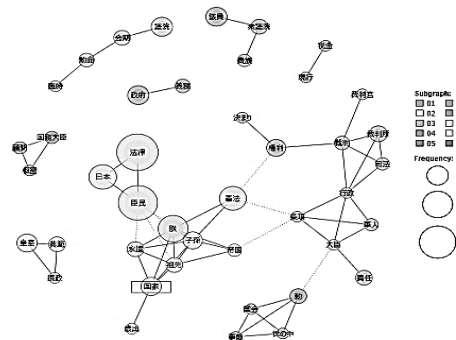


図2 大日本帝国憲法

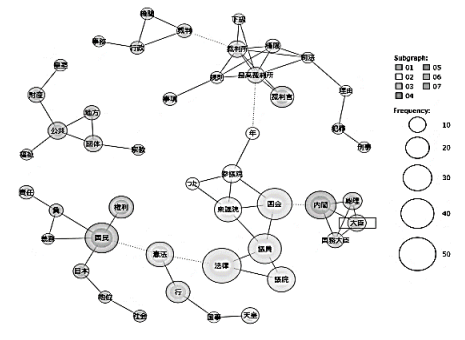


図3 日本国憲法

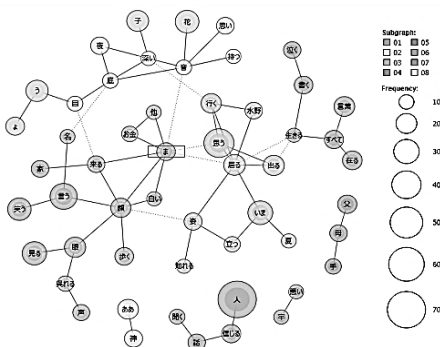


図4 太宰治 前期

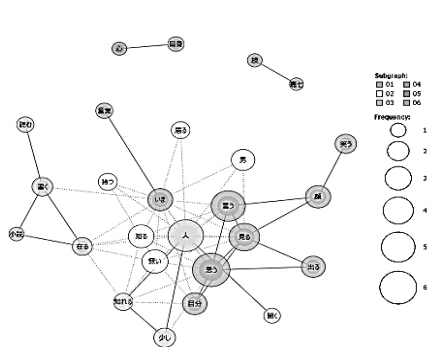


図5 太宰治 中期

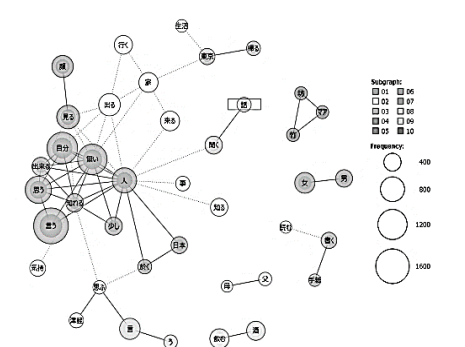


図6 太宰治 後期