

# オープンキャンパスアンケートデータのテキストマイニング

小棹 理子<sup>a</sup> 関 祐太郎<sup>b</sup>

<sup>a</sup> 湘北短期大学情報メディア学科 <sup>b</sup> 早稲田大学大学院 先進理工学研究科

## 【抄録】

テキストデータを対象とするテキストマイニング手法を用いて、本学のオープンキャンパス来場者（高校生）に対するアンケート調査の回答部分のうちの感想・コメントなど、自由記述部分の分析を行った。昨年同様に、6月と7月の来場者データを対象とし、分析結果の年次変化を検討した。6月と7月では来場者の進学意思の決定度合いが異なることと、学科ごとに印象が違うことは昨年と同じ傾向が認められた。また、効果的なイベントなどを抽出するのに本手法が有効であることが示唆された。

## 【キーワード】

テキストマイニング    オープンキャンパス    単語頻度解析    特徴語抽出

## 1. はじめに

テキストマイニング<sup>[1]</sup>は、データマイニング手法<sup>[2]</sup>の一つであり、「文章の中に埋もれている共通性や相関関係を発掘し、有用な知見を獲得する分析手法」として、アンケート調査などにおいて大量のフリーコメント部分を客観的に、かつ効率よく分析するのに用いられている。筆者らは、データマイニング手法を用いて、データベースのアクセスログの解析を行い、アソシエーション分析によりエコ技術に関する一般動向調査を行ってきた<sup>[3, 4]</sup>。また、同手法を湘北短期大学のWebサイトのアクセスログデータに適用することにより、学科によりサイト閲覧者が大きく異なる挙動をすることを発見し、そこから閲覧者の興味の違

いが読み取れることを示した<sup>[5]</sup>。

そこで、2008年の6月と7月の本学オープンキャンパス来場者のアンケート回答の感想・コメントをデータとして用い、テキストマイニングをフリーコメント部分の分析に適用した。その結果、データマイニング分析から得られた知見と同じく学科ごとに印象が異なることが確認でき、また、6月と7月では来場者の本学への進学意思が異なる可能性があることが示唆された<sup>[6]</sup>。

今回は、上記知見が年ごとに変化するのかどうかを検討するために、平成21年度の6月と7月のオープンキャンパスデータを対象としたテキストマイニング分析を行い、前年度に得られた知見との差異を検討した。その結果を報告する。

## 2. テキストマイニングの実施

全体的な処理の流れはすでに報告済みである<sup>[6]</sup>。

---

<連絡先>

小棹 理子    ozao@shohoku.ac.jp

形態素解析は、文字列を文法的に意味のある単位である構成要素に分割し、各要素の文法的素性(品詞など)を決定する。構文解析は、要素間の文法的・意味的關係を示し、係受け構造を見出だす。このような概念抽出を行うことによって新たな定型情報を得て多次元解析を行う。解析ソフトは、前回同様(株)数理システムのText Mining Studio Ver3.1を用いた。本ソフトの機能である類義語辞書登録、原文参照を活用しながら今回は下記に絞り、前年との比較を試みた。

- ① 基本情報分析
- ② 単語頻度解析
- ③ 特徴分析
- ④ ことばネットワーク

### 3. オープンキャンパスアンケート回答の解析

#### 3.1 基本情報

オープンキャンパス(OC)に来場した高校生に対するアンケートのフリーコメント部分の解析を行った。アンケート回答は、(A)2009年6月6日と(B)2009年7月25日(模擬授業を実施)に得られたものである。来場者数の学科別内訳を表1に示す。この表から、昨年度(2008年)と比較して7月の来場者数が減っていることがわかる。とくに保育は昨年の63%であるが、その理由は不明である。

表2はオープンキャンパスの情報源の複数回答結果である。6月ではホームページを閲覧しての来場者が圧倒的に多いが、7月では入学案内・募集要項を情報源にしている生徒が多い。本学を第一志望とする来場者の割合は、6月で

表1 アンケート回答を得たオープンキャンパス実施日(2009年)の来場者内訳

実施日	6月6日	7月25日
情報メディア	28	29
生活プロデュース	69	49
総合ビジネス	62	56
保育	88	55
合計	247	189

表2 オープンキャンパスの情報源(複数選択可)

情報源	6月6日	7月25日
湘北のホームページを見て	99	49
先輩・友人から	67	39
本学の入学案内・募集要項で	45	59
高校内での進学説明会で	40	22
オープンキャンパスのポスターを見て	22	33
受験雑誌で	12	1
高校の先生から	12	18
進学相談会	6	6
その他	7	10

40% (=97/245) であるのに対し、7月では49% (=95/193) に達しており、本学への志望意思の固まった生徒の割合が高くなっている。この結果は昨年と同様である。また、Webサイトは、広く浅くオープンキャンパス来場者を獲得するのに重要な役割を果たしていると言えよう。

### 3.2 単語頻度解析

フリーコメント部分の品種別出現回数を表3に示す。名詞が多いのは一般的な傾向であり、形容詞・副詞も通常の頻度で現れている。表4にはテキストマイニングの基本である単語頻度解析結果を示した。ここでは6月と7月の解析結果を比較しており、6月の頻度の高い順に順序を合わせてある。図1.2に、人数で規格化した頻度分布を示す。「良い」「楽しい」は昨年同様、二強であり、大きな変化はない。

動詞が比較的多いが、「わかる」や「来る」が大きな割合を占めている。6月では保育学科の「入る+したい(入りたい)」、「来る+したい(来たい)」が目をひく。形容詞は感性的な文章であることの表れであることが多いが、保育学科の「凄い」は突出している。また、昨年度と同じ傾向であることも大変興味深い。「凄い」は、保育学科希望者ではポジティブな感情を強調して表すために用いている傾向が読み取れる。「凄くやさしい」「すごく楽しい(6件)」「すごくいい」「凄く湘北に通いたい」「すごく明るい」「すごくきれい」「すごい元気がいい」などがその例である。さらに、7月では、保育学科の「気持ち」がたいへん特徴的であるが、原文参照すると、入学意思を強く表す文中に登格する。以下はその例である。「入りたい気持ちが強くなった」「うれしい気持ちになりました」「入学したい気持ちがより一層強まりました」「ますます入学したい気持ちが増えました」「行きたいという気持ちが大きくなって」「入りたい!という気

持ちが大きくなった」。

「優しい」は人に対する印象を表すことが多いが、6月は保育、7月は総合ビジネスと保育で多く見られる。同様に、人に対して「親切」を使う例がしばしば見られるが、6月の生活プロデュース学科と7月の総合ビジネスで比較的多く使用されている。この2学科において、「人」を主体とする感想が多い傾向は昨年と同様である。

表3 フリーコメント部分の品詞出現回数

品詞	6月6日		7月25日	
名詞	1413	51.9%	739	49.1%
動詞	512	18.8%	231	15.3%
形容詞	385	14.1%	234	15.5%
副詞	313	11.5%	192	12.8%
連体詞	22	0.8%	14	0.9%
接続詞	35	1.3%	16	1.1%
感動詞	45	1.7%	79	5.2%

表4 単語頻度解析結果 (6月と7月を比較)

	品詞	6月					6月正規化後					7月					7月正規化後				
		全体	情報	生活	総合	保育	欠損	情報	生活	総合	保育	全体	情報	生活	総合	保育	欠損	情報	生活	総合	保育
良い	形容詞	91	12	20	21	37	1	0.3	0.2	0.2	0.3	43	4	10	10	17	2	0.2	0.2	0.2	0.4
楽しい	形容詞	74	9	16	16	33	0	0.3	0.2	0.2	0.3	66	7	6	17	35	1	0.2	0.1	0.2	0.5
わかる+しやすい	動詞	55	5	19	15	15	1	0.2	0.3	0.3	0.2	25	1	6	8	8	2	0.1	0.3	0.3	0.3
説明	名詞	53	5	18	13	16	1	0.2	0.3	0.3	0.2	23	2	6	6	7	2	0.2	0.3	0.3	0.3
凄い	形容詞	44	4	9	5	26	0	0.2	0.2	0.1	0.5	19	3	1	3	11	1	0.3	0.1	0.1	0.5
きれい	名詞	38	2	5	11	18	2	0.1	0.1	0.3	0.4	13	1	4	3	4	1	0.1	0.3	0.2	0.3
来る	動詞	38	5	12	5	14	2	0.3	0.3	0.1	0.3	16	2	1	4	8	1	0.2	0.1	0.2	0.5
先生	名詞	37	4	12	11	10	0	0.2	0.3	0.3	0.2	12	1	2	3	5	1	0.2	0.2	0.2	0.4
オープンキャンパス	名詞	35	3	9	9	13	1	0.2	0.2	0.3	0.3	17	3	4	7	3	0	0.3	0.2	0.3	0.1
学校	名詞	35	1	9	9	15	1	0.1	0.3	0.3	0.4	15	2	4	0	9	0	0.2	0.3	0	0.5
美味しい	形容詞	35	5	10	5	15	0	0.3	0.3	0.1	0.3	13	1	4	6	2	0	0.1	0.3	0.4	0.1
思う	動詞	26	4	6	8	8	0	0.3	0.2	0.3	0.2	4	0	0	2	2	0	0	0	0.5	0.5
自分	名詞	25	3	7	5	10	0	0.3	0.3	0.2	0.3	8	1	0	5	2	0	0.2	0	0.6	0.2
校舎	名詞	24	1	4	7	11	1	0.1	0.2	0.3	0.4	5	0	1	2	1	1	0	0.3	0.5	0.2
優しい	形容詞	23	0	4	5	14	0	0	0.2	0.3	0.5	15	0	3	6	6	0	0	0.2	0.4	0.4
参加	名詞	20	1	4	2	13	0	0.1	0.2	0.1	0.5	10	1	4	3	2	0	0.2	0.4	0.3	0.2
湘北	名詞	19	1	2	4	12	0	0.1	0.1	0.2	0.5	18	2	2	7	6	1	0.2	0.1	0.4	0.3
先生方	名詞	19	1	6	4	7	1	0.1	0.3	0.2	0.3	8	1	1	4	2	0	0.2	0.1	0.4	0.2
入る+したい	動詞	18	1	4	4	9	0	0.1	0.2	0.2	0.4	9	0	4	2	3	0	0	0.5	0.2	0.3
雰囲気	名詞	18	1	4	7	6	0	0.1	0.2	0.4	0.2	11	0	3	3	5	0	0	0.3	0.3	0.4
来る+したい	動詞	18	2	4	3	9	0	0.3	0.2	0.2	0.4	9	1	3	2	3	0	0.2	0.3	0.2	0.3
親切	名詞	17	1	7	2	7	0	0.1	0.4	0.1	0.3	15	0	2	7	6	0	0	0.1	0.5	0.4
湘北祭	名詞	16	0	0	2	14	0	0	0	0.2	0.8	0	0	0	0	0	0	0	0	0	0
見る	動詞	15	1	3	5	6	0	0.2	0.2	0.4	0.3	3	0	0	0	2	1	0	0	0	1
聞く	動詞	15	2	4	3	6	0	0.3	0.2	0.2	0.3	5	0	0	2	3	0	0	0	0.4	0.6
わかる	動詞	14	2	7	2	3	0	0.3	0.4	0.1	0.1	4	0	0	2	2	0	0	0	0.5	0.5
学食	名詞	14	0	4	4	6	0	0	0.3	0.3	0.4	8	1	1	5	1	0	0.2	0.1	0.5	0.1
学生	名詞	14	0	5	6	3	0	0	0.4	0.5	0.2	6	0	1	2	3	0	0	0.2	0.3	0.5
話	名詞	14	0	4	4	6	0	0	0.3	0.3	0.4	5	1	1	1	2	0	0.3	0.2	0.2	0.3
学生スタッフ	名詞	12	1	2	3	6	0	0.2	0.2	0.3	0.4	5	0	2	1	2	0	0	0.4	0.2	0.4
気持ち	名詞	12	0	2	2	8	0	0	0.2	0.2	0.6	6	0	1	1	4	0	0	0.2	0.2	0.7

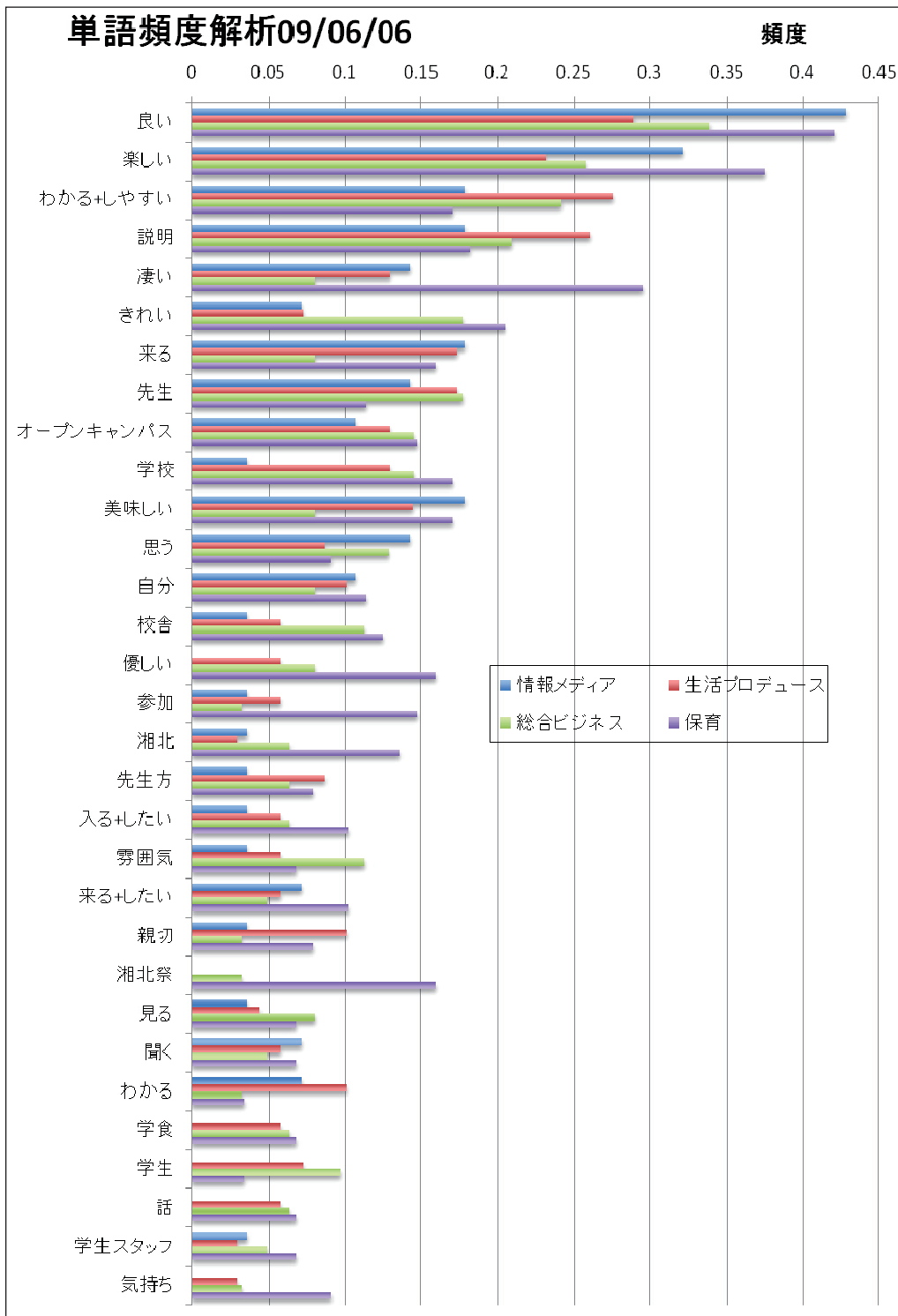


図 1 単語頻度解析 (学科別、人数で規格化 6 月)

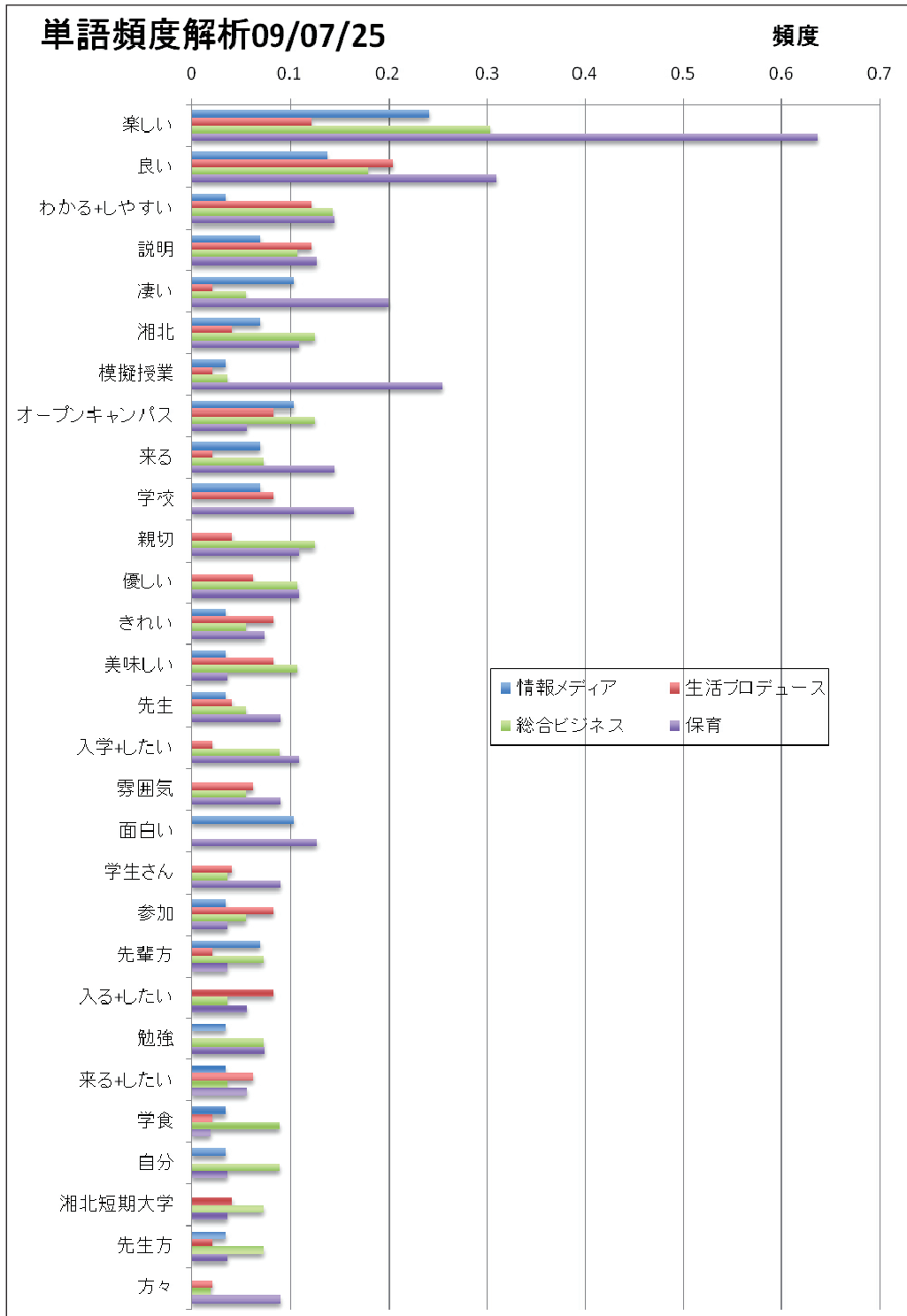


図2 単語頻度解析 (学科別、人数で規格化7月)

### 3.3 特徴分析 (特徴語抽出)

特徴分析では、データの属性分布と異なる分布を持つ単語や係受けを特徴としてとらえることができる。つまり、学科ごとに頻度が異なるので、たとえば、「楽しい」、というコメントが学科により偏りがあるかどうかはそのままでは判断できないが、ここでは指標値を用いて定量化することによって特徴的なことばを抽出可能であるということである。特に、7月には模擬授業があったので、学科ごとに特徴が出るはずである。特徴（一般からの偏り）を定量的に表すためにはさまざまな指標値があるが、ここでは、前年度同様補完類似度<sup>[6]</sup>を用いている。

表5.6と図3～図6に情報メディア学科、生活プロデュース学科、総合ビジネス学科、保育学科の特徴語抽出結果を示す。

6月の情報メディア学科は「良い」が特徴語となっているが、「良かったです」「来て良かった」「知れて良かった」という漠然とした感想と「就職率が良い」、という2つのタイプの「良い」があるのが他学科と異なる点である。一方、生活プロデュース学科と総合ビジネス学科では、「わかりやすい」が特徴語として抽出されている。とくに生活プロデュースでは、さらに「説明」と「わかる」が特徴的である。この2学科においては、「先生」が登場するのも特徴である。この結果は、昨年度の傾向<sup>[6]</sup>と同様で興味深い。保育学科では、他学科と異なり、「湘北祭」が際立っているが、湘北祭におけるパフォーマンスの映像が与えるインパクトが大きかったことを物語っている。

7月のコメントの中では総合ビジネス学科と保育学科において模擬授業に関する記述が特徴として抽出されている。生活プロデュース学科では、サークルに関する詳細説明が好印象であったことがうかがえる。全体印象を表す形容詞が第一の特徴語と抽出されている情報メディア学科と保育学

科では、前者が「面白い」であるのに対し、後者が「楽しい」であり、この点も興味深い。

表5 特徴分析 (6月、学科ごと)

情報メディア	指標値	生活プロデュース	指標値	総合ビジネス	指標値	保育	指標値
良い	13.887	わかる+しやすい	15.005	わかる+しやすい	7.589	湘北祭	13.570
楽しい	8.623	説明	13.726	雰囲気	7.565	凄い	12.206
美味しい	6.695	わかる	8.957	生徒	7.428	参加	7.923
3度目	6.418	興味	8.772	先生	7.314	優しい	7.202
大学生	6.418	先生	8.238	学生	7.234	湘北	6.823
パソコン	6.106	来る	7.691	フィールド	7.097	映像	6.596
思う	5.979	参加+したい	7.492	きれい	6.790	先輩	5.875
来る	5.760	親切	7.315	勉強	6.378	SE入試	5.496
合う	5.483	コース	7.308	思う	5.797	ピアノ	5.496
充実	5.483	資格	7.308	総合ビジネス学科	5.716	気持ち	5.155
方々	5.483	勉強	5.850	学科	5.522	案内	4.587
安心	5.171	イベント	5.481	楽しむ+できる	5.191	ナビ	4.397
文字列欠損	4.733	パンフレット	5.481	行く	5.191	教える	4.397
やさしい	4.548	経験	5.481	授業	5.191	卒業生	4.397
強い	4.236	やる+したい	5.118	仲	4.667	湘北短期大学	4.056
設備	3.925	資料	4.933	校舎	4.416	人	3.866
ITフィールド	3.209	内容	4.933	見る	4.279	感じる	3.677
いる+ない	3.209	美味しい	4.585	入学+したい	4.142	魅力	3.490
うける+したい	3.209	学生さん	4.386	学校見学	3.811	お話	3.487
がんばる	3.209	学生	4.208	気軽	3.811	広い	3.487
しよ的	3.209			言葉	3.811		
しんせつ+したい	3.209			仕事	3.811		
せつび	3.209			仲良し	3.811		
ぴったり	3.209						
みる+したい	3.209						
メディアデザインフィールド	3.209						
回れる	3.209						
基本要項	3.209						
機械	3.209						
気持ちよい	3.209						
系統	3.209						
経験+できる	3.209						
建物	3.209						
見つける	3.209						
在学生	3.209						
鮭いくら丼	3.209						
受け入れる	3.209						
集中+できる	3.209						
笑	3.209						
情報メディア学科	3.209						
食べる+したい	3.209						
生徒方	3.209						
全日程出席	3.209						
入る+?	3.209						
入れる	3.209						
幅広い	3.209						
分野	3.209						
目指す	3.209						
料理	3.209						
量	3.209						



オープンキャンパスアンケートデータのテキストマイニング

表 6 特徴分析 (7月、学科ごと)

情報メディア	指標値	生活プロデュース	指標値	総合ビジネス	指標値	保育	指標値
面白い	7.549	サークル	8.093	ミニ講座	6.364	楽しい	14.197
1回	6.037	入る+したい	6.224	親切	6.108	模擬授業	12.886
オープンキャンパス	5.741	良い	5.982	学食	6.069	凄い	5.957
校内	5.735	入る	5.953	自分	6.069	カエル	5.828
楽しい	5.463	参加	5.757	美味しい	5.146	学校	5.342
凄い	5.139	説明	4.898	オープンキャンパス	4.851	面白い	4.727
先輩方	4.530	きれい	4.355	湘北	4.222	体操	4.662
学校	2.723	美味しい	4.355	優しい	3.889	方々	4.112
来る	2.422	わかる+しやすい	3.963	湘北短期大学	3.849	学ぶ+したい	3.804
感じ	2.416	来る+したい	3.617	先生方	3.849	授業	3.804
広い	2.416	学校	3.421	入学+したい	3.555	学生スタッフさん	2.946
行く	2.416	行く	3.346	先輩方	3.221	気持ち	2.946
前回	2.416	可愛い	2.879	勉強	3.221	1日	2.639
2回	2.115	学生スタッフ	2.879	わかる+しやすい	2.042	感じ	2.639
くわしい	2.115	雰囲気	2.683	わかる	1.925	湘北ナビ	2.639
話	2.115	オープンキャンパス	2.486	スタッフ	1.925	来る	2.461
湘北	1.819	学生スタッフさん	2.411	楽しむ+できる	1.925	入学+したい	1.845
良い	1.529	聞く+できる	2.411	思う	1.925	聞く	1.781
学食	1.211	湘北短期大学	1.477	2回	1.296	明るい	1.781
自分	1.211	優しい	0.814	くわしい	1.296	学生さん	1.538
先生方	1.211	1回	0.738	校舎	1.296	学生	0.923
勉強	0.910	3回	0.738	校内	1.296	嬉しい	0.923
来る+したい	0.910	広い	0.738	聞く	1.296	聞く+できる	0.923
参加	0.608	湘北ナビ	0.738	話しかける	1.296	雰囲気	0.680
説明	0.313	前回	0.738	学生	0.668	3回	0.615
先生	0.006	大学	0.738	嬉しい	0.668	わかる	0.615
		学生さん	0.542	参加	0.373	スタッフ	0.615
		2回	0.271			楽しむ+できる	0.615
		くわしい	0.271			思う	0.615
		学ぶ+したい	0.271			先輩たち	0.615
		校舎	0.271			勉強	0.372
		校内	0.271				
		話	0.271				
		話しかける	0.271				

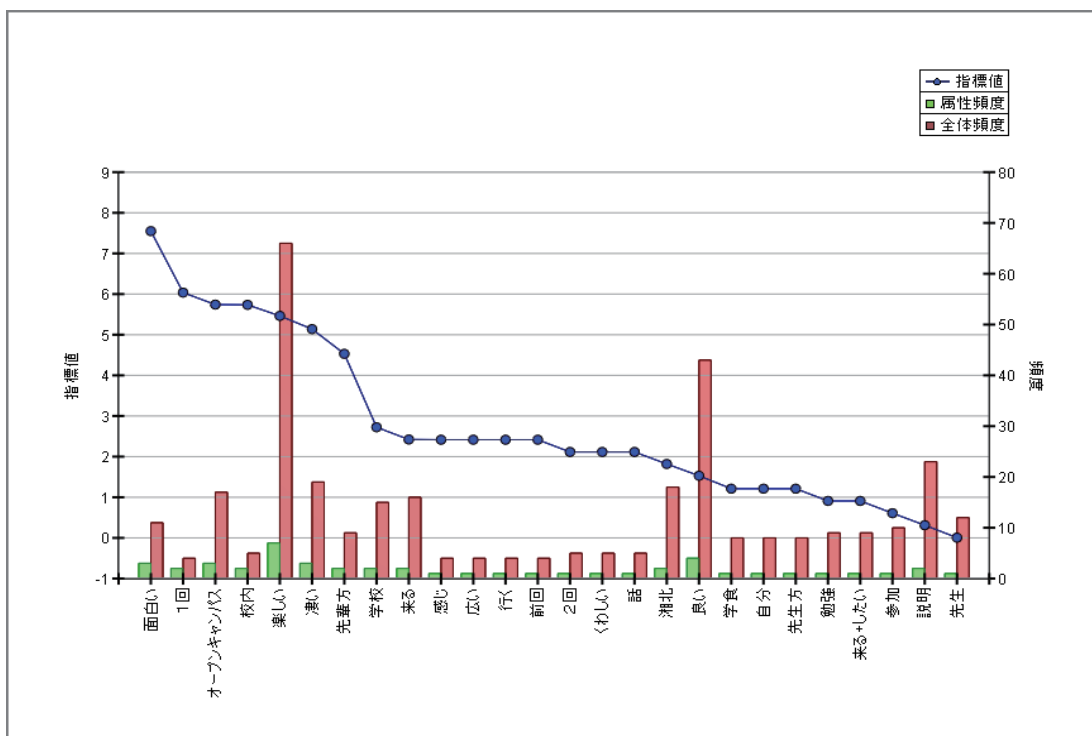
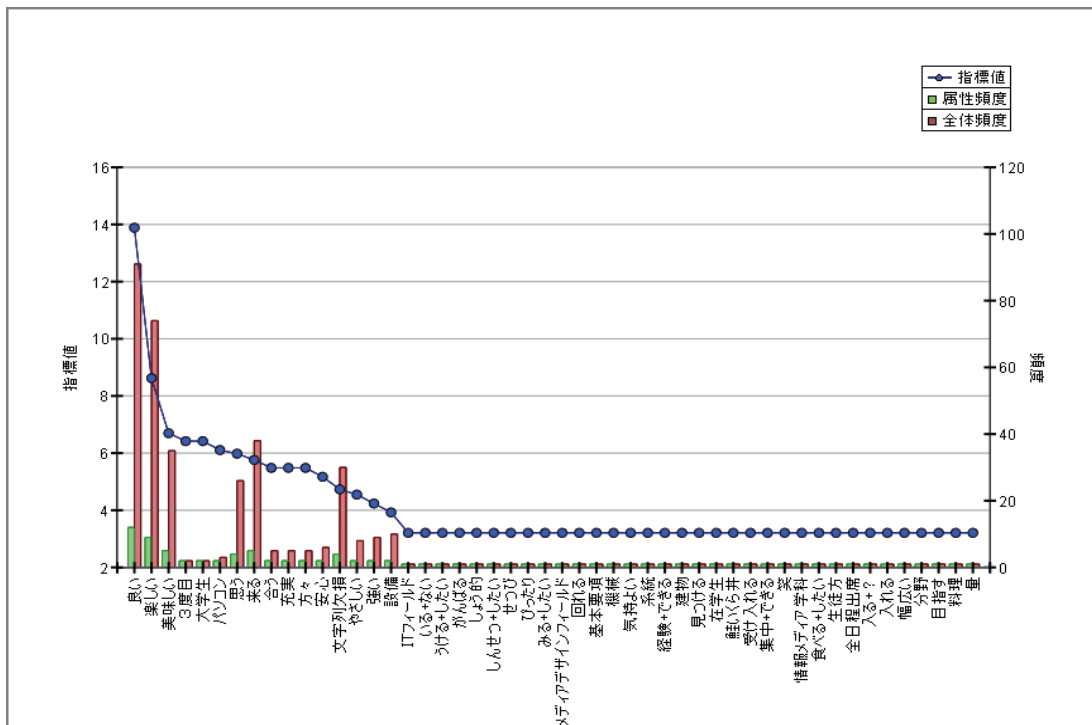


図3 特徴語抽出 (情報メディア学科) 上: 2009年6月6日、下: 7月25日

オープンキャンパスアンケートデータのテキストマイニング

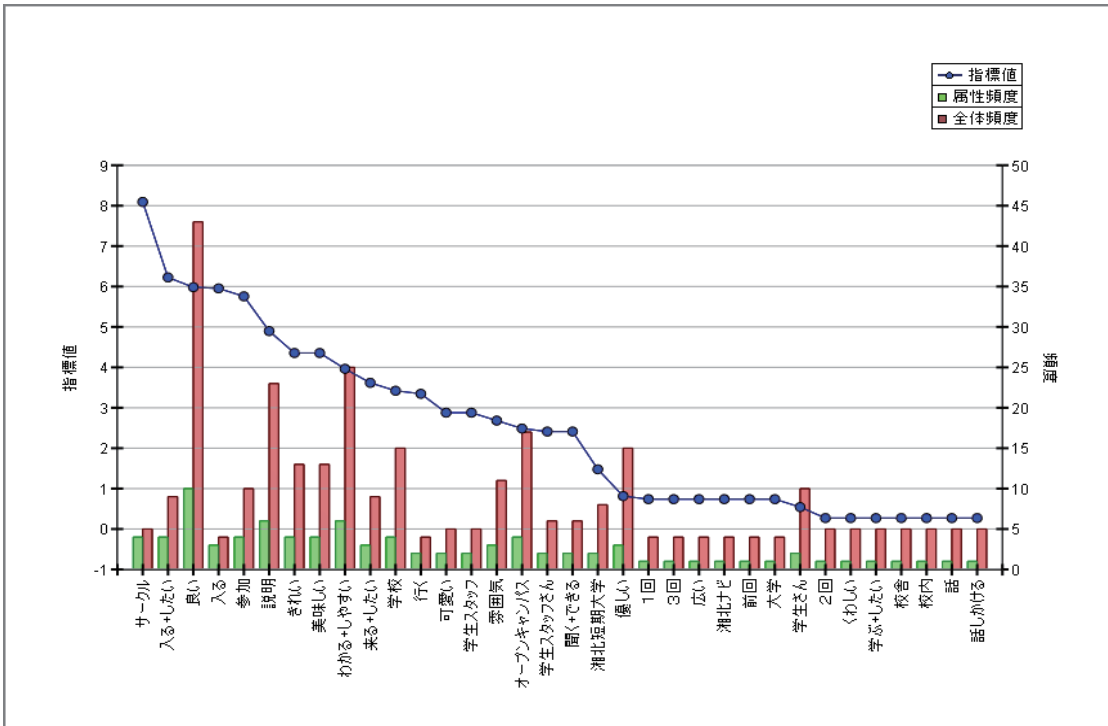
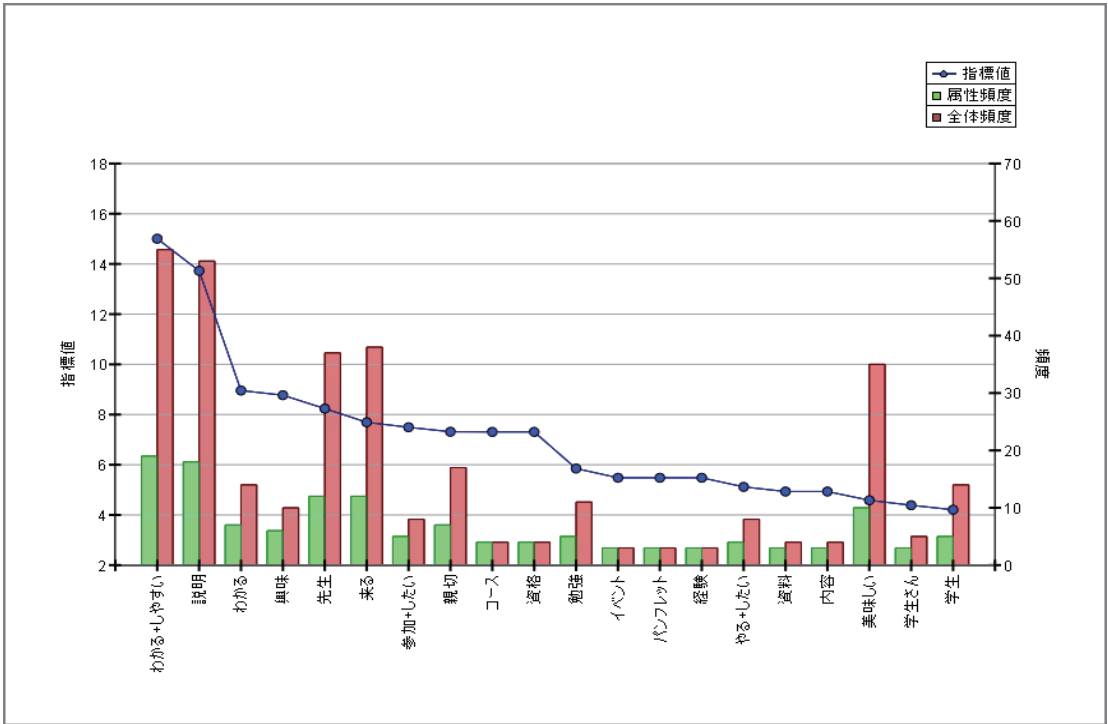


図 4 特徴語抽出 (生活プロデュース学科) 上: 2009年6月6日、下: 7月25日

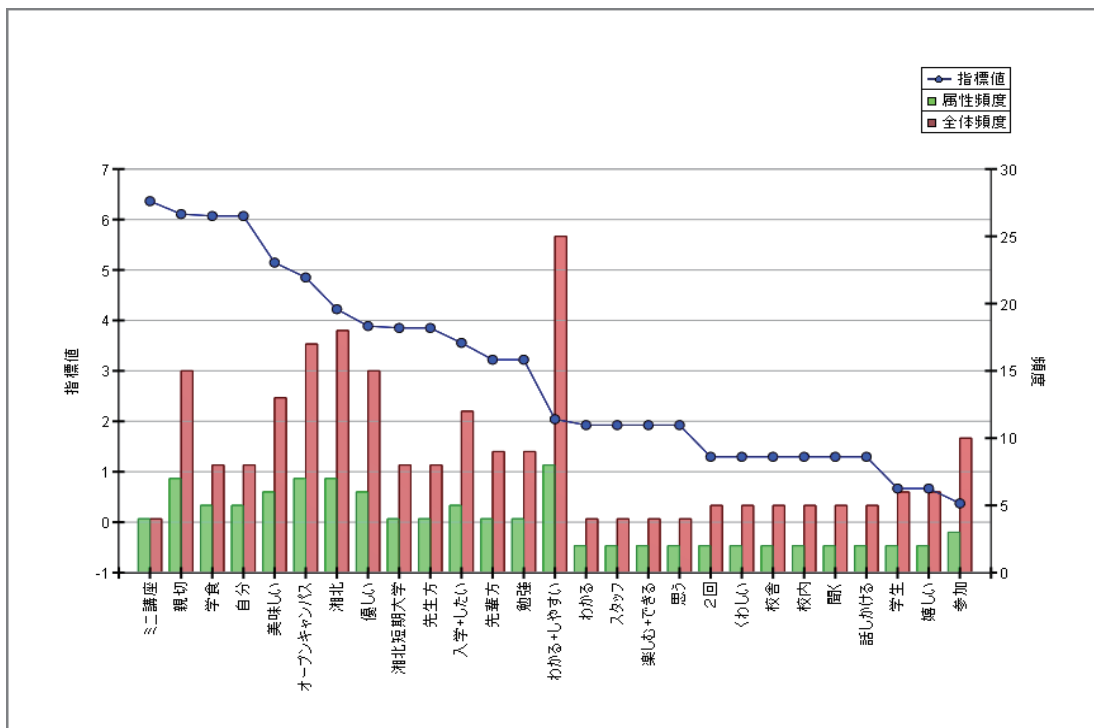
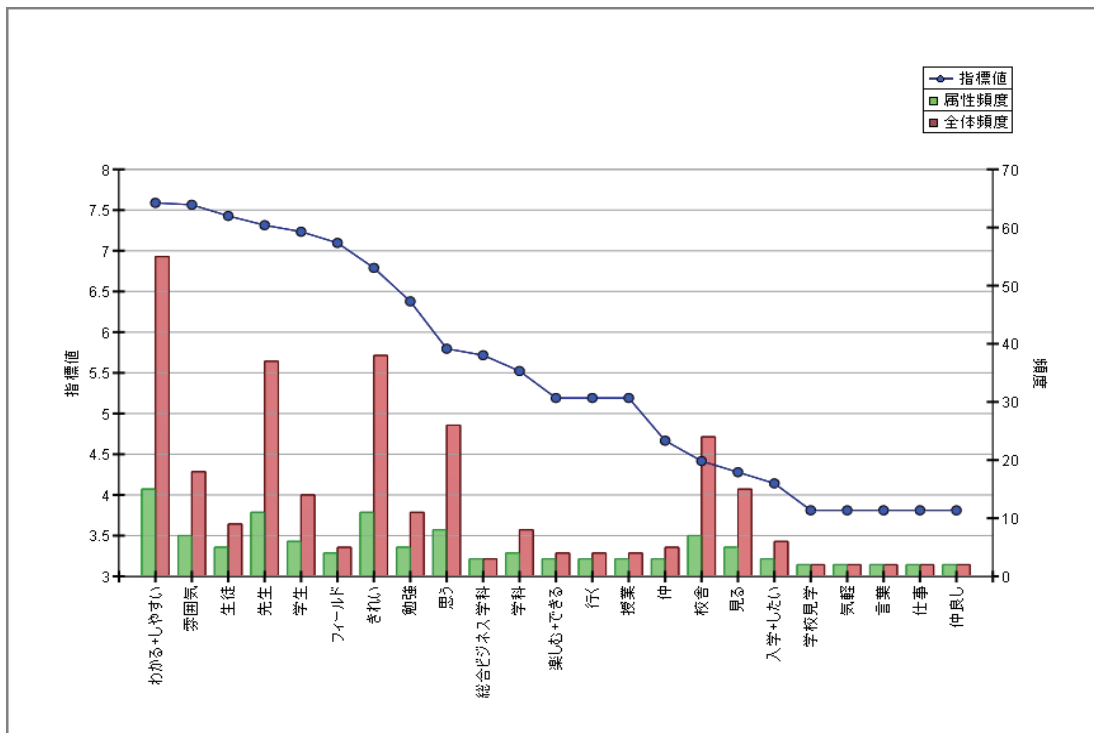


図5 特徴語抽出(総合ビジネス学科) 上:2009年6月6日、下:7月25日

オープンキャンパスアンケートデータのテキストマイニング

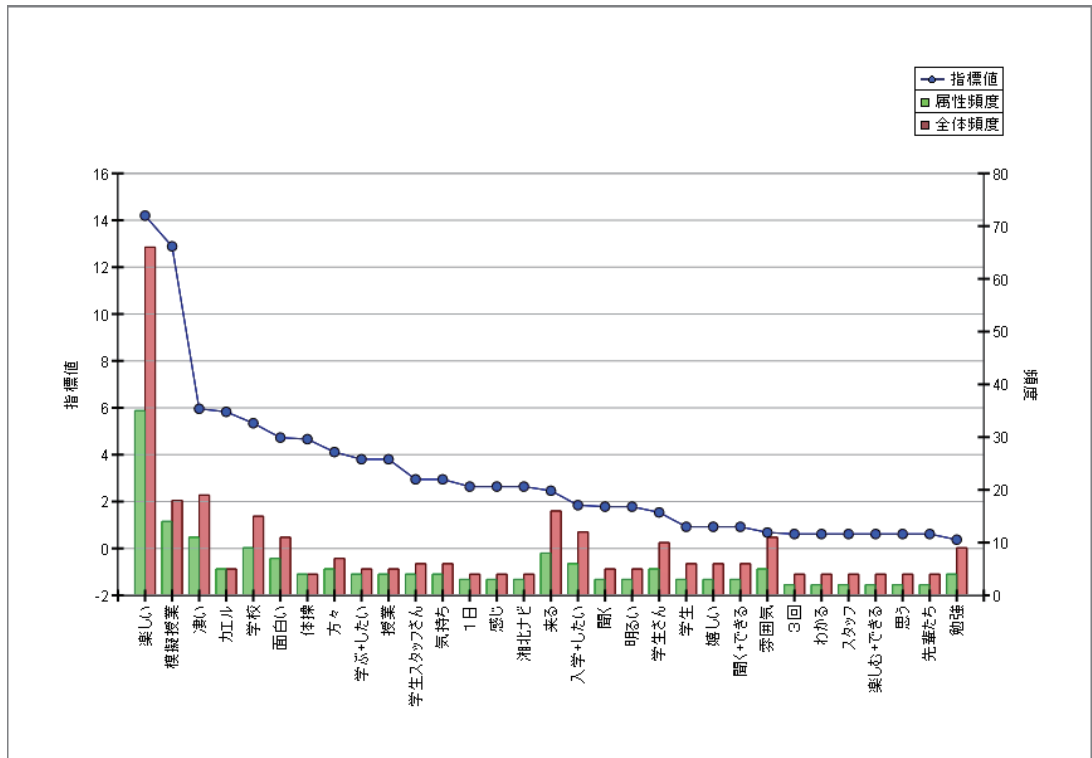
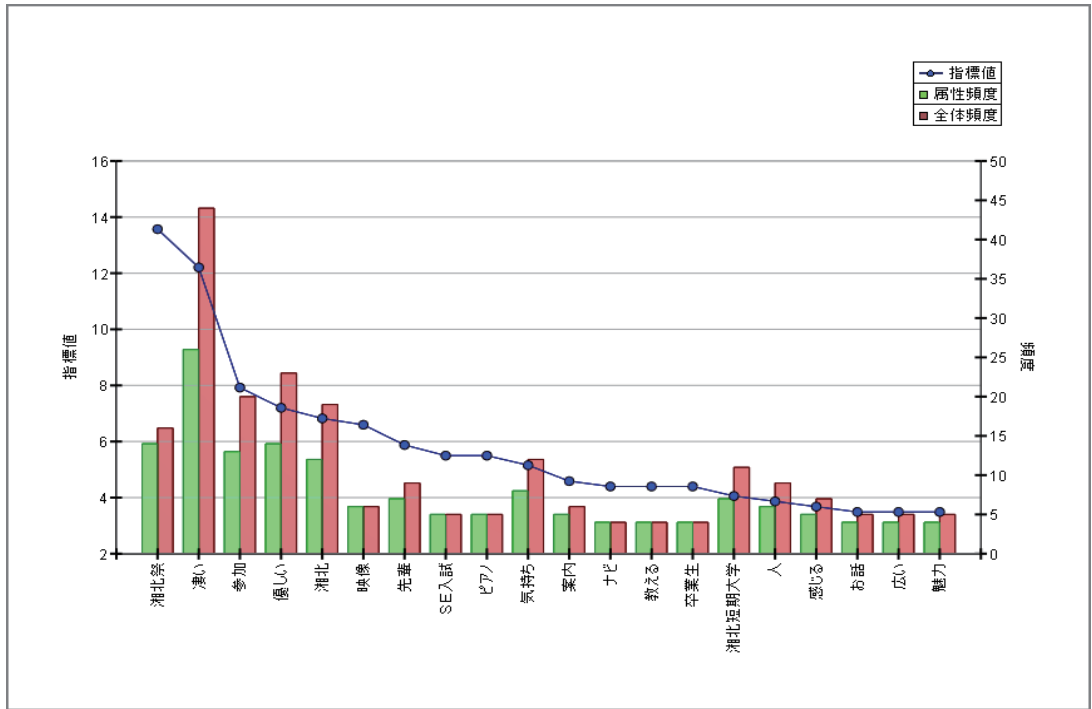


図6 特徴語抽出 (保育学科) 上: 2009年6月6日、下: 7月25日

### 3.4 ことばネットワーク

コメントは全体的に本学に対して「ポジティブ」に語られていることは明らかであるが、ことばとことばの関係（係受け解析）をネットワーク図で表現することによって、さらにことばと属性の共起分析よりイメージ分析が可能になる。また、係受け頻度と属性の頻度分析も可能である。つまり、話題の中心とその属性の関係がより明確になる。7月の全コメントを対象としたことばネットワークを示す（図7）。矢印で共起関係（名詞-形容詞、形容動詞、動詞）を表し、太さで頻度（支持度）を示す。ここでは、最低信頼度が50、出現回数が2

回以上の共起ルールを抽出している。

「楽しい」ということばに対して、多くの共起関係があることがわかる。この中で興味深いものとして「忙しい」→「楽しい」が少なからず存在することが挙げられる。原文を参照すると、「大学は忙しい」であることが分かった一方、「学生は楽しそう」で、大学の生活が正確に伝わっていることがわかる。「楽しい」と「良い」に共起関係があるが、この2つのことばの仲介となっている「パンフレット」「湘北ナビ→安心、優しい」の存在も興味深い。

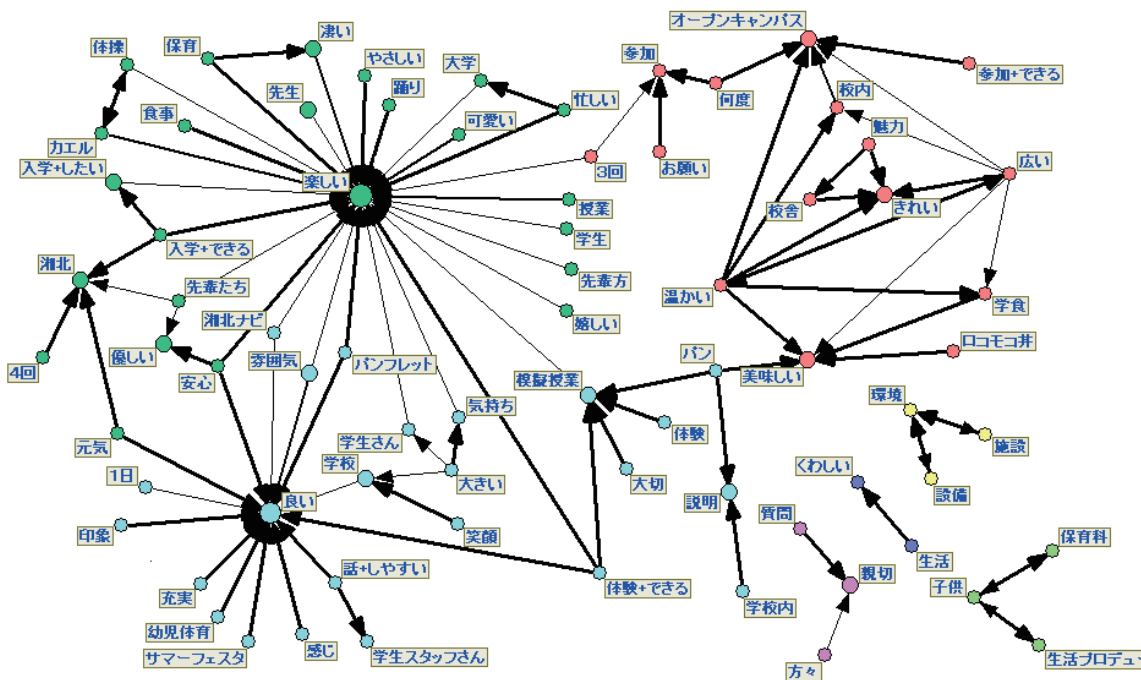


図7 ことばネットワーク図

### 4. おわりに

平成21年度の研究では、平成20年度で試みた6.7月のオープンキャンパスにおけるアンケートのフリーコメント部分のテキストマイニングを、21

年度アンケートの回答の分析にも適用した。この結果、以下が確認できた。

- ①6月の来場者の生徒の6割以上は本学への進学は未定であるが、7月の参加生徒の約半数が本学を第一志望に考えている。

- ②学科に対する印象は昨年と類似した傾向を示す。生活プロデュース・総合ビジネス学科では、「説明がわかりやすい」、や「先生」、が学科の特徴を表しており、保育学科希望者は「凄い」や「気持ち」ということばを用いることで、入学意欲の強さを表す傾向がある。情報メディア学科では、「面白い」ということばで「楽しい」よりも知的好奇心が喚起されたことが表現されていると思われる一方、「親切」や「優しい」という「人」に関する印象を表す語が少ない。
- ③7月のオープンキャンパスは模擬授業が実施されていたので、学科ごとに異なる特徴語が抽出されており、本分析が有効であることが確認できた。
- ④保育学科の「湘北祭」という語が抽出されたことで、湘北祭におけるパフォーマンス映像の放映が来場者に効果的に作用したことがわかった。この例から、イベント等の効果の測定にも利用できることが示唆された。

- 文堂 (2005)
- 3 R. Ozao, T. Sawaguchi, H. Ishida, M. Iji, T. Furuyama, Y. Shinohara, K. Yamada, K. Halada, "Eco-MCPS: A newly developed web-based database for eco-materials in Japan", Mater. Trans, 48 (12) pp. 3043-3049 (2007) .
- 4 H. Ishida, R. Ozao, T. Utsumi, Y. Shinohara, K. Halada, Y. Nishimoto, "Trends in Eco-materials and Products as Observed through Studies on a Web Database, Eco-MCPS", Trans. MRS-J, 34(2) pp.249-252 (2009) .
- 5 小棹理子・石田英弥、「湘北短期大学ウェブサイトのアクセスログから見られる閲覧者の各学科への興味の違い」、湘北紀要 29、pp1-6 (2008)
- 6 小棹理子・石田英弥、「テキストマイニングのアンケート解析への応用の試み」、湘北紀要 30、pp83-95 (2009)

## 謝辞

平成21年度湘北短期大学助成金を得て実施された研究結果の一部を報告したものです。ここに謝意を表します。

## 注

- 1 那須川哲哉、「テキストマイニングを使う技術／作る技術」(東京電機大学出版局 (2006)
- 2 マイケル・J.A. ベリー & ゴードン・S. リノフ (著)、江原 淳、上野 勉、河野 順一、佐藤 栄作、朝稲 努 (訳)、「データマイニング手法—営業、マーケティング、CRM のための顧客分析」 海

## Text Mining for Successful Open Campus Events

OZAO Riko SEKI Yutaro

### **[abstract]**

Text data mining was applied to text data obtained as comments from the response to the questionnaires given to open campus participants. The analyzed results were compared with those obtained last year. Surprisingly, the results showed tendencies similar to those of the last year; the ratio of the potential applicants change from June to July, and the every department left its own impression on the high school students.

### **[key words]**

Text data mining, text clustering, concept/entity extraction, sentiment analysis, public relations