

テキストマイニングによる自由記述の分析手法の提案

— データの概要提示の簡素化による業務改善を目指して —

専門支援部研修課 長期研修員 長田 裕介

1 主題設定の理由

長期研修員として各種研修の運営補助をしており、研修後に提出される研修アンケートや、演習で作成された研修成果物を目にする機会が多い。研修アンケートには4件法による研修満足度及びその理由を記述する箇所がある。研修担当者は全文を精読し、概要を把握する作業を行うが、それには多くの時間と労力がかかること及び、恣意的になりやすいことが考えられる。そこで、研修アンケートの内容満足の理由を効率的かつ客観的にまとめることができないかと考えた。また、研修成果物は、研修後に掲示・ギャラリーウォーク・保管するだけでなく、記入されたテキストを整理し、研修員の思考の深化、視点の変化等を読み取ることで、研修改善の一助になるのではないかと考えた。

現在、多くの人がスマートフォンやタブレット端末を使用しており、各種研修や学校現場においてアンケート等の電子化が進んでおり、紙媒体の記述をテキスト入力し、データを作成する過程が簡略化されつつある。

そこで、既にあるテキストデータを要約し、概要を示すテキストマイニング¹の手順を提案することで、静岡県総合教育センター主催研修（以下「センター研修」という。）の改善だけでなく、学校における様々なアンケートにおいても利用でき、学校行事や授業におけるPDCAサイクルの構築に寄与し、業務改善に役立てられると考え、主題を設定した。

2 研究の目的

テキストデータにおける語句の傾向や関係性を分析することにより、自由記述を効率的・客観的に要約・提示する手法を提案し、業務改善に資することを目的とする。

3 研究の方法

テキストデータの計量分析用ソフトウェア「KH Coder²」及び形態素解析ソフトウェア「Chasen³」を使用し、センター研修における研修アンケート及び研修成果物であるグループワーク付箋紙のテキストデータの分析を行う。

4 研究の内容

(1) 分析手法の提案

ア 語の抽出と頻出語の確認

はじめに、形態素解析ソフトウェア「Chasen」を使用し、テキストを語に分解する。意味が同一であるが漢字と平仮名、送り仮名の違いによる表記の仕方が異なる語が

表1 除外品詞名とその例

KH coder 内の品詞名	例
名詞B	まち、ゆとり、ひとり
動詞B	する、ある、わかる
形容詞B	ない、よい、きびしい
副詞B	もっと、あまり
否定助動詞	ない、ぬ
形容詞（非自立）	ほしい、よい、やすい

あるため、元のテキストを全て確認し、統一する。その後、「KH Coder」のデフォルト設定である、「名詞B」「動詞B」「形容詞B」「副詞B」「否定助動詞」「形容詞（非自立）」を除外する。該当の品詞名とその例を表1に示す。なお、「動詞B」のような「B」を付した品詞名は、平仮名だけで構成されている語を表しており、どのようなテキストの中でも出現するような一般的な語が多く、内容を反映しないことが多い。

ごくわずかな回数しか出現していない語は統計的な分析には利用しにくいため除外する。目安として、語を出現回数の降順に並び替え、上位から30位に達した出現回数を分析の最少出現回数とし、ここまでの頻出語を分析対象とする。その後、頻出語のみの分析では、解釈の容易性という点で難が残り、語と語のつながりは全く把握できないため、複数の語が同一テキストの言語的環境において生起しているかを意味する共起関係を調べる。

イ 共起分析

(7) Jaccard 係数

共起関係の強さの尺度として Jaccard 係数を適用した共起ネットワーク分析により、語同士のつながりを把握し、より詳細な情報を抽出する^[1]。Jaccard 係数とは、ある語の組合せに対する積集合を和集合で除したものであり、整理すると図1のようになる^[2]。つまり、Jaccard 係数は特定の二つの語のどちらか若しくは両方が用いられたテキストのうち、両方が用いられているテキストの割合を表しており、0から1の間の値をとる。値が大きいほど二つの語が両方使われているテキストの割合が大きく、それぞれの語がテキストに用いられている二つの集合

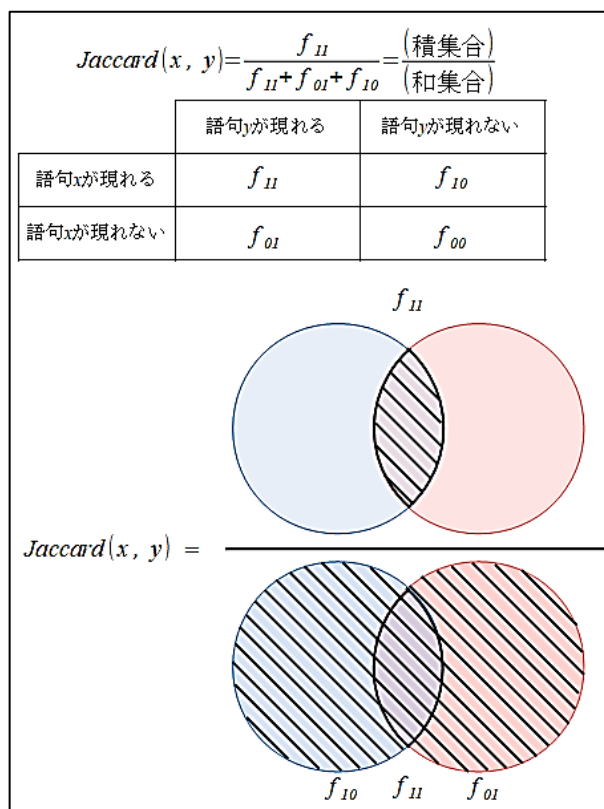


図1 Jaccard 係数

の類似度は高いということになる。Jaccard 係数の値については 0.1 であれば関連がある、0.2 であれば強い関連があるという目安で判断するものとする。

(4) 共起ネットワーク

出現回数が多い語の中で、特定の語との結びつきの基準として $Jaccard \geq 0.2$ を共起ネットワーク分析の基準とする。ただし、調査対象によっては、この条件を満たす共起関係が一定数あるとは限らない。描画されるネットワーク自体が少な

い場合、全体像を把握することは困難であるため、その場合については出現回数が上位 30 位程度の語かつ Jaccard 係数を降順に並び替えた内の上位 20 位までを示すものとする。

図 2 は共起ネットワークの例を示している。共起ネットワークについては、単に語がお互いに近くに布置されているというだけでは、それらの語の間に強い共起関係があることを意味しない。重要なのは線で結ばれているかどうかであって、近くに布置されているだけで線で結ばれていなければ、強い共起関係はない点に注意が必要である。

共起ネットワークに関する他の説明を以下に示す。

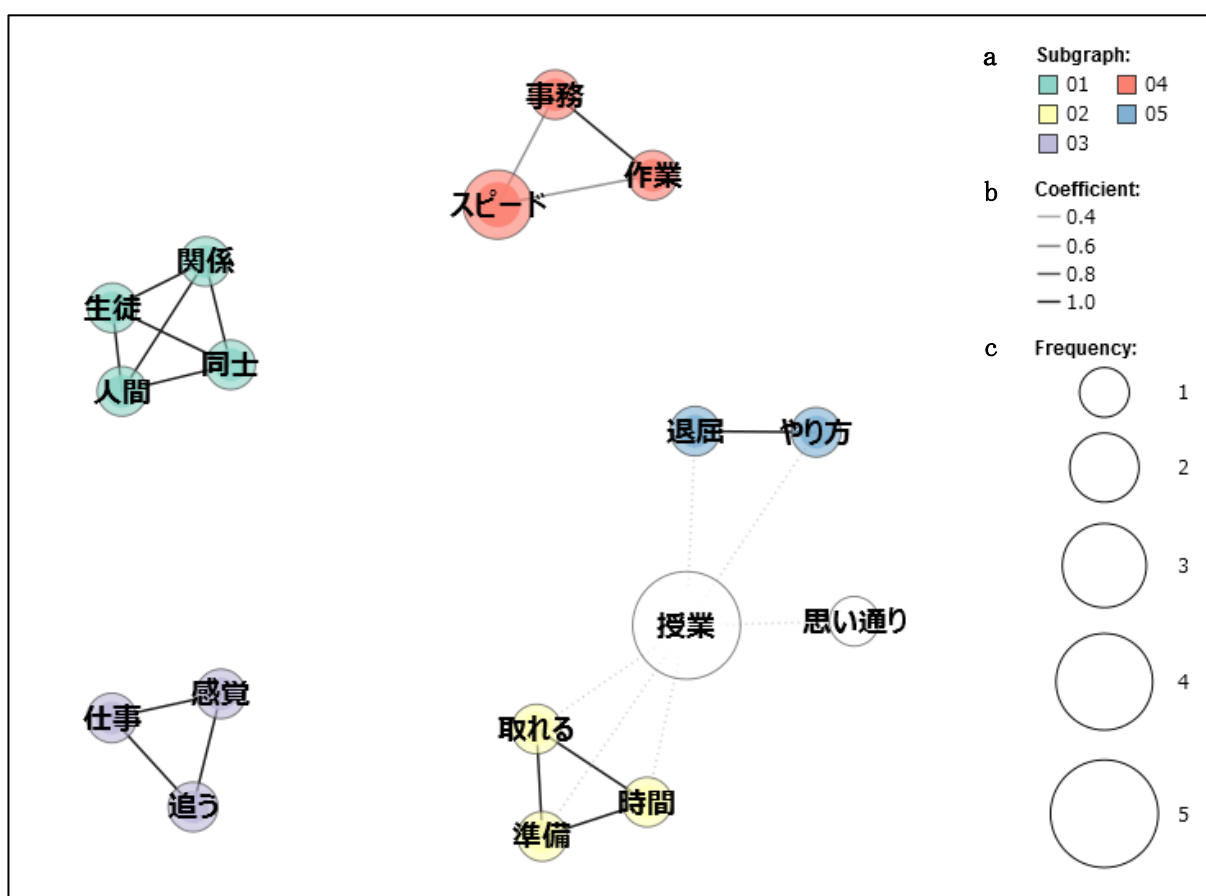


図 2 共起ネットワーク例

a カラー

比較的強くお互いに結びついている語を自動的に検出してグループ分けを行い、その結果を色分けによって示している。背景が白で、丸い囲み枠が黒色であれば、他の語とグループを形成していない語であることを意味している。

b 共起の程度と線の太さ

相対的に太く描画されているほど、共起の程度は強い。

c 出現語と語の大きさ

出現数の多い語ほど大きい円で描画されている。

(ウ) 意味の抽出

図2の共起ネットワークを確認し、元のテキストに戻り、抽出された語の組合せがどのように用いられているのかを全て確認し、要約する。

(イ) 提案する分析手順

本来、「KH Coder」では分析者が作成した基準（コーディングルール）に従って語やテキストを分類する Dictionary-based アプローチと、共通する語を多く含むテキストのグループを、多変量解析によって自動的に発見・分類する Correlational アプローチを併用した接合アプローチを用い、計量的分析を行う^[1]。しかし、本研究では、Correlational アプローチのみを用いることで、分析者のもつ理論や問題意識の影響を極力受けない形で、テキストデータを扱う。なお、抜粋については、テキストをそのまま抜粋したものは“ ”内に記し、分析者による補足をく)内に記す。抽出語については下線を付している。以上に述べた分析手順を図示すると図3のようになる。

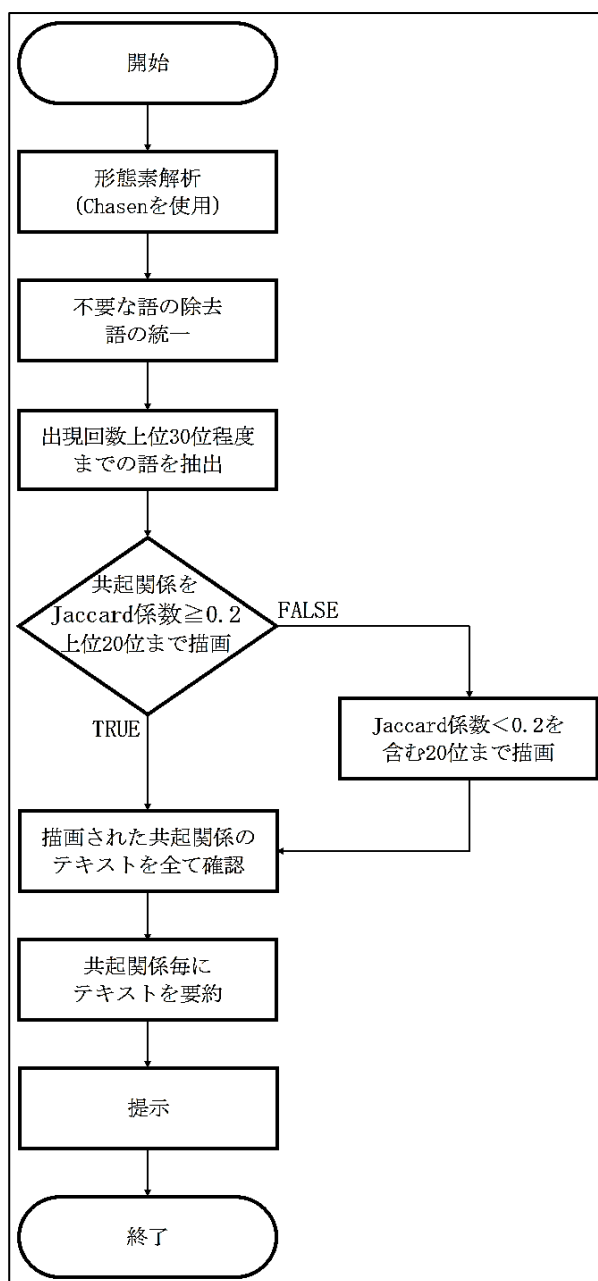


図3 分析手順のフローチャート

(2) 研修アンケートの分析

ア 平成31年度第1回公立高等学校中堅教諭等資質向上研修

(7) 研修アンケートの概要

第1回公立高等学校中堅教諭等資質向上研修（以下「高校中堅研」という。）が、「学校組織活性化のためのマネジメントスキルを理解する」、「学習指導、生徒指導、教職員の勤務服务等の専門的な知識・技能に関する理解を深める」、「学校組織におけるミドル

表2 平成31年度第1回高校中堅研センター研修内容

回・期日等	主たる研修内容
第1回 6月6日 (木) 7日 (金)	①開講式 ②オリエンテーション ③教育法規 ④学校組織マネジメントの基礎 ⑤学校内外の環境分析 ⑥学校と警察の連携 ⑦カリキュラムマネジメントの導入 ⑧教科別研修

リーダーとして、学校運営や職場の人材育成に貢献する意欲を高める」ことを目標に実施された(表2)。1日目の開講式後に実施されたオリエンテーションについて、研修アンケート106枚のテキストデータ(内容満足の理由)を分析対象とした。

(イ) 語の抽出と頻出語の確認

「KH Coder」を用いて前処理を実行し、テキストの単純集計を行った結果、145の文が確認された。総抽出語数2,688語(異なり語数520語)から「助詞」、「助動詞」、「名詞B」、「動詞B」、「形容詞B」、「副詞B」、「否定助動詞」、「形容詞(非自立)」を除外し、最終的に分析に使用される語数は833語(異なり語数352語)となった。その中で上位42語とその出現回数及び出現割合を表3に示した。上位42語が全833語のうち、どれくらいの比率を占めるのかを分析したところ、上位42語で全体の4割超を占めており、全833語を使わなくても、研修アンケートの概要を把握できると考えた。

表3 平成31年度第1回高校中堅研修アンケート オリエンテーション 内容満足の理由 頻出語

抽出語	出現回数	割合	抽出語	出現回数	割合
研修	25	3.0%	ポイント	6	0.7%
振り返る	22	2.6%	機会	6	0.7%
自分	21	2.5%	資料	6	0.7%
ミドル	19	2.3%	持つ	6	0.7%
思う	19	2.3%	大人	6	0.7%
リーダー	18	2.2%	年	6	0.7%
理解	15	1.8%	聞く	6	0.7%
感じる	12	1.4%	パワー	5	0.6%
今	11	1.3%	育つ	5	0.6%
中堅	11	1.3%	恩	5	0.6%
成長	10	1.2%	学ぶ	5	0.6%
内容	10	1.2%	教員	5	0.6%
年間	10	1.2%	原則	5	0.6%
良い	9	1.1%	原理	5	0.6%
改めて	8	1.0%	講義	5	0.6%
考える	8	1.0%	仕事	5	0.6%
先生	8	1.0%	思える	5	0.6%
話	8	1.0%	時間	5	0.6%
気持ち	7	0.8%	自覚	5	0.6%
求める	7	0.8%	説明	5	0.6%
経験	7	0.8%	目的	5	0.6%
			記述文総数145 使用語総数833		

に分析に使用される語数は833語(異なり語数352語)となった。その中で上位42語とその出現回数及び出現割合を表3に示した。上位42語が全833語のうち、どれくらいの比率を占めるのかを分析したところ、上位42語で全体の4割超を占めており、全833語を使わなくても、研修アンケートの概要を把握できると考えた。

(ウ) 共起分析

上位42語について、共起分析を行い、共起ネットワークを作成した(図4)。

Jaccard \geq 0.2かつ上位20位までの語の共起関係だけを表示した。

(エ) 意味の抽出

図4について、真上から語のグループを時計回りに確認し、アンケートの内容満足の理由に影響を与えた語を探る。図4の「ミドルー自覚」を含むグループで共起関係が生じた全テキストを表4に示した。表4から、テキストを要約し、「ミドルリーダーとしての自覚を持ったから」という意味を抽出した。他のグループに

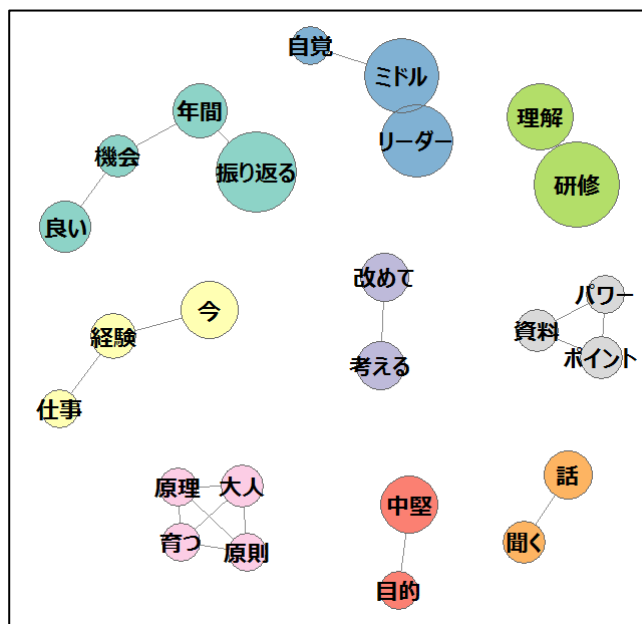


図4 平成31年度第1回高校中堅研修アンケート オリエンテーション 内容満足の理由 共起ネットワーク Jaccard係数上位20位(0.23)まで

についても同様の操作を行い、意味を抽出した。図4の共起ネットワークから抽出した結果を表5に示す。表の各列は左から語の組合せ、Jaccard係数、意味を抽出した要約結果である。表5の「要約不可」とは、同一グループのテキストにおいて、意味のまとまりが見出せず、意味を抽出できなかったグループである。例として、「話一聞く」の全テキストを表6に示す。

(オ) 考察

研修満足度は4から1を選択する4件法で、満足度が高いほど値が大きい。全体としては4が70.3%、3が26.7%、2が3.0%であり、満足度は高い。要約文からも講義内容及び自身の理解度に満足している様子が見られる。

特定の語について、「Chasen」においては2語以上で分割され、「KH Coder」の前処理の段階において複合語扱いをしていなければ、別の語として扱われる。今回の事例においては、「ミドルリーダー」、「パワーポイント」、「中堅研」、「原理原則」がこれに該当する。しかし、そのような語については、出現回数も多く、特定の語との組合せで使用されることが多いため共起関係が生起し、共起ネットワーク上に出現することにより、傾向として読み取ることができる。このように、前処理の段階において複合語を設定しないことにより、作業を効率化するとともに、分析者の恣意的な視点の介入を避けることができると推察される。

表4 図4「ミドルー自覚」を含むグループから抽出した全テキスト

<p>ミドルー自覚 (0.28)</p>	<ul style="list-style-type: none"> ・ “<u>ミドルリーダー</u>としての自覚意識が生まれました。” ・ “<u>自身がミドルリーダー</u>を担う年齢になったという自覚ができました。” ・ “この10年間で改めて振り返ることができたとともに、<u>ミドルリーダー</u>という立場になるという自覚につながったため。” ・ “<u>ミドルリーダー</u>としての自覚を持つことができました。” ・ “<u>「ミドル」</u>には「中心」の意味もあるので、<u>自覚</u>を持とうと思った。”
<p>ミドルーリーダー (0.89)</p>	<ul style="list-style-type: none"> ・ “やはり<u>「ミドルリーダー」</u>という言葉に責任を感じ、もっと頑張らねばと思いを新たにした。” ・ “<u>ミドルリーダー</u>としての自覚意識が生まれました。” ・ “<u>ミドルリーダー</u>として、気の引き締まる思いを持ちました。” ・ “<u>ミドルリーダー</u>という言葉は荷が重く感じますが、大人が育つ原理の話にあった背伸びの経験の大切さを理解したので、少しずつ<u>ミドル</u>としての役割を担っていると良いと思いました。” ・ “<u>自身がミドルリーダー</u>を担う年齢になったという自覚ができました。” ・ “<u>「ミドルリーダーの育成」</u>という研修の趣旨をしっかりと理解できたから。” ・ “<u>中堅研の目的、ミドルリーダー</u>として求められていることがよく理解できました。” ・ “自分が<u>ミドルリーダー</u>として何が求められていてこれからどうしていかなくてはいけないか把握することができた。” ・ “2年間、<u>ミドルリーダー</u>として、どのような取組をしていけば良いか、その指針としていきます。” ・ “<u>中堅研とは、ミドルリーダー研修</u>であるという、今回の研修を含めて、自分がどのような研修を行っていけば良いかという目標、目的を把握することができたからです。” ・ “<u>ミドルリーダー</u>としての意識が高まった。” ・ “<u>ミドルリーダー</u>がどのようなポジションか、求められていること、考えられることなど、項目でまとめられていて、わかりやすかったから。” ・ “この10年間で改めて振り返ることができたとともに、<u>ミドルリーダー</u>という立場になるという自覚につながったため。” ・ “<u>ミドルリーダー</u>として、さらに学ぶことで学校運営や生徒を育てる能力を身に付けることができると知ったから。” ・ “<u>ミドルリーダー</u>としての自覚を持つことができました。” ・ “<u>ミドルリーダー</u>として、自分のことだけでなく若手の同僚にも伝えて、ともに成長していきたいと感じました。” ・ “<u>ミドルリーダー</u>に求められることを確認できた。”

表5 平成31年度第1回高校中堅研 研修アンケート オリエンテーション
内容満足の原因 共起ネットワーク 要約

ミドルリーダー (0.89)	ミドルリーダーとしての <u>自覚</u> を持ったから。
ミドル自覚 (0.28)	
理解-研修 (0.33)	研修の目的・趣旨を <u>理解</u> できたから。
パワーポイント (0.80)	
資料-パワー (0.43)	要約不可
資料-ポイント (0.38)	
改めて-考える (0.23)	要約不可
話-聞く (0.27)	要約不可
中堅-目的 (0.23)	中堅研の <u>目的</u> を理解できたから。
育つ-原理 (1.00)	
大人-育つ (0.83)	
大人-原理 (0.83)	
育つ-原則 (0.67)	大人が <u>育つ原理原則</u> をよく理解できたから。
原理-原則 (0.67)	
大人-原則 (0.57)	
経験-仕事 (0.33)	
経験-今 (0.23)	これまでの <u>仕事上の実践経験</u> で今があることに気付いたから。
振り返る-年間 (0.45)	
機会-良い (0.25)	10 <u>年間</u> を <u>振り返る良い機会</u> となったから。
年間-機会 (0.23)	

表6 図4「話-聞く」より抽出した全テキスト (要約不可の例)

話-聞く (0.27)	<ul style="list-style-type: none"> ・ “話を<u>聞</u>いているだけで楽しかったです。” ・ “免許更新の<u>話</u>も思い出して<u>聞</u>いた。” ・ “新鮮な気持ちで<u>話</u>を<u>聞</u>くことができ、非常に良かった。”
-------------	--

イ 平成31年度第3回公立高等学校中堅教諭等資質向上研修

(7) 研修アンケートの概要

第1回と同じ目標で、第3回高校中堅研が実施された(表7)。なお、共起ネットワーク分析において、Jaccard \geq 0.2ではJaccard係数を降順に並び替えた内の上位20位までが表示されず、Jaccard $<$ 0.2も含めた上位20位までを表示する事例として、講義「特別支援教育」を扱う。研修アンケート104枚のテキストデータ(内容満足の原因)を分析対象とした。

表7 平成31年度第3回
高校中堅研 センター研修内容

回・期日等	主たる研修内容
第3回 11月12日 (火)	①特別支援教育 ②教育相談 ③選択演習

(4) 語の抽出と頻出語の確認

「KH Coder」を用いて前処理を実行し、テキストの単純集計を行った結果、163の文が確認された。総抽出語数3,339語(異なり語数588語)から「助詞」、「助動詞」、「名詞B」、「動詞B」、「形容詞B」、「副詞B」、「否定助動詞」、「形容詞(非自立)」を除外し、最終的に分析に使用される語数は1,038語(異なり語数388語)となった。その中で上位40語とその出現回数及び出現割合を表8に示した。上位40語が全1,038語のうちの5割超を占めており、全1,038語を使わなくても、概要を把握できると考えた。

(ウ) 共起分析

上位 40 語について、共起分析を行い、共起ネットワークを作成した（図 5）。Jaccard ≥ 0.2 の語の組合せだけでは上位 20 位までを表示できなかったため、Jaccard < 0.2 も含めた上位 20 位までの共起関係を表示した。

(エ) 意味の抽出

図 5 について、真上から語のグループを時計回りに確認し、アンケートの内容満足の原因に影響を与えた語を探る。共起関係が生じた元のテキストに戻り、意味を抽出した。なお、「理解」については、他のどのグループとも色が異なるため、要約しない。図 5 の共起ネットワークから抽出した結果を表 9 に示す。各列は左から語の組合せ、Jaccard 係数、意味を抽出した要約結果である。

表 8 平成 31 年度第 3 回高校中堅研修アンケート 講義「特別支援教育」内容満足の原因 頻出語

抽出語	出現回数	割合	抽出語	出現回数	割合
生徒	65	6.3%	学校	8	0.8%
支援	43	4.1%	指導	7	0.7%
理解	29	2.8%	接す	7	0.7%
特別	28	2.7%	全て	7	0.7%
考える	27	2.6%	大切	6	0.6%
必要	26	2.5%	知る	6	0.6%
困る	24	2.3%	良い	6	0.6%
思う	21	2.0%	ADHD	5	0.5%
教育	18	1.7%	クラス	5	0.5%
内容	16	1.5%	確認	5	0.5%
感じる	15	1.4%	環境	5	0.5%
自分	13	1.3%	気持ち	5	0.5%
対応	13	1.3%	講義	5	0.5%
改めて	12	1.2%	子ども	5	0.5%
具体	11	1.1%	実践	5	0.5%
障害	11	1.1%	振り返る	5	0.5%
多い	10	1.0%	捉える	5	0.5%
配慮	10	1.0%	伝える	5	0.5%
発達	9	0.9%	認識	5	0.5%
学ぶ	8	0.8%	本校	5	0.5%
記述文総数 163		使用語総数 1,038			

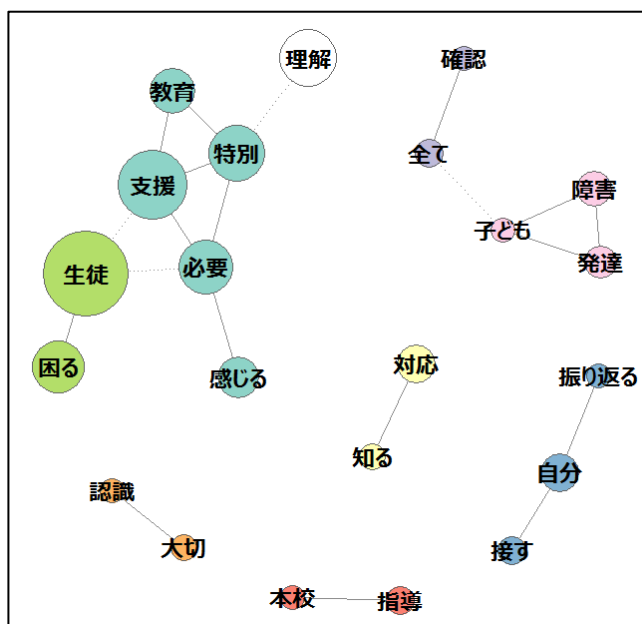


図 5 平成 31 年度第 3 回高校中堅研修アンケート 講義「特別支援教育」内容満足の原因 共起ネットワーク Jaccard 係数上位 20 位 (0.19) まで

表 9 平成 31 年度第 3 回高校中堅研修アンケート 講義「特別支援教育」内容満足の原因 共起ネットワーク 要約

支援－特別 (0.67)	
特別－教育 (0.48)	
支援－教育 (0.36)	特別支援教育の <u>必要性</u> を感じたから。
特別－必要 (0.27)	
必要－感じる (0.22)	
全て－確認 (0.20)	<u>全て</u> の子どもに対する支援の <u>必要性</u> を再確認したから。
障害－発達 (0.90)	
発達－子ども (0.27)	<u>発達障害の子ども</u> の視点に立った演習がためになったから。
障害－子ども (0.25)	
対応－知る (0.19)	<u>対応を知る</u> ことができたから。
自分－接す (0.27)	<u>自分</u> の子どもへの <u>接し方</u> を振り返り、反省する機会となったから。
自分－振り返る (0.21)	
指導－本校 (0.20)	要約不可
大切－認識 (0.22)	相手の立場に立って考えることの <u>大切さ</u> を認識したから。
生徒－困る (0.26)	「 <u>困った生徒</u> 」ではなく「 <u>困っている生徒</u> 」という視点を理解できたから。

(オ) 考察

全体としては4が59.6%、3が30.3%、2が10.1%であり、満足度は高い。要約文からも講義内容及び自身の理解度に満足している様子が見られる。「特別支援教育」という語については、複合語に設定しなくても分析の過程で確認することができ、また、講義で特に印象に残ったと思われる「「困った生徒」ではなく「困っている生徒」という記述についても、元のテキストに戻って確認し、要約の際にそのまま利用した。

(3) グループワーク付箋紙 平成30年度第1回・第10回公立高等学校初任者研修（5月、2月）

ア 分析対象設定の理由

平成29年度実務研修員が第1回公立高等学校初任者研修（以下「高校初任研」という。）、第1回高校中堅研及び教育行政新任職員研修Ⅰ期の研修成果物（グループワーク付箋紙）を対象に形態素解析及び分類後の記述に関する統計的検定を行った。その後の課題として分析の継続や年度内変化を読み取る定点観測の必要性を述べているため、昨年度の初任研のグループワーク付箋紙に対して、本研究が提案する分析手法を用い、年度内変化を読み取ることにした^[3]。

イ グループワーク付箋紙の概要

高校初任研では、センター研修が年間10回、合計14日間実施された（表10）。センター研修以外では、校内研修及び所属校が企画する課題研究による研修がある。高校初任研では「教育に対する理解を深め、教育公務員としての自覚を高めるとともに、自己の特性や課題を把握し、実践的指導力の向上を図る」ことを目的に、様々な研修プログラムが実施された。第1回高校初任研においては研究協議「1か月を

表10 平成30年度高校初任研 センター研修内容

回・期日等	主な研修内容	回・期日等	主な研修内容
第1回 5月9日 (水)	①オリエンテーション ②開講式 ③教職員の服務 ④メンタルヘルス ⑤ソーシャルスキル ⑥1か月の振り返り	第5回 9月11日 (火) 12日 (水)	①生徒指導 ②学校と警察の連携 ③教育相談 ④ファシリテーション ⑤生徒指導
第2回 5月31日 (木)	①教科の目標・内容（学習指導要領） ②学習指導・教科指導の在り方 ③授業づくりの実際	第6回 教科ごと1日実施	①校長講話 ②研究授業参観 ③学習指導・教科指導 ④授業づくりの実際
第3回 6月28日 (木) 29日 (金)	①教育監講話 ②人権教育 ③部活動指導 ④学校行事等の組織的運営 ⑤第4回研修の事前研修 ⑥特別支援教育 ⑦生涯学習 ⑧授業の基礎技術	第7回 10月23日 (火)	①教育の情報化 ②キャリア教育 ③ホームルーム経営
第4回 8月1日 (水) ～3日 (金)	①体験的活動（共同生活、炊飯実習、キャンドルファイア実習等） ②特別活動（学校行事） ③身体表現活動 ④環境教育 ⑤4か月の振り返り ⑥教員のモラル	第8回 地区ごと1日実施	①特別支援学校訪問 ②校長講話
		第9回 1月18日 (金)	①総合的な学習の時間 ②特別活動 ③学習指導・教科指導
		第10回 2月1日 (金)	①静岡県の高校教育 ②1年間の振り返り ③閉講式

振り返る」における「課題と感じていること」を自由記述したグループワーク付箋紙 308 枚のテキストデータを、第 10 回高校初任研においては研究協議「1 年間の振り返る」における「今後の決意」記述の中で課題が記述されているグループワーク付箋紙 580 枚のテキストデータを分析対象とした。

ウ 語の抽出と頻出語の確認

(7) 平成 30 年度第 1 回高校初任研

「KH Coder」を用いてテキストの単純集計を行った結果、付箋紙枚数と同じ 308 の文が確認された。総抽出語数 2,401 語（異なり語数 591 語）から「助詞」、「助動詞」、「名詞 B」、「動詞 B」、「形容詞 B」、「副詞 B」、「否定助動詞」、「形容詞（非自立）」を除外し、最終的に分析に使用される語数は 1,022 語（異なり語数 403 語）となった。その中で上位 35 語とその出現回数を表 11 に示した。

(4) 平成 30 年度第 10 回高校初任研

同様に、「KH Coder」を用いてテキストの単純集計を行った結果、付箋紙枚数と同じ 580 の文が確認された。総抽出語数 2,764 語（異なり語数 706 語）から「助詞」、「助動詞」、「名詞 B」、「動詞 B」、「形容詞 B」、「副詞 B」、「否定助動詞」、「形容詞（非自立）」を除外し、最終的に分析に使用される語数は 1,454 語（異なり語数 547 語）となった。その中で上位 32 語とその出現回数を表 12 に示した。

表 11 平成 30 年度
第 1 回高校初任研 研究協議
「1 か月を振り返る」 頻出語

抽出語	出現回数	抽出語	出現回数
生徒	82	テスト	7
授業	78	知識	7
指導	36	バランス	6
活動	25	顧問	6
部	25	事務	6
仕事	21	先生	6
時間	21	展開	6
対応	12	配分	6
研究	11	名前	6
距離	10	感じる	5
教科	10	関係	5
教材	10	厳しい	5
準備	10	遅い	5
自分	9	分掌	5
担任	9	聞く	5
クラス	8	方法	5
少ない	8	優先	5
内容	8		
記述文総数 308		使用語総数 1,022	

表 12 平成 30 年度
第 10 回高校初任研 研究協議
「1 年間の振り返る」 頻出語

抽出語	出現回数	抽出語	出現回数
生徒	109	成長	10
指導	65	対話	10
授業	51	距離	9
ICT	22	研究	9
活動	21	持つ	9
教員	21	実態	9
活用	19	ルール	8
部	18	計画	8
学ぶ	14	考える	8
理解	14	知識	8
教材	13	コミュニケーション	7
時間	13	見る	7
自分	13	主体	7
連携	13	身	7
学習	12	評価	7
目標	12	明確	7
記述文総数 580		使用語総数 1,454	

エ 共起分析

表 11、表 12 の頻出語について、共起分析を行い、共起ネットワークを作成した（図 6・図 7）。ともに Jaccard ≥ 0.2 の語の組合せだけでは上位 20 位までを表示できなかつたため、それぞれ Jaccard < 0.2 も含めた上位 20 位までを表示した。

オ 意味の抽出

それぞれの図について実施段階におけるの課題意識を探るため、真上から語のグ

カ 考察

共起ネットワークから要約して抽出した意味を比較するため、第1回と第10回の高校初任研における課題を4つの要素に整理し、表15に示した。まず、両方で確認された課題として、「教材研究の不足」、「部活動指導」、「生徒指導」がある。これらの課題は、経験によって解消される部分もあるが、短期間で解決する課題ではないと考えられる。次に、第1回では見られたが第10回では見られない「授業準備の時間の不足」、「教科に関する知識が少ない」、「授業内の時間配分」、「顧問としての自分の役割」、「担任の先生との会話が少ない」、「分掌業務等との仕事のバランス」、「仕事の優先順位の決め方」については、概ね解消された課題と考えられる。最後に、第10回で新たに現れた「授業で考えさせる時間を計画すること」、「ICTの活用」、「主体的・対話的で深い学びの実現」、「学習指導」、「生徒理解」、「生徒との距離感」、「教員〈同僚〉との連携」、「目標の明確化」については、経験による視点の多角化・多面化により新たに獲得した課題又は第1回で表れた課題がより深化した課題と考えられ、今後更に充実した研修が求められている。グループワーク付箋紙に書かれた情報を分析することにより、研修員の視点の変化等を読み取ることができた。

表15 平成30年度高校初任研
第1回から第10回への課題意識の変化
第1回 第10回

第1回	第10回
学習指導についての課題	
<ul style="list-style-type: none"> ・教材研究の不足 ・授業準備の時間の不足 ・教科に関する知識が少ない ・授業内の時間配分 	<ul style="list-style-type: none"> ・教材研究の不足 ・授業で考えさせる時間を計画すること ・ICTの活用 ・主体的・対話的で深い学びの実現 ・学習指導
部活動指導についての課題	
<ul style="list-style-type: none"> ・部活動指導 ・顧問としての自分の役割 	<ul style="list-style-type: none"> ・部活動〈指導〉
生徒との関わり方についての課題	
<ul style="list-style-type: none"> ・生徒指導 	<ul style="list-style-type: none"> ・生徒指導 ・生徒理解 ・生徒との距離感
働き方についての課題	
<ul style="list-style-type: none"> ・担任の先生との会話が少ない ・分掌業務等との仕事のバランス ・仕事の優先順位の決め方 	<ul style="list-style-type: none"> ・教員〈同僚〉との連携 ・目標の明確化

5 研究のまとめ

(1) 研究の成果

提案した分析手法を用いることにより、多くの研修員のテキストに見られた内容について簡便に全体像を把握することができた。また、分析者が全文精読後に概要をまとめる場合、分析者によってまとめが異なるが、この分析手法においては誰が分析し

ても共起ネットワークまでは同様の結果となり、客観性が保障される。全文精読によって少数ながらも貴重な意見を確認することは当然大切であるが、本手法はテキストの概要を簡便に把握するためには有効な手段であると考えられる。また、場合によっては、把握した概要から新たな問題を発見し、Dictionary-based アプローチを使用することにより、分析者のもつ理論や問題意識を操作化することも必要であると考えられる。以下、成果についてまとめる。

- ・テキストの概要を客観的な視点で把握することができる手法を提案した。センター研修の改善につなげるだけでなく、学校現場においても、評価アンケートや授業アンケート等の分析で利用することが可能である。
- ・初任者の意識の考察については、課題であった同一年度の同一対象研修を分析することで、経験による意識の変化を読み取ることができた。また、研修改善の手立てとして、演習における研修成果物を分析することの有用性を再確認した。

(2) 今後の課題

本研究において、使用語については出現回数上位 30 位程度、共起関係については Jaccard 係数上位 20 位までを分析対象としたが、基準の妥当性を含めた分析精度の向上が課題である。さらに、分析によって得られたデータ概要の正当性の検証方法も課題であり、さらなる事例の蓄積が必要と考えられる。

【参考文献・参考 URL】

- [1] 樋口耕一. 社会調査のための計量テキスト分析 内容分析の継承と発展を目指して. ナカニシヤ出版. 2014
- [2] 株式会社 Faber Company. “【技術解説】集合の類似度 (Jaccard 係数, Dice 係数, Simpson 係数)”. ミエルカ AI - 【法人向け】自然言語処理技術を活用した、ソリューション提供. https://mieruca-ai.com/ai/jaccard_dice_simpson/. (参照 2020-01-14).
- [3] 戸塚優季. “研修改善のための研修成果物及び研修事前課題の分析-センターと学校での学びを繋ぐための活用を目指して-”. 静岡県総合教育センター. 2018

-
- 1 通常の文章からなるデータを語や文節で区切り、それらの出現回数や共起の相関、出現傾向、時系列などから統計学、パターン認識、人工知能等のデータ解析の技法を適用することで有用な情報を取り出す、テキストデータの分析方法である。
 - 2 テキスト型データの計量テキスト分析もしくはテキストマイニングのためのフリーソフトウェア。各種の検索を行うことができ、どんな語が多く出現していたのかを頻度表から見ることができる。さらに、多変量解析によって、共起関係がある語のグループを見ることで、データ中に含まれるコンセプトを探索できる。本研究においては ver. 3. Alpha. 17h (2019 年 10 月 10 日公開) を使用した。
 - 3 奈良先端科学技術大学院大学情報科学研究科自然言語処理学講座松本研究室が開発した形態素解析ツールのフリーソフトウェア。多くの研究者が利用可能なように、使用者によって品詞体系、単語認定基準などを容易に変更できるよう配慮されている。