

AO 入試選考書類のテキストマイニング

吉村 幸 (長崎大学 アドミッションセンター)

本研究では、AO 入試 1 次選考に提出された書類の分類並びにそれらと諸属性との関係、評定との関係を探索的に調べ、そのことにより出願者や評価者の特性を知ることがを目的に、選考書類 3 種のテキストマイニングを行なった。各選考書類はいくつかのタイプに分類され、それらと志願者の諸属性との間にいくつか連関関係がみられることがわかった。

1 はじめに

吉村 (2006), 吉村・南部 (2007) は、AO 入試における書類選考データを分析することで、その方法のあり方をより洗練させることは重要であることを指摘した。しかし、選考書類の内容にまで言及するものではなかった。

島田 (2007) は AO 入試の合格者 79 名の「志望理由書」を何がどう書かれたか、という観点から分析した。島田は書類を一件ずつ読み、事前に設定した項目が含まれているかどうか、あるいはそれらがどのような順序で書かれたかを詳細に検討した。その結果に基づき島田は、志望理由書の内容や構成の画一化の著しさを指摘するとともに、選考資料としての存在意義に疑念を呈している。

本研究の関心も、やはり選考書類に何が書かれているかにある。さらに、それらにはどのような特徴があるか、受験生の他の属性との関係、評定者の評定結果との関連にも関心がある。これをテキストマイニングの手法で機械的に大雑把にとらえようとするのが本研究の目的である。なお、用いる方法の性質上「どう書かれているか」は本研究では扱えるテーマではなく、関心の外となる。

2 方法

2.1 分析対象

200X 年に実施された N 大学教育学部 AO 入試志願者の一部である 105 名が提出した 3 種類の選考書類 - 志望理由書 (800 字), 自己アピール (800 字), 理想の教師像 (600 字前後) - をそれぞれ分析対象とした。

これらの各書類を 7 名の評価者が A, B, C の 3 段階で評価をしている。

2.2 分析方法

以下の分析には WordMiner.1.1 を用いた。このソフトウェアは、日本語を分かち書きし、文章を構成する要素 (必ずしも語ではない) やキーワードを抽出し、各文章 (サンプル) × 構成要素の頻度行列に基づいて対応分析、クラスタリングを行なう。手法の詳細については、同ソフトウェアのマニュアル、あるいは Lebart et.al. (1998) や大隅ら (1994) を参照して頂きたい。

まず、3 種それぞれの書類を分かち書きし、キーワードを抽出する (同ソフトの機能によるもの)。次に、抽出されたキーワードを若干編集し (削除や置換)、出現頻度 5 以上のキーワードを用い、サンプル × キーワードの出現頻度に基づくクロス表を作成した。このクロス表を元に対応分析を行い、得られた

15軸までのスコアを用い、サンプルのクラスタリングを行った。

それぞれのクロス表のサイズ並びに15軸までの累積寄与率は、志望理由書(105×399, 29.39%), 自己アピール(105×419, 28.54%), 理想の教師像(105×299, 33.18%)であった。

2.3 選考書類3種のクラスタリング

種々の指標と解釈のしやすさを参考に探索的にクラスタリングを行い、各選考書類を次に示すように分類した。名づけるにあたっては、各クラスタに特徴的に出現する(他のクラスタに比べ多く出現する)ワードを参考にした。以下、各選考書類のクラスタを記す。なお各クラスタ名後の()内の数はクラスタのサイズである。

「志望理由書」は6つのクラスタに分類した。それぞれのクラスタに現れた特徴は、

- 1) 複式学級の実習(12), 2) 障害児教育(10), 3) 教員志望の強さ(24), 4) 幼稚園教諭に関心(34), 5) 採用試験対策(12), 6) 吹奏楽部(13), であった。

「自己アピール」は9つのクラスタに分類した。それぞれのクラスタに現れた特徴は、

- 1) コンクールでの受賞をアピール(9), 2) 英語の学習をアピール(7), 3) 困難にくじけなかったことをアピール(17), 4) 吹奏楽部での活動をアピール(4), 5) 体育系部活動での活躍をアピール(16), 6) 忍耐力、リーダーシップをアピール(14), 7) リーダーシップ、協調性をアピール(8), 8) 障害児、障害者との関わり、養護学校教諭になりたいことをアピール(8), 9) 生徒会活動をアピール(22), であった。

「理想の教師像」は7クラスタに分類した。それぞれのクラスタに現れた特徴は、

- 1) 担任や部活の顧問(12), 2) 周囲との連携(9), 3) 子どもの目線にたった実践(8), 4) 子どもに夢を与える、専門性を身に付ける(23),

- 5) 子どもに笑顔が絶えない, 子ども同士が向き合う(31), 6) 学力をつける, 魅力ある授業(16), 7) 愛情をもって子どもに接する(6), であった。

3 結果

3.1 志望理由クラスタと諸属性の連関

志望理由クラスタと希望校種との間に連関がみられた(図1) ($\chi^2=71.41$, $df=20$, $p<0.001$)。

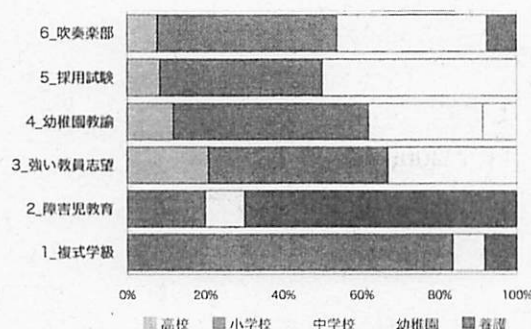


図1 志望理由クラスタと希望校種との連関

複式学級という言葉を支望理由書に書いた者は希望校種を小学校とする者に圧倒的に多い。教員採用試験対策に言及したものは、小中学校を希望するものがそのほとんどを占める。当然ではあるが、志望理由で障害児教育に言及したものの大部分が養護学校を希望している。

3.2 自己アピールクラスタと諸属性の連関

自己アピールクラスタと性別との間に連関がみられた(図2) ($\chi^2=19.26$, $df=8$, $p=0.014$)。男子では体育会系の活動, 女子では困難を克服したというアピールが特に目立つ。

3.3 理想の教師像クラスタと諸属性の連関

教師像クラスタと希望校種との間に連関がみられた(図3) ($\chi^2=40.57$, $df=24$, $p=0.019$)。

また、出身高校の地域との間にも連関がみられた(図4) ($\chi^2=81.55, df=60, p=0.003$)。

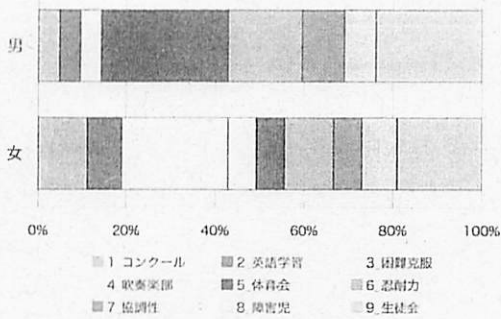


図2 自己アピールクラスと性別との連関

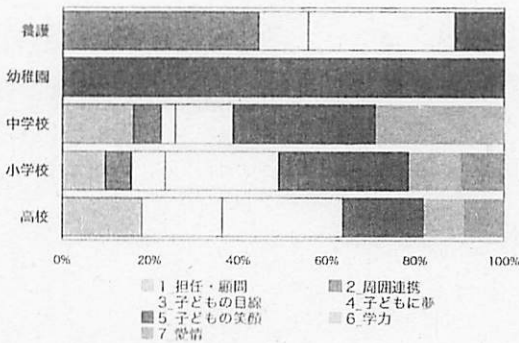


図3 教師像クラスと希望校種との連関

希望校種によって理想とする教師像が異なるのは自然なことだろう。図に表れたように、幼稚園では「子どもの笑顔」、養護学校では「周囲との連携」、中学校では「学力」がそれぞれ特徴的である。

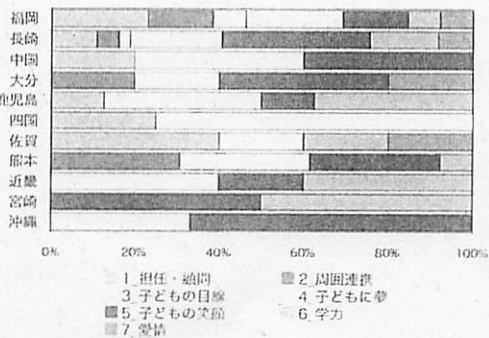


図4 教師像クラスと地域との連関

長崎と沖縄で「子どもの笑顔」、鹿児島で「学力」、熊本では「周囲との連携」「子どもの目線」がそれぞれ特徴的である。それぞれの都道府県の学校教育の志願者が感じてきたその都道府県の特徴を表しているのかもしれない。

3.4 クラスタ同士の連関

3種の書類のクラスタ同士の連関を調べてみたところ、志望理由と理想の教師像との間に($\chi^2=81.98, df=30, p=0.008$)、また志望理由と自己アピールとの間にそれぞれ連関がみられた(図5, 図6) ($\chi^2=73.43, df=40, p=0.001$)。

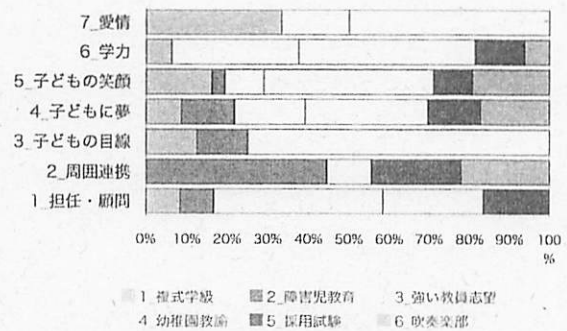


図5 志望理由と理想の教員像との連関

志望理由で障害児教育に言及したものは、周囲との連携を重視していることがわかる。教員志望を強く述べたものの理想は「子どもの目線で」という特徴もみられる。

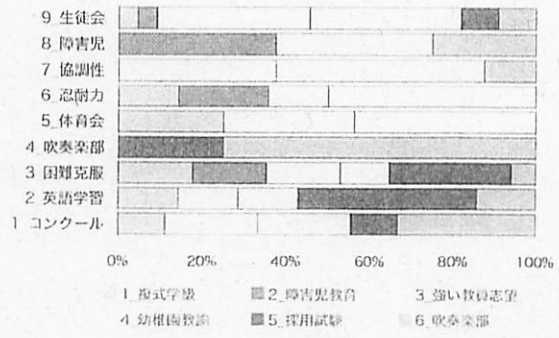


図6 志望理由と自己アピールとの連関

志望理由で吹奏楽に言及したものが自己アピールで吹奏楽に言及するのは当然であろう。志望理由で幼稚園教諭への言及が特に目立ったものが、体育会、忍耐力、協調性、障害児、生徒会、などいくつかのクラスタに比較的多くみられる。幼稚園教諭も志すものは、活動的な人が多いのだろうか。

ただ、複数の書類間の連関があまりにも強すぎる場合、得られる情報が重複するため複数の書類を書かせる意味がなくなる。何を目的にどんな情報を得るために書類を書かせるかをあらかじめ明確にしておくことが重要である。

3.5 各クラスタと評価の連関

どの書類でもクラスタと合否の連関がみられなかった。これは選考書類がどのようなタイプのものであるかが合否に影響しないことを意味する。AO入試ではいくつかの材料を総合的に判断して合否を判定するので、どれかひとつの書類が合否と強く連関することは好ましくない。この結果は妥当だと言えよう。

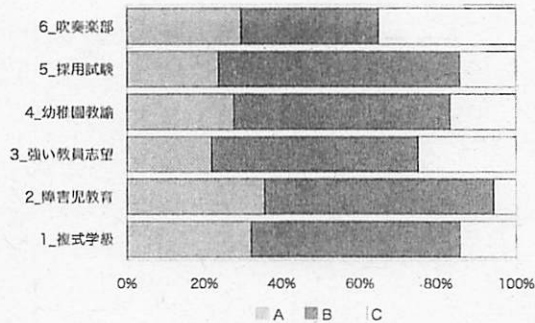


図7 志望理由クラスタと評価

図7は志望理由クラスタと7名の評価全体(A,B,C)の関係を表したものであるが、この2変数には連関がみられた($\chi^2=36.48$, $df=10$, $p<.0001$)。図に示されたようにクラスタ1(複式学級)、2(障害児教育)でやや評価が高いように見える。この結果が、「志望理由にこのような内容を盛り込めば評

価が高くなる」と高校や予備校などの指導に利用されると困るが、しかし一方で、どのような内容が評価されやすいかがある程度は明確である証拠でもある。

4 おわりに

本研究では、AO入試における選考書類3種のテキストマイニングを行った。各選考書類はいくつかのタイプに分類することができ、それらのタイプと種々の属性との間には、いくつか連関関係がみられた。

テキストマイニングという言葉を用いたが、実際にはそれほど深くマイニングしたわけではない。非常に浅い処理しかしていない。分かち書きされた結果をみて、辞書を丁寧に編集し、探索を繰り返せば、さらに有用な情報が得られるかもしれない。どこまでマイニングするかによって得られるものも異なる。今回に限って言えばマイニングにかけたコストが少なかったという反省がある。

参考文献

Lebart, Salem, and Berry (1998). "Exploring Textual Data", Kluwer Academic Publishers.
 大隅昇・L.ルバル・A.モリノウ・K.M.ワーウィック・馬場康維 (1994). 記述的多変量解析法, 日科技連.
 島田康行 (2007). 「AO入試「志望理由書」の研究—何がどう書かれたか—」, 入研第2回大会研究発表予稿集, 43-49.
 吉村宰 (2006). 「長崎大学AO入試における書類選考データの分析」, 入研協第1回大会研究発表予稿集, 33-36.
 吉村宰・南部広孝 (2007). 「AO入試による入学者の追跡調査—選抜方法改善の観点から—」, 入研協第2回大会研究発表予稿集, 189-192.