

各大学経済学部におけるアドミッション・ポリシーの テキストマイニングによる分析

齋藤朗宏（北九州市立大学経済学部）

近年、大学においてディプロマ、カリキュラム、アドミッションという3つのポリシーの重要性が増している。一方で、ポリシーに対する共通した理解は、いまだ得られていないように見える。そこで、本研究では、テキストマイニングの手法を用いて各大学経済学部のアドミッション・ポリシーを分析し、そのあり方についての現状を確認した。その結果、入学時の能力、入学後のビジョンが主たる内容であることが確認できた。

1 はじめに

1.1 アドミッション・ポリシー

平成20年12月の中央教育審議会においてとりまとめられた「学士課程教育の構築に向けて」(答申)において、「学位授与の方針(ディプロマ・ポリシー)」、「教育課程編成・実施の方針(カリキュラム・ポリシー)」、「入学者受入れの方針(アドミッション・ポリシー)」という3つのポリシーの重視が提言された。また、平成22年6月の学校教育法施行規則一部改正により、大学が公表すべき情報が明確化され、上記の内容を含めた各種情報を、インターネットを含めて広く公開することが定められたことで、3つのポリシーの重要度が増している。特にアドミッション・ポリシーは、平成11年の答申においてその制定が提言されていることもあり、早い段階で普及が進んでおり、2012年3月段階で各大学経済学部のウェブサイトを確認したところ、掲載されていない大学は存在しなかった。

これを受け、鳴野他(2004)では、各大学にアンケート調査を行い、制定されているアドミッション・ポリシーの内容、策定水準(大学全体のポリシーか、学部レベルか、学科レベルかなど)等に関する分析を行っている。その結果をまとめたものとして、鈴木他(2005)がある。この調査の結果、記載率の高い内容は、

順に「基礎学力」、「目的意識」、「興味関心」、「教科学力」であった。これに引き続き、鳴野他(2006)では、高等学校側から見たアドミッション・ポリシーに注目し、高校生、また、高校教員に対する調査を行っている。

しかし、このアドミッション・ポリシーにはいくつかの問題がある。一つには、鳴野他(2006)で指摘された、高校側に対する認知度の低さが挙げられる。この調査によると、アドミッション・ポリシーという言葉を知っていた高校生は全体の13.8%、志望校のアドミッション・ポリシーを知っていると回答した高校生は8.1%に過ぎなかった。第二に、鳴野他(2004)で指摘されているように、アドミッション・ポリシーの本来の意味は各大学が求める学生像であるにもかかわらず、教育理念や目標との違いが不明確になっている点も挙げられる。実際に制定されているのを見ると、アドミッション・ポリシーとして、教育理念・目標を掲げるケースもあれば、求める人材像を掲げるケースもあり、高校の時に学習して欲しい内容を掲げる場合もあれば、入試において課す科目を述べることもあるというように、アドミッション・ポリシーという言葉の意味そのものについて、大学間で共通した理解が得られている訳ではないということが確認できる。

1.2 テキストマイニング

こういった言語情報を分析する手法として、近年、テキストマイニングと呼ばれる研究分野が発展している。テキストマイニングとは、膨大なテキスト（文書）情報の中から有用な情報を掘り出す（マイニング）ことで、定型化されていないテキストデータを、一定のルールに従って定型化して整理し、データマイニングの手法を用いながら、相関関係などの定量分析を行う手法である。

特に最近では、樋口(2004)のKH coderをはじめ、松村・三浦(2009)のTinyTextMiner、金(2009)のMLTPに見られるように、フリーのソフトウェアも豊富となっており、応用研究も多くなっている。大学入試の分野に絞っても、アメリカにおいて用いられている小論文採点システム e-rater を参考に、文体、論理構成、内容の観点から日本語小論文を自動的に採点するシステム Jess を作成、公開している石岡・亀田(2003)や、教育学部のA0入試の選考書類について内容のクラスタリングを行い、志望理由と希望校種の連関など、内容と属性の関連性を調べている吉村(2009)などがあり、応用例は少なくない。

1.3 研究目的

本研究では、このテキストマイニングの手法を用い、日本の大学の経済学部におけるアドミッション・ポリシーの現状について分析を行う。この分析を通して、アドミッション・ポリシーに対する共通理解の構築と、策定、見直しの上での材料を提供することを目的とする。特に、鳴野他(2004)との比較によって、経済学部の特徴を得ることが出来ると考えられる。

2 方法

2.1 データベースの作成

研究に先立ち、大学の公式ウェブサイトを利用してアドミッション・ポリシーの抽出を

行い、データベースを作成した。鳴野他(2004)では、予め内容に関する選択肢を準備し、大学側がそれに当てはまっているかどうかを回答しているが、本研究では、前提となる知識を置かず、制定された文章そのものを分析対象としている。この点が大きな違いであると考えられる。こうすることで、同じ文面であったとしても、アンケートの回答者によって、内容に関する回答が異なるという可能性をなくすことが出来る。

文章を抽出するにあたっては、以下のような条件を設定した。

1. ある大学に、「経済学部(政治経済学部も含む)」という名称の学部が存在する場合。
 - (ア) 学部にアドミッション・ポリシーがあればその内容を抽出する。
 - (イ) 学部にはなく、「経済学科」などにアドミッション・ポリシーがあるのであれば、その内容を抽出する。
 - (ウ) 学部にも学科にもアドミッション・ポリシーがないのであれば、抽出の対象外とする。
2. 他の学問分野との複合的な学部名称である場合。
 - (ア) 「経済学科」などの学科があり、そこにアドミッション・ポリシーが制定されているのであれば、その内容を抽出する。
 - (イ) 学部全体にのみ制定されているのであれば、抽出の対象外とする。
3. 単一学部の大学の場合
 - (ア) 経済学部単独なのであれば、大学のアドミッション・ポリシーを経済学部のもののみとする。
 - (イ) 複合的な学部なのであれば、2と同様に考える。
4. 入試制度ごとにアドミッション・ポリシーが制定され、学部学科全体では制

定されていない場合には、抽出の対象外とする。これは、入試制度特有のアドミッション・ポリシーが、入試制度に準拠した内容に偏っていることが多く、5でも述べる内容が限定されるという問題点が危惧されるためである。

5. アドミッション・ポリシーとして書かれている文書の内容のうち、求める人材像、並びに高校のうちに学習しておいて欲しい内容、課外活動や資格などについて述べている部分のみを抽出対象とする。これらについて述べている部分がない大学は抽出の対象外とする。教育理念や目標、入試の内容は、述べている大学に限られている上、内容が極めて限定され、分析上の悪影響がある可能性が否定できないためである。

以上のような条件でアドミッション・ポリシーの抽出を行った結果、国立23大学、公立13大学、私立94大学の合計130大学がデータベースに登録され、分析の対象となった。尚、データベース作成は2012年3月から5月にかけて行った。

2.2 頻度集計

諸々の分析に先立ち、出現単語の頻度集計を行った。集計、その後の分析には、KH Coder 2.beta.28b(樋口, 2004)を用いた。集計に際しては、分析の対象となる品詞を名詞に限定し、「経済」や「経済学部」、「学生」、「大学」、「学科」といった単語は分析の対象外とした。これらの単語は、たとえば「〇〇大学経済学部では、××というような人材(学生)を求めています」というような記述に用いられることが多く、解釈上の意味が殆どないからである。集計には、一般的に単語の出現した回数(Term Frequency, TF)を求める方法と、単語の出現した文書の数(Document Frequency, DF)を求める方法がある。即ち、一つの大学で「知識」という

単語が2回出現したとき、それを2回と数えるのがTF, 1大学分ということで1回と数えるのがDFである。本研究においては、集計は基本的にDFとした。

ただし、この集計のみでは、当たり前の単語のみが抽出され、キーワードとしての意味をなさない可能性がある。そこで、サーチエンジン Google を用いて単語の検索を行い、発見されたページ数をもってその単語の一般性と考える、所謂 Google DF と、データベース全体の TF を利用して TFIDF を求め、単純集計による出現大学数との比較を行った。Google DF は、同社で提供する「Google カスタム検索」の API を利用し、ブラウザから検索することなく、単語のリストから一括して得ている。この手順、設定等は付録の通りとなっている。Google DF を利用した単語 w の TFIDF は以下の通りである(Bekkerman & McCallum, 2005)。

$$TFIDF(w) = \frac{TF(w)}{\ln GoogleDF(w)} \quad (1)$$

2.3 内容分析

TFIDF の値で上位となった38単語を分析対象として、単語の分類を目的としたクラスター分析を行った。単語間の距離には Jaccard 距離を、クラスターの併合には Ward 法を用いた。

次に、クラスター分析の結果から確認された主要な単語の分類を用いて、国公立大学間での分類ごとの出現頻度の違いを確認した。

最後に、国公立間での違いをより詳細に確認するために、国公立大分類と、TFIDF の値で上位60の単語との間でのコレスポネンス分析を実施した。

3 結果

3.1 頻度集計

単語について、TFIDF の値が大きい順に上位35単語を並べたものが表1である。TF は

前述の通りデータベース全体の出現頻度である。GoogleDFは、「Google カスタム検索」によるヒット数であり、TFIDFは、(1)式を元に算出したTFIDFの大きさである。

表 1 単語の出現大学数等の集計

単語	大学数	TF	GoogleDF	TFIDF
社会	97	229	5480000	14.76
意欲	76	114	4850000	7.41
問題	66	105	4690000	6.84
知識	60	103	5620000	6.63
関心	72	97	5320000	6.26
能力	60	95	5360000	6.13
基礎	51	70	4250000	4.59
地域	36	63	1180000	4.51
学力	36	47	306000	3.72
身	40	55	4890000	3.57
学習	31	49	1110000	3.52
専門	34	53	5160000	3.43
活動	24	45	2190000	3.08
論理	26	36	331000	2.83
解決	38	43	4300000	2.82
理解	30	43	4800000	2.80
貢献	33	40	4550000	2.61
教育	25	36	988000	2.61
国際	29	38	2390000	2.59
思考	28	39	4880000	2.53
企業	21	36	3220000	2.40
分野	22	31	563000	2.34
積極	28	34	3760000	2.25
活躍	23	33	5020000	2.14
学校	22	31	2820000	2.09
興味	26	32	5150000	2.07
科目	17	26	336000	2.04
情報	20	31	5160000	2.01
数学	26	30	4420000	1.96
自分	22	30	5460000	1.93
世界	26	28	4020000	1.84
基本	22	24	590000	1.81
英語	23	27	4430000	1.76
コミュニケーション	23	25	3970000	1.65
表現	19	24	5560000	1.55

特に上位に意欲、関心といった単語が並んでいる点からも、学習に対する意欲や興味関心を重視していることが確認できる。中でも社会問題に関する興味関心を問うていると思われる大学が多いようだ。目的意識、興味関

心に対する重要性は鳴野他(2004)で見出された結果とも合致しているが、特に社会が重視される点が経済学部の特徴と考えられる。一方で学力については、基礎という単語が上位にある点からも、基礎的な学力のみ要求していることが確認できる。中でも、数学が上位にある点は注目される。社会については、科目としての社会科と、社会問題という意味での社会が混在しているため、明確な結果とは言えない。

また、これら以外では、地域や国際といった単語が上位に並び、これらが大きなテーマになっていることが示唆される。

TFIDFと大学数との違いに注目すると、大学数では19番目である論理がTFIDFでは14番、23番目である教育が18番、32番目である企業が21番である点が注目される。これらのキーワードは、ある特定の大学で繰り返し言及される、あるいは一般的な文書においては出現頻度があまり高くないなどが原因となって、経済学部のアドミッション・ポリシーにおいて特徴的なキーワードとなっている。ただ、全体的には数十万～数百万件のヒット数となる単語ばかりであり、TFを基準にすると、極端に大きく順位が変わることはなかった。

3.2 クラスタ分析

単語の分類を目的としたクラスタ分析の結果は以下の通りである。

1. 学校, 科目, 英語, 数学, 国語
2. 教育, 基本, 歴史, 学習, 理解
3. 自分, コミュニケーション, 表現
4. 論理, 思考
5. 基礎, 学力
6. 意欲, 知識, 関心, 能力, 社会, 問題, 身, 解決
7. 活躍, 専門, 地域, 貢献, 国際, 企業
8. 情報, 分析
9. 分野, 積極, 世界, 活動, 興味

第1クラスターは学ぶべき科目について言及しているものと考えられる。第2クラスターは第1クラスターとの関わりで、理解しておくべき内容について言及しているものと考えられる。同様に、第3クラスターは表現力、第4クラスターは思考力、第5クラスターは基礎的学力について言及している。第6クラスターはどんな知識、能力を身につけたいと考えているかという学ぶ意欲に言及しているものと考えられる。第7クラスターは卒業後にどう活躍したいのかについてであろう。以下、第8クラスターは情報にかかわる内容、第9クラスターは主に在学中にどう活動したいと考えているのかとなっている。

以上の結果から、より大きくまとめるのであれば、「入学前に何をしておくべきか」、「入学してからどうしたいか」が主な内容であり、入学前については「学習内容・科目」、「思考力」、「コミュニケーション能力」、入学後については「学習・活動への意欲」、「将来へのビジョン」が主たる内容になっているということが確認できる。これらの内容は、概ね鳴野他(2004)の内容カテゴリの一部分と言えるが、表現力、将来へのビジョンはカテゴリとして存在していない。これらは、経済学部としての特色であるのか、近年記載が増えた内容であるのかは不明であり、検討の必要がある。

3.2 内容分類別集計

クラスター分析の結果から類型を確認して6種類に整理し直し、国公立大学の分類別に集計を行った。その結果は表2の通りである。尚、それぞれの内容の有無については、

キーワードを利用するなどの方法はとらず、実際にアドミッション・ポリシーの内容を読んだ上で割り振っている。また、「学力」は学力についてアドミッション・ポリシーの中で言及しているか否かを指す一方で、「科目」は特定科目に対する言及があるかを意味している。そのため、一部の例外を除き、学力でカウントされていれば科目でもカウントされている。

ここから、表1などで見てきた通り、学習に対する意欲、興味関心が最も重視されていることは明らかである。将来への展望や学力についても、やはり多くの大学で言及されていることがわかる。

逆に、思考力と個々の科目についての言及には大きな違いがあった。特に思考力は、国立大学では過半数の大学が言及している一方で、私立大学では僅か7%の大学でしか言及していない。言及している7%にしても、「論理的思考力を身につけたいと思っている者」のような、意志に関するものであり、入学前に身につけておくべき能力として要求しているケースは稀であった。個々の科目にかかわる言及も含めて、能力に関する突っ込んだ要求は国立大学で多いことがわかる。

国立大学には思考力に関する記載が多く、私立大学では少ないという点は先行研究と同様であったが、公立大学は国立大学同様半数近くの大学で思考力に関する記述があったと指摘されていた一方で、本調査の結果では、寧ろ公立大学に近くほぼ記載が見られなかった点が大きく異なっている。

表2 内容別、国公立大分類別集計

	意欲	展望	表現力	思考力	学力	科目
国立	22 (96%)	15 (65%)	9 (39%)	13 (56%)	13 (56%)	7 (30%)
公立	12 (92%)	6 (46%)	4 (31%)	2 (15%)	7 (54%)	1 (8%)
私立	91 (97%)	46 (49%)	15 (16%)	7 (7%)	38 (40%)	20 (8%)
合計	125 (96%)	67 (52%)	28 (22%)	22 (17%)	58 (45%)	28 (22%)

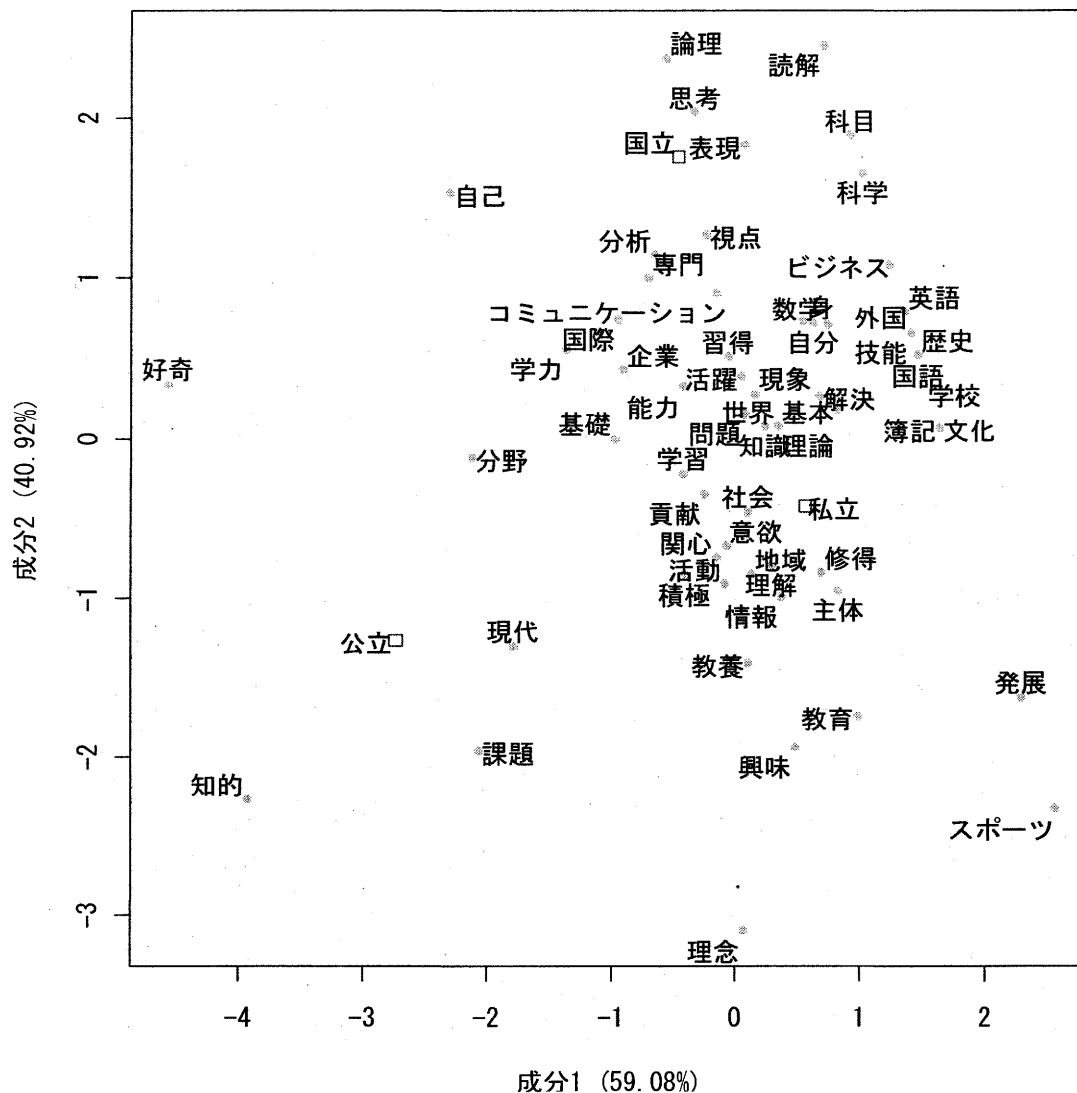


図 1 国公立大分類別×単語のコレスポネンス分析

3.3 コレスポネンス分析

国公立大分類と TFIDF の値で上位 60 の単語との間でコレスポネンス分析を行った結果が図 1 である。第 2 軸正の方向には国立大学が布置され、論理、思考、表現、読解といった各種能力に関する単語が同時に布置されている。一方で右下方向には、私立大学が布置され、スポーツ、興味、関心、簿記といった単語が同時に布置されている。ここから、国立大学に特徴的な、抽象的な能力に関する要求、私立大学に特徴的な、各種資格やスポーツなど、より具体的、実際的な物事に関する要求という特徴が見て取れる。

一方で公立大学を示す左下方向については、あまり明確な結果とは言えなかった。左下方向に布置される単語にはあまり一貫性がないようである。元々公立大学はサンプル数が少ないため、齋藤・豊田(2004)にも見られるように結果の安定性が期待できず、解釈の難しい結果になってしまった可能性が考えられる。

4 まとめ

頻度集計、クラスター分析を通して、アドミッション・ポリシーの内容について、9 種類の基本的な類型が確認できた。これらの内

容を確認すると、入学前の段階で必要なものとして、身につけておくべき学力、思考力等の一般的能力を、また、入学後にきちんと勉学に取り組むために必要な条件として、専門分野に関する興味関心、卒業後の進路に対するビジョンを述べていることがわかる。中でも興味関心は単語の出現頻度が極めて高く、入学者に求める必要条件として、一般的な理解が得られていることが窺える。ただ、大半の大学が何らかの形で掲げているということは、言及したとしても独自性を出し、他大学との差別化を図るのは難しい内容であるとも言える。この結果は、先行研究を裏付けるものと言える一方で、特に社会問題に関する記載が多かった点に経済学部の特徴を見ることが出来た。

これらの結果を元に実際のアドミッション・ポリシーを見直すことで、自分たちで定めた内容が、内容的にどの部分で不足しているのか、また、特徴的な部分はどこであるのかの自己分析が可能になると思われる。

国公私立大学別の集計、コレスポネン分析の結果からは、より抽象的な能力を要求する国立大学と、課外活動や資格なども重視する私立大学というように、内容面に大きな違いがあることが確認された。ただ、この内容面の違いが何に起因しているのかについては、受験生の学力、大学の歴史、大学のあり方など様々な可能性が考えられ、現時点では不明である。

今後の課題としては、カリキュラム・ポリシーやディプロマ・ポリシーと、今回分析したアドミッション・ポリシーとの比較、学部間でのアドミッション・ポリシーのあり方の違いについての分析などが考えられる。また、原因のわからなかった国立大学と私立大学との間での違いについても分析を加えていく必要があるだろう。

文献

- Bekkerman, R. and McCallum, A. (2005). “Disambiguating Web appearances of people in a social network”, Proceedings of the 14th international conference on World Wide Web, 463-470.
- 樋口耕一(2004). 「テキスト型データの計量的分析:2つのアプローチの峻別と統合」『理論と方法』19(1), 101-115.
- 石岡恒憲・亀田雅之(2003). 「コンピュータによる小論文の自動採点システム Jess の試作」『計算機統計学』16(1), 3-19.
- 金 明哲(2009). 「テキストデータの統計科学入門」岩波書店.
- 松村真宏・三浦麻子(2009). 「人文・社会科学のためのテキストマイニング」誠信書房.
- 文部科学省 中央教育審議会(1999). 『初等中等教育と高等教育との接続の改善について(答申)』(平成20年12月24日).
- 文部科学省 中央教育審議会(2008). 『学士課程教育の構築に向けて(答申)』(平成11年12月16日).
- 齋藤朗宏・豊田秀樹(2004). 「コレスポネン分析における布置の精度」『オペレーションズ・リサーチ』49(3), 168-173.
- 嶋野英彦他(2004). 「高等学校における進学情報の利活用とアドミッション・ポリシー」大学入試センター研究開発部共同研究報告書.
- 嶋野英彦他(2006). 「アドミッション・ポリシーと入学受入方策」大学入試センター研究開発部共同研究報告書.
- 鈴木規夫他(2005). 「『入学者受入方針等に関する調査』結果の概要」『大学入試研究ジャーナル』15, 19-24.
- 吉村 幸(2009). 「A0 入試選考書類のテキストマイニング」『大学入試研究ジャーナル』19, 157-160.

付録

Google DF 算出のための Google カスタム検索の設定等は以下の通りである。尚、今回は、検索の実行から結果取得、必要な箇所の切り出しについては Perl を用いた。

1. Google アカウントでログインし、以下の URL からカスタム検索エンジンを作成する。検索するサイトには、今回は *.com, *.net, *.org, *.jp と指定した(後で、検索対象を Web 全体と設定している)。たとえば Wikipedia のように、特定のサイトをデータベースとして用いたいのであれば、ここで指定するとよい。言語は日本語、名前は適当で問題ない。

<http://www.google.co.jp/cse/manage/creacr>

2. コントロールパネル→基本から、検索エンジン ID を確認、記録する。また、追加したサイトの検索方法を、「追加したサイトを重視して、ウェブ全体を検索する」と設定する。特定のサイトをデータベースとするのであれば、この部分の指定は必要ない。
3. 以下の URL から API Access を選択し、Simple API Access 中にある API key を記録する。

<https://code.google.com/apis/console>

4. 記録した検索エンジン ID と API key を利用して、以下の形で検索を行い、JSON 形式の結果を得る。

https://www.googleapis.com/customsearch/v1?key=APIキー&cx=検索エンジンID=lang_ja&prettyPrint=false&q=検索キーワード

5. 得られた結果のうち、totalResults 部分を適当な方法で切り出し記録する。この値を Google DF とする。