

〔研究ノート〕

## テキストマイニングによる 短期海外研修の自由記述の分析

飯塚 雄一  
ケイン・エレナ  
小玉 容子  
松本 亥智江

1. 目的
2. 研究方法
  - (1)対象
  - (2)方法
3. 結果—頻度分析—
  - (1)単語頻度解析
  - (2)係り受け頻度解析
  - (3)文章分類
  - (4)評判分析
  - (5)対応分析
  - (6)希望、不満の分析
  - (7)特徴分析
4. 考察

### 1. 目的

近年、学生の短期海外研修への応募が減少傾向である。高い研修費用をどうするかという検討も重要であるが、どうすれば学生の短期海外研修へ興味や関心を高めてやるのか方策を検討する必要がある。1996年から出雲キャンパスで始まった語学・看護学海外研修及び浜田キャンパス、松江キャンパスでの研修に対する学生の感想文（自由記述）を資料として検討する。

本研究では、短期海外研修の自由記述の報告書を、最近の新しい手法であるテキストマイニングによって検討した。複数年度にわたる大量の自由記述感想文の中から有用な情報を選別し取り出すために、テキストマイニングによって大量の発言内容をコンピュータにより処理すれば、代表的な感想や傾向を把握することができる。また処理された言語は数量化されているので様々な分析にかけることができる。これを活用することで、学生たちの希望や不満などを把握することもできる。また学生たちがどのような収穫を得たかを探

ることできる。

そこでこれまでに浜田、松江、出雲の各キャンパスで海外研修へ参加した学生たちがどのような感想を述べているかを調べる。つまり出雲キャンパス、松江キャンパス、浜田キャンパスの既存の海外研修報告書の自由記述感想を、使われている単語という側面からイメージを分析する。更に年度やキャンパスによる違いがあるかなども調べる。海外情報がマスコミなどであふれている現状で、実際に海外へ行くことから得られるものが何かを探り、これを学生への動機づけの一助として活用できないか検討する。

## 2. 研究方法

### (1)対象

鳥根県立大学、短大部出雲キャンパス、松江キャンパスで実施された短期海外研修の報告書を対象とする。松江キャンパス1991年度（ワシントン州エレンズバーグとシアトル）、鳥根県立大学は2007年度（カリフォルニア州モントレイ）、出雲キャンパスは1996、1998、2000、2003、2006、2007の各年度（ワシントン州ワナチとシアトル）の報告書を取りあげた。なお、松江キャンパスの記録は、テキストマイニングの分析対象年度1991年に加え、1992、1997、1998、2002、2007、2008の各年度（ワシントン州エレンズバーグとシアトル他）の報告書も参考資料とした。但し、2007年度と2008年度は、学年暦の変更により研修期間が25日間から19日間となり、報告書（日誌）枚数もそれに従い減っている。3キャンパス間の報告書の数（語数）には入手可能性の違いにより大きなばらつきができたり、参加者は女性が多数である、などのため、本報告ではキャンパス、性による違いよりも、主に、全体的な傾向を把握することにした。

### (2)方法

質的データ解析はテキストマイニングのソフトウェアであるTMStudio 3.0（数理システム）を用い分析した。テキストマイニングでは、記述された文章を分解し、分解した記述語ひとつひとつを変数とみなし、数量データと同じように扱っている。すなわち、自由記述文から得られたテキスト型データをまず分かち書きし、単語（構成要素）に分ける（形態素解析）。例えば、「私はアメリカに行きました」という自然言語文に形態素解析を実行した場合、「私（名詞）、は（助詞）、アメリカ（名詞）、に（助詞）、行き（動詞）、まし（助動詞）、た（助動詞）」のようになる。さらに互いに依存関係（係り受け関係）にある文節の組を作り（構文解析）、特徴表現分析や文章分類、評判分析などを行った。

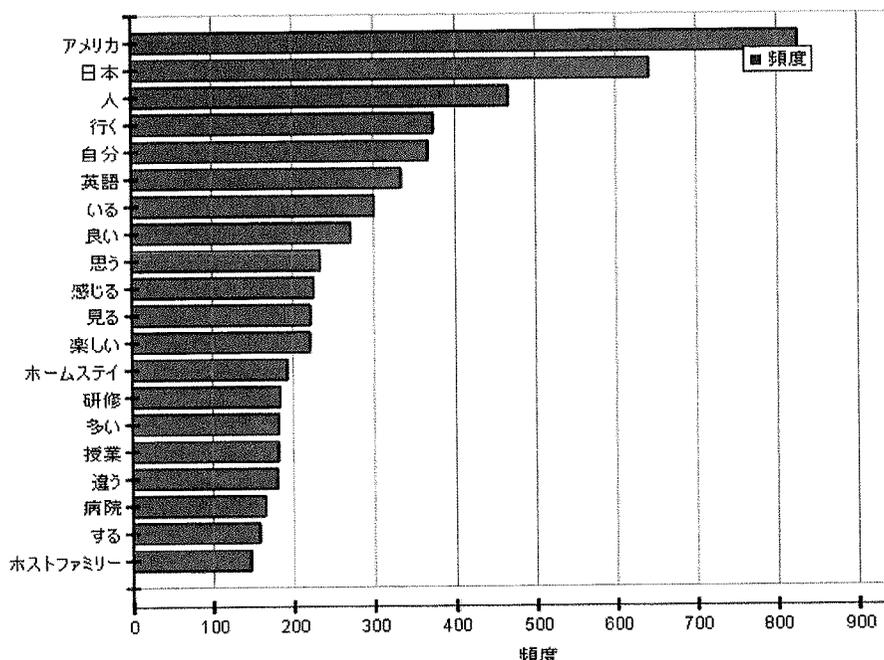
## 3. 結果—頻度分析—

### (1)単語頻度解析

テキスト情報の品詞別出現回数では、頻出順に、名詞、動詞、副詞、形容詞、などで感動詞も多かった。

まず、報告書にどのような単語が多く出現しているかをみるために、単語頻度解析を行った（図1）。すべての報告書を合わせて最も頻度の多い単語は、「アメリカ」、「日本」、「人」、「行く」、「自分」、「英語」などの単語が多く出ていることがわかる。

図1 単語出現頻度



この頻度解析をもとに、松江キャンパスの記録で使用されている使用頻度の高い語を使用回数という面からさらに比較検討した。表1はワード (Word) の検索機能を用い、使用回数を年度別にまとめたものである。使用回数が2桁になっているものを取り上げたが、比較検討のため、同様の語で使用頻度の低いものも表に入れている。

17年間にわたっているが、時の経過による使用語の特徴的变化は見られず、多少の偏りは参加者の属性によると考えられる。感情を表す語としては「楽しい・嬉しい」が、感想を表す語としては「良かった」が高い頻度で用いられている。さらに、研修中の活動としては、「授業」の他に、「話す」「買う」「食べる」ことへの関心の高さがみられる。また、「体験」と結びつくと考えられる「貴重な」は、まず「初めて」体験し、「貴重な」という価値を計る言葉の使用回数は低くなっている。「すごく」という副詞の使用回数の高さも、実体験が心、気持ちに与える刺激の高さを示していると考えられる。

表1 単語出現頻度 (松江キャンパス)

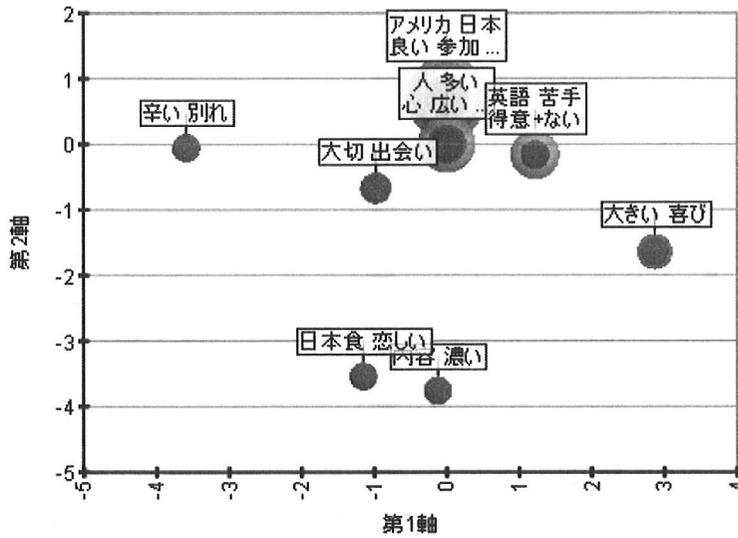
	1991	1992	1997	1998	2002	2007	2008
英語 English	4	6	4	8	15	10	6
アメリカ (人)	16	36	6	11	30	7	14
日本 (語)	10	30	13	22	27	15	19
授業	20	35	4	16	27	16	13
楽しい	15	13	13	13	31	24	22
うれし (嬉し) かった	11	7	2	6	11	3	11

良 (よ) かった、良 (よ) い	18	16	11	11	19	20	20
おもしろかった+ 興味深かった+感動	2+3+2	5+1+1	0+0+0	3+0+0	1+4+7	0+1+2	0+0+2
すごく	17	17	5	14	33	19	9
大きい (な) (でかい, big)	6	6	2	3	15	5	6
小さい (な)	4	0	0	2	3	2	1
いろんな+いっぱい	4+4	3+4	5+4	0+7	8+5	1+2	6+5
貴重な (大切な)	0	1	1	1	2	1	1
経験 (体験)	2	6	4	3	5	5	5
初めて	5	10	9	5	9	6	1
話す (した)	22	15	17	20	39	10	25
買う (買い物、 ショッピング)	23	26	19	26	36	8	16
食べる (食事など)	38	24	20	30	40	13	21
店	8	13	9	7	11	5	4
歩く	6	18	1	11	12	0	2
疲れ (て、る)	2	5	11	8	7	2	3
寝て	7	7	12	11	3	6	9
ありがとう (感謝)	1	2	7	0	8	5	6
友達	9	4	3	11	6	3	2
もっと	2	2	0	0	9	6	6
泣く (涙) + 悲しい	1+0	7+4	4+3	0+0	1+0	1+0	3+4

## (2)係り受け頻度解析

単語頻度解析では、大体のイメージをつかむために1語に絞ってみた。さらに詳細に内容をみるため互いに依存関係(係り受け関係)にある文節の組を作る。感想文の中に表れている係り受け表現について、係り元単語と係り先単語の頻度を求め、対応分析によりデータを2次元上に配置してみた結果である(図2)。対応分析では関連のあるものは、近い点に配置される。「参加は良いもので、喜びも大きく、出会いが大切である」、そして、「別れは辛い」、「英語が苦手」というメッセージが読み取れる。

図2 係り—受けの関係



(3)文章分類

次に、感想文の中のすべての文を、言葉の使われ方が似ているもの同士の5つに分類してみた。文章分類とは、使われている単語に応じて、テキストの行もしくは文章をクラスターに分類し、クラスターごとの特徴単語を表示することである。分類1は「自分が行って見るのは良い」というクラスター（図3）、分類2は、「アメリカへ自分が行く」というクラスター（図4）、分類3は、「英語の授業やホームステイは楽しい」というクラスター（図5）、分類4は、「人」などの単語を多く含む文章の集まり（図6）、分類5は「日本とアメリカの病院は違う」というクラスター（図7）があった。

図3 文章分類1—自分

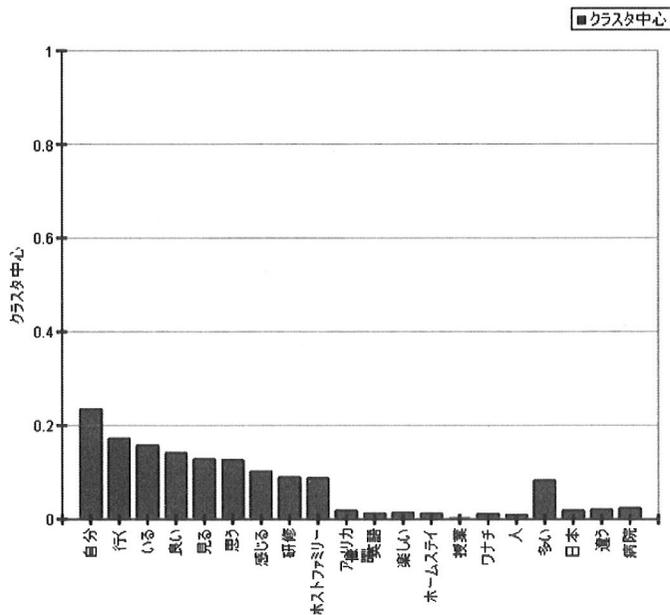


図4 文章分類2—アメリカ

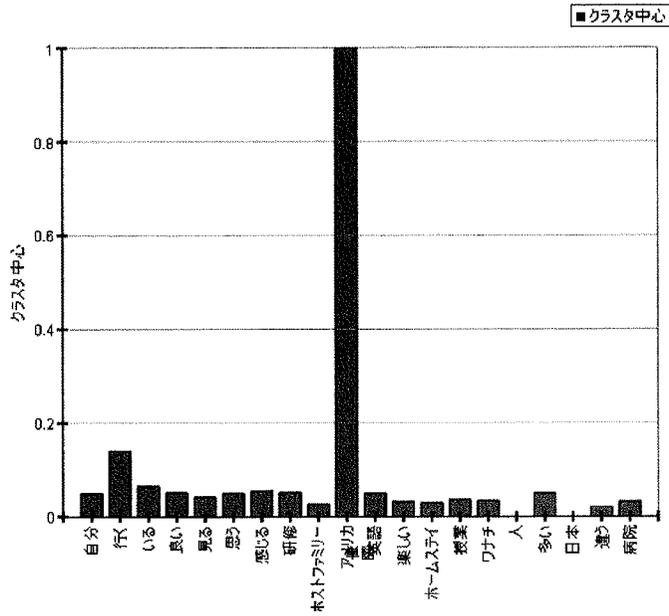


図5 文章分類3—英語

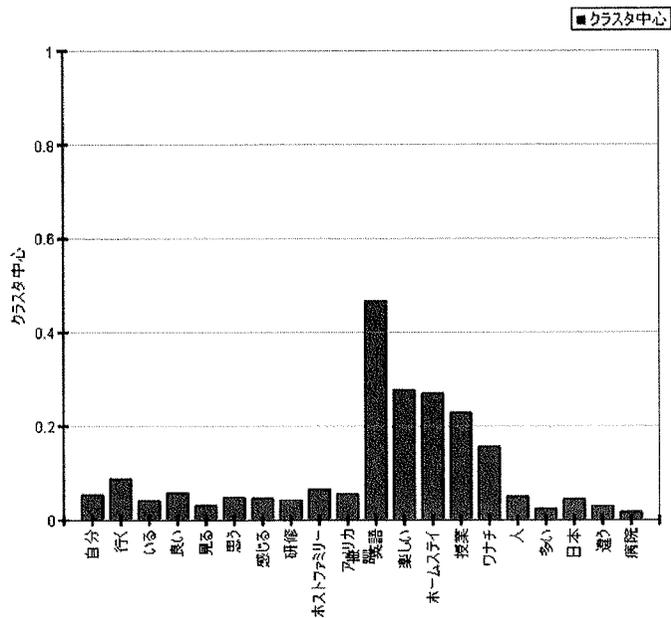


図6 文章分類4—一人

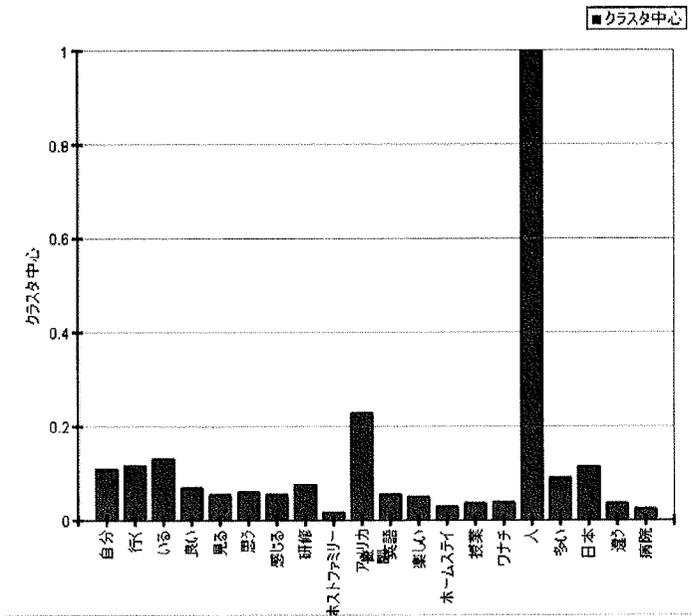
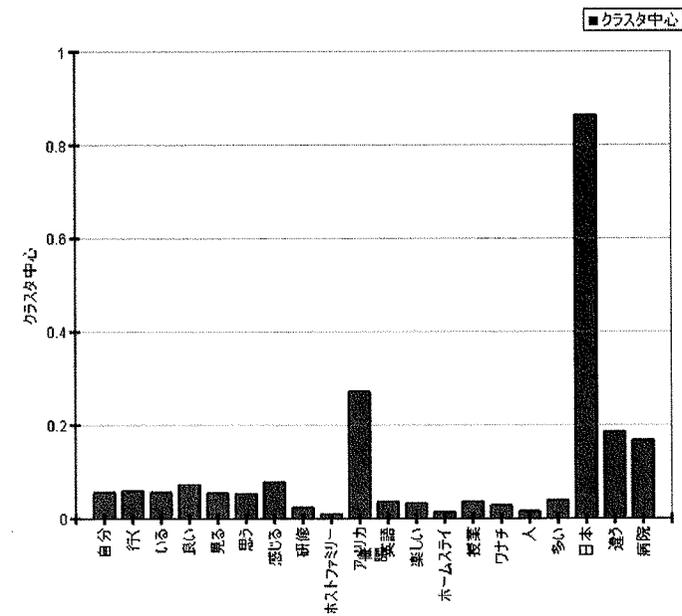


図7 文章分類5—日本



(4) 評判分析

次に、良いイメージで語られている単語、悪いイメージで語られている単語を調べてみた（評判抽出）。これは、単語に対して好意的な表現や非好意的な表現それぞれ語られた回数をカウントし、それをもとに、好評語・不評語のランキングを作成する。つまり単語

に対して、好意的な表現、否定的な表現、それぞれで語られた回数を数え、それをもとに単語の、肯定度、否定度の順位を表した（図8、9）。人、体験、思い出、アメリカ、などが肯定的にみられ、別れ、英語、英語力等の語は否定的にみられている。

図8 好評語ランキング

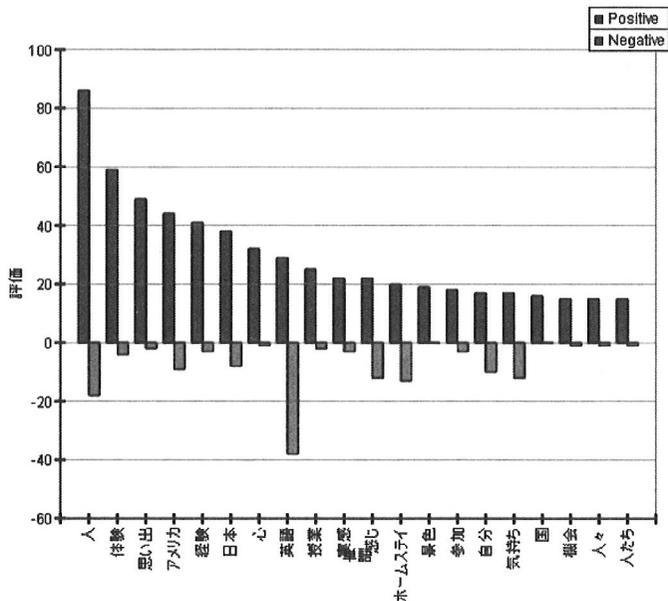


図9 不評語ランキング

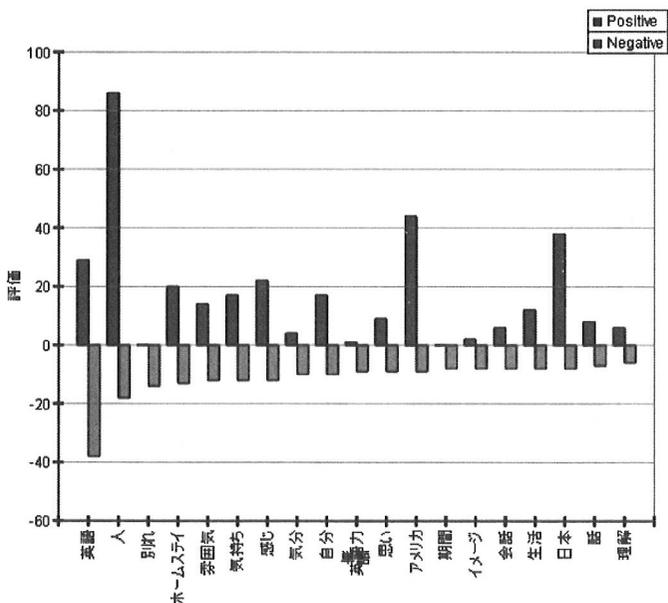


図9において「英語」は、不評語として一番にランクされているが、好評語でも8番目にランクされていて、その両方で肯定的にも否定的にも表現されている。図2では、「英

語は苦手」と言う言葉の使い方がされているように、苦手意識を持たれている英語だが、一方、図5の文章分類では、「英語、ホームステイ、授業」などのクラスタに「楽しい」が含まれている。これらの結果は、実際の記録を見ると、学生の思いの複雑さや変化を知ることができる。例えば、松江キャンパスの第一回目授業はそれぞれ次のように記録されている。

7/24/1992：Language とCultureの第一回目の授業では、挨拶や自己紹介をするが、英語の独特の表現も学び、「それらは文章としても、言葉的にも難しいものではない。調子が良く、歌のようになっている。…Cultureでは、歌を教えてもらった後ゲームをした…(先生の) カーラもキャシーもとてもいい人だった。これならやっていけそうだ。」

7/15/1998：今日は初めての授業があった…初めは何をするんだろうとドキドキしていたけど、なかなかきさくな先生で、みんなでギターに合わせて歌を歌ったりしてとても楽しかった。

7/13/2001：今日が授業開始の日で、Taylor先生のCultureの授業は発音の練習をしたり、アメリカ合衆国の地図を見て各州について学んだりしました…宿題が3つ出され、少し大変だとは思ったけど、頑張りたいと思います。Knight先生のLanguage では、基礎的なことをわかりやすく教えてもらい、家族の紹介などをしました。

7/15/2002：今日は初めての授業でした。緊張したけどとても楽しい授業でした。Language の授業は先生の配られたプリントを見て、ちょっとした会話とかをしました。後、先生の弾かれるギターに合わせて歌を歌ったりして、とても楽しかったです。

8/04/2008：初めての授業はとても楽しかった！サラという先生がとても優しく気さくな人で、みんなリラックス&エンジョイして授業が受けられたと思う。

図8、9の「ホームステイ」も好評語、不評語それぞれで、「英語」と同じタイプの受け止められ方をしている。それは、初めての家で、どのように過ごすのか不安を抱えてスタートしたホームステイが、2泊3日で素晴らしい経験として残ってくることを表してもいる。その実例としていくつかの記録を以下に示す。

7/26/1991：ホームステイ；いつものように9：00から授業を受ける。今日は、二人の先生ともにホームステイの事について話してくれた。ホストファミリーにどんな事を聞いたらいいかとか、家具の名称。これでみんなも少しは、不安が解消したことだろう。4：45に寮の1階の集会所みたいなところに集まって、ホストファミリーに会いに行った。その時、私は、不安と緊張で胸が張り裂けんばかりだった。唯一、私のところは、ホームステイをする人が、私をあわせて3人いるので、それが救いだと思った。

7/28/1991：ホームステイを終えた後の記録；…ファミリー・パーティだったらしくすごく多くの親戚の人たちに会った。ビーチを見下ろしての昼食は、楽しかった。それぞれの、家庭が作ったごちそうをセルフサービスで食べた。皆さん、仲がとてもよく、私達まで家族の一員とみなしてくれたし、もちろん、私もその一員になった感じがした。Williams Familyはすごく居心地が良かった。娘のような気がするほど良かった。2泊3日はとっても短かった。1週間Home Stayでも良かったと思う。



図11 要望の頻度

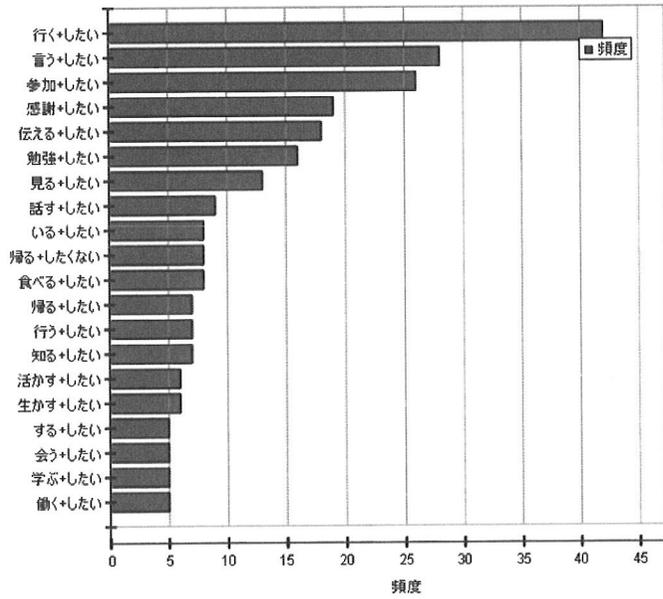
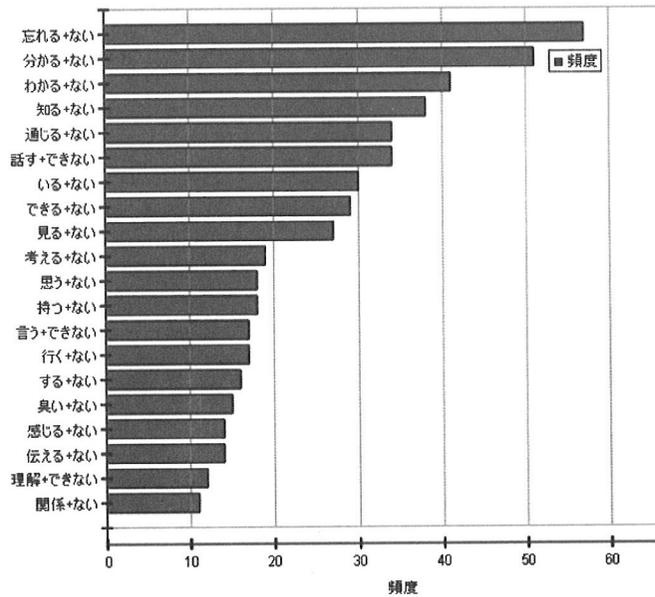


図12 困難、不満等の頻度



(7)特徴分析

次に、各年度別に、特徴的に出現していた単語を抽出してみた。まず、松江キャンパス(1991年度)は、「エレンズバーグ、CWU、Yakima、アメリカ人学生」(図13)などであり、浜田キャンパス2007年度は「日本、アメリカ、自分、国、文化」(図14)などであった。

次に出雲キャンパスの1996年度は「アメリカ、日本、車椅子」など、1998年度は「アメリカ、ワナチ、病院、日本など」(図15)、2000年度は「ワナチ、クライアント、講義」など、2003年度は「アメリカ、日本、専攻科」など、2006年度は「ワナチ、研修、楽しい」など、2007年度は「笑い、良い、行く、楽しい」などが特徴的であるといえる。

図13 松江キャンパス (1991年度) の特徴語

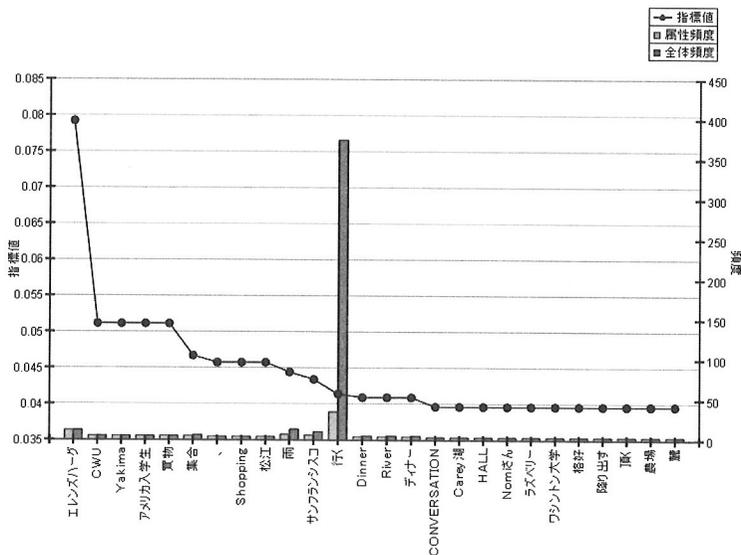


図14 浜田キャンパス (2007年度) の特徴語

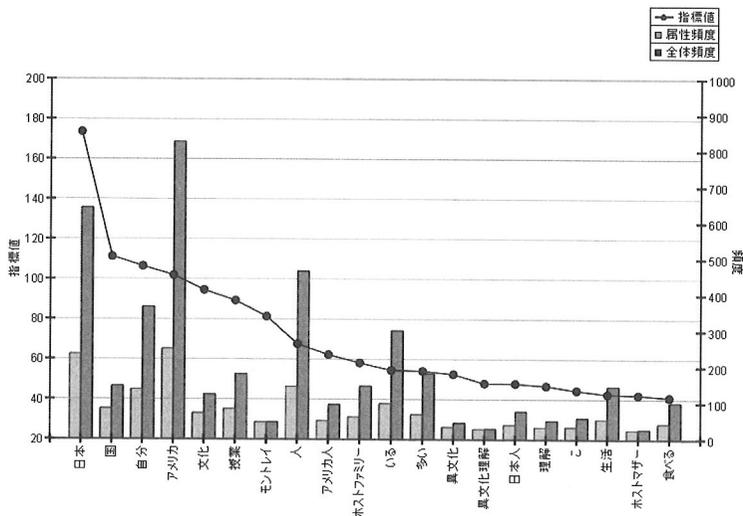
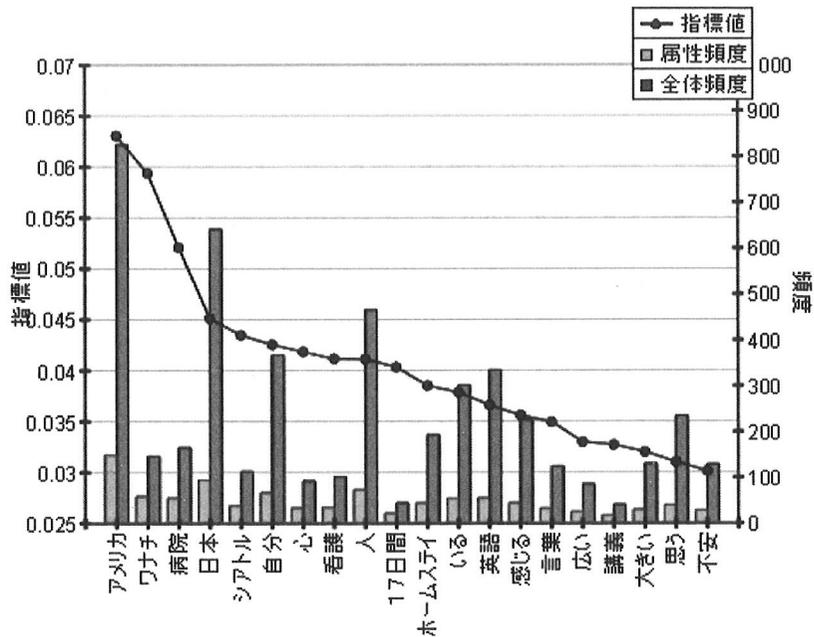


図15 出雲キャンパス（1998年度）の特徴語



#### 4. 考察

単語頻度解析によると、すべての報告書を合わせて頻度の多い単語は、「アメリカ」、「日本」、「人」、「英語」などであった。アメリカ研修後の報告書なので、これらの言葉が上位に来るのはうなずける。また、単語の品詞頻度解析の結果、名詞、動詞、副詞、形容詞などに次いで、感動詞の頻度の多いのが特徴的であった。これは学生が海外研修によって感動的な体験をしてきたということが推測できる。海外での実地体験をすることが、現在最も欠けていると思われる「感動を与える教育」になっていると考えられる。学生の活動面に関係する語を追加的に拾い上げてみると、授業、ホームステイなどの「話す」という英語に関わるとされる活動の他に、異文化体験として、大学寮での生活、ダイニングでの食事、町へ出たの食事や買い物などが挙げられる。特に松江キャンパスの記録では、表1に示されたように、「買う（買い物）」、「食べる（食事）」などは、「話す」と共に、上位を占めており、実際にアメリカへ行くことで、学生が食事や買い物を通して、アメリカでの生活を積極的に満喫していることがうかがえる。これらは、単に「英語」の情報だけでは得られない「経験・体験」の大きな部分を構成している。英語が苦手だから参加を見合わせるのではなく、やはり、心配するより参加をうながし、「だまされたと思って、感動するからとにかくいってみなさい」などと言って動機づけてやるのがよいのかもしれない。係り受け分析からも「参加しては良かった、喜びも大きかった」、そして「別れが辛かった」などというメッセージが代表的なものであった。このように、単語分析の結果は直接的な国際交流、実地体験の重要性を物語っていると思われる。

次に自由記述文に散在している好意的な印象と非好意的な印象を選別するために評判分析を行った。海外研修を終えて、多数の学生が、楽しかった、きてよかった、よい体験をした、などといった、満足感や達成感、喜びなどの肯定的感情の表現をしていた。良いイ

メージで語られている単語には、アメリカ、アメリカ人などがあげられている。しかし非好意的なイメージで語られている単語に、英語や英語力が挙げられていることは注目に値する（図8、図9）。アメリカやアメリカ人に、好意を持ってどんどん話してみたいのに、自分の思うように、意思伝達できない不満、いらだちが表れているようである。例えば、「ホームステイで切実に感じたのが、自分の英語力の低さだった。電子手帳と筆記用具と紙をいつも持ち歩き、わからなければ筆談したりして、わかるまでには疲労も感じたが、通じたときの喜びはとても大きかった。もう少し英語力があればもっといろいろなことが話せたのにと残念にも思う。」というような感想が多くみられた。要望、不満の分析結果（図11、12）からも、積極的な意欲が示されているにもかかわらず、短期間には英語力が伸びない不満も示されている。Ward, Bochner & Furnham (2001) も、短期海外旅行は、愉快で、楽しく、面白い経験であるが、同時にまたストレスをも伴う経験であると述べている（p.130）。このジレンマを解決してやるのが、今後の課題であると言えよう。つまり、海外研修に行く前に、ある程度の英語力を身に付けさせていく必要があると思われる。

報告書の分析から、たとえ短期間のあっても、異文化との出会い、交流が、学生に様々な喜びや感動をもたらしていることが明らかにされた。

## 文献

Ward, C., Bochner, S., & Furnham, A. 2001 *The psychology of culture shock* (2<sup>nd</sup> ed.)

London : Routledge

教理システム *Text Mining Studio* 操作マニュアル (v.3.0)

キーワード：短期海外研修 自由記述 テキストマイニング 単語頻度解析

(IZUKA Yuichi, Eleanor KANE, KODAMA Yoko, and MATSUMOTO Ichie)

## 要約

短期海外研修に対する感想文の単語頻度解析の結果、「アメリカ」、「日本」、「とても」、「人」、「英語」などの感想が得られた。また文章分類の結果「自分」、「アメリカ」、「英語」、「人」、「日本」の5つのクラスターがみられた。全般的にみて、学生たちの短期海外研修の評価としては良好であったと思われる。特徴的な印象語であった英語や英語力に対して肯定的、否定的な相反する評価がありアンビバレントな傾向がみられ、英語に対して好意的な反応がある反面、思うように英語を使えないいらだちがうかがえた。

## A Text-mining Approach to Free Writing concerning Short-term Study Abroad Programs

Yuichi IIZUKA, Eleanor KANE, Yoko KODAMA, & Ichie MATSUMOTO

Using text mining techniques, text data were analyzed from the University of Shimane's three campuses: Hamada, Izumo, and Matsue. The data consisted of 11 years of students' final reports written in Japanese concerning their experiences during short-term study abroad programs in the United States, dating from 1991 to 2008. The reports were scanned and then analyzed using the text mining software, TMStudio 3.0 to reveal word frequencies and both collocations and colligations. It was found that America, Japan, person, go, myself, and English were the most frequent lexical items and their collocations and colligations were then noted and classified as positive or negative. In general, student evaluations of short-term study abroad were favorable for lexical items such as person, experience, memory, and America. However, items such as understand, know, be understood, speak, and express generally colligated with negative particles, suggesting that future short-term study abroad programs include more language study prior to departure in order to reduce student anxiety. The data also revealed student anxiety concerning study abroad, which frequently clustered with items suggesting that later this anxiety was replaced by enjoyment and satisfaction. With the recent decline in students enrolling in short-term study abroad it is hoped that this data will be helpful to teachers in assuaging student anxiety before departure and thus increase the number of students participating in short-term study abroad.

Key words: short-term study abroad, free writing, text mining, word frequency analysis

