

教育・研究のための自由記述アンケートデータ分析入門:

SPSS Text Analytics for Surveys を用いて

山西 博之
関西外国語大学

概要

質的な言語データ（例えば、自由記述アンケートのデータ）は様々な情報が含まれていることから、教育・研究のための貴重な情報源といえるが、回答者数が多くなると分析に多大な時間と労力が要求される。そのような質的な言語データの分析に有効な手段が、ソフトウェアを使用したテキストマイニングである。本稿では、授業に関する自由記述アンケートデータを分析対象として、テキストマイニング用のソフトウェアである SPSS Text Analytics for Surveys を用いた分析の手順を示す。データ収集、分析、結果提示、解釈の各段階の手順を紹介し、特にデータ収集と分析においては、留意すべき事項を挙げる。

キーワード: SPSS Text Analytics, テキストマイニング, 自由記述アンケート

1. はじめに

著者の勤務している大学（勤務校）においても、学生による「授業評価」なるものが行われている（その名称や在り方の是非を議論するだけでも、もう1つの論考になりそうなので、それについてはここでは触れない）。現時点での勤務校のFD（Faculty Development）においては、その授業評価結果（17項目からなる5件法でのアンケート調査と1項目の自由記述アンケート調査）のうち、第17番目の項目である「この授業に対するあなたの総合評価は」（1: 非常に悪い, 2: あまり良くない, 3: 普通, 4: 良い, 5: 大変良い）の結果がもっとも重要視されているようである。この授業評価はマークシート形式で行われ、評価結果は教学担当部署による集計を経た上で返却される。少しだけ具体的な話をすると、4の「良い」と5の「大変良い」の回答の割合（パーセンテージ）の合計で80%以上を目指すことが、教員の努力目標になっている。このことについては、大学やそのFD組織の考え方があるため、学内において議論を進めていく事柄であるが、著者が指摘したいことは、そのような数値の影に隠れている重要な情報源があるのではないかと、ということである。

それは、上にも述べた「自由記述アンケート」項目の結果である。勤務校の授業評価における自由記述回答欄は、マークシート形式のアンケート調査用紙の裏面にあり、質問

項目は「その他、この授業についてのコメントがあれば、ここに書いてください」というものであり、例として「この授業で良かったと思うこと」「この授業で改善して欲しいと思うこと」「その他思うこと」が挙げられている。この項目に関する著者の考えは第 3.1 節にて「データ収集の際の留意点」として後述するが、この項目はスキャナで取り込みイメージ処理された上で、分析などは経ず、そのまま授業担当者のもとに返却される。この自由記述は、担当する学生の数が増えれば当然ながら膨大な数になる。数値で処理されたマークシート形式での評価項目と異なり、その結果を集計することはままならない。しかし、この自由記述こそ、学生が教員の授業に対して感じた事柄を表す「生の声」であり、その結果に真剣に向き合えば、今後の教育・研究活動の改善につながるものと期待される。

2. テキストマイニング

前節では、ややオフィシャルな授業評価における自由記述アンケートを例にあげたが、外国語教育に携わる我々は、ほかにも自由記述データのような「質的な言語データ」(Qualitative Verbal Data) を入手する機会が多いものと考えられる。授業評価の自由記述アンケートに近い例では、日常的な授業の改善のための学生の質問(リアクションペーパーなど)から得られた自由記述データもあるだろうし、研究対象としての英作文や英文和訳などの文字データも、質的な言語データと呼べるだろう。これらに対する分析方法は多岐に及び、そのままの形で解釈することもできれば、例えば頻度データとしてカイ二乗検定などの統計手法で分析することもできる。これらの方法に加えて、近年の技術的な進歩に伴い、特に注目されているのがテキストマイニング(Text Mining) と呼ばれる分析方法である(石田, 2008)。マイニング(Mining)とは文字通り、採掘(Mine)することであり、テキストマイニングとは、前述した「数値の影に隠れている重要な情報源」である質的な言語データ(テキスト)の中に埋もれている情報を掘り起こし、活用するための方法である。テキストマイニングにおいては、まず得られた言語データを品詞や活用といった形態素レベルや主語と述語の係り受けなどの構文レベルで解析し、その結果を分析や分類して提示するといった作業が行われる。

テキストマイニングにおける、このような一連の作業を行うためのソフトウェアは、有料のものから無料のものまで様々なものが開発されている(本稿では、日本語を主たる分析対象言語とするソフトウェアを挙げる)。有料のソフトウェアの一例を挙げると、総合的なデータマイニング用ソフトウェアである SPSS 社(現 IBM 社)の SPSS Modeler(以前、Clementine と呼ばれていたもの)は、Text Mining for Clementine というアドオンを加えることでテキストマイニングが可能である。しかしながら、SPSS Modeler は、個人での使用よりも企業等での顧客データの分析を主な使用用途としているようである(このような企業での顧客データ分析等を想定したソフトウェアには、数理システム社の Text Mining Studio などがある)。一方、教育現場で教員が個人で使用することを想定して

いる有料のソフトウェアとしては、Just Systems 社の MiningAssistant / R.2 というものがある。MiningAssistant / R.2 は、他のソフトウェアと併せたパッケージ (TRUSTIA / R.2) の形で販売されている。このパッケージには、教育現場の教職員や学生を対象とするアカデミック版も用意されている。無料のテキストマイニングのソフトウェアとしては、形態素解析のための MeCab (和布蕪) というソフトウェアと、これも無料の統計ソフトウェアの R を組み合わせた方法 (RMeCab) がある (RMeCab に関しては、石田 (2008) および石田基広氏のウェブサイト (<http://rmecab.jp/wiki/index.php?RMeCab>) に詳しいので、そちらを参照されたい)。

これらのソフトウェアは、有料のものであれば価格 (主に企業等での使用を想定しているものは、個人での購入には躊躇する金額であるかも知れない)、使用できる PC や OS の種類、扱う事ができるデータの種類、得られる分析結果の多様性、あるいは使用する際に生じた疑問に対するサポート体制などがそれぞれ異なるため、自らの教育・研究の目的等に応じて、選択 (あるいは使い分け) することになる。

本稿で取り上げるソフトウェアは、SPSS 社 (現 IBM 社) の SPSS Text Analytics for Surveys (以下、SPSS Text Analytics) である。このソフトウェアは SPSS Modeler の機能のうち、テキストマイニングに特化したものといえる。ソフトウェアのウェブサイト (http://www.spss.co.jp/software/pasw_tafs/) によると、「アンケートの自由回答をかんたんに、そしてすばやくコード化、視覚化を可能に」し「自由回答の分析価値を高めるテキストマイニングツール」であり、以下の3点が特徴として挙げられている。

- 1 専門ツールの迷うことのない直感的で明快な操作性
- 2 圧倒的な語彙数を誇る標準辞書と更新の簡便さ
- 3 Excel や IBM SPSS Statistics へ簡単エクスポート

著者はテキストマイニングの専門家ではないので、明快な操作性をもつ「専門ツール」であること、形態素解析や構文解析の辞書が充実していること、そして普段使用している SPSS Statistics と連携していることに魅力を感じ、このソフトウェアを選択した。本稿では、このソフトウェアの使い方を紹介しつつ、授業に関する自由記述のアンケート結果に対して、データ収集、分析、結果提示、解釈の各段階での手順を挙げるものとする。特に、データ収集と分析においては留意事項を述べる。

3. 自由記述アンケートデータの分析

3.1 対象データ・データ収集方法

本稿では、上述の通り、授業に関する自由記述のアンケート結果を分析対象とする。今回用いるデータは、田中・山西 (印刷中) において報告された、CALL 教室 (254 名)

と普通教室（77名）での英語音声指導に関する自由記述アンケート結果である（Appendix 参照）。この自由記述アンケートの回答人数は、計 331 名であり、さらに、後述の通り自由記述の項目数は 4 項目であるため、手作業での集計や分析には労力が要求されるといえる。田中・山西論文では、このデータに対する分析結果としては、紙幅の都合上 Appendix に挙げた図による結果提示と、それらに対する解釈を示すに留まったが、本稿ではこの結果を得るに至ったプロセスを示していくものとする。

授業の詳細な説明は割愛するが、田中・山西論文の概略としては、英語を専攻していない大学 1 年生に対して、共通教科書によるリスニング指導に加え、著者らが準備した Sound Focus と呼ぶ英語音声学・音韻論的な解説と演習を含んだ教材による指導を半期（15 回）行い、その効果検証をテストによるリスニングパフォーマンスの変化とともに自由記述アンケートによっても測定しようとした試みである。授業は著者の前任校で行ったものであるため、本稿冒頭の勤務校の事例とは若干異なるものの、大学がオフィシャルな形で用意している「授業アンケート」（前任校での「授業評価」の名称）では全体的な授業の感想（良かった点、改善点）を尋ねることができるのみである。そのため、特定の新しい指導や各々の授業科目に特化した指導に関する学生の感想を得ることによって、教育・研究の改善を目指すならば、漠然とした全体的な質問ではなく、より具体的かつ多角的な質問項目を複数作るべきである。これは、著者が専門とするライティングなどのパフォーマンステストにおいて行われる総合的（全体的）評価と分析的評価の関係と同様に、診断や改善という目的においては、総合的な評価よりも分析的な評価の方が適しているためである（Weigle, 2002）。²

データ収集の際の留意点：データ収集は、複数の質問項目により、具体的かつ多角的に行う

なお、実際に用いられた質問項目は以下の 4 項目であった。データ収集は半期 15 回の授業の最終回に、大学の「授業アンケート」実施に先立って、手書きによる記述方式で行った。得られた手書きのデータは個人情報に十分に注意しつつ、プライバシーマーク取得のデータ入力業者に依頼し、³ 図 1 のように、エクセル（XLS）形式でテキストデータ化した（1 行目には変数名を入力しておく、後の分析時に便利である）。

- a. 役に立ったと思う教材は何ですか。また、どのように役立ちましたか。具体的に書いてください。
- b. 授業で紹介・練習した学習方法（シャドーウィング、ディクテーション、ロールプレイなど）で役に立ったものは何ですか。また、どのように役立ちましたか。具体的に書いてください。
- c. 教室環境（DVD、CD 機材、プロジェクタ、パソコン（ただし、教室の広さやエアコ

ンの性能については除く)) はどうでしたか。

d. 学習環境 (Moodle やインターネットの使用, グループ学習・ペア活動) はどうでしたか。

A	B	C	D	E	F
連番	クラス	FA-1	FA-2	FA-3	FA-4
Q_edu-29	CALL	自分ではあんまりリスニングシャドーウイングとかは、予 快適でした。でも、板書の			グループとかと話して、で
Q_edu-30	CALL	教科書の会話が役に立つシャドーウイングによって聞 広くて、快適でした			ペア活動が多くて、習った
Q_edu-31	CALL	moodleでのリスニングがリシャドーウイングです。発音 問題なかったと思います。ペア活動がとても面白く、			
Q_edu-32	CALL	Sound Focusは何度も聞い シャドーウイング、CDのは、プロジェクトはみやすくてよ			ネット上でリスニングがで
Q_edu-33	CALL	教科書のTOEICのリスニングロールプレイでコミュニケー 前のパソコンで表示されて			とてもやりやすく、仲が深
Q_edu-34	CALL	Sound Focusがリスニング シャドーウイングがスキ、英 ぼっちり。			いいと思う。
Q_edu-35	CALL	Sound Focusが、発音やリシャドーウイングをして、少 スライドが見えにくかった。インターネットの発音のペ			
Q_edu-36	CALL	Sound Focusがリスニング ディクテーションがリスニン ホットボードに書いた字			Moodleでいつでも英語が
Q_edu-37	CALL	TOEICを受けようと思っていない		後ろからは明るいので見	ペアの活動を増やし、別の
Q_edu-38	CALL	Sound Focusが役に立った シャドーウイングは早く読 む			席が決まっているので目
Q_edu-39	CALL	Sound Focusで、少しは英語の発音やアクセントにふ パソコンを使って、聞いてし			ペアでやったり、グループ
Q_edu-40	CALL	教科書の会話文が、楽しリスニングが少しましにな ともよかった。			友達ができてよかった。
Q_edu-41	CALL	単語テストがスピーキング 文章を聞くことでリスニング 充実していた			グループ学習の時パソコン
Q_edu-42	CALL	Sound Focus。発音に興味 Sound Focus リスニング; 問題なかったと思います。グループ活動は、しなくて			
Q_edu-43	CALL	TOEICのリスニングがよか ロールプレイ			いいと思う
Q_edu-44	CALL	単語が本文で使われてい シャドーウイングが発音力		プロジェクトは大きくて見	何度も発音を聞くことが出
Q_edu-45	CALL	Sound Focusで聞き取りの シャドーウイングなど、声に プロジェクトはわかりやす			ペアでの練習はとも乗
Q_sci-1	CALL	リスニングにはまだ自信は シャドーウイングは面白か			ヘッドホンを使用すること
Q_sci-2	CALL	単語がよく覚えれた よく英語がききとれる		良かった	良かった
Q_sci-3	CALL	Sound focusで、英語の発 発音の発声時に、正しい発			時々Moodle内のテストな

図 1. 分析に用いたデータ (一部)

3.2 分析手順

SPSS Text Analytics を起動し、ツールバーの「ファイル」→「新規プロジェクト」→「Excel ファイル」から、上述のテキストデータ化されたエクセル形式のファイルを開くと、図 2 の「変数の選択」というダイアログがあらわれるので、「一意の ID」に連番 (これは学籍番号等でも可) を、「自由記述式のテキスト」に自由記述アンケートの回答 (今回のデータでは FA-1~FA-4 の 4 項目) を選択する。

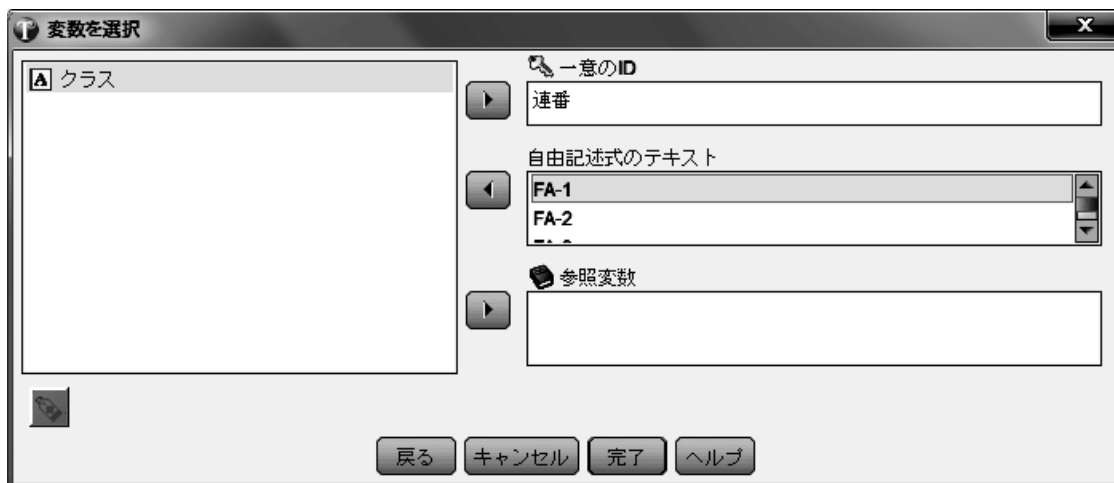


図 2. 「変数の選択」ダイアログ

そうすると、図 3 のように、回答項目 (FA-1~FA-4) または全体データ (プロジェクト全体) の回答データが表示される。この状態では、「カテゴリ」列には情報が入っていないが、Appendix に示したような結果を得るためには、このカテゴリ情報を得ることが不可欠である。そのため、以下、CALL 教室 (254 名) における質問 FA-1 (「a. 役に立ったと思う教材は何ですか。また、どのように役立ちましたか。具体的に書いてください。」) への回答データを例に、カテゴリ情報を得るための手順を説明する。

Id	回答	カテゴリ
Q_Fri1-1	教科書の会話	
Q_Fri1-2	単語の勉強で単語力がついた。	
Q_Fri1-3	教科書 (普段使ってた本) リスニング力向上	
Q_Fri1-4	身近にあるPCを使うため、気軽にリスニングができた。	
Q_Fri1-5	ToEICのリスニング練習	
Q_Fri1-6	日々のスピーキングが会話において役に立った。	
Q_Fri1-7	Sound Focusがよかった。	
Q_Fri1-8	教科書やSound Focusがリスニング能力に役立ちました。	

図 3. 「回答データ」表示

カテゴリ情報を得るためには、まず、「キーワード」の抽出という作業を行うことになる。このキーワードの抽出は、第 2 節で上述した形態素解析や構文解析に相当する作業である。図 4 の左上にある「抽出」というアイコンをクリックすると、SPSS Text Analytics が持つ辞書に基づいて自動的に抽出作業が完了し、図 4 の左ウィンドウにはキーワードごとの抽出結果 (実際には、品詞に応じて色分けがなされている) が、右ウィンドウには回答データに抽出結果が反映されたものが表示される。

Id	回答
1 Q_Fri1-1	教科書の会話
2 Q_Fri1-2	単語の勉強で単語力がついた。
3 Q_Fri1-3	教科書 (普段使ってた本) リスニング力向上
4 Q_Fri1-4	身近にあるPCを使うため、気軽にリスニングができた。
5 Q_Fri1-5	ToEICのリスニング練習
6 Q_Fri1-6	日々のスピーキングが会話において役に立った。
7 Q_Fri1-7	Sound Focusがよかった。
8 Q_Fri1-8	教科書やSound Focusがリスニング能力に役立ちました。
9 Q_Fri1-9	教科書 - <名詞> リスニング力の向上に役立ったと思

図 4. 形態素解析 (「キーワード」の抽出) 結果

次に、カテゴリの作成を行う。ツールバーの「カテゴリ」→「カテゴリの作成」と進むと、「言語学ベース」と「頻度ベース」の2種類のカテゴリ作成オプションが用意されている。「言語学ベース」は、図5のダイアログのタイトルにあるように、「言語学的手法に基づくカテゴリの作成」方法である。具体的には、共通の文字列が含まれているキーワード同士を1つのカテゴリにまとめる方法である。結果は図6のように、「リスニング」という語が含まれているキーワードが、「リスニング」という名称のカテゴリとして統合されている。

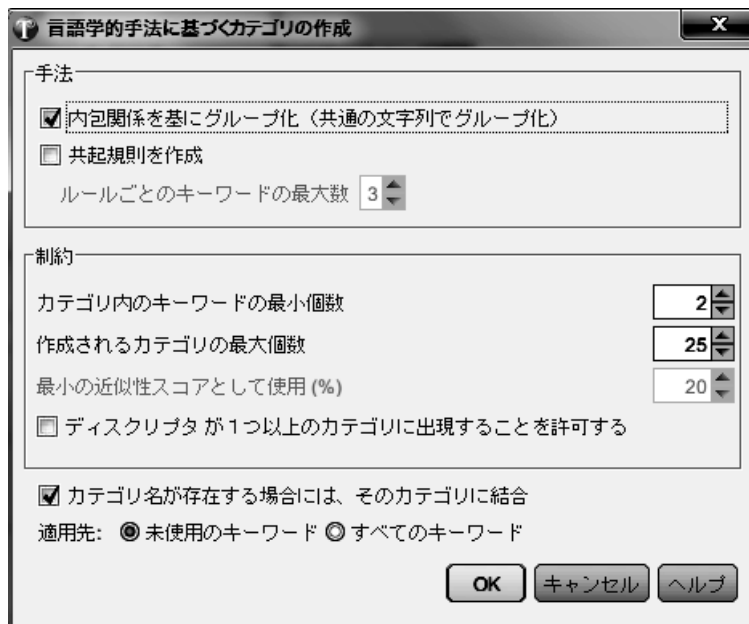


図5. 「言語学ベース」のカテゴリ作成



図6. 「言語学ベース」のカテゴリ作成結果

一方の「頻度ベース」は、「出現頻度に基づくカテゴリの作成」方法である。ダイアログにあるように、出現頻度の下限を設定し、それ以上の頻度であらわれたキーワードをカテゴリとする（それ以下の頻度のキーワードはカテゴリとしない）方法である。そのため、図8のように、それぞれのカテゴリは1つのキーワードから構成されることとなる。

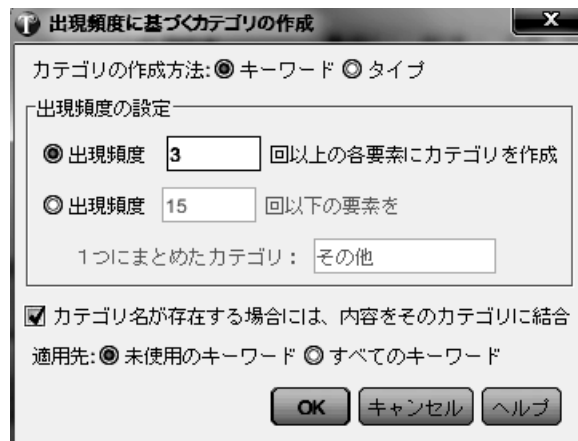


図7. 「頻度ベース」のカテゴリ作成

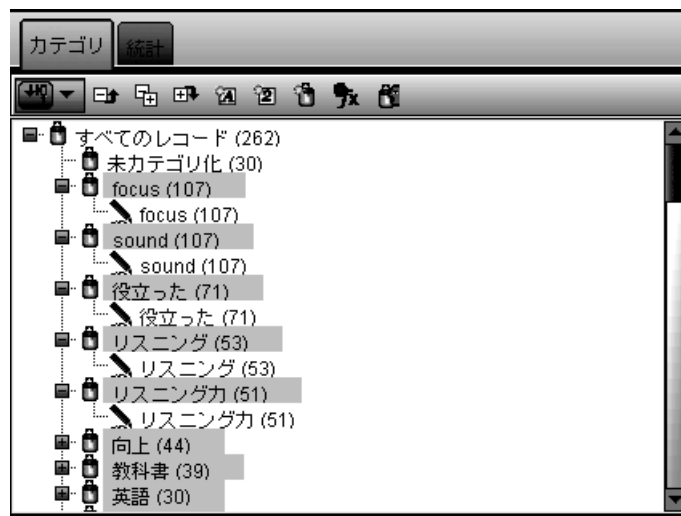


図8. 「頻度ベース」のカテゴリ作成結果

これらの2つの方法のうち、どちらがより優れているかということは一概にいえない。「言語学ベース」の方法は、自動的に複数のキーワードからなるカテゴリを作成してくれるため、一見優れた方法に見える。しかし、図6に示されているように、今回の自由記述データにおいてもっとも見たかった結果、かつ、もっとも高頻度であらわれているキーワードである Sound Focus（キーワードとしては、Sound と Focus に分かれて抽出されている）が、カテゴリとして作成されていないことが分かる。これは、この語が辞書に含まれ

ていないことが主因であると考えられるが、この Sound Focus のように独自の名称を与えた指導法を用いることは教育現場では多いといえるため、「言語学ベース」のカテゴリ作成結果をそのまま使うことができない場合があることには留意されたい。また、「頻度ベース」の結果は、抽出されたキーワードが頻度順に並んでいるだけなので、同じような内容のキーワード（例えば、図 8 の「リスニング」と「リスニング力」）が実質的に同じものを指しているにもかかわらず、異なるカテゴリとなってしまふ。特に、得られた回答数が多い場合には、出現頻度の下限を 3 回程度に設定しても、非常に多くのカテゴリが作成されてしまうことには留意する必要がある。参考までに、この時点での「言語学ベース」と「頻度ベース」におけるカテゴリ間の係り受けの関係を、「サークルレイアウト」という形で図示した（図 9 と図 10）。これを見ると、図 9（「言語学ベース」）は肝心の Sound Focus が含まれていない点と「力」や「気」など抽象的なカテゴリが作成されている点で解釈が難しく、図 10（「頻度ベース」）に至っては解釈が不可能なほど多くのカテゴリが作成されていることが分かる。

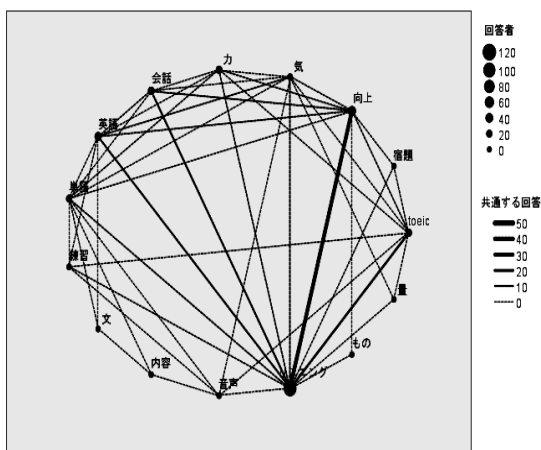


図 9. 言語学ベースのカテゴリの図示結果

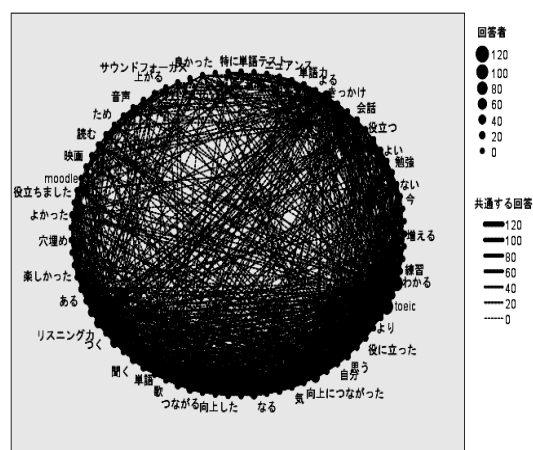


図 10. 頻度ベースのカテゴリの図示結果

ここから指摘できることは、テキストマイニングのソフトウェア（本稿では SPSS Text Analytics）で自動的に行われる分析のみでは、適切な分析結果を得ることは難しい場合がある、ということである。そこで必要になってくる作業が、分析者としての視点を加えながら、より適切な分析結果を得ることを目指すということである。この作業は、数量的な統計解析においても同様に行われていることである。例えば、探索的因子分析における因子の推定方法（主因子法、最尤法、etc.）の選択や因子軸の回転方法（斜交回転、直交回転）の選択といった作業は、上述したカテゴリの作成方法（言語学ベース、頻度ベース）の選択に対応するものであるといえる。また、因子分析における因子数の決定や分析において不適格な項目の除外といった作業は、これから述べるカテゴリ数の決定やカテゴ

リに加えるキーワードの取捨選択といった、カテゴリの再編成の作業と同種のものであるといえる。

データ分析の際の留意点:ソフトウェア任せにせず、分析者の視点を持ち、より適切な分析結果を得るように努める

以下は、田中・山西論文で実際にカテゴリ情報を得るために行った手順である。得られた自由記述データの回答数やデータの性質によって、適切な手順は異なってくるものと考えられるので、あくまで分析の1つの例と考えて頂きたい。まず、SPSS Text Analyticsによるカテゴリ作成方法としては「頻度ベース」を採用した。「言語学ベース」を用いなかった理由は、Sound Focus等の重要な情報の取りこぼしを避けるためである。「頻度ベース」の出現頻度の下限は15回とした。これは、図10のように多量のカテゴリが作成されることを避けるためである。ただし、CALL教室の254名の回答データといっても、すべての重要なキーワードが15回以上の頻度であるとは限らない。そのため、15回以上という頻度で大きくフィルターをかけた上で、すべてのキーワードに分析者自身で目を通し、必要であると思われるものは、頻度が低くとも、すくい上げるという作業を行った。⁴

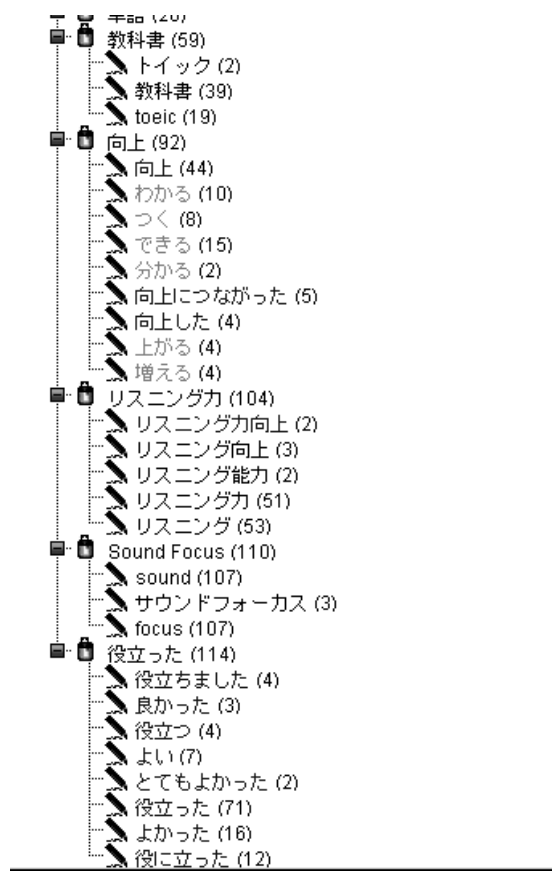


図 11. カテゴリの再編成結果

このカテゴリ再編成の具体例（実際の論文で用いたカテゴリ）は、図 11 に示した通りである。SPSS Text Analytics では、一度ソフトウェアまかせで作成したカテゴリを柔軟に再編成するための機能が付いている。例えば、「教科書」というカテゴリ（「頻度ベース」で自動的に作成されたもの）の中に、「教科書」（頻度 39）に加え、「トイック」（頻度 2）と「toeic」（頻度 19）というキーワード（どちらも自由記述の原文のまま）を追加した。これは、授業における TOEIC®は、共通教科書に付属のエクササイズとして指導されたためである。その結果、「教科書」カテゴリの頻度は 59 となった（3 つのキーワードの頻度を足すと 60 になるが、ある学生の回答内で「教科書」と「toeic」が共起していたため重複分はカテゴリの頻度にはカウントされていない）。また、「向上」というカテゴリには、「向上」という語が含まれたキーワードのみでなく、「わかる／分かる」や「できる」といった英語力の向上に関連するキーワードも含める形で統合した。また、Sound Focus というカテゴリを新規作成し、その中には「Sound」、「Focus」、そして「サウンドフォーカス」というキーワードが含まれるよう再編成した（「Sound」と「Focus」はともに頻度 107 であるが、共起しているため、「サウンドフォーカス」の頻度 3 と併せて、カテゴリの頻度は 110 となっている）。

このようなカテゴリの再編成作業は、絶えず、図 12 に示されたようなローデータ（回答原データ）を参照することで、キーワードに否定語がつくことで逆の意味になっていないか（例えば、ローデータを見ると「役立った」ではなく「役立ったとはいえない」であったなど）を確認しながら進めた。

	Id	回答	カテゴリ
1	Q_Fri1-1	教科書の会話	教科書 会話
2	Q_Fri1-2	単語の勉強で単語力がついた。	向上 単語
3	Q_Fri1-3	教科書（普段使ってた本）リスニング力向上 教科書 - <名詞>	教科書 向上 リスニング力
4	Q_Fri1-4	身近にあるPCを使うため、気軽にリスニングができた。	向上 リスニング力
5	Q_Fri1-5	Toeicのリスニング練習	教科書
6	Q_Fri1-6	日々のスピーキングが会話において役に立った。	スピーキング 役立った 会話
7	Q_Fri1-7	Sound Focusがよかった。	役立った Sound Focus
8	Q_Fri1-8	教科書やSound Focusがリスニング能力に役立ちました。	教科書 リスニング力 役立った Sound Focus

図 12. 回答データとカテゴリ再編成結果の確認画面

3.3 結果の提示

結果の提示に関しては、用いるソフトウェアの機能や考え方によって異なる。本稿で用いている SPSS Text Analytics は、上述した製品ウェブサイトの「アンケートの自由回答をかんとんに、そしてすばやくコード化、視覚化を可能に」するという文言の通り、分かりやすい形で結果を視覚化するという考え方を優先しているようである。用意されている「視覚化」オプションは、カテゴリの頻度を示す「カテゴリ棒グラフ」、カテゴリ間の係り受けの関係を図で示す「カテゴリ Web」、カテゴリ Web の内訳を表で示す「カテゴリ Web テーブル」である。この中でも「カテゴリ Web」には、図示方法として「サークルレイアウト」「有向レイアウト」「グリッドレイアウト」「ネットワークレイアウト」の 4 種類が用意されているため、この「カテゴリ Web」による結果提示がもっとも分かりやすい視覚化機能であるといえる。4 つのレイアウトのどれを選択するかは、データの性質や見やすさ（あるいは好み）等に応じて異なるため、実際のデータの結果をもっとも良くあらわすと考えられるレイアウトを選択すると良いだろう。

3.4 結果の解釈

結果の解釈は、図の形で提示した結果だけに対しても行うことができる。図示した情報は、これまで述べてきた手順によって、回答データが圧縮された結果であるといえるためである（そのようにいえるためにも、第 3.2 節で挙げた 2 つの留意事項が重要である）。田中・山西論文では、以下の引用のような解釈を行った。なお、引用中の図 5 は、本稿では Appendix を指す。

結果は、一連のグラフ（図 5）に示した。グラフの基本的な解釈の方法は、表現ごとの回答数（●で表現；グラフごとに回答数が多いほど相対的に大きく表現される）および表現同士の共通性（実線および点線で表現；共通性が強いほど相対的に太く表現される）を見ることで行われる。例えば、「役立ったと思う教材」に関する質問 a の結果は、CALL 教室、普通教室ともに「Sound Focus」、リスニング力、「向上」、 「役立った」という回答が多く、それぞれの共通性も強い。そのため、両環境とも「Sound Focus がリスニング力の向上に役立った」と学生に認識されていたと考えられる。このような解釈をそれぞれの質問において行った。（田中・山西、印刷中）

また、可能であれば、自由記述アンケートの回答の分析結果だけでなく、同時に得た数値データ（本稿冒頭に挙げた「授業評価」の例であれば、17 項目からなる 5 件法でのアンケート調査項目、田中・山西論文であれば、テストによるリスニングパフォーマンス

の測定結果)の分析結果を併せて解釈することが、総合的な授業の振り返りのためには、より望ましいと考えられる。

4. まとめ

本稿では、質的な言語データを取り扱う方法として、ソフトウェアを利用したテキストマイニングを取り上げ、実際の授業に関する自由記述アンケートの回答データをもとに分析の手順や結果の提示・解釈手順を示した。本稿では著者の所有する SPSS Text Analytics を用いたため、その他のソフトウェアを所有している読者の方にはそのままでは適用できない部分もあるだろうが、データの収集と分析における注意事項に関しては、汎用性の高いものとなるように心がけた。

最後に、このようなテキストマイニングによって、何ができ、何ができないかについての私見を簡潔に記しておきたい。まず、できること(得意なこと)としては、大量の自由記述アンケート結果の処理が可能になるということである。その際の留意事項はあるものの、このメリットは大きいといえる。できないこと(苦手なこと)は、グラウンデッド・セオリーや KJ 法といった「人の手による」質的な言語データの分析方法のような、ダイナミックかつ繊細な分析を行うことである。また、回答データ数や回答内のバリエーションが少なくなると、十分な結果の提示を行うことは難しい(Appendix の質問 c の普通教室の図を参照)。このできること、できないことに関する建設的な解決策が、清水(n.d.)によって提案されているので紹介したい。清水は、テキストマイニングとグラウンデッド・セオリーや KJ 法、それぞれの得手不得手を認めた上で、両者を組み合わせた分析を行うことを提案している。清水の提案にもあるように、テキストマイニングは手法としての限界もあるが、適切に使用すると有効な分析方法であることは間違いない。本稿が、そのような適切な使用に少しでも役立てば幸いである。

注

1. なお、SPSS Text Analytics は、教育機関向け(アカデミック版)を年数回のキャンペーン時期に購入すると、比較的入手しやすい価格で購入できることを付記しておきたい。
2. もちろん、目的によっては、全体的(総合的)な自由記述こそが分析対象となる場合もあるだろう。あるいは、全体的な項目と分析的な項目によって得られた自由記述データを比較することも有意義であろう。
3. プライバシーマーク制度とは、「個人情報について適切な保護措置を講ずる体制を整備している事業者等を認定」するものである。自由記述アンケート等の個人情報を含んだデータの入力を業者に依頼するときには、このプライバシーマークを取得しているか業者であるか確認することが望ましい。
4. 田中・山西論文では、数量データに対する統計的な解析も含め、データ分析は基本的に第 2 著者である本稿の著者が行ったが、このカテゴリ再編成の作業は、第 1 著者と共同で、協議を行いつつ実施した。

謝辞

本稿は、外国語教育メディア学会（LET）関西支部メソドロジー研究部会，2010年度第2回研究会（2010年10月2日，関西大学）にて著者が発表した内容に基づき，フロアからの質疑・コメントの内容を反映させ執筆したものである。発表時に有益なコメントを寄せてくださった皆様に感謝申し上げたい。また，本稿で用いた自由記述データの分析や引用に使用した論文の執筆は，大阪医科大学の田中英理氏と共同で行ったものである。田中氏には本稿執筆においても助言をいただいたことをここに記して感謝する。

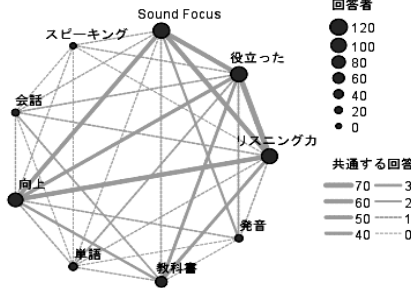
参考文献

- 石田基広 (2008). 『Rによるテキストマイニング入門』 森北出版株式会社.
- 清水裕士 (n.d.). 「Word Miner を用いたテキストマイニング」 テキストマイニング講座 発表資料. Retrieved from <http://www.geocities.jp/simizu706/wordminer.pdf>
- 田中英理・山西博之 (印刷中). 「英語音声学・音韻論的特徴の習得を目指した授業の効果検証」 *JALT Journal*, 33.
- Weigle, S. C. (2002). *Assessing writing*. Cambridge: Cambridge University Press.

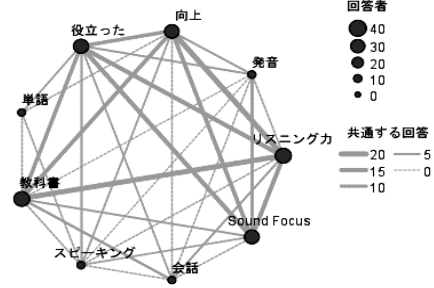
Appendix: 結果の図（サークルレイアウト）による提示例（田中・山西，印刷中より）

a. 役に立ったと思う教材は何ですか。また、どのように役立ちましたか。具体的に書いてください。

CALL 教室

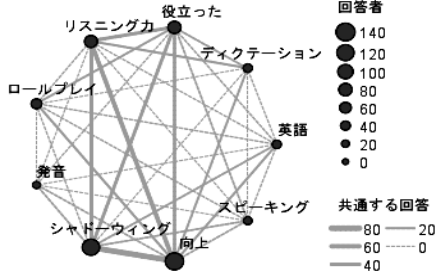


普通教室

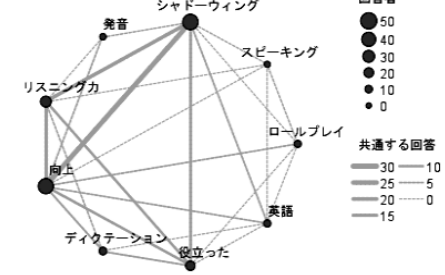


b. 授業で紹介・練習した学習方法（シャドーウィング、ディクテーション、ロールプレイなど）で役に立ったものは何ですか。また、どのように役立ちましたか。具体的に書いてください。

CALL 教室

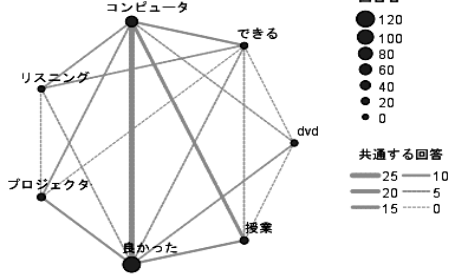


普通教室

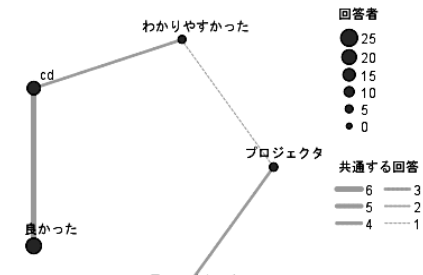


c. 教室環境（DVD、CD 機材、プロジェクタ、パソコン（ただし、教室の広さやエアコンの性能については除く））はどうでしたか。

CALL 教室

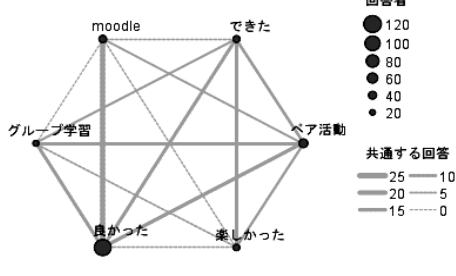


普通教室



d. 学習環境（Moodle やインターネットの使用、グループ学習・ペア活動）はどうでしたか。

CALL 教室



普通教室

