

授業分析におけるデータサイエンス活用の可能性

—潜在的意味解析を用いた逐語記録の分節わけの試み—

*坂 本 將 暢

1. はじめに
2. 先行研究
3. テキストマイニングによるデータ解析
 - 3-1. データ
 - 3-2. 潜在的意味解析
 - 3-3. 解析手順
4. 分析結果
 - 4-1. 分節のわけ目となる発言
 - 4-2. 授業の構造
5. 考察
6. まとめと今後の課題

1. はじめに

本研究の目的は、データサイエンスを授業分析に適用し、新しい知見を得ることが可能か否かを模索するという大枠の中で、とりわけ今回は、テキストマイニングの一手法である潜在的意味解析（LSA：Latent Semantic Analysis/ Index）を用いた分節わけを試みる。

データサイエンスとは、統計やアルゴリズムを用いて社会に有益な知見を導き出そうとするアプローチである。データサイエンスの広がり背景には、ブームとしては下火にあった人工知能（AI）やディープラーニング（深層学習）が、コンピュータ処理の速さ、情報量の多さ（ビッグデータ）、そしてプログラミングのし易さの影響を受けて、2020年度までの科学技術基本計画（内閣府2016）に「世界に先駆けた『超スマート社会』の実現」として「Society 5.0」が記された頃から再燃してきたことが挙げられる。つまり、データサイエンスは、特別な知識を有した人や特定の研究者だけが扱えるものではなく、それに興味を持つ人ならば誰にでも扱える身近なものになったことで、言葉と

してだけでなく、データの取得・処理において、我々の身近なものになった。

データサイエンスは情報分野に限ったものではない。例えば、大学にはデータサイエンスのコースや授業が設定されたり、小・中・高校生にはスタディ・ログの蓄積および活用にもとづく個別最適化学習が実施されたりすることから、今後、教育分野に身を置く我々に対して、その意義や活用方法を考えさせる事柄になろう。筆者の立場としては、無批判にこのような状況を受け入れるのではなく、そのような時代の到来を受受しつつも一歩引いて、授業分析の立場でデータサイエンスの適用の可能性や困難性を明らかにする立場をとる。

ところで、筆者が取り組んでいる授業分析は、重松鷹泰と上田薫が1950年代に始めて以来、約70年間名古屋大学教育方法学研究室が取り組んでいる方法論で、「教育の科学化」を志向した逐語記録に基づいた分析によって、子どもの思考過程や学びの可能性を明らかにするものである。授業中のやり取りは「教室談話」という行為として位置づけられることが多く、教師による問いかけの研究（Hargie 1978, Wilen & Clegg 1986）、教師の主導性に関する研究（Becker 1953, Pearson and Dole 1987, Johnson 1994）、授業のパターン化の研究

* 名古屋大学大学院教員

(Flanders 1965, Bellack 1968), あるいは教室の中の政治性に関する研究 (Bloome, Harris & Ludlum 1991, Grant 1996, Damron 1996, Lipscomb 2019)が行われている。これらの研究と授業分析は、授業中のやり取りを分析することでは共通しているものの、決定的な違いの一つに、逐語記録を用いるか否かが挙げられる。授業分析において逐語記録を用いる意義は、「学問的にリベラルな態度をつらぬく」(柴田2019, p.26) ことにある。柴田 (2019) は逐語記録について、「事実を残したあるいは写し取った〈物=オブジェクト〉あり『データ』である」とした上で、「読みながら、自らのあり方を問い直す(反省する)きっかけが生まれ」、「ぶつかっても跳ね返す壁として、あるいは自らを映し出す鏡として、私の外側に歴史と存在し」、「読めば読むほど、あらたな意味(切り口)を我らに提示する」ものに位置づけている。言葉として表れた子どもの考えや主張を手がかりに、その背景にある日常生活の経験やそれから得た価値観、それらそのものからやそれらを繋ぎ合わせて導き出した考え、あるいは、矛盾を含むこともある論理を解明するための根拠になり得る。分析者は単に記録を読めばいいのではなく、ときには子どもの様子を想像したり発言の背景を推測したりすることや、自身の経験や価値観等と対峙することが求められ、主体的かつ対話的な態度が求められる。

逐語記録に記されるのは教師と子どもの発言であるが、教師や子どもの発言回数、発言に含まれる語数、単語の出現回数、教師の立ち位置(座標)など、授業の中にはデータサイエンスで扱うことが可能なデータがある。また、データサイエンスによって、特徴を浮かび上がらせたり、分類したり、予測したりすることが可能である。本研究では、上述したことを踏まえて、分析可能なデータのうち、逐語記録に含まれる語の出現に注目し、発言を分類して、授業分析の最初の作業である分節わけを行う。

分節わけは、授業を構造的に捉えたり、その構造を形成するための各部分を持つ力やそれに働く力を考察したりするために、授業をいくつかの「段落」にわけることである。重松(1961)は「一時間の授業の中にいくつかの段落のあることは、授業を見たり、実施したことがある人には周知のことである。授業者自身がここまでは導入段階で、ここまでは展開の段階、ここからあとはまとめの段階であると、意識している場合も多い。」(p.62)と、指導案の過程と位置づけられる区別の存在を示しつつも、「導入・展開・評価、とか予備・教授・整理とかいう、いわゆる教授の段階をもって、分節とするのでは、粗大にすぎ、また授業者側に

のみ考察が偏っていて、分節を分ける趣旨にそわないのである。」(pp.62-63)と述べている。授業は上に挙げた3つ(導入、展開、まとめ)のみに分けることができず、授業の話題、教師の問い、子どもの思考の変化などを分節わけの手がかりにして、分析者が豊かに分けることの重要性を示唆している。そうであることは同時に、授業分析の課題を示している。特定の「分節わけの解」はなく、何度も逐語記録を深く読み込んだり熱心に協議したりして「妥当な分け目」を探るため、構造的な把握一つをとっても、容易な作業とは言えない。ましてや、十分な分析視点を有していない初学者にとってみれば、自身の分節わけが、適当か不適当かが定かではないため足踏み状態になり、分析が先に進められないことも起きる。

筆者の研究課題の根底には、分析者が主体的かつ対話的に逐語記録を読み、その第一段階に分節わけが位置づけられているのにも関わらず、その困難さを経験やセンスではなく、データサイエンスの力を借りて、効果的な分節わけができないかを追究することにある。本稿は、この研究課題の第一歩として、テキストマイニングの一手法である潜在的意味解析を用いて、分節わけの有効な手がかりを得られるかどうかを試みるものである。

2. 先行研究

上述したように逐語記録には教師と子どもの発言が記されており、教師や子どもの発言回数、発言に含まれる語数、単語の出現回数を分析した研究がある。例えば、大谷と八田(1987)は、授業分析支援システムの一機能として、教師と子どもの1分毎の発言量の推移をグラフ化する機能を開発している。このグラフは、授業記録を作成した後に、発言文字数を数え、■(教師)と□(子ども)を積み上げて、時系列に並べたものである。柴田(1997)は授業記録に含まれている学習内容に関わる事象・事象を示す語を特徴語として抽出し、それを累積相対度数グラフで出現パターンを明らかにしている。また、柴田(1999)は、上述した語の出現パターンをもとに、授業記録を分割する方法を開発し、それをデンドログラムで表現し、分節わけの手がかりとしている。

このような逐語記録を用いた研究に、本研究ではテキストマイニングを適用するのだが、これまでのテキストマイニング研究を大別すると、「特徴抽出」と「分類」に分けられる。「特徴抽出」では、文書を形態素解析して、特徴語の出現頻度を算出する。それを著者や年代など、さまざまな切り口で整理することで、文書

の特徴を見出す。李ら（2017）は、日本国内の約50年の医学・生命科学に関する文献を対象に、論文題目に使用される語の傾向、年代別の語の特徴などをテキストマイニングで分析している。坂本（2017）は、世界授業研究学会（WALS）の発表題目に使用されている語の分析し、主要な国・地域の授業研究の動向を明らかにしている。また、辛（2017）は、論文がTEXを使って作成されることを前提にした論文要約システムを開発している。ファイルに記述されるマークアップのうち、節、参考文献、アブストラクトのマークアップに注目し、各節から重要文を抽出し、それを結合して要約を作成するシステムである。「分類」では、文書を形態素解析した後に、語の出現の仕方などの特徴から、文書同士の距離（類似度）を計算する。その数値が大きいほど文書同士が似ていなく、逆に数値が小さいほど文書同士が似ていることを表す。分類する際は、クラスター分析をしてデンドログラムで表示したり、多次元尺度法を用いて平面にプロットしたりする。例えば、小野ら（2004）は、病院情報システムに書かれた約36,000件の入院中の状況を対象に分析し、疾患を分類している。そして、形態素解析で抽出された語同士の類似度を、 $tf \cdot idf$ 法で重み付けしたベクトルを用いて算出し、その値を用いて最大全域木で分類（デンドログラムの作成）している。

近年では、文書がいくつかの潜在的なトピックから生成されていると仮定したトピックモデルを用いた研究もなされている。とくに「分類」で使用されることが多く、トピックを構成していると想定されるいくつかの単語の出現から、対象とする文書がおもに何について記述されているのかを推定するモデルである。

松河ら（2017）は、約6万件の大学の授業評価アンケートを対象にトピック分析を行い、そこに書かれているトピックを明らかにしている。トピック分析で適当な解を得るためには、事前にプログラムに少なくとも数百個の学習をさせておく必要があったり、トピック数を調整する必要があるが、コストや質の面で、上で挙げたこれまでのテキストマイニングと比べて高いとは言えない。

しかし、もし授業分析の初学者が見出し難いトピックや、見落としてしまう逐語記録に記されている語を探索的に発見することができたり、授業分析の経験者と類似した分類がトピック分析によってできたりするのであれば、授業分析を進める上で十分に貢献できるだろう。

3. テキストマイニングによるデータ解析

3-1. データ

本研究で分析対象とするデータは、社会科の初志をつらぬく会が発行している「考える子ども」に掲載されている「新しい世の中へ～東京オリンピックから考える戦後～」(本間2015)という授業の逐語記録である。本授業は、全6時間のうちの第5時間目の授業で、「東京オリンピック開催の意味について考え、その後の国際社会における日本の役割に目を向ける」というテーマで取り組んだものである。それまでに子どもたちは、終戦直後の写真と20年後の写真を見比べたり、人々の生活や町の様子の変化を家族へ聞き取ったり、戦後20年間の出来事を調べて年表にまとめたりしてきた。本授業では、前時の授業で挙がった、「多くの感動があつて国民が元気になったから、安いと思う」「経済も元気になったからお金をかける意味はあったと思う」「オリンピックをするのに国家予算の3分の1と同じ額のお金を使っている。やりすぎじゃないかな。」「何でそんなにお金を使ってオリンピックをやったのかな?」という子どもの意見を引き継ぎ、「なぜ、一兆円もお金を使って東京オリンピックを開催したのか」というテーマで話し合いをした。

発言は123発言である。多くの授業では250から300発言あることを考えると、半分くらいの発言数であるが、子どもが自身の意見を話しているため、一発言の発言量（文字数）は他の授業と比べると多い。分析の詳細は後で述べるが、本研究で用いる潜在的意味解析では、小説や演説、あるいは授業評価アンケートのように、一定程度のボリュームを持つ文書ではなく、教師や子どもの一発言を一文書とみなす。授業特有の端的な発言も分析対象であるが、例えば「はい」や「いいえ」のように、トピックになり得る語が含まれていない場合には、分析の過程で捨象されることがある。このように、本研究で本逐語記録を選んだ理由は、発言数こそ少ないが一発言の発言量が多く、発言の中にトピックになり得る語が含まれていると判断したからである。

3-2. 潜在的意味解析

本研究では、授業記録の中の特徴語の出現を手がかりに各発言を分類する際に、潜在的意味解析（LSA：Latent Semantic Analysis/ Index）を用いる。本研究で用いる潜在的意味解析は、トピック分析の1つで、語をグルーピングして、文章中の情報量を圧縮する技術である。トピック分析はほかに、EMアルゴリズム

を用いる確率的潜在的意味解析法 (PLSI: Probabilistic Latent Semantic Indexing) と、ベイズ推定を用いる潜在ディリクレ分布法 (LDA: Latent Dirichlet Allocation) がある。LSA は、圧縮する次元数の設定がアドホックに決定できること、1つの文書にトピックが1つしか存在しないことを前提にしているため、アンケートやレポートに書かれたことを分類する際には向いていない。一方、PLSI と LDA は、機械学習と呼ばれる分析事例を用意したり、例えば10,000の分析対象のうちの約1割を教師データとして用いたりする必要があるのである。そのため、本研究のように、一回性の授業の発言を分析対象とする研究では、PLSI や LDA よりも LSA が適していると考える。

このような特徴のある LSA の使い方は、分類の対象となる複数の文書に含まれる特徴的な語 (テキストマイニングでは「ターム」を用いる) が各文書に含まれているか否かを、文書数×ターム数の行列で表し、この行列を特異値分解 (SVD: Singular Value Decomposition) する。例えば、Doc01「自動車は技術アピールになる」、Doc02「車は技術アピールになる」、Doc03「自動車は白いのが好き」、Doc04「果物が好き」という4文書を分析する場合、各文書にタームが含まれているか否かを、表1のように表すことができる (行列 A)。これは文書数が4つ、解析するためのターム数が7つあることを意味し、文書とタームがクロスしたところに記している「1」は言葉が出現していること、「0」は出現していないことを意味する。このベクトルを特異値分解 ($A=UDV^t$) すると、図1のようになる。このときの U の値を k 次元に圧縮することによって、トピックを絞り込む。例として2次元に圧縮し、x-y 座標上にプロットすると、図2のようになる。この図から、左上に Doc01 と Doc02、右下に Doc03 と Doc04 があり、大きく2つのグループに分けられるこ

表1：各文書における語の出現の有無

Word	Doc01	Doc02	Doc03	Doc04
自動車	1	0	1	0
技術	1	1	0	0
アピール	1	1	0	0
車	0	1	0	0
白	0	0	1	0
好き	0	0	1	1
果物	0	0	0	1

とがわかる。一般的には「自動車」や「車」のことを記した Doc01, Doc02, Doc03 が1つのグループを作りそうだが、テキストマイニングの結果では、「技術アピール」のことを記した Doc01 と Doc02 と、「好き」なものを記した Doc03 と Doc04 とに分けられたと解釈するのが適当だろう。このような4つの短文ではテキストマイニングの効果は十分に発揮されず、いくつものタームの出現の組み合わせがあるような複雑なデータを解析する際には効果を発揮すると考えられる。

3-3. 解析手順

本研究では、潜在的意味解析を行うために統計パッケージ R を用いるが、その前処理のためには表計算ソフトを用いる。具体的な手順について、表2に示す。

まず、分析者は発言記録から気になる言葉や注目する言葉を選定する。本研究では、「学習」「1兆円」「お金」「オリンピック」「使い過ぎ」「アジア」「敗戦」「アピール」「復興」「開催」「比べ」「平和」「余裕」「戦争」

左特異U=	[.1]	[.2]	[.3]	[.4]
[1.]	-0.45804	-0.34787	-0.51415	-0.29840
[2.]	-0.57093	0.27288	0.08436	-0.11625
[3.]	-0.57093	0.27288	0.08436	-0.11625
[4.]	-0.26738	0.21665	0.30737	0.68652
[5.]	-0.15449	-0.40410	-0.29114	0.50436
[6.]	-0.20173	-0.67012	0.35433	0.12760
[7.]	-0.04724	-0.26601	0.64547	-0.37677
右特異V=	[.1]	[.2]	[.3]	[.4]
[1.]	-0.69689	0.10549	-0.27754	-0.65283
[2.]	-0.61384	0.40642	0.38254	0.55830
[3.]	-0.35468	-0.75807	-0.36235	0.41016
[4.]	-0.10844	-0.49902	0.80333	-0.30640
特異値D=	[.1]	[.2]	[.3]	[.4]
[1]	2.29578	1.87593	1.24457	0.81323

図1：例の特異値分解の結果

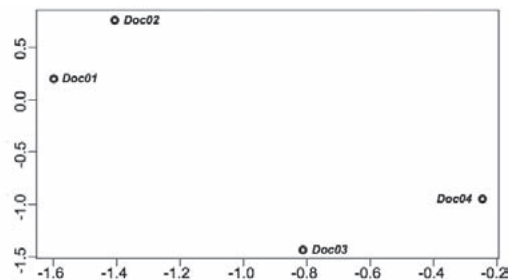


図2：左特異 U を2次元に圧縮した結果

「観光客」「スタジアム」「新幹線」「建物」「技術」「意見」「製品」「原爆」「高速道路」「競技場」「文化」「空港」「選手村」「必要」「役立(つ)」「生活」「記事」「読(む)」の32個を選定した。

表計算ソフトに書いた発言記録の1行目にそれらの言葉を書き、発言とクロスする部分に、各発言に含まれる気になる言葉や注目する言葉の出現回数を記す(表2内の1のAとイ)。すべての語の出現回数を記した後に、例えば「はい、〇〇さん」と指名している発言や、「はい」と返事をしている発言のように、気になる言葉や注目する言葉が出現しなかった発言は分析の対象から外す。これによって分析対象の発言が、122発言から76発言になった(表2内の1のウ)。

上で作成したデータから発言者と発言を削除し、さらに行列を入れ替えたデータを作成して、統計パッケージRに読み込ませる(表2内の2)。統計パッケージRにはsvd()関数があり、これを用いると特異値分解を行うことができる(表2内の3)。本研究では語の出現の有無を手がかりに発言同士の類似性を明らかにするため、特異値分解された要素(左特異(文書)ベクトル, 特異値, 右特異(言葉)ベクトル)のうち、左特異ベクトルをk次元に圧縮した近似行列を算出することで関連性の強い言葉同士が近似値を取る。本研究では、視認のための表作成時のプロットのしやすさと、原点からの距離計算の単純さを考慮して、圧縮する次元を2とした(表2内の4と5)。

さらに、分節の分け目を把握するために、原点からの距離を算出し、距離の遠い順に並び変えた(表2内の6)。

4. 分析結果

4-1. 分節のわけ目となる発言

発言記録を8分節に分けられると想定した場合、原点からの距離が遠い順に上位7発言を取り上げると、発言3, 発言14, 発言25, 発言29, 発言47, 発言69, そして発言118だった。これらの発言について説明する。

発言3は教師が「東京オリンピックね。なかなか凄かったね。先生も映像は初めて見ました。で、見てみて、この間の授業でもやったけれども、なかなかいい大会だったなというのは、みなさんも自分たちで調べてきて分かっていると思いますが」と前回までの授業内容を振り返り、「(オリンピック開催のために1兆円を使ったことが)でもやっぱり使い過ぎなんじゃないかということがあると思いますので、なんでそれでもオリンピックを開催したのかなというのを、みなさんで考えていきたいと思います。」と本時の課題を発言した個所である。

発言14は教師が「さっきのSkさんの意見とは違うの？復興のアピールというのとはまたちょっと違う？技術のアピール。技術って、例えばどんな技術なの？これ、このままだとちょっと曖昧なだけけど。」と、発言5と7で出てきた意見(復興していることのアピール)と、発言13で出てきた意見(技術の高さのアピール)の相違に注目させている箇所である。

発言25は、技術のアピールに傾き始めたクラスの見解に対して教師が「みんなどう思った？あの傘って日本の技術アピールになったかな？日本の傘ってすごいなって感じだった？それでも技術かな？(略)傘って

表2：本研究の解析手順

-
- 1) 自立語のみを抽出し、それらの頻度からベクトルを作成
 - (ア) 行に言葉(ターム), 列に発言(文書)をとる
 - (イ) 出現回数を書く

Excelの計算式: $=\text{LEN}(\$C2)-\text{LEN}(\text{SUBSTITUTE}(\$C2,\text{D\$1},""))/\text{LEN}(\text{D\$1})$
 - (ウ) 0の行列がないように掃除をする
 - 2) 1のベクトルを書いたCSVファイルを読み込む
 - 3) svd()関数を用いて、特異値分解を行う
 - (ア) svdの要素は, \$u (document vector), \$d (特異値), \$v (word vector) で確認することができる
 - 4) $\text{TDk}=\text{UTK}*\text{TD}$ によってk次元に圧縮する
 - (ア) $t(\text{test.svd}\$u[1:2])$ によって2次元に圧縮する
 - 5) x-y座標上にプロットする
 - 6) 座標(0,0)から距離が遠い順に、発言を並べる
 - 7) 分節あるいは場面の変わり目とみなす
-

技術かな？」と子どもに確認をしながら、ある種のゆさぶりをかけている発言である。

発言29は、児童Skが「空港」と「選手村」を本授業で初めて用いながら「(略) オリンピックを開催するにあたって、まず空港を作らなければいけないし、選手がいっぱい来るから選手村も作らなきゃいけないし、観光客が増えるなら、東京から大阪へ行く新幹線も作らなきゃいけないし、鉄道も作らなければいけないし、(略) 自然といっぱいお金を使わなければいけないから…」とオリンピックを開催するのにあたって、整備すべきインフラの必要性和、そのためのお金の使用の必要性を述べている発言である。

発言47は、教師が発言45でオリンピックのために作ったものが、いまの子どもたちの生活に役立っているかを考えさせる発言をした後に、児童Roが「(略) ここから博多とか大阪とかまで行く時に、新幹線がなかったら、電車とかはあったから、電車で乗り換えとかして行かなきゃいけないから時間もお金もかかったと思うから、(略) オリンピックで作られたものが、今の生活に役立っていると思います。」と、新幹線を例に具体的に「生活」に「役立」っていることを述べている発言である。

発言69は、教師の「(略) みんなアピールっていうのはすごく言ってるね。意見の最初の方の人は。(略) アピールじゃないよ。必要なもの、今でも使うものをたくさん作ったから、それはすごく意味があったんじゃないかなって言うふうに考えている人もいました。どうでしょう？これに対してまだ言ってないことありますか？(略)」という発言である。これは、授業のはじめに出た、「1兆円」もの「お金」が使われて整備されたインフラは、「技術」の「アピール」のためだという主張と、その後の出た、「生活」のために「必要」なものだという主張を整理しつつも、ほかの見方がないかを子どもたちに探っている発言である。これによって、話し合いの内容は、比較的現在の話しに傾倒していたものから、聖火ランナーや「戦争」や「平和」について言及され、それまでに子どもたちが学習してきた戦後の核心的なものへと移り変わるきっかけとなった。

発言118は、教師が「(略) 一人ひとりね、この意味っていうのをね、考えてほしいなってますが(略) オリンピックでこれだけ出来ましたってアピールしたとして、その後じゃあ日本はどうしたんだろうね。オリンピックここまでよく出来ましたで終わり？」と子どもたちに最後の問いかけをし、次回の学習課題に引き継いでいる。

4-2. 授業の構造

上の分析結果をもとに本授業を8分節に区切り、各分節に小見出しを付けると、次のようになる。

- 第1分節：発言 T1～S2「本時の学習課題の確認」
- 第2分節：発言 T3～S13「敗戦からの復興アピールのために1兆円をかけて開催した東京オリンピック」
- 第3分節：発言 T14～S24「技術アピールとしての東京オリンピック」
- 第4分節：発言 T25～T28「話し合いを深めるための教師のゆさぶり」
- 第5分節：発言 S29～S46「東京オリンピックの開催に伴って整備されたインフラ」
- 第6分節：発言 S47～S68「生活に役立つ観点での1兆円をかけたインフラ整備の意義」
- 第7分節：発言 T69～S117「戦争と平和の観点での東京オリンピック開催の意義」
- 第8分節：発言 T118～S123「東京オリンピック後の日本の成長と次の授業の学習課題の提示」

これを踏まえた分節構造図を図3に示す。分節2と3は、復興アピールと技術アピールで異なるが、いずれもアピールという点で共通している。第3分節での技術は、第5分節でおもに新幹線、競技場、空港、選手村という具体的な建造物で表現される。それは第6分節に引き継がれ、生活に役立つという観点で意義を話し合う。第7分節は、第2・3分節で話し合われたアピールと、第5・6分節で話し合われたインフラを踏襲しつつも、戦争と平和の観点で東京オリンピック開催の意義を話し合っており、とくに第5分節までに出てきた観光客から派生した選手や記者にも焦点を当てている。そして、第8分節では、オリンピック後のアピールの中でも、第2・3分節で話し合われた復興、技術の発展、国としての成長に焦点を当てた学習課題が設定された。

5. 考察

上述したように、本研究では、潜在的意味解析を用いて2次元に圧縮し、x-y座標にプロットし、原点からの距離が遠い順に並べ、類似度の低い発言から分割する方法を提案し、実際の授業記録に適用した。とくに本稿では、分析に必要と思われる32語を用いて発言を分類した。32語で、ある程度の分類が可能で、分節わけをする際の一助になる見通しを得るには至ったが、分類の鍵となる言葉の数と、次元の問題に目を向ける

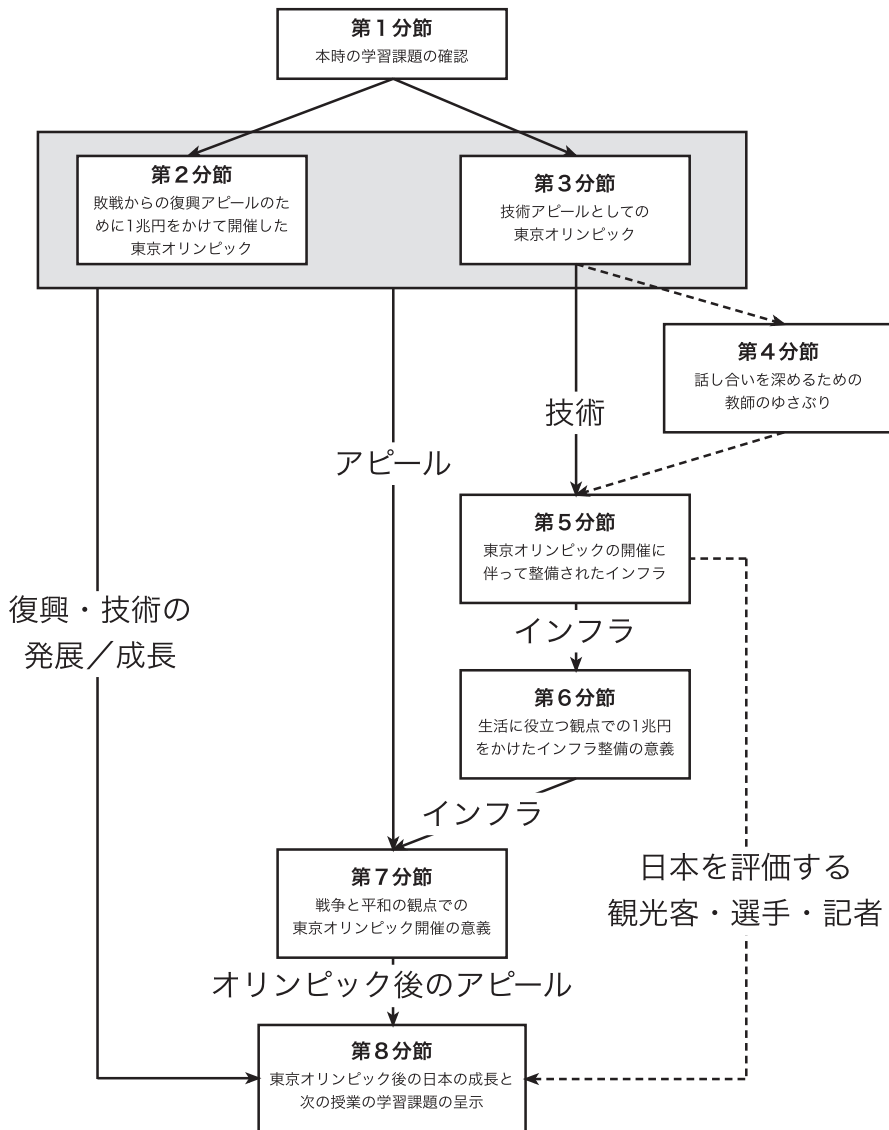


図3：本授業の分節構造図

必要があろう。

そこで、本研究では、発言記録の頻出の上位20語の場合と、上位10語の場合の分節わけの位置を、同様の方法で算出する。分析に必要と思われる32語と、授業記録の頻出上位20語は概ね似ており、その違いは、「思う」（1位）、「考える」（18位）、「分かる」（20位）のような授業の特徴的な語というより、一般的な語が20位までに入り、特徴的ではあるが出現回数の少ない語が外れた。

また、32語の言葉を対象に、3次元に圧縮した場合の分節わけの位置を、同様の方法で算出する。分節の

数を8と想定して、原点からの距離が遠い順から7個を取り上げ、発言記録を区切った結果を図4に示す。発言番号横の数字は、原点からの距離の順位である。

図4から、1) すべての実験において、発言69が最初のわけ目になること、2) 32語、上位20語、上位10語のいずれも、原点からの距離に並び替えたときの順に違いはあるが、分節の区切り位置は同じであること（3番目と4番目、6番目と7番目が入れ替わっているのみ）、そして3) 3次元に圧縮したものは、発言47と発言118で分かれず、発言96と発言100で分かれていることがわかる。

授業分析におけるデータサイエンス活用の可能性

注目語：32語 次元数：2	注目語：頻出上位20語 次元数：2	注目語：頻出上位10語 次元数：2	注目語：32語 次元数：3
<p>1 T-1：今日の学習問題はなんでしたか？</p> <p>3 T-3：東京オリンピックね。(略) お金か1兆円(略) 使い過ぎなんじゃないか(略) 考えている意見、どうですか？ そのか-5：他国へのアピールだと思ふなぞき-10：平和で観客余剰があるということをみせたかったんじゃないか しよや-12：観光客に日本のことを知ってもらえる はると-13：日本の技術を知らせたい</p> <p>14 T-14：そのかさんの意見とは違うの？(略) 技術って例えばどんな技術なの？ りよま-15：新幹線とかモノレールとか、日産スタジオムとか(略) そういう技術 たつぎ-19：(略) 技術のアピールするために、車を作って輸出している ここね-21：観光客が増えて(略) プラスって言うか、いいこと</p> <p>25 T-25：あの着て日本の技術のアピールになったかな？(略) 着て技術かな？ そのか-27：学習問題とはズレちゃう</p> <p>29 そのか-29：空席を作らなければいけないし選手も作らなきゃいけないな。東京から大阪へ行く新幹線も作らなきゃいけないし鉄道も作らなければいけない T-35：ちょっと待って、一回切るよ。しよや-39：新幹線ができたことで4時間短縮 T-42：これは技術だね(略) 2週間のためにお金使ったの？ おんふ-43：別にオリンピックのために作ったものではない そのか-44：別に1兆円使ってもいい T-45：みんなの生活に役立つの？</p> <p>47 りお-47：もし今、新幹線がなかったとしたら(略) 時間もお金もかかったと思う なぞき-49：便利なものが増えた りん-50：家族とまで遠くに行ったりすることが増えたみたい こうた-52：2020年にあるオリンピックにもホテルとカワリのホテルが絶対役立つ こうた-54：最初は1兆円は使いすぎだと思っただけど、少しは使い過ぎでもいいかなと ゆうたろう-60：最初は使い過ぎかな(略) 1兆円使っちゃった はると-62：1兆円使っても、すぐ戻ってくる こうた-64：そんな簡単に1兆円は手に入るお金じゃない</p> <p>69 T-69：(略) 全世界に放映されたんだよね。テレビでね。でも、アピールじゃないよ。今でも使うのだからちゃんと作った意味があったんじゃないかと考えている人もいました。 このか-72：坂井雅則さんなんですけど、東京オリンピックの開会式で、最終聖火ランナーを務めることで このか-74：この人は1945年8月6日に生まれたので戦争をもうしないでっていう風に言いたかったのかなと そのか-77：広島原爆が落とされた日 T-80：みんな、このこと知らなかったよね このか-83：原爆が落とされた日に生まれるのは、私だったら嫌で、二度と戦争はしないっていう思いは当然 なぞき-87：走るだけで戦争をしないというメッセージにはなりにくいんじゃないか なぞき-91：(略) 色んな被害に日本は遭って、こまで復興したけど、こんなにみんなが明るくて(略) 平和な国が一番いいよねという感じで、日本全体が戦争をしないみたいなアピールにはなかったかな そのか-92：走るだけでは平和のメッセージは分らないけど、オリンピックをやると自体、平和な国じゃないといけないから(略) 聖火ランナーから戦争をしないイメージはしない T-95：当時の新聞記事を見てください T-100：選手一人ひとりの声を聴んでみてください はると-102：東京オリンピックがすごいっていう意見の人が多くて思いました T-109：日本がアピールしたかったところが伝わってるのかな？</p> <p>118 T-118：一人一人ね、この意味って言うのを考えて欲しいな(略) 一方で、オリンピック後のことを考えて(略) オリンピックここまでよかったですね終わ たつぎ-121：オリンピックが1964年なんですけど(略) 車の生産台数が994万台くらいに増えている</p> <p>122</p>	<p>1 T-1：今日の学習問題はなんでしたか？</p> <p>3 T-3：東京オリンピックね。(略) お金か1兆円(略) 使い過ぎなんじゃないか(略) 考えている意見、どうですか？ そのか-5：他国へのアピールだと思ふなぞき-10：平和で観客余剰があるということをみせたかったんじゃないか しよや-12：観光客に日本のことを知ってもらえる はると-13：日本の技術を知らせたい</p> <p>14 T-14：そのかさんの意見とは違うの？(略) 技術って例えばどんな技術なの？ りよま-15：新幹線とかモノレールとか、日産スタジオムとか(略) そういう技術 たつぎ-19：(略) 技術のアピールするために、車を作って輸出している ここね-21：観光客が増えて(略) プラスって言うか、いいこと</p> <p>25 T-25：あの着て日本の技術のアピールになったかな？(略) 着て技術かな？ そのか-27：学習問題とはズレちゃう</p> <p>29 そのか-29：空席を作らなければいけないし選手も作らなきゃいけないな。東京から大阪へ行く新幹線も作らなきゃいけないし鉄道も作らなければいけない T-35：ちょっと待って、一回切るよ。しよや-39：新幹線ができたことで4時間短縮 T-42：これは技術だね(略) 2週間のためにお金使ったの？ おんふ-43：別にオリンピックのために作ったものではない そのか-44：別に1兆円使ってもいい T-45：みんなの生活に役立つの？</p> <p>47 りお-47：もし今、新幹線がなかったとしたら(略) 時間もお金もかかったと思う なぞき-49：便利なものが増えた りん-50：家族とまで遠くに行ったりすることが増えたみたい こうた-52：2020年にあるオリンピックにもホテルとカワリのホテルが絶対役立つ こうた-54：最初は1兆円は使いすぎだと思っただけど、少しは使い過ぎでもいいかなと ゆうたろう-60：最初は使い過ぎかな(略) 1兆円使っちゃった はると-62：1兆円使っても、すぐ戻ってくる こうた-64：そんな簡単に1兆円は手に入るお金じゃない</p> <p>69 T-69：(略) 全世界に放映されたんだよね。テレビでね。でも、アピールじゃないよ。今でも使うのだからちゃんと作った意味があったんじゃないかと考えている人もいました。 このか-72：坂井雅則さんなんですけど、東京オリンピックの開会式で、最終聖火ランナーを務めることで このか-74：この人は1945年8月6日に生まれたので戦争をもうしないでっていう風に言いたかったのかなと そのか-77：広島原爆が落とされた日 T-80：みんな、このこと知らなかったよね このか-83：原爆が落とされた日に生まれるのは、私だったら嫌で、二度と戦争はしないっていう思いは当然 なぞき-87：走るだけで戦争をしないというメッセージにはなりにくいんじゃないか なぞき-91：(略) 色んな被害に日本は遭って、こまで復興したけど、こんなにみんなが明るくて(略) 平和な国が一番いいよねという感じで、日本全体が戦争をしないみたいなアピールにはなかったかな そのか-92：走るだけでは平和のメッセージは分らないけど、オリンピックをやると自体、平和な国じゃないといけないから(略) 聖火ランナーから戦争をしないイメージはしない T-95：当時の新聞記事を見てください T-100：選手一人ひとりの声を聴んでみてください はると-102：東京オリンピックがすごいっていう意見の人が多くて思いました T-109：日本がアピールしたかったところが伝わってるのかな？</p> <p>118 T-118：一人一人ね、この意味って言うのを考えて欲しいな(略) 一方で、オリンピック後のことを考えて(略) オリンピックここまでよかったですね終わ たつぎ-121：オリンピックが1964年なんですけど(略) 車の生産台数が994万台くらいに増えている</p> <p>122</p>	<p>1 T-1：今日の学習問題はなんでしたか？</p> <p>3 T-3：東京オリンピックね。(略) お金か1兆円(略) 使い過ぎなんじゃないか(略) 考えている意見、どうですか？ そのか-5：他国へのアピールだと思ふなぞき-10：平和で観客余剰があるということをみせたかったんじゃないか しよや-12：観光客に日本のことを知ってもらえる はると-13：日本の技術を知らせたい</p> <p>14 T-14：そのかさんの意見とは違うの？(略) 技術って例えばどんな技術なの？ りよま-15：新幹線とかモノレールとか、日産スタジオムとか(略) そういう技術 たつぎ-19：(略) 技術のアピールするために、車を作って輸出している ここね-21：観光客が増えて(略) プラスって言うか、いいこと</p> <p>25 T-25：あの着て日本の技術のアピールになったかな？(略) 着て技術かな？ そのか-27：学習問題とはズレちゃう</p> <p>29 そのか-29：空席を作らなければいけないし選手も作らなきゃいけないな。東京から大阪へ行く新幹線も作らなきゃいけないし鉄道も作らなければいけない T-35：ちょっと待って、一回切るよ。しよや-39：新幹線ができたことで4時間短縮 T-42：これは技術だね(略) 2週間のためにお金使ったの？ おんふ-43：別にオリンピックのために作ったものではない そのか-44：別に1兆円使ってもいい T-45：みんなの生活に役立つの？</p> <p>47 りお-47：もし今、新幹線がなかったとしたら(略) 時間もお金もかかったと思う なぞき-49：便利なものが増えた りん-50：家族とまで遠くに行ったりすることが増えたみたい こうた-52：2020年にあるオリンピックにもホテルとカワリのホテルが絶対役立つ こうた-54：最初は1兆円は使いすぎだと思っただけど、少しは使い過ぎでもいいかなと ゆうたろう-60：最初は使い過ぎかな(略) 1兆円使っちゃった はると-62：1兆円使っても、すぐ戻ってくる こうた-64：そんな簡単に1兆円は手に入るお金じゃない</p> <p>69 T-69：(略) 全世界に放映されたんだよね。テレビでね。でも、アピールじゃないよ。今でも使うのだからちゃんと作った意味があったんじゃないかと考えている人もいました。 このか-72：坂井雅則さんなんですけど、東京オリンピックの開会式で、最終聖火ランナーを務めることで このか-74：この人は1945年8月6日に生まれたので戦争をもうしないでっていう風に言いたかったのかなと そのか-77：広島原爆が落とされた日 T-80：みんな、このこと知らなかったよね このか-83：原爆が落とされた日に生まれるのは、私だったら嫌で、二度と戦争はしないっていう思いは当然 なぞき-87：走るだけで戦争をしないというメッセージにはなりにくいんじゃないか なぞき-91：(略) 色んな被害に日本は遭って、こまで復興したけど、こんなにみんなが明るくて(略) 平和な国が一番いいよねという感じで、日本全体が戦争をしないみたいなアピールにはなかったかな そのか-92：走るだけでは平和のメッセージは分らないけど、オリンピックをやると自体、平和な国じゃないといけないから(略) 聖火ランナーから戦争をしないイメージはしない T-95：当時の新聞記事を見てください T-100：選手一人ひとりの声を聴んでみてください はると-102：東京オリンピックがすごいっていう意見の人が多くて思いました T-109：日本がアピールしたかったところが伝わってるのかな？</p> <p>118 T-118：一人一人ね、この意味って言うのを考えて欲しいな(略) 一方で、オリンピック後のことを考えて(略) オリンピックここまでよかったですね終わ たつぎ-121：オリンピックが1964年なんですけど(略) 車の生産台数が994万台くらいに増えている</p> <p>122</p>	<p>1 T-1：今日の学習問題はなんでしたか？</p> <p>3 T-3：東京オリンピックね。(略) お金か1兆円(略) 使い過ぎなんじゃないか(略) 考えている意見、どうですか？ そのか-5：他国へのアピールだと思ふなぞき-10：平和で観客余剰があるということをみせたかったんじゃないか しよや-12：観光客に日本のことを知ってもらえる はると-13：日本の技術を知らせたい</p> <p>14 T-14：そのかさんの意見とは違うの？(略) 技術って例えばどんな技術なの？ りよま-15：新幹線とかモノレールとか、日産スタジオムとか(略) そういう技術 たつぎ-19：(略) 技術のアピールするために、車を作って輸出している ここね-21：観光客が増えて(略) プラスって言うか、いいこと</p> <p>25 T-25：あの着て日本の技術のアピールになったかな？(略) 着て技術かな？ そのか-27：学習問題とはズレちゃう</p> <p>29 そのか-29：空席を作らなければいけないし選手も作らなきゃいけないな。東京から大阪へ行く新幹線も作らなきゃいけないし鉄道も作らなければいけない T-35：ちょっと待って、一回切るよ。しよや-39：新幹線ができたことで4時間短縮 T-42：これは技術だね(略) 2週間のためにお金使ったの？ おんふ-43：別にオリンピックのために作ったものではない そのか-44：別に1兆円使ってもいい T-45：みんなの生活に役立つの？</p> <p>47 りお-47：もし今、新幹線がなかったとしたら(略) 時間もお金もかかったと思う なぞき-49：便利なものが増えた りん-50：家族とまで遠くに行ったりすることが増えたみたい こうた-52：2020年にあるオリンピックにもホテルとカワリのホテルが絶対役立つ こうた-54：最初は1兆円は使いすぎだと思っただけど、少しは使い過ぎでもいいかなと ゆうたろう-60：最初は使い過ぎかな(略) 1兆円使っちゃった はると-62：1兆円使っても、すぐ戻ってくる こうた-64：そんな簡単に1兆円は手に入るお金じゃない</p> <p>69 T-69：(略) 全世界に放映されたんだよね。テレビでね。でも、アピールじゃないよ。今でも使うのだからちゃんと作った意味があったんじゃないかと考えている人もいました。 このか-72：坂井雅則さんなんですけど、東京オリンピックの開会式で、最終聖火ランナーを務めることで このか-74：この人は1945年8月6日に生まれたので戦争をもうしないでっていう風に言いたかったのかなと そのか-77：広島原爆が落とされた日 T-80：みんな、このこと知らなかったよね このか-83：原爆が落とされた日に生まれるのは、私だったら嫌で、二度と戦争はしないっていう思いは当然 なぞき-87：走るだけで戦争をしないというメッセージにはなりにくいんじゃないか なぞき-91：(略) 色んな被害に日本は遭って、こまで復興したけど、こんなにみんなが明るくて(略) 平和な国が一番いいよねという感じで、日本全体が戦争をしないみたいなアピールにはなかったかな そのか-92：走るだけでは平和のメッセージは分らないけど、オリンピックをやると自体、平和な国じゃないといけないから(略) 聖火ランナーから戦争をしないイメージはしない T-95：当時の新聞記事を見てください T-100：選手一人ひとりの声を聴んでみてください はると-102：東京オリンピックがすごいっていう意見の人が多くて思いました T-109：日本がアピールしたかったところが伝わってるのかな？</p> <p>96 T-96：ちょっと読んでみてください。5分くらいです</p> <p>100 はると-102：東京オリンピックがすごいっていう意見の人が多くて思いました T-109：日本がアピールしたかったところが伝わってるのかな？ T-113：平和っていう部分はどうかかな なぞき-114：こうやってインテリジェンタに答えてくれていると自体が(略) 平和だと思ふ T-118：一人一人ね、この意味って言うのを考えて欲しいな(略) 一方で、オリンピック後のことを考えて(略) オリンピックここまでよかったですね終わ たつぎ-121：オリンピックが1964年なんですけど(略) 車の生産台数が994万台くらいに増えている</p>

図4：実験的な分析の結果

1のことは、発言69が、それまでに話し合っていたアピールやインフラの話から、戦争と平和の話に変わった箇所であり、本授業が大きく話題転換した箇所であるからと考えられる。そのため、今回は8つの分節に分けたが、もしそれよりも多くても少なくても、同様に発言69が最初のわけ目になると想像される。次に2のことから、分析者が読んで受ける逐語記録に対する印象は、上位10語に収まる高頻出の語であると考えられる。上述したように、「思う」が高頻出の語であるが、それ以外は共通していることが影響していると考えられる。しかし、原点からの距離が各実験で異なることから、分節数を8以外の数字に設定した場合、分節のわけ目が異なる可能性があると思われる。そして3は、2次元に圧縮したもの（32語、上位20語、上位10語）の6番目と7番目に原点からの距離が遠い発言で分かれず、3次元に圧縮したものでは、教師が新聞記事を読ませている箇所（発言96）と、第5分節までに出てきた観光客から派生した選手や記者に注目させている箇所（発言100）で分かれていることを表している。これは、2次元に圧縮したものでは、東京オリンピックによって整備されたインフラそのものの話として、その意義を分けた結果を出したが、3次元に圧縮したものでは「整備されたインフラとその意義」のように統合され、次元が増えることで注目語の「記事」や「読（む）」がトピックとしてもたらされた結果と考えられる。

これらの明らかになったことと、潜在的意味解析を用いることを踏まえて、次の観点で本研究を考察する。

〈授業の構造的な把握〉

上で示したように、発言29と発言47を除いては、教師の発言が分節のわけ目になっている。このことから、子どもが新しい話題を出しているのではなく、教師が授業の主導を握っている様子がうかがえる。しかしながら、本授業の主流と言える「アピール」を児童Skが最初に用いていることと、授業が大きく転換する第7分節の前の第5分節（発言29から）と第6分節（発言47から）のきっかけを児童が作っていることを考慮すると、本授業は、教師の授業意図に沿いつつも、子どもが授業を構成していると考えられる。さらに分析を深めることで、授業分析が取り組むべき「中核的関連の考察」（重松1961, p.22）や「思考体制の動きの追究」（重松1961, p.27）をすることの可能性を有していると考えられる。

〈分析者の主体的で対話的な態度〉

上で、逐語記録を読む意義として「学問的にリベラ

ルな態度をつらぬく」（柴田2019, p.26）を記した。そのために、ときには子どもの様子を想像したり発言の背景を推測したりすることや、自身の経験や価値観等と対峙することが求められ、主体的かつ対話的な態度が求められることも記した。本研究では、授業分析の分節わけを支援する方法として、テキストマイニングの潜在的意味解析を行い、その結果にもとづいた分節わけと分節構造図を作成した。その際に、結果として出力された分節のわけ目を用いて、ただ分けるのではなく逐語記録に戻って各分節を説明したり、発言間の関連にも目を向けて構造を明らかにしたりした。教師や子どもの発言回数、発言に含まれる語数、単語の出現回数のように、《量》を算出して比較したり傾向を明らかにしたりする単純なデータ分析とは異なり、《質》を探索的に追究するための手がかりに位置づけることで、分析者の主体的で対話的な態度は一定程度担保できるのではないかと考えられる。

〈分析対象の一回性〉

八田（1963）が教師について記す中で、授業のことを「動いているその中において、一つ一つが歴史的一回生起的現象としておこってくるもの」（p.91）と述べているように、教科や単元は繰り返されるとしても、《その授業》は一回限りである。さらには、昨日のある児童は、今日の児童とも別に捉えることがあり、一回限りの《その授業》は極めて「一」である。一方、多くのデータを収集したり、もともとあるビッグデータを用いたりして統計処理をし、特徴や傾向を解析するデータサイエンスでは、「多」を「一」と見なす。したがって、逐語記録をいくつも収集・蓄積してデータサイエンスで解析したとしても、その結果がどの程度ほかの授業に意義のあるものになるかは、疑問が残る。ただし、教師の問いかけに子どもがどのように応答したかを蓄積し、子どもの可能性を引き出す教師の問いがAIから明らかになるとしたらデータサイエンスの意味があるが、極めて「一」の《その授業》からAIが答えを出すためには、数百、数千の類似した教師の問いかけを蓄積する必要がある。そのように考えると、授業の一回性という性質と、データサイエンスの状況は矛盾関係にあると言える。この点を考慮した、授業分析へのデータサイエンスの適用を検討する必要がある。

6. まとめと今後の課題

本研究では、テキストマイニングの一手法である潜在的意味解析を用いた分節わけを試みた。この方法は本来、発言と出現した語のマトリクスを特異値分解

し、似た発言毎に分類する手法であるが、本研究では、2次元に圧縮した際の座標と原点からの距離に注目し、その距離の遠い順（ほかと似ていない順）に発言をランキングし、順に発言を区切り、それを分節のわけ目とした。

分類の指標となる語を、分析者が設定するものとして32語を分析的に選択し、2次元に圧縮した後、8分節に分けると仮定して、距離の遠い順に並べた上位7発言を分節わけの箇所と設定した。ただし一般的には、分類の指標となる語は、形態素解析をし、高頻度で出現する語を設定する機会が多いため、上で述べた、分析者が設定した場合との違いを比較し、検証した。また、より多様に分類することが可能になるかどうかを検証するために、圧縮する次元を3次元に設定して比較した。その結果、分節わけの手がかりとなる原点からの距離に多少の違いはあるものの、概ね、いずれの設定も似た結果になり、本研究で対象とした授業記録に限って言えば、分析者が語や語数、次元数を比較的自由に設定できると考えられる。

今回は分析対象にしていないが、授業分析の初学者や経験者を対象に、分節わけの共通や相違を明らかにすることに加えて、本研究で用いた手法との共通や相違を明らかにすることは今後の課題である。上で述べたように、余ほどの特異な語を分類の指標となる語に設定しない限り、似た結果になるのだとしたら、テキストマイニングを用いて授業記録を分析する意義そのものを検討する必要がある。もし人間による分析と機械による分析と違いが「時間」や「労力」であるならば、授業分析の経験を積むことに注目すれば、分析観点の取得や豊かな経験を得られない点で「損」をすることが出てくるだろう。例えば、3日かけていた分節わけが、テキストマイニングによって1日で済んだとしても、空いた2日を分析的に実りある日として我々は過ごすことができるだろうかという問いが生じる。一方で、多忙な小・中・高校の教員が授業を分析しようとする場合、「時間」や「労力」の削減は歓迎される事柄であり、テキストマイニングのような分析は貢献するだろう。授業分析の根本的な意義を、その役割や状況に場合分けて検討することは、今後の課題である。

本研究で分析対象にした授業記録は、一回性の授業を文字起こししたものである。一方、テキストマイニングで分析対象にするテキストの多くは、犯罪にかかわる文書を対象にテキストマイニングを適用し、筆者の識別の有効性を検討した財津・金（2015）のように、長年に渡る様々な文書を用いたり、上で記した、松河

ら（2017）のように、同じような授業を数年に渡って繰り返して行い、そこで収集した授業評価アンケートの自由記述を用いたりする。「授業」という点では本研究と同じであるが、授業評価アンケートの自由記述の場合、テーマを絞った書き言葉であるため、話し言葉である逐語記録のような多様な表現がされにくく、また、一回性の特徴が表れにくい。したがって、複数年に渡った研究が比較的容易である。当然、同一の単元で、単元全体の逐語記録を収集し、それを分析対象にすれば、教室毎の特徴を明らかにすることはできるだろう。しかし、そうすることで、教師の問いの工夫による子どもの発言の相違、子どもの生活経験の相違とそれによる発言の相違など、授業の有する個性が、「一」で捉えられなくなる。逐語記録に表れる特徴でも、人間の読みや解釈では見落とされる事柄を明らかにすることができれば、言い換えれば、本研究で試みたように、テキストマイニングによって、分類し難い外れ値を一回の授業記録からでも見出すことができれば、授業が有する個性を保ちつつ、授業分析にとっての有益な情報が得られよう。そのような手法を開発することも、今後の課題である。

〔参考文献〕

- Becker, H. S. (1953) The Teacher in the Authority System of the Public School. *The Journal of Educational Sociology*, Vol. 27, No. 3, pp. 128-141.
- Bellack, A. A. (1968) *Methods for Observing Classroom Behavior of Teachers and Students*.
- Bloome, D., Harris, O. L., & Ludlum, D. E. (1991) Reading and writing as sociocultural activities: Politics and pedagogy in the classroom. *Topics in Language Disorders*, 11 (3), pp. 14-27.
- Damron, J. C. (1996) *Instructor Personality and the Politics of the Classroom*. Unpublished manuscript (http://krypton.mnsu.edu/~br8520zh/Damron_politics.html).
- Flanders, N. A. (1965) *Teacher Influence, Pupil Attitudes and Achievement*. Washington, DC.: US Government Printing Office.
- Grant, R. W. (1996) The Ethics of Talk: Classroom Conversation and Democratic Politics. *Teachers College Record*, Vol. 97, No. 3, pp. 470-482.
- Hargie, O. D. W. (1978) The Importance of Teacher Questions in the Classroom. *Educational Research*, Volume 20, Issue 2, pp. 99-102.
- 八田昭平 (1963) 授業分析の立場と視点. 重松鷹泰・上

- 田薫・八田昭平編, 授業分析の理論と実際. pp. 52-121, 黎明書房, 愛知.
- 本間宏志 (2015) 小六・社会科 子どもが社会的事象を自分とのかかわりでとらえ, 主体的に追及する授業をどう創ったらよいか ―新しい世の中へ～東京オリンピックから考える戦後～. 考える子ども, No. 366, pp. 71-86.
- Johnson, V. G. (1994) Student teachers' conceptions of classroom control. *The Journal of Educational Research*, 88 (2), pp. 109-117.
- 李慧瑛・下高原理恵・緒方重光 (2017) 知識創出支援ツールとしてのテキストマイニングの強みと弱み. *情報の科学と技術*, 67巻, 12号, pp. 643-649.
- Lipscomb, M. (2019) Politics in the classroom. *Nursing Philosophy*. 20 (3), pp. 1-8.
- 松河秀哉・大山牧子・根岸千悠・新居佳子・岩崎千晶・堀田博史 (2017) トピックモデルを用いた授業評価アンケートの自由記述の分析. *日本教育工学会論文誌*, 41巻, 3号, pp. 233-244.
- 内閣府 (2016) 第5期科学技術基本計画. (最終閲覧日: 2020年1月27日)
<https://www8.cao.go.jp/cstp/kihonkeikaku/5honbun.pdf>
- 小野大樹・高林克日己・鈴木隆弘・横井英人・井宮淳・里村洋一 (2004) テキストマイニングによる退院サマリー自動分類の試み. *医療情報学*, 24(1), pp. 35-44.
- Pearson, P. D. & Dole, J. A. (1987) Explicit comprehension instruction: A review of research and a new conceptualization of instruction. *The Elementary School Journal*, 88 (2), pp. 151-165.
- 坂本将暢 (2017) 世界授業研究学会の研究動向 ―発表題目の計量的分析を通して―. 日本教育工学会監修・小柳和喜雄・柴田好章編著, *教育工学会選書 II Lesson Study*, pp. 209-225, ミネルヴァ書房, 東京.
- 柴田好章 (2019) 一人一人の子どもに応じた教育実践を展開するために ―授業記録にもとづく授業研究による教師の発達―. 考える子ども, No. 392, pp. 22-28.
- 重松鷹泰 (1961) 授業分析の方法. 明治図書, 東京.
- 辛沅夏 (2017) セグメント構造に基づく学术论文の自動要約. 北陸先端科学技術大学院大学修士論文, (最終閲覧日: 2020年1月27日)
<https://dspace.jaist.ac.jp/dspace/bitstream/10119/14148/5/paper.pdf>
- Wilens, W. W. & Clegg Jr. A. A. (1986) Effective Questions and Questioning: A Research Review. *Theory & Research in Social Education*, Volume 14, Issue 2, pp. 153-161.
- 財津亘・金明哲 (2015) テキストマイニングを用いた犯罪に関わる文書の筆者識別. *法科学技術*, 20(1), pp. 1-14.

The Possibility of Utilizing Data Science in Lesson Analysis: Attempts with the Segment Division of Transcripts using Latent Semantic Analysis

Masanobu SAKAMOTO*

The purpose of this study is to explore whether it is possible to apply data science to Lesson Analysis, expanding current knowledge in the field. In particular, an attempt at a segment division using Latent Semantic Analysis, one method of text mining, is the purpose of this study.

The Lesson Analysis that the author is currently performing is one of other methodologies initiated by Shigematsu and Ueda in the 1950s. It has continued to be utilized in the Method of Education at Nagoya University for approximately 70 years. It requires an analysis using a transcript that potentially reveals the child's thought and learning processes.

In this study the appearance of words contained in a transcript, and the subsequent performance of a segment division is the focus and first task of Lesson Analysis. In order to understand the lesson structure and consider each part of the structure, a lesson is divided into “paragraphs.” This is referred to as a segment division. Because there is no specific “solution for the segment division,” it is essential to explore a “reasonable parting” by reading the transcript deeply and perhaps numerous times, as well as discussing it passionately. The creation of such a lesson plan is not an easy task.

Of course, whether a single segment is appropriate or not is questionable. For example, beginners without sufficient analysis skills may cease to make individual investigations toward successful discussions, resulting in an inactive group session, which does little to promote their studies. Thus, this research attempts to ascertain whether beginners can make segment divisions effectively by using data science. As a first step in this research subject, the author attempts to see whether valid clues can be obtained in a segment division using Latent Semantic Analysis.

Data used in this study is in the form of a transcript of a lesson of social studies. The lesson has 123 remarks. The remark in the number is about half of the number of classes. Since the student is talking about their own opinion, the number of characters in one speech is often compared to other lessons. The reason the author chose this particular transcript is that it contains various topics in the remarks, and because there are many remarks, even if there is only one remark, the number of remarks is less. Analysis procedure performs a singular value decomposition using statistical software that calculates the strengths of the degree of association of utterances. The value is plotted in a two-dimensional plane, with segments divided from the distance farthest from the origin. The analysis selected in this study was a characteristic 32 words. Also, in order to compare and verify this analysis, the author compared the results analyzed by 20 words and 10 words being the most frequent, with the results analyzed by setting the three-dimensional dimensions of compressing.

As a result, some of differences in the distance from the origin, related to the clue of the segment division, generally found a result resembling the settings. As far as transcript used in this study, the Lesson Analysis is considered to be set words, number of words, or number of dimensions a, utilized relatively freely.

* Associate Professor, Graduate School of Education and Human Development, Nagoya University