

テキストマイニングによる自殺に関する文献研究

CiNii における検討

芝崎 恵一

和光大学 社会文化総合研究科 発達・教育臨床論コース

キーワード（5語以内）

テキストマイニング 自殺 時代変遷 タイトル分析 CiNii 論文

text mining, suicide, chronological change, title analysis, academic papers of CiNii

1. 問題

世界保健機関（WHO, 2016）によると、全世界で年間 80 万人が自殺によって亡くなっている^[5]。どのように自殺を防止するかに関するガイドラインも作成されており、自殺の予防は世界的な課題であるといえる^[6]。日本では、年間の自殺者は 20000 人を超え、自殺率は 20 人／10 万人を超えている^[4]。自殺防止のための NPO 法人も存在し、国民の関心があるといえる^[3]。よって、日本において自殺は重要な、解決すべき課題であるといえる。

しかし、自殺に関する体系的な文献研究は多くは見られない。石井（2003）は 1997 年から 2003 年の自殺の研究についての文献研究を行っているが、それ以前とそれ以降の年の文献についての研究は多くは無い^[2]。本研究では、CiNii における文献について、テキストマイニングを用いて研究を行うことで、自殺の研究がどのように変遷しているかを明らかにすることを試みた。

2. 目的

本研究の目的は、1948 年から 2016 年の自殺に関する論文のタイトルを分析することで、自殺の研究がどのように変遷したかを明らかにすることである。

3. 方法

3-1. 分析対象

データベース CiNii の検索結果を分析対象とした。キーワードを「自殺」と設定し、論文検索を行った。本検索結果から確認できた最初期の論文は 1892 年であったが、1940 年から 1947 年は対象論文が存在しないなど、1947 年以前の論文は一部が散逸しており、統計

的分析に不適當であると推測されるものも多かったため、1948年から2016年までを対象とした。ここで、2017年は年度の途中であることから本分析の対象外とした。さらに、論文の著者と出版年が不明であるものも本分析の対象外とした。

3-2. 調査時期

2017年7月23日に検索を行った結果を分析対象とした。

3-3. 分析方法

Text Mining Studio Ver.6.0.3を用いて、テキストマイニングの分析による論文タイトルの分析を行なった。

分析については基本情報、単語頻度解析、係り受け頻度解析、対応バブル分析、特徴語抽出の順に行った。なお、分析を行う際に1948年から2016年の年について、それぞれ1950年代以前、1960年代、1970年代、1980年代、1990年代、2000年代、2010年代(2016年まで)の7つの属性を分け、それぞれの分析を行った。以上の7つの属性については年代と命名し、分析を行った。

4. 結果

4-1. 分析対象

CiNiiにおいて「自殺」をキーワードに設定したところ、検索結果は9866件であった。このうち、1947年以前と年代が不明な41件、2017年の論文である131件、著者が不明であった1443件を分析対象から除外した。その結果、8251件が分析対象となった。

4-2. 基本情報

基本情報の結果については表1に示した。分析対象の総論文数は8251件であった。一論文当たりの平均文字数は40.1文字であった。さらに、総文章数については8794文、平均文章長は37.6文、延べ単語数は62183単語、単語種別数は18999種であった。

表1 自殺論文の基本情報

	項目	値
1	総行数	8251
2	平均行長(文字数)	40.1
3	総文章数	8794
4	平均文章長(文字数)	37.6
5	延べ単語数	62183
6	単語種別数	18999

4-3. 年代別推移

年代別推移の結果については図1に示した。1950年代以前は157件、1960年代は167件、1970年代は311件、1980年代は549件、1990年代は1018件、2000年代は3416件、2010年代(2016年まで)は2633件であった。

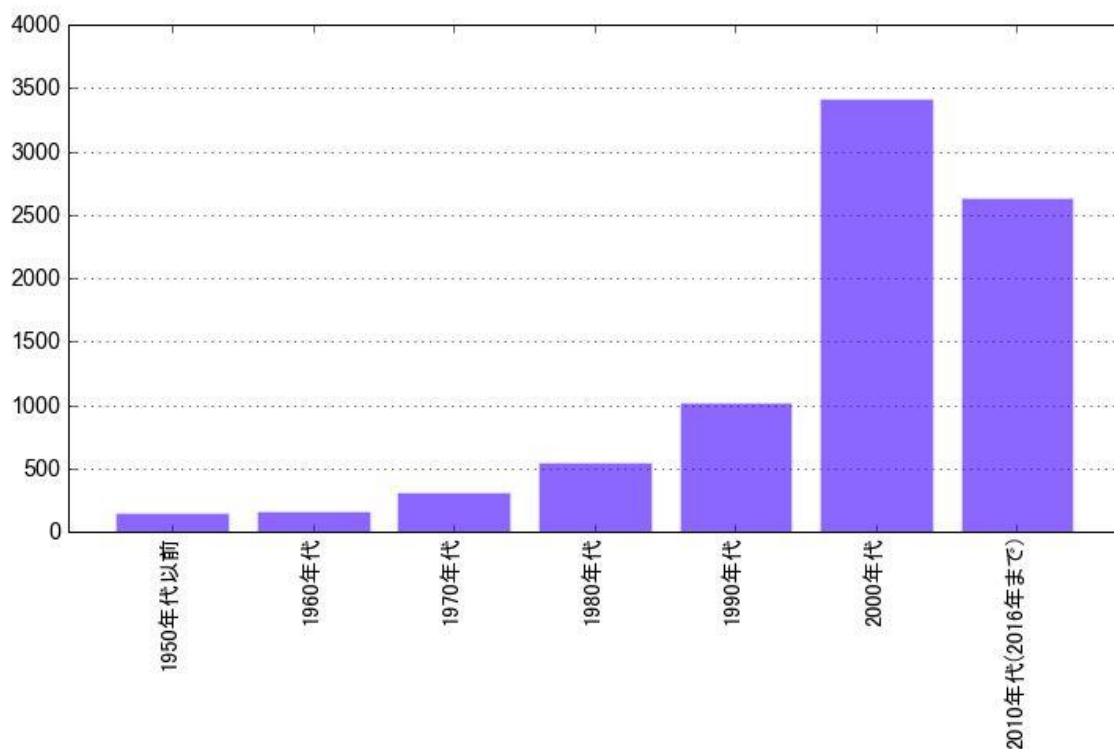


図1 自殺論文の年代別推移

4-4. 単語頻度解析

名詞について単語頻度解析を行なった。その結果については図 2、図 3 に示した。単語頻度分析については、「自殺」と「特集」が多く含まれていたため、それらについては一致しない条件を付け、さらに数詞については含まない条件に設定し分析を行った。単語頻度については 32 件抽出された。

図 2 に示したように、上位 10 単語では、自殺予防が 501 件、過労自殺が 360 件、自殺対策が 295 件、考えるが 271 件、過労死が 246 件、検討が 244 件、自殺企図が 244 件、うつ病が 243 件、子どもが 242 件、いじめが 233 件であった。

図 3 では、それぞれの単語における年代別の頻度を示した。ほとんどの単語が、1990 年代以降にみられたが、自殺予防、過労自殺、自殺対策、うつ病などについては 2000 年代から急増していた。

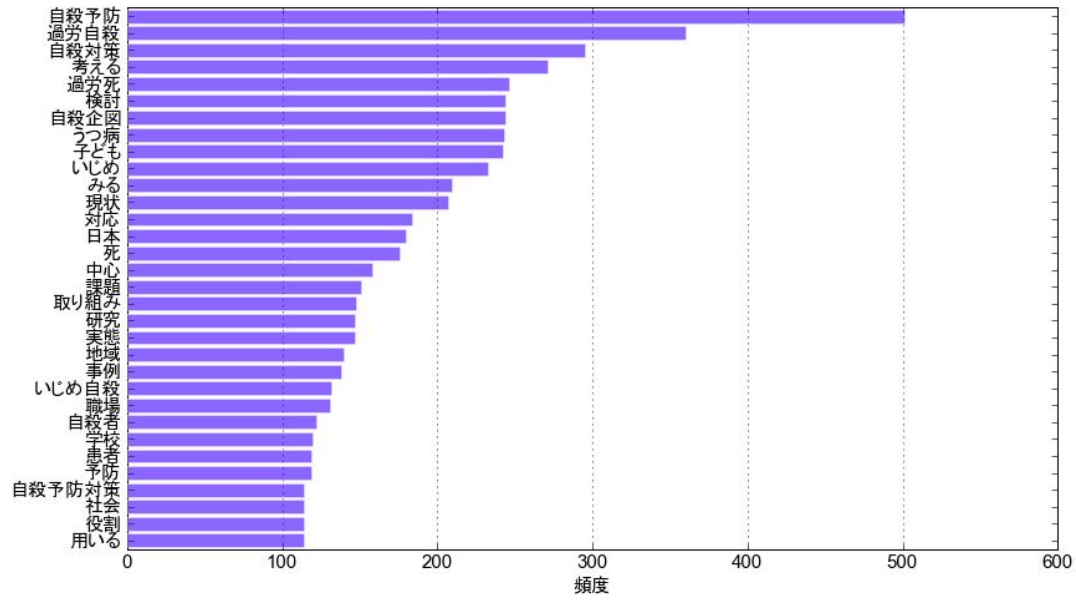


図2 自殺論文の単語頻度

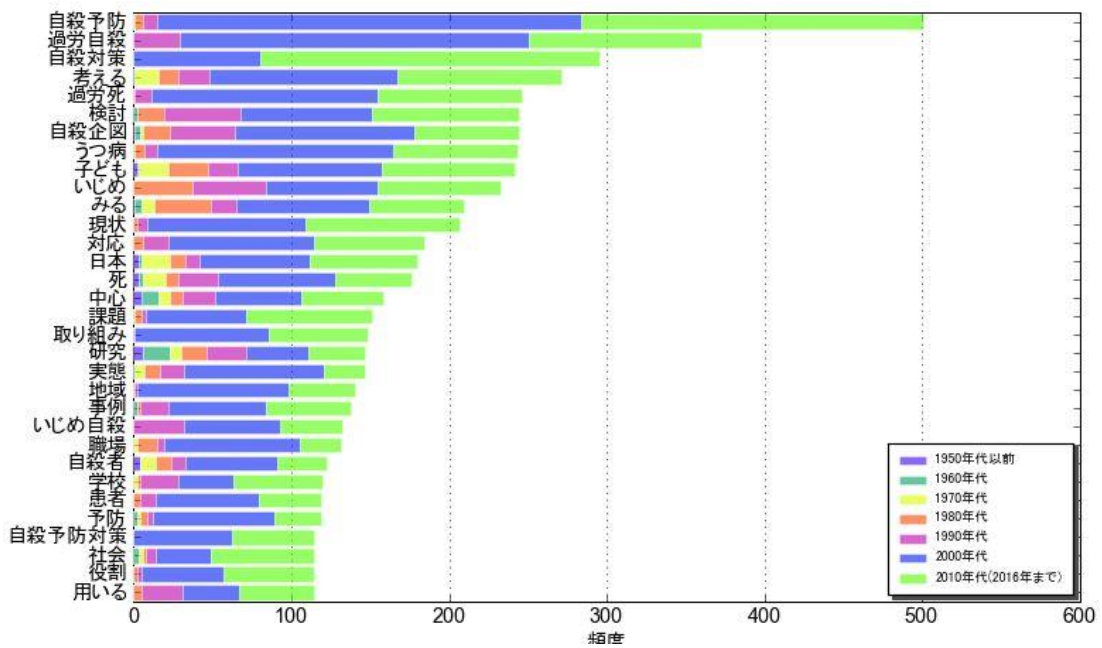


図3 自殺論文の年代別単語頻度

4-5. 係り受け頻度解析

係り受け頻度分析の結果については、図4に示した。図より、いじめ—自殺は1980年代以降、自殺—特集も1980年代に見られた。さらに、自殺—予防や取り組み—特集は2000年代以降に顕著に見られた。一方で、体罰—自殺は1980年代に顕著であり、作家—自殺は1970年代に顕著であった。

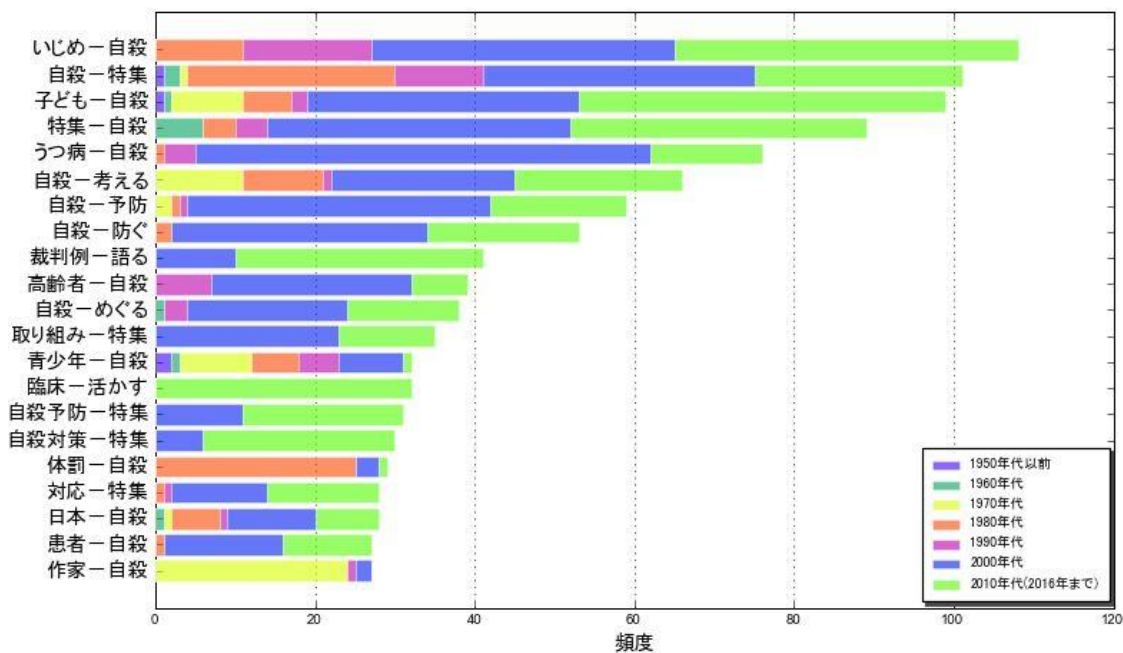


図4 自殺論文の係り受け頻度解析

4-6. 対応バブル分析

対応バブル分析の結果については図5に示した。図に示されたように、2000年代以降に自殺予防、過労自殺といった単語との結びつきが強く見られた。

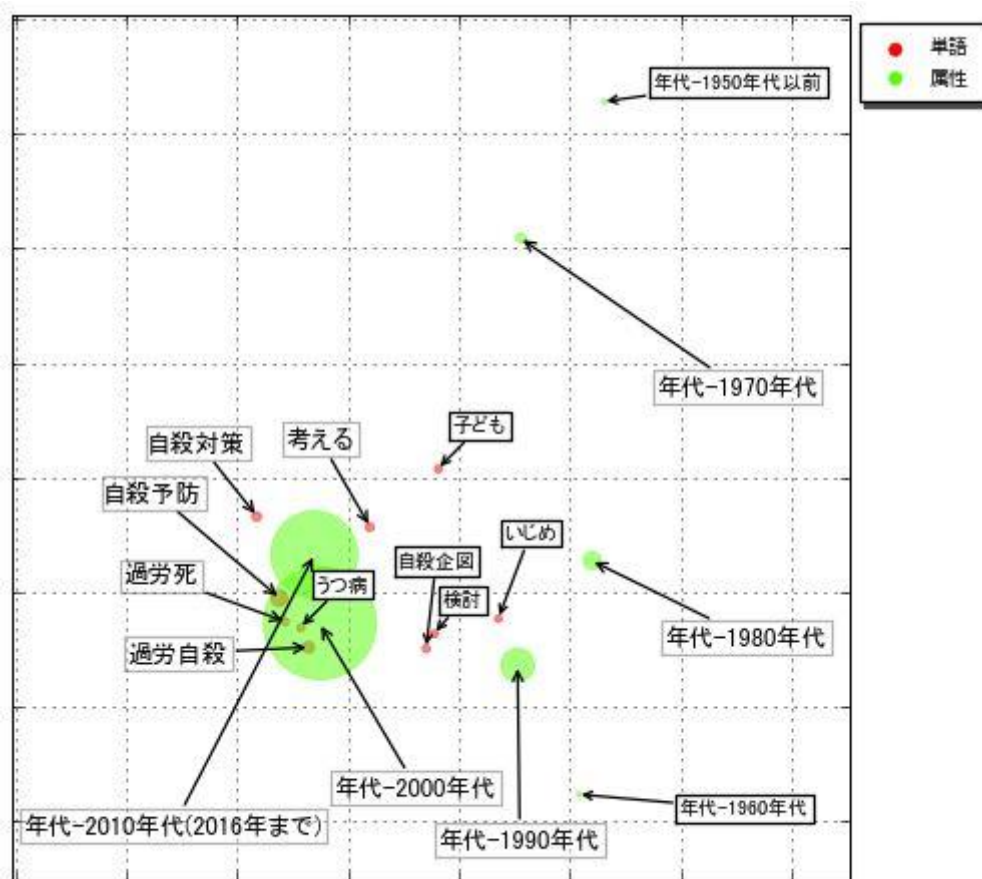


図5 各年代と頻出単語との関係を示す対応バブル分析

4-7. 特徴語抽出

各年代の頻度上位 20 位を対象とした特徴語抽出を行った。その結果については図 6 から図 12 に示した。1950 年代、1960 年代については自殺の研究が大きく特徴的に見られた。1970 年代については特集や作家といった単語が見られた。1980 年代についてはいじめ、体罰といった単語が見られた。1990 年代については、いじめ自殺や学校といった単語が見られた。2000 年代については過労自殺、自殺予防、うつ病といった単語が見られた。2010 年代 (2016 年まで) については自殺対策といった単語が特徴語として見られた。

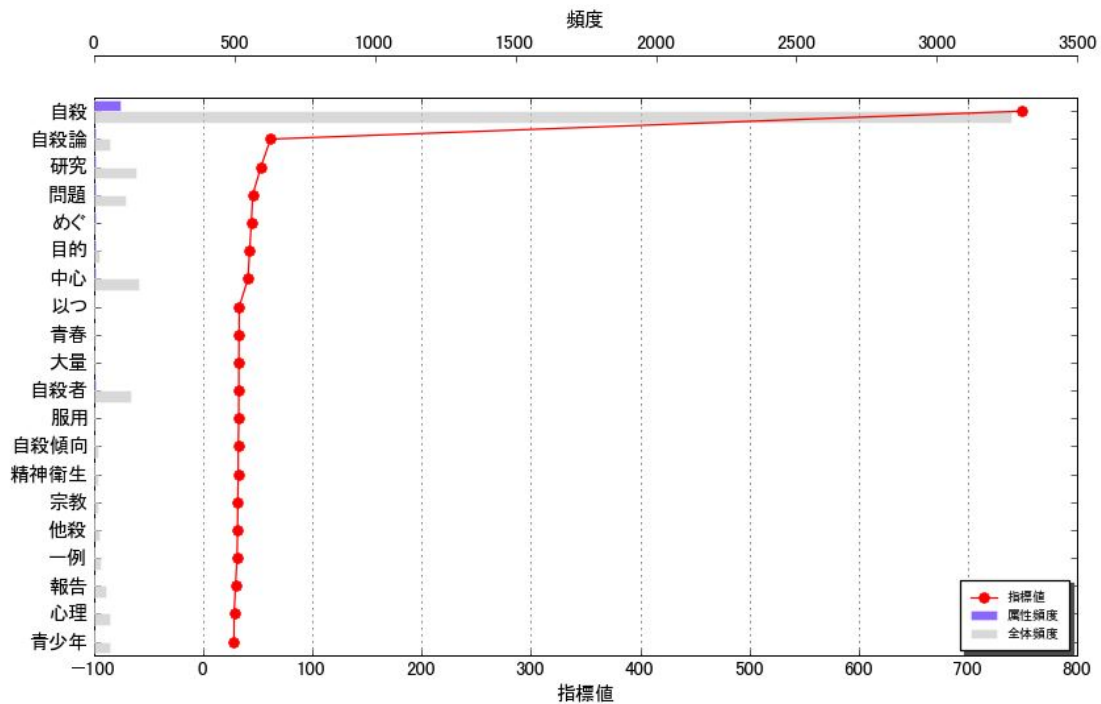


図 6 自殺論文の 1950 年代以前における特徴語分析結果

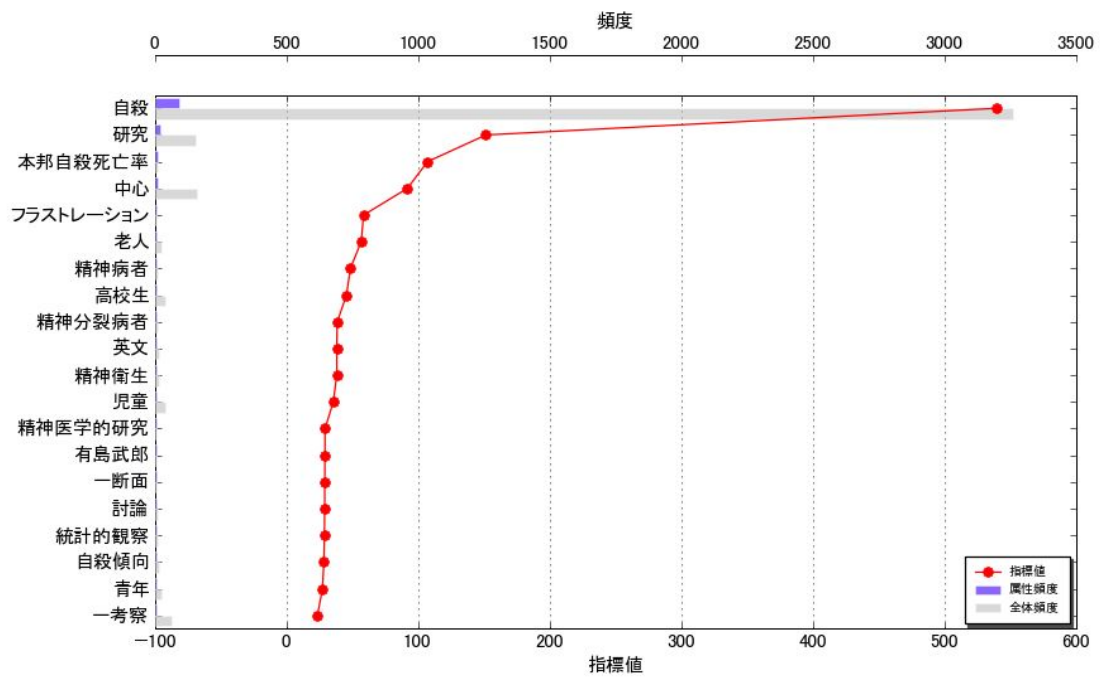


図7 自殺論文の1960年代における特徴語分析結果

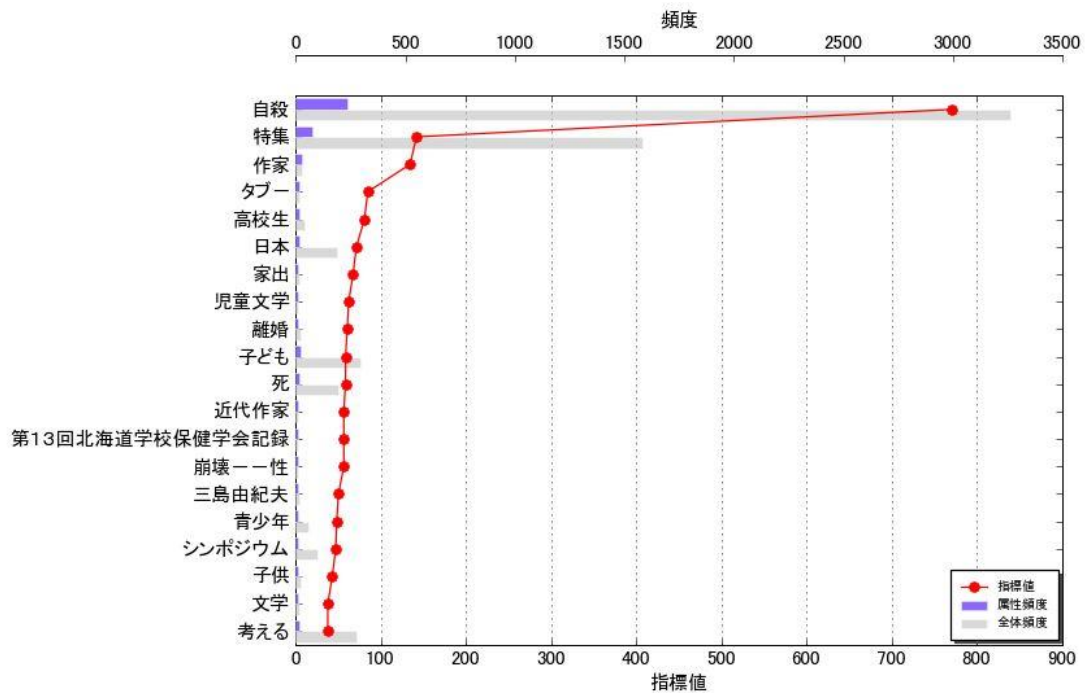


図8 自殺論文の1970年代における特徴語分析結果

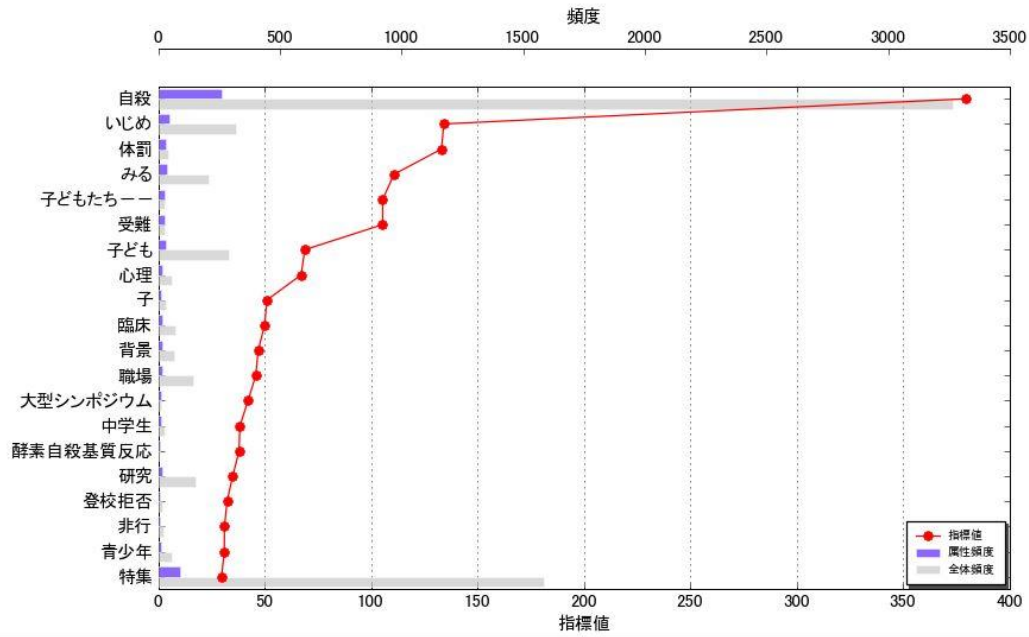


図9 自殺論文の1980年代における特徴語分析結果

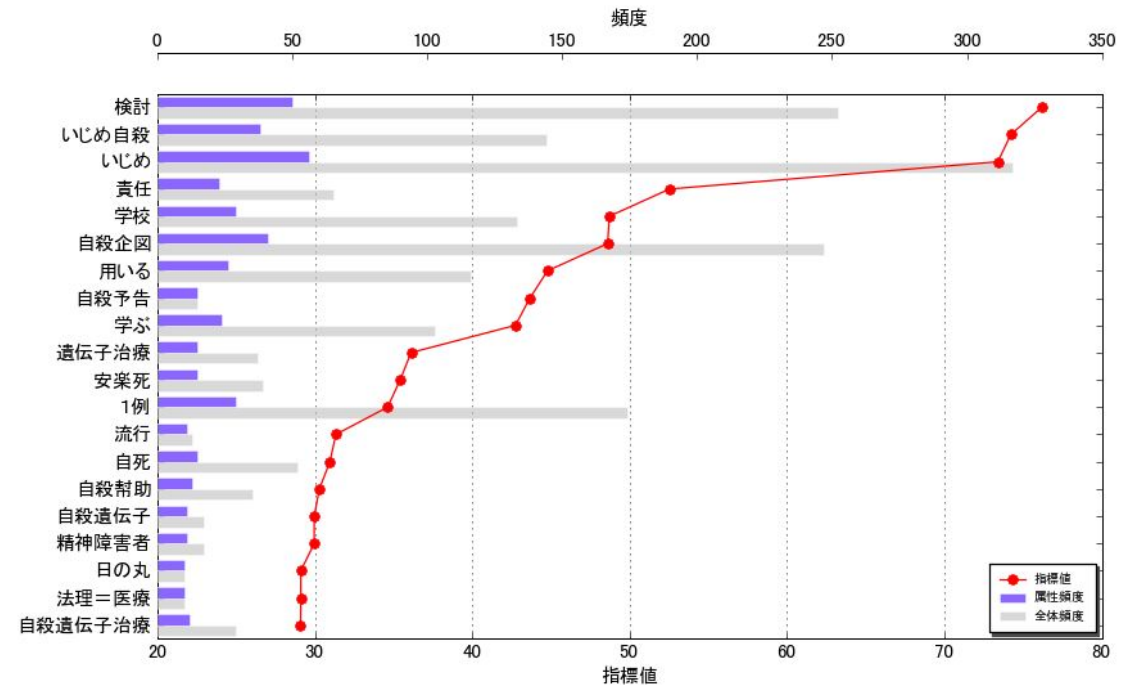


図10 自殺論文の1990年代における特徴語分析結果

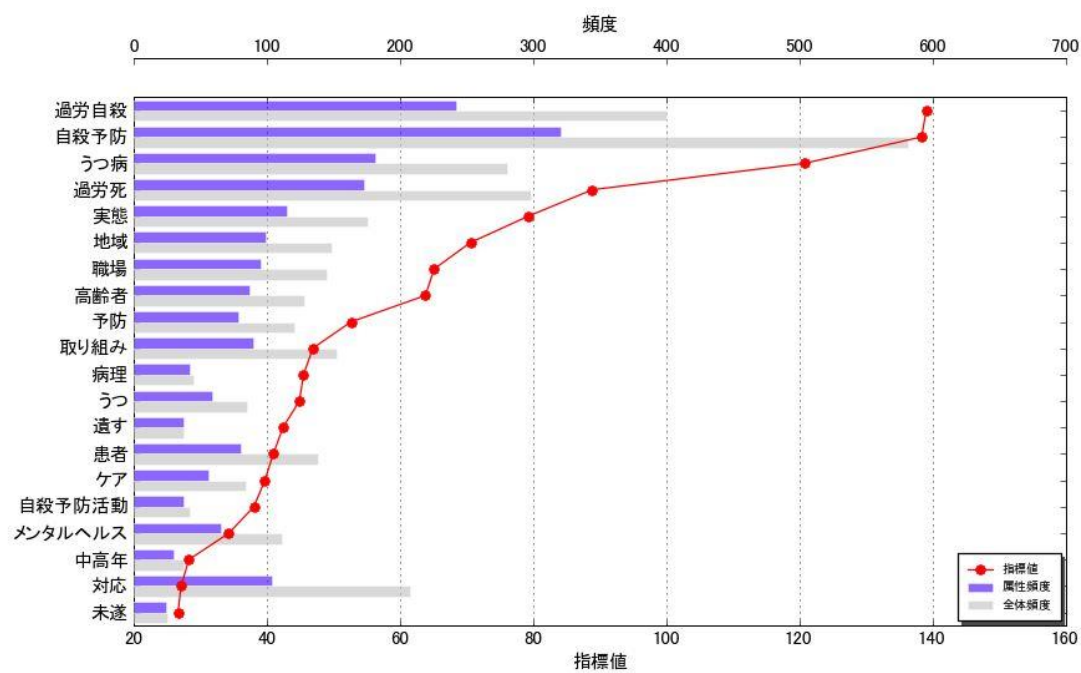


図 11 自殺論文の 2000 年代における特徴語分析結果

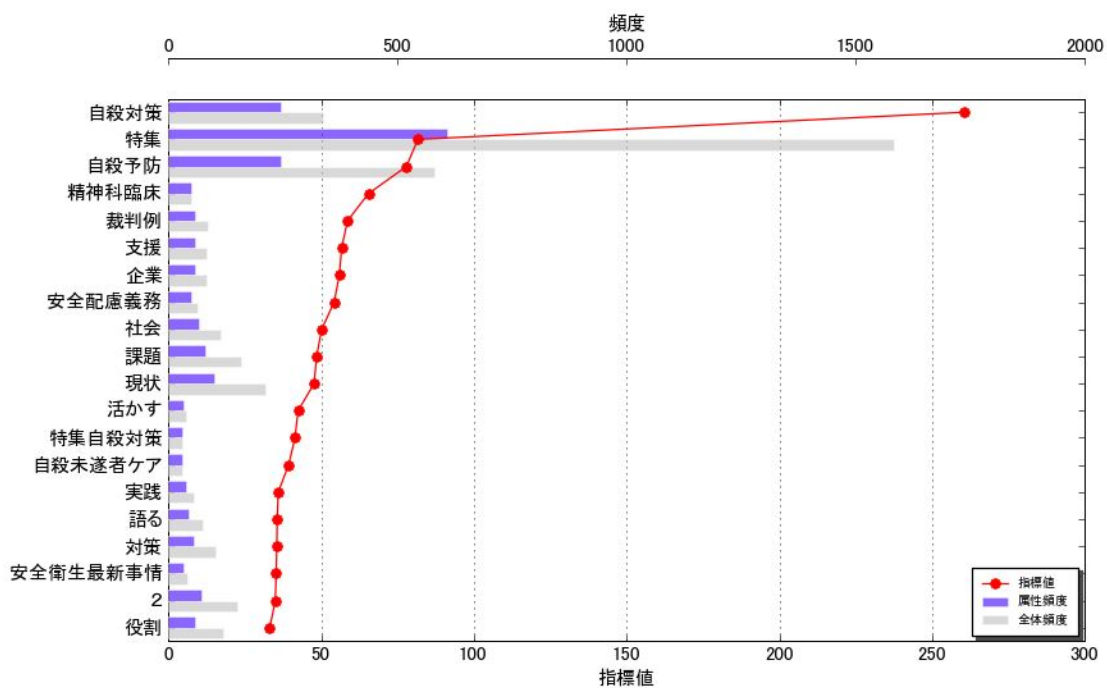


図 12 自殺論文の 2010 年代 (2016 年まで) における特徴語分析結果

5. 考察

5-1. 本研究により明らかになった点

本研究では、第一に、自殺者の急増とともに、自殺に関する研究論文も図1の年代別推移に見られるように増加したことが明らかになった。厚生労働省(2016)によると、21897人が2016年に自殺で亡くなっている^[1]。自殺者は図13のように、1997年から1998年にかけて急増し、2012年に30000人を下回ったものの、依然として20000人を超えており、世界的に見ても高い水準にある^[5]。1997年から1998年の急増以降は「急増・恒常性・若年化」が自殺の傾向として見られるが、この自殺の急増と恒常性により、論文数も同様に急増したと考えられる^[4]。

第二に、自殺の論文の内容については、図3の年代別単語頻度、図5の対応バブル分析、図11、12の特徴語分析の結果に見られるように、2000年代以降に自殺対策、自殺予防などの自殺への取り組みに関する論文が増えていることが明らかとなった。これについては、2006年に自殺対策基本法が制定されたことにより、自殺対策関係予算が内閣府によって計上されたことが原因と考えられる。厚生労働省(2016)によると、2016年度の自殺対策関係予算は約793億円である^[1]。これらは自殺予防、自殺対策に利用されており、実際に2010年以降に自殺者の人数は減少している。これらの予算が組み込まれ、研究にも用いられたことが自殺対策や自殺予防の論文が増えた原因であると考えられる。

第三に、自殺の論文の内容について、2000年代以降は過労自殺や過労死などの過労による自殺の論文についても増加している。これについても、自殺対策関係予算が影響していると考えられる。厚生労働省(2016)によると、中小企業再生支援協議会事業に約584億円が用いられており、予算793億円のうちの多くはこの部分に集中している^[1]。中小企業再生支援協議会事業費によって対策される自殺には過労死が考えられる。このことが、これらの論文数の増加に影響したと考えられる。

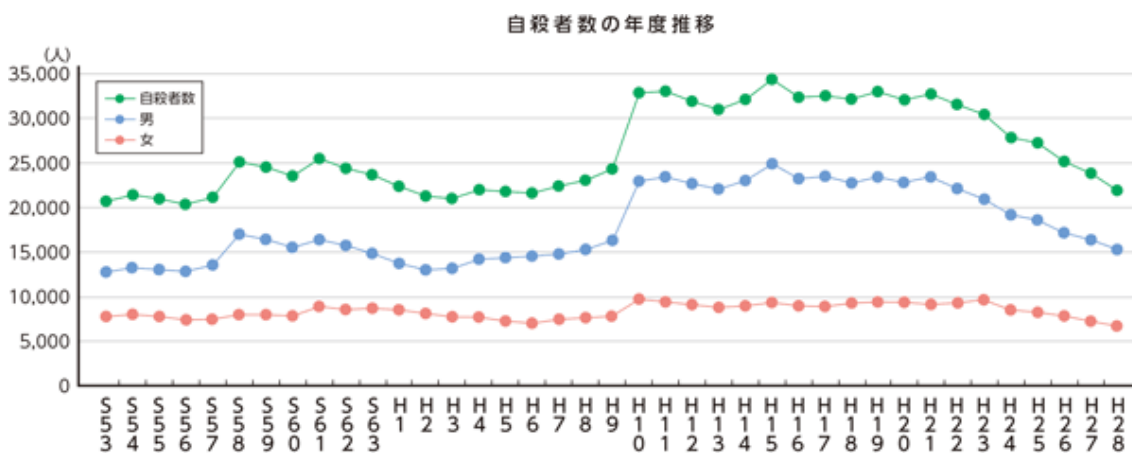


図13 自殺者数の年次推移

第四に、図 4 の係り受け頻度解析の結果に見られるように、1980 年代はいじめと自殺、体罰と自殺に関連した研究が多くあることが明らかとなった。これは、1970 年代後半から 1980 年代前半にかけては校内暴力の全盛期とされているため、その時期にいじめ、体罰と自殺について関連した論文が多く書かれたと推測される。一方で、1970 年代は作家と自殺の結びつきが強いが、これは 1970 年に三島由紀夫、1972 年に川端康成が自殺によって亡くなるなど、著名な作家の自殺が影響したものであると考えられる。

5-2. 本研究の限界と今後の課題

本研究の問題点として、データベースを CiNii しか用いていないこと、自殺以外の単語によってタイトルが書かれている自殺に関連する論文が抽出できていないこと、年代の区切りが根拠の無いものであることが挙げられる。第一に、データベースに用いた検索ツールが CiNii だけである点について述べる。本研究では、医中誌や Google Scholar といった他の検索ツールによるデータは用いていない。これにより、分析に用いられていない論文が存在する可能性がある。次に、自殺以外の単語で検索を行っていないという問題点について述べる。自殺と同等の意味を持つ単語として、自死や自害などが考えられる。特に、自死は、「自死遺族」といった表現で論文のタイトルに用いられることが多くある。これらの単語については本研究では分析されていないことは問題点として挙げられる。最後に、年代の区切りについては、本研究では 1960 年代以降は 10 年ごとの区切りとした。しかし、自殺者の 1997 年の急増、2006 年の自殺対策基本法制定といった、自殺に関する大きな動向があった年で区切るということも考えられる。本研究では区切りについては 10 年ごとという見やすさ、利便性を考慮しており、必ずしも適切な区切りではない可能性も考えられる。

以上のような問題点が研究の限界として存在するものの、本研究は自殺の論文に関して体系的に分析した結果、自殺の研究の変遷について明らかにしているといえる。今後も自殺の研究は増加していくと予想されるが、本研究の用いた方法によってこれまでどのような自殺の研究の変遷があったかを議論しうる一考察を与えたと推測される。

6. 謝辞

学生研究奨励賞の原稿作成にあたり、Text Mining Studio Ver.6.0.3 を使用させていただきました数理システム様に感謝申し上げます。また、本原稿を作成するに当たり、和光大学のいとうたけひこ教授の丁寧なご指導を頂きましたことに感謝申し上げます。

7. 参考文献

- [1] 厚生労働省 (2016). 自殺の統計：各年の状況
http://www.mhlw.go.jp/stf/seisakunitsuite/bunya/hukushi_kaigo/shougaishahukushi/jisatsu/jisatsu_year.html

- [2] 石井 敏弘 (2003). 自殺に関する研究の現状：国内 保健医療科学 **52**, 261-271.

- [3] 再チャレンジ東京 (2017). 自殺防止国民運動
<http://www.jigyo-saisei.com/kokuminundou-top.html>

- [4] 澤田 康幸・上田 路子・松林 哲也 (2013). 自殺のない社会へ 有斐閣

- [5] World Health Organization. (2016). *Suicide data*.
http://www.who.int/mental_health/prevention/suicide/suicideprevent/en/

- [6] World Health Organization. (2014). *Preventing suicide: A global imperative*.
http://www.who.int/mental_health/suicide-prevention/world_report_2014/en/