

中学校理科における恐竜の復元を単元とした評価の 検討：テキストマイニングソフトウェアを用いた 知識の構造化の分析を通して

著者	田村 響太郎, 石上 靖芳
雑誌名	静岡大学教育学部研究報告. 教科教育学篇
巻	52
ページ	56-70
発行年	2020-12
出版者	静岡大学大学院教育学領域
URL	http://doi.org/10.14945/00027846

中学校理科における恐竜の復元を単元とした評価の検討

ーテキストマイニングソフトウェアを用いた

知識の構造化の分析を通してー

Consideration of the Evaluation of the Unit of Dinosaur Restorations in junior high school science
- Analysis of Knowledge Structuring by Text Mining Software-

田村 響太郎¹

石上 靖芳²

Kyotaro Tamura

Yasuyoshi Ishigami

（令和 2 年 11 月 30 日受理）

ABSTRACT

The purpose of this study is to examine the effectiveness of analysis by text mining using KHCoder as an evaluation of a unit. A unit called "Reconstruction of Dinosaurs" was held in 9 hours in science at a junior high school. As an evaluation, KHCoder was used to analyze and compare the summary description about the method of "restoration of dinosaurs" at the beginning of the unit and after the unit ended. First, as a result of comparing the list of frequently-used words after the start of the unit, the number of types of words increased about four times and the number of extracted items also increased about three times after the end of the unit. Also, from the results of comparing the co-occurrence network diagrams after the beginning of the unit, the number of connected words increased after the unit and the number of word groups also increased. Furthermore, the viewpoint of the method of "Dinosaur restoration" increased from the result of the comparison of the coding analysis at the beginning and after the unit. From the above, it was shown that unit evaluation using text mining software is effective.

1. 問題の所在と研究の目的

平成 29 年中学校学習指導要領の改訂により、中学校において「主体的・対話的で深い学び」となる授業を通じて「何が身に付いたか」（学習評価の充実）が挙げられ、そのための授業改善を促している。中学校の理科においても、自然現象に関する問題を見いだし、問題解決のための方法の立案、実験・観察などの結果からの考察、問題の解決を通じて、その単元や授業の中で、どのようなことを学んだのかを明確にしていく必要がある。そのために、テキストマイニングとして生徒が記述した文章に含まれる言葉の数や言葉の関係を分析し、その結果から中学校の授業における生徒の学びの検証を行う必要がある。

テキストマイニングとしてソフトウェアを用いた単元の評価を行った研究が進められている。石川・小野塚・良知・奥井・得丸(2016)では、小学校において豚の飼育と業者に引き渡しまでの体験をさせ、そのはじめの段階と中間の段階と最後の段階において児童にインタビュー調査

1 袋井市立袋井中学校

2 教職大学院系列

を行い、児童の答えについての傾向を分析している。mecab, cabocha, ttm(Tiny Text Miner), R というテキストマイニングソフトウェアを用いて多かった言葉を抽出し、その関係を階層的クラスタ分析^{*1}によってグループ分けして、生命を扱うことに対する気持ちの変化を全体の傾向として抽出している。堀井・奥田(2018)は、大学3年生の保健体育科教育法という講義における模擬授業実践後の自己省察のために、IBM SPSS Text Analytics for Surveys 4.0.1 というテキストマイニングソフトウェアを用いて学生の自由記述について分析し、多かった言葉を抽出し共起関係図をもとに言葉のグループを見だし、全体の傾向としての省察のレベルを見だししている。

このようにテキストマイニングソフトウェアが単元の評価として用いられているが、もっとも多く使われているのが樋口耕一(2014)によるKHCoderというソフトウェアである。吉村・南美(2018)は、多言語を扱う授業によって生徒たちが言語に対してどのような意識を持つようになるのかを明らかにすることを目的として、中学校の英語の授業において全6時間の単元の授業実践し、授業の中で最も印象に残った授業を1つ選び、その理由を感想として自由記述させ、その記述についてKHCoderを用いて分析している。その結果、自由記述からの頻出語を150語抽出し、共起ネットワーク図^{*2}からつながりの強い語を確認し、そのつながりの強さを示す生徒の記述例を挙げている。このように、生徒が何を重要としていたのかを分析するためにKHCoderを用いている。桑原・藤田・阿部・菊池・佐藤・西岡・倉持(2018)は、中学校の家庭科の授業において、バナナケーキの調理実習を行い、そこで得られた調理における衛生面の注意点を記述させ、次にハンバーグの調理実習を行い、肉料理における衛生名の注意点を記述させ、どのように衛生面の注意事項が生かされているのかを、KHCoderを用いて階層的クラスタ分析を実施して、言葉のまとまりを抽出から生徒の学びを分析している。奥村・熊野(2016)では、高等学校の生物Ⅱの「環境」や「生態系とその保全」の単元において、7時間で調査から発表までを行い、それを8つの項目で評価し、生徒の発表直後とビデオで振り返った後の評価として記述した内容についてKHCoderを用いて分析している。全体の傾向として、「発表」という言葉に対するコンコーダンス分析^{*3}や共起ネットワーク図の比較により、「発表」という言葉を取りまく言葉の変化を見だし、ビデオで振り返ることを評価している。

本研究の対照となる中学校の理科においてもKHCoderを用いたテキストマイニングによる分析を通じて単元の評価を行う研究が進められている。小川・高林・池野・竹本・平田・松本(2019)では、授業評価のためにKHCoderを用いて、中学校2年生の「電流と磁界」の単元開発の効果の検証を行っている。生徒の感想記述をもとにいくつかの言葉のまとまりの関連性をKHCoderで分析し、コーディング分析^{*4}の結果、「科学の内容」と「日常生活」、「日常生活」と「有用感」というグループの関連性が高かったことから、実社会・実生活との関連を志向する真正の学習論に着目した中学校理科の授業が生徒の科学や理科学習に対する態度の改善に一定の効果があるとしている。

このようにKHCoderを用いてテキストマイニングの分析を通じて、集団全体の傾向を分析し、授業や単元の評価を行うことが可能となると考えられる。しかし、単元のはじめと終わりの比較をしている研究や、どのように知識が構造化したのかを分析した研究は多くは見られない。そこで本研究では、中学校の理科における「恐竜の復元」という開発した単元において、KHCoderを用いたテキストマイニングによる分析を行い、単元のはじめと単元終了後の生徒のまとめの記述の分析結果を比較することで、単元の評価について検討する。そしてこれらを通じて、

KHCoder を用いたテキストマイニングによる分析についての単元の評価の可能性を検討することを本研究の目的とする。

2. 研究方法

(1) 単元と授業実践の概要

中学校の理科において、科学的な思考力の育成を目指して思考スキル活用を位置づけた「恐竜の復元」という単元を開発し、中学校の2年生193名の6クラスに対して授業実践を行った。そのときの単元内容をまとめたものが表1である。この単元において、単元の目標を「化石から産出地の情報を収集して骨の部位や恐竜の特徴を見だし、それらの情報をもとに恐竜の種類を推定すること」として、恐竜の研究者が研究した道筋をたどるように単元を構成した。第1時では「恐竜の復元」の方法を予想し、第2時では恐竜そのものの特徴を捉えられるように恐竜の図鑑づくりを通して恐竜を分類した。第3時ではヒトや恐竜の骨格の理解や一部の恐竜の骨から部位や恐竜の種類を推定を行った。第4時では産出状況からその化石が堆積した当時の環境を推定した。第5時ではボーンベッド^{*5}と呼ばれる恐竜の化石が大量に出てくる地層のスケッチから恐竜の骨を見だして恐竜を復元した。第6時では第5時で復元した恐竜や産出状況から恐竜の骨が堆積した当時の環境を推定した。第7時ではそれまでの学習した「恐竜の復元」の結果や復元の仕方について振り返ることを通じて「恐竜の復元」の方法の見直しを行い、第8時ではポストテストAとして学習した内容をもとに「恐竜の復元」を個人で行うパフォーマンス課題を行い、第9時として単元終了後から1か月後にポストテストBとして生徒の

表1 本研究における授業実践の「恐竜の復元」の単元の内容のまとめ

時数	目 標	活動内容
第1時	単元の導入として恐竜に興味をもつ	「恐竜の復元」の方法の予想
第2時	恐竜そのものの理解と見通し	恐竜の図鑑づくりによる恐竜の分類
第3時	恐竜の骨格や部位の理解と種類の推定	恐竜の一部の骨から部位や種類の推定
第4時	現在の堆積状況の理解と当時の環境の推定	現在の堆積状況を理解し化石の産出状況をもとにした当時の環境の推定
第5時	ボーンベッドからの恐竜の復元	ボーンベッドの恐竜の部位や種類の推定
第6時	ボーンベッドの当時の環境の推定	産出状況や恐竜の情報をもとにした当時の環境の推定
第7時	これまでの学雌雄の振り返りと「恐竜の復元」の方法を見いだす	復元した恐竜や環境の推定結果についての検討を通じた「恐竜の復元」の方法の見直し
第8時	ポストテストA	個人で他のボーンベッドの「恐竜の復元」
第9時	ポストテストB	個人で恐竜の骨の組み立てを含めた「恐竜の復元」

「恐竜の復元」の方法の予想にも多かった恐竜の骨の組み立てを含めた「恐竜の復元」を個人で行うパフォーマンス課題を実施した。

(2) KHCoder を用いたテキストマイニングの分析方法

本研究では先行研究を踏まえ、テキストマイニングソフトウェアである KHCoder を用いて、単元のはじめと終了後の生徒の記述についての分析を行う。KHCoder を用いた分析の例として、「恐竜の復元とは骨を組み立てることである」という文章の「復元」に注目した場合に、単語の位置に関して、「恐竜」という言葉が「復元」という言葉の左の1つめにあり、「骨」という言葉は「復元」という言葉の右の1つめにあるというものである。このように言葉の位置関係から言葉と言葉の距離が求められ、多数の記述のデータからそのデータ全体の言葉の関係性を見いだすものとなっている。その関係性からまとまりになったものをサブグラフと呼び、言葉のグループとして、サブグラフごとに色分けしてネットワークとして描かれたものが共起ネットワークと呼ばれるものとなっている。

本研究では、生徒の単元のはじめの「恐竜の復元」についての予想をした記述のデータと単元終了後の「恐竜の復元」についてのまとめの記述のデータを出席していた全生徒の記述を Excel ファイルに入力し、それぞれのデータの分析を行った結果を比較することで、学年全体の言葉の関係性の傾向を検証する。具体的には、以下の A～G のように行う。

- A. ワークシートの自由記述を Excel ファイルの各行に1人ずつ入力し、明らかな書き方の違いや誤字・脱字といった部分を修正した後に、KHCoder に読み込ませる。
- B. 得られた自由記述を単語に分解し、全体的な傾向(頻出語)を確認する。単語の抽出に際しては、「ボーンベッド」と「カマラサウルス」といった複数の語に分解されるところを1つの語として抽出するように指定したり、骨を「組み立て」ることに関して「組み立てる」と「組み立てて」のように表現の違う語を同一の語の回数として抽出されるように指定したりして、単元のはじめの第1時と単元終了後の第9時のまとめの記述についての頻出語の一覧をつくる。
- C. 学年全体の単元のはじめの第1時と単元終了後の第9時のまとめの記述から抽出した言葉の比較を行うために、KHCoder を用いて共起ネットワーク図をつくる。第1時と第9時の共起ネットワーク図を比較する。
- D. 単元のはじめの第1時と単元終了後の第9時の生徒の記述の変化を調べるために、共起ネットワーク図の言葉の集まりをもとに、抽出した代表的な解答をしている生徒が記述した第1時と第9時の内容を比較する。
- E. 単元ははじめの第1時と終了後の第9時のまとめの記述から、単元を通じて恐竜の復元に必要な視点の関連の変化を分析するために、恐竜の復元に関する言葉のグループにまとめ、その名前としてコード名をつけ、コーディング分析を行い、第1時と第9時の結果を比較する。
- F. コード間の類似度行列^{*6}を用いて Jaccard 係数^{*7}による各コード間の関連を調べ、単元ははじめの第1時と終了後の第9時のまとめの記述におけるコードの関連の度合いを比較する。
- G. 以上の分析を通じて、単元を通じた学年全体の学びを明らかにする。

3. 結果と考察

(1) KHCoder を用いた単元のはじめと終了後の頻出語の比較の結果

生徒の記述から生徒の学習した内容における言葉のつながりについて詳しく分析するために、KHCoder を用いて生徒のワークシートの記述から「恐竜の復元」に関する言葉の関係性について分析した。単元のはじめの第1時と単元終了後の第9時のまとめの部分に記述された内容における語彙の出現回数について KHCoder を用いて分析した。

第1時については、177人のまとめの回答データをKHCoderに読み込ませたところ、総抽出語数は2786語、異なり語数は339個、文は181文、集計単位H5^{*8}は178であった。データ全体を通じてどのような語が多く出現したのかを確認するために、言葉の出現回数を6回以上のものを対象として抽出したところ、表2の通りであった。上位には「復元(132個)」、「恐竜(123個)」、「骨(103個)」、「部位(47個)」、「形(40個)」などの名詞の他、「調べる(62個)」などの動詞、「大きい(20個)」などの形容詞の単語が見られた。

第9時については、182人のまとめの回答データをKHCoderに読み込ませたところ、総抽出語数は9798語、異なり語数は864個、文は616文、集計単位H5は182であった。データ全体を通じてどのような語が多く出現したのかを確認するために、言葉の出現回数を6回以上のものを対象として抽出したところ、表3の通りであった。上位には「恐竜(419個)」、「骨(222個)」

表2 第1時のまとめの回答記述データから抽出した6回以上の頻出語24語

抽出語	出現回数	抽出語	出現回数
復元	132	環境	14
恐竜	124	種類	14
骨	103	考える	14
調べる	62	分かる	10
部位	47	発掘	9
組み立て(る)	40	分ける	9
形	40	見つける	8
大きい	20	予想	7
特徴	17	見る	7
化石	16	部分	6
場所	15	きれい	6
肉	15	分類	6

(注)強制抽出語:「組み立て」「ボーンベッド」

表3 第9時のまとめの回答記述データから抽出した6回以上の頻出語99語

抽出語	出現回数	抽出語	出現回数	抽出語	出現回数	抽出語	出現回数
恐竜	419	考える	25	食べ物	12	授業	8
骨	222	たくさん	24	部分	12	人	8
復元	154	調べる	24	分ける	12	数	8
化石	79	違う	23	情報	11	生物	8
形	79	いろいろ	22	食	11	昔	8
種類	65	川	21	人間	11	体長	8
時代	64	名前	21	発見	11	動物	8
特徴	55	地質	20	流す	11	肉	8
環境	52	組み立て	19	1つ	10	埋まる	8
方法	52	歩行	19	学ぶ	10	やり方	7
肉食	50	難しい	18	骨格	10	英語	7
場所	48	分類	18	自然	10	楽しい	7
分かる	48	似る	17	首	10	子ども	7
大きい	46	生息	17	推測	10	状態	7
草食	45	知る	17	知れる	10	水	7
仕方	37	予想	17	地層	10	生態	7
多い	32	産出	16	ジュラ紀	9	大切	7
年代	32	思う	15	見た目	9	大変	7
食べる	29	産地	14	見つける	9	細かい	6
植物	28	生きる	14	足	9	参考	6
堆積	28	長い	13	それぞれ	8	想像	6
当時	28	ボーンベッ	12	ストーリー	8	他	6
部位	28	見分ける	12	見つかる	8	体	6
見る	27	住む	12	死ぬ	8	必要	6
バラバラ	25	少し	12	視点	8		

(注)強制抽出語:「組み立て」「ボーンベッド」

「復元(154個)」、「化石(79個)」、「形(79個)」、「種類(65個)」、「時代(64個)」などの名詞の他、「分かる(48個)」、「食べる(29個)」などの動詞、「大きい(46個)」などの形容詞の単語が見られた。

第1時のまとめの記述のデータと第9時のまとめの記述のデータから得られた頻出語を比較

すると、頻出語が第1時は24語に対して第9時は99語と約4倍になり、第9時の方が大きく増加した。このことから生徒の知識が変容したと考えられる。

(2) KHCoder を用いた第1時と第9時の共起ネットワーク図の比較の結果

抽出した言葉のつながりから生徒の「恐竜の復元」について学習した内容の知識が構造化されたことを調べるために、「恐竜の復元」についての単元はじめの第1時のまとめの記述と終了後の第9時のまとめの記述をもとに KHCoder を用いて共起ネットワーク図を作成した。共起ネットワーク図は、抽出した言葉の前後の回数を集計し、その集計が多いものから順位付けして数の多さを円の大きさと表し、サブグラフとして関係する言葉のグループごとに色分けして言葉のつながりを可視化するものである。

単元はじめの第1時のまとめをもとに分析した共起ネットワーク図が図1、第9時のまとめをもとに分析した共起ネットワーク図が図2である。第1時では「恐竜」、「復元」、「骨」、「部位」、「調べる」、「特徴」、「見る」という①のサブグラフ、「形」、「大きい」、「種類」、「分ける」、「考える」、「場所」という②のサブグラフ、「組み立て」、「化石」、「肉」、「見つける」、「きれい」、「環境」、「発掘」、「分類」という③のサブグラフとして分類が可能となった。

単元終了後の第9時では、「恐竜」、「復元」、「骨」、「種類」、「形」、「方法」、「大きい」、「時代」、「化石」、「環境」、「特徴」、「肉食」、「草食」、「分かる」、「考える」、「当時」、「部位」という①のサブグラフ、「部分」、「歩行」、「見分ける」、「長い」、「首」、「数」、「生きる」、「足」、「違う」、「子ども」、「食べ物」という②のサブグラフ、「堆積」、「ボーンベッド」、「バラバラ」、「川」、「体長」、「流す」、「水」、「産地」、「調べる」、「食べる」、「分ける」、「埋まる」、「見た目」、「知る」という③のサブグラフ、「植物」、「多い」、「たくさん」、「似る」、「1つ」、「見つける」、「楽しい」、「少し」、「人間」、「授業」、「参考」という④のサブグラフ、「地質」、「年代」、「産出」、「それぞれ」、「名前」という⑤のサブグラフ、「予想」、「思う」、「地層」、「死ぬ」、「肉」、「動物」、「やり方」、「生物」、「体」、「他」、「生態」、「大変」、「必要」、「住む」、「見つかる」、「情報」、「想像」、「ストーリー」、「大切」という⑥のサブグラフとして分類が可能となった。

第1時のまとめの記述のデータをもとにした図1と第9時のまとめの記述のデータをもとにした図2を比較した結果として、第9時の方がより多くの言葉が抽出され、より多くの言葉と深い関係性を築いていると考えられ、第1時より第9時の言葉のまとまりの方がより多くの言葉を用いて、多くの説明をしている。また、単元終了後の第9時では、第1時よりも太い線になったことや言葉を結ぶ線の数が増加したことから、より深いネットワークになったと考えられる。具体的には、第9時の①のサブグラフにおいて第1時の①のサブグラフにはなかった「環境」、「肉食」、「草食」といった言葉が増え、第1時にはなかった「バラバラ」、「堆積」、「川」、「ボーンベッド」といった言葉を含む③のサブグラフや「年代」、「地質」、「産出」といった⑤のサブグラフが第9時のまとめの記述から抽出された。

生徒の記述の変化を調べるために、抽出した X, Y, Z の3名の生徒(XYZ:女子)の第1時と第9時のまとめの記述を比較したものが表4である。Xは第1時では「恐竜の復元」の方法の予

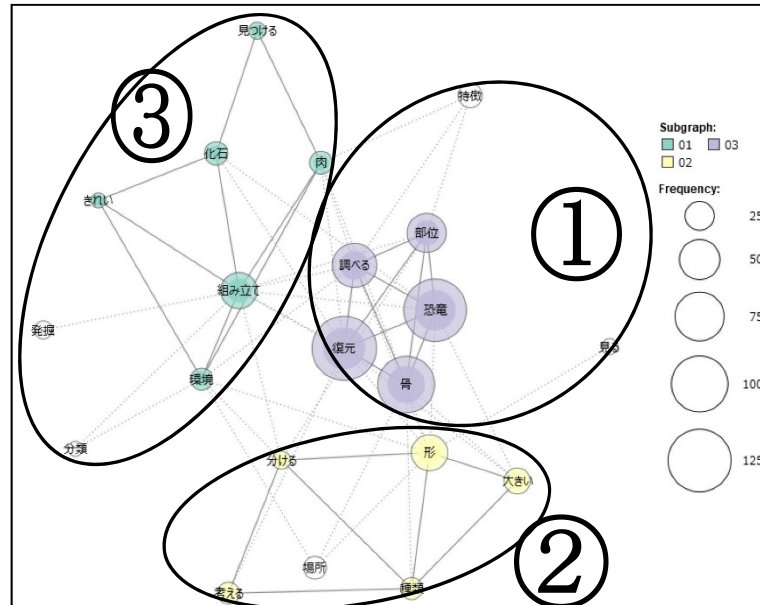


図1 第1時のまとめにおける「恐竜の復元」に関する言葉の共起ネットワーク図

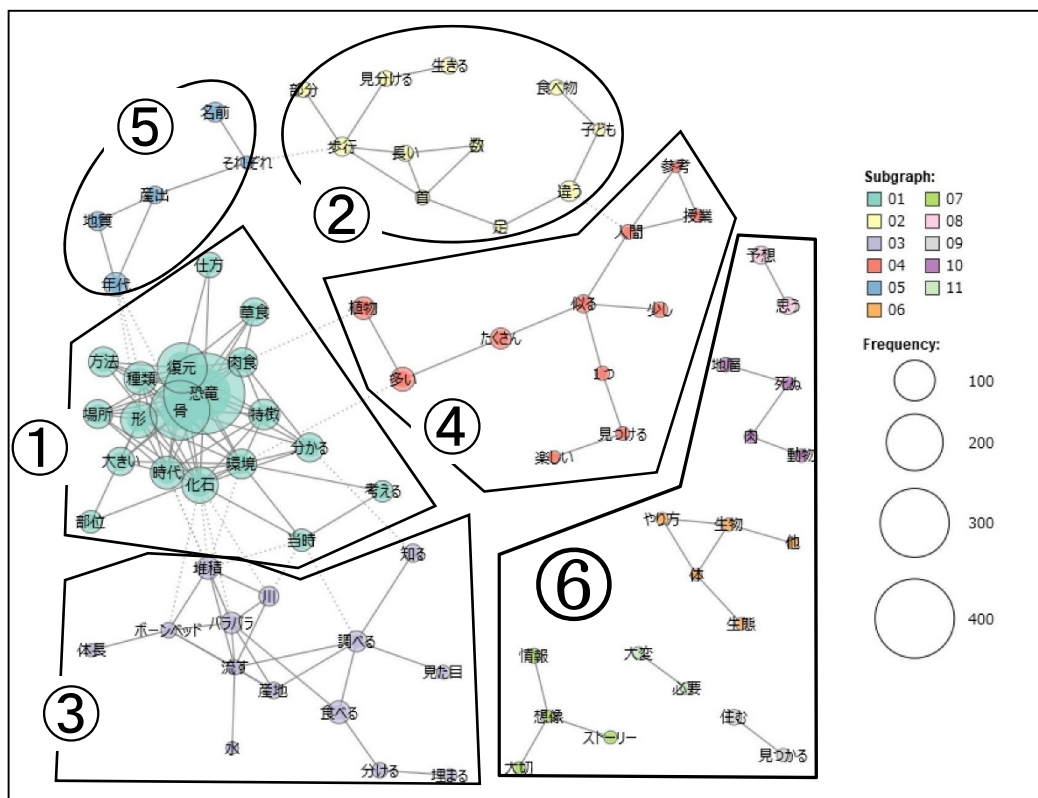


図2 第9時のまとめの記述から抽出した言葉による共起ネットワーク図

想として、「恐竜の骨の部位を調べる」と記述していた。第9時では「恐竜の復元」の方法のまとめとして「恐竜は骨ができた地質年代、食べる物、産出地、骨などで多少は測定できる。地質年代、骨の種類、産出地、その環境、生息していた場所、骨の形、DNA、食べる物、首の長さ、しっぽの長さ、足が何足歩行、骨の部位などのデータで推測できること。恐竜にもそれぞれ

表 4 抽出した生徒の第 1 時と第 9 時のまとめの記述の比較

	第 1 時のまとめの記述	第 9 時のまとめの記述
X	恐竜 ^① の骨 ^① の部位 ^① を調べる ^①	恐竜 ^① は骨 ^① ができた地質 ^⑤ 年代 ^⑤ 、食べる ^③ 物、産出 ^⑤ 地、骨 ^① などで多少は測定できる。地質 ^⑤ 年代 ^⑤ 、骨 ^① の種類 ^① 、産出 ^⑤ 地、その環境 ^① 、生息していた場所 ^① 、骨 ^① の形 ^① 、DNA、食べる ^③ 物、首 ^② の長さ ^② 、しっぽの長さ ^② 、足 ^② が何足歩行 ^② 、骨 ^① の部位 ^① などのデータで推測できること。恐竜 ^① にもそれぞれ産出 ^⑤ 地が違い、形 ^① も違う ^② し、食べる ^③ 物もそれぞれ違う ^② ということがわかった。
Y	様々な資料(骨 ^① や DNA など)から予想し復元 ^① する	かつての環境 ^① や生息地などを参考にする。そこからのストーリーも考える ^① 。地質 ^⑤ の年代 ^⑤ でいつの時代 ^⑤ の恐竜 ^① なのか考える ^① 。同じ地域、地層 ^⑥ で特徴 ^① が似ているもの、形 ^① の骨 ^① を組み合わせる。人間 ^④ や他 ^⑥ の生物 ^⑥ の骨 ^① の形 ^① を参考 ^④ にどの部位 ^① の骨 ^① なのか考える ^① 。時代 ^① 、場所 ^① からいくつか恐竜 ^① の候補を出してそこからさらにしぼっていくなどをする恐竜 ^① の復元 ^① ができる。
Z	恐竜 ^① は化石 ^③ を見つけて骨 ^① の観察をして肉 ^③ などをつけて復元 ^① する	いろいろな恐竜 ^① がいること。恐竜 ^① の種類 ^① は年代 ^⑤ 、食べる ^③ 物、産地 ^③ などいろいろな視点で分けられること。骨 ^① の形 ^① はいろいろある。骨盤の数で恐竜 ^① が何体 ^⑥ いるかわかる。骨 ^① の部位 ^① にわけて復元 ^① する。復元 ^① するときには産地 ^③ や年代 ^⑤ を調べる ^③ 。恐竜 ^① の骨 ^① は必ずしもそのままの恐竜 ^① の形 ^① で残っているとは限らない。骨 ^① がバラバラ ^③ のときは川 ^③ に運ばれた。一緒に埋まっていた化石 ^① からも情報 ^⑥ を得て復元 ^① できる。

れ産出地が違い、形も違うし、食べる物もそれぞれ違うということがわかった。」と記述していた。X は第 1 時では①のサブグラフの言葉の 4 種類の 4 つを用いて表現していたが、第 9 時では①のサブグラフの 5 種類を 13 個、②のサブグラフの 5 種類の 8 つ、③のサブグラフの 2 種類の 2 つ、⑤のサブグラフの 4 種類の 7 つが増加した。加わった言葉として、①のサブグラフでは「形」と「環境」が加わり、②のサブグラフの「首」、「長さ」、「足」や③のサブグラフの「食べる」、⑤のサブグラフの「地質」、「年代」、「産出」という言葉が加わっていた。Y も Z も同様に第 1 時よりも第 9 時の方がサブグラフの数、言葉の種類、記述数が増加していた。これらのことから「恐竜の復元」の単元を通じて生徒が学習した知識の情報量が増加し、知識の変容が確認されたと考えられる。さらに、知識としての言葉が結びつき、より深いネットワークとなって構造化されていると考えられる。

(3) 第 1 時と第 9 時の「恐竜の復元」の方法の視点に関するコーディング分析の比較の結果

単元を通じて「恐竜の復元」の方法に、どのような視点が増加したのかを分析するために、さらに KHCoder を用いてコーディング分析を行い、コーディングの結果と類似度行列によるコード間の関連の強さを調べた。

コーディングとは、ある言葉のグループに名前をつけることであり、コーディング分析とはそのグループについて分析することである。コーディング結果とは、そのグループに含まれる言葉がある文の中に 1 つでもあれば 1 つカウントされ、調べたい文全体に対して、そのグループが含まれる割合をパーセントで算出するものである。また、類似度行列とはあるグループとあるグループの言葉が 1 つ以上含まれる文が、文全体のうちどのくらいの割合となるのかを小

数で示したものである。例えば「恐竜の復元に必要な視点は骨の形から部位を推定することや産出した地層の地質年代を調べることである。」という文があり、言葉のグループ①として「恐竜」と「復元」が含まれると決めてあった時、この文に対してグループ①が1つ集計される。文全体の中で3割の文にグループ①の言葉が1つ以上入った文があると30%と表示される。さらに、類似度行列によるコード間の関連の強さの調べ方は、「恐竜」と「復元」を含むグループ①と「形」、「骨」を含むグループ②としたときに、この例文の中には「恐竜」と「骨」が含まれていることからグループ①とグループ②の言葉が両方含まれているとして1つ集計される。このように調べたい文全体に対して集計し、文全体のうち、50%が①と②のグループ含む場合、0.500と算出される。このようにして調べたい文全体に対して集計すると、全体の中のその言葉が含まれる文の割合が算出でき、コードとコードの関連として示すことができる。小川他(2019)では、共起ネットワークのサブグラフごとにコード化し、そのコード間の関連の強さを算出している。

そこで本研究では、「恐竜の復元」とその視点としての具体的な方法に関する言葉を表3の第9時のまとめの記述の抽出語から選び、必要だと考えられる視点ごとにグループにコード名をつけて、コーディング・ルールとしてまとめた。そのコーディング・ルールをまとめたものが表5である。「①恐竜の復元」として「恐竜」、「復元」、「推定」といった言葉のグループにコード名をつけ中心となる言葉のグループとした。続いて「恐竜の復元」に必要な視点のグループとして、「②骨の情報」として「骨」、「大きさ」、「形」、「部位」、「首」といった言葉のグループ、「③産出地」として「産地」、「地層」、「場所」といった言葉のグループ、「④地質年代」として「時代」、「産出」、「地層」、「年代」という言葉のグループ、「⑤恐竜の特徴」として「食べる」、「肉食」、「草食」といった言葉のグループ、「⑥当時の環境」として「堆積」、「ボーンベッド」、「川」、「流す」、「水」、「バラバラ」といった言葉のグループにコード名をつけ、「恐竜の復元」の方法の視点となるコードとした。

第1時と第9時のまとめの記述からコーディングした結果が表6である。「①恐竜の復元」のコードが第1時は89.27%、第9時は92.31%ともっとも多くなった。次に「②骨の情報」のコードが第1時は72.88%、第9時は79.67%と多かった。このことから「恐竜の復元」において恐竜の種類を推定するために、骨から得られる情報をもとにしていたと考えられる。「④地質年代」のコードは第1時に4.52%であったが、第9時では46.70%と約10倍に割合が増加した。次に「⑥当時の環境」のコードは第1時は7.34%であったが、第9時では39.01%と約5倍に割合が増加した。「③産出地」のコードは第1時には9.60%であったが、第9時では35.16%と約4倍に割合が増加した。「⑤恐竜の特徴」のコードは第1時では10.73%であったが、第9時では38.46%と約4倍に割合が増加した。これらのことから「恐竜の復元」を通じて、「恐竜の復元」の方法の視点として、「③産出地」、「④地質年代」、「⑤恐竜の特徴」、「⑥当時の環境」という視点が増えたと解釈できる。

さらに、「恐竜の復元」の方法に必要な視点が中心となる「①恐竜の復元」の言葉のコードに対する関連性の変化を分析するため、類似度行列を用いて、6つのコード間の関連の強さを分析した。そのコード間の関連の強さをある言葉が含まれている割合としてのJaccard係数によって類似度行列の分析を行った。第1時のまとめの記述から測定した結果が表7、第9時のまとめの記述から測定した結果が表8である。どの関連の強さも高くなった。その中でも、中心となる「①恐竜の復元」と方法の視点となる「②骨の情報」の関連が第1時では0.698と第

表5 復元の方法についてのコーディング・ルール

コード名	対応する語
①恐竜の復元	恐竜, 復元, 種類, 名前, 推測, 推定
②骨の情報	骨, 大きさ, 形, 部位, 首, 足, 数, 部分, 見た目, 化石, 体長
③産出地	産地, 地層, 場所, 産出
④地質年代	時代, 産出, 地層, 年代
⑤恐竜の特徴	食べる, 肉食, 草食, 歩行, 肉, 生態
⑥当時の環境	堆積, ボーンベッド, 川, 流す, 水, バラバラ, 埋まる, 分ける, 死ぬ, 植物, 多い

(注)強制抽出語:「組み立て」「ボーンベッド」

表6 復元の方法についてのコーディングの結果

コード名	第1時		第9時	
	頻度	割合	頻度	割合
①恐竜の復元	158	89.27%	168	92.31%
②骨の情報	129	72.88%	145	79.67%
③産出地	17	9.60%	64	35.16%
④地質年代	8	4.52%	85	46.70%
⑤恐竜の特徴	19	10.73%	70	38.46%
⑥当時の環境	13	7.34%	71	39.01%

表7 第1時の復元の方法についての Jaccard 係数による各コード間の関連

	①恐竜の復元	②骨の情報	③産出地	④地質年代	⑤恐竜の特徴	⑥当時の環境
①恐竜の復元	-					
②骨の情報	0.698	-				
③産出地	0.101	0.115	-			
④地質年代	0.044	0.046	0.25	-		
⑤恐竜の特徴	0.113	0.121	0.029	0.038	-	
⑥当時の環境	0.075	0.092	0.111	0	0.067	-

表8 第9時の復元の方法についての Jaccard 係数による各コード間の関連

	①恐竜の復元	②骨の情報	③産出地	④地質年代	⑤恐竜の特徴	⑥当時の環境
①恐竜の復元	-					
②骨の情報	0.758	-				
③産出地	0.365	0.384	-			
④地質年代	0.471	0.494	0.505	-		
⑤恐竜の特徴	0.384	0.405	0.354	0.36	-	
⑥当時の環境	0.39	0.44	0.298	0.357	0.382	-

1時の中で最も高く、第9時では0.758となり第9時の中でもっとも多かった。このことから「恐竜の復元」のために骨の情報を用いていたと考えられる。続いて「②骨の情報」と「④地質年代」の関連が第1時では0.046であったが第9時では0.494と約10倍に増加し、「①恐竜の復元」と「④地質年代」の関連が第1時では0.044であったが第9時では0.471と約10倍に増加し、増加した割合がもっとも高かった。このことから「恐竜の復元」において骨の情報だけではなく地質年代についても用いるようになったと考えられる。他のコード間の関連の強さもすべて割合が増加した。

これらのことから「①恐竜の復元」の方法の視点として、「②骨の情報」から復元するだけでなく、「③産出地」、「④地質年代」、「⑤恐竜の特徴」、「⑥当時の環境」という視点が加わって復元したと考えられ、「恐竜の復元」の方法についての知識が関連を強めていたことから、「恐竜の復元」に関する知識が拡大し構造化されたことが示唆される。

4. KHCoder を用いた分析の総合考察

KHCoder を用いて第1時と第9時のまとめの記述をもとにした分析の結果の比較により、第9時の方が多くの言葉を抽出できたことから、知識が増えたという変容を確認することができた。

図2の学年全体の第9時のまとめの記述をもとにした共起ネットワーク図から、6つの言葉のサブグラフを確認することができた。このサブグラフは、それぞれの語の言葉の距離が近いことからサブグラフとしてまとめられると考えられることから、それぞれのサブグラフが1つの文章を表していると解釈できる。このことから、学年全体の傾向として各サブグラフから「恐竜の復元」についての記述を再現したものが表9である。この記述はサブグラフに含まれる頻出語を関連している順に並べて言葉をつなげて再現したもので、太字はサブグラフに含まれる頻出語である。サブグラフ①に含まれる言葉をつなぎ合わせると、「恐竜の復元の方法として骨

表9 サブグラフに含まれる言葉から再現した生徒の記述の内容

サブグラフの番号	サブグラフに含まれる言葉から復元した生徒の記述の内容
①	恐竜の復元の方法として骨の形や大きさから部位や種類を見だし、肉食や草食などの特徴が分かる。また、化石がとれる場所や時代から復元する。それらから当時の環境について考える。
②	恐竜の部分を見分けると生きるための歩行や首が長い恐竜の数、足、食べ物、子どもなどが違う。
③	恐竜が産地によってボーンベッドなどでバラバラなのは、川で水が流すことが原因で堆積して埋まる。見た目や体長を調べることや食べることによって分ける。
④	植物が多い、たくさんあったと考えられ、授業で学習したことを参考に、人間に少し似ることがわかった。恐竜を1つ見つけることが楽しい。
⑤	産出した地質年代が恐竜の種類によってそれぞれ決まっていて、そこから恐竜の名前が分かる。
⑥	予想して思うことは、動物で肉がなくなり死ぬと地層になる。やり方として他の生物の体や生態を調べる。復元に必要なことは大変なことである。恐竜が住むところがわかれば化石が見つかる。それらから情報をもとに想像し、ストーリーを考えることが大切である。

の形や大きさから部位や種類を見だし、肉食や草食などの特徴が分かる。また、化石がとれる場所や時代から復元する。それらから当時の環境について考える。」と再現できる。他のサブグラフでも同様に文を再現することができた。これらの文の内容から、学年全体としての記述の傾向として、生徒の学びを確認することができると考えられる。

第1時と第9時のまとめの記述の「恐竜の復元」の方法についての視点のコーディング分析の結果の比較から、「恐竜の復元」の方法の視点のコードとして「②骨の情報」、「③産出地」、「④地質年代」、「⑤恐竜の特徴」、「⑥当時の環境」というコードの割合が増加したことを通じて、単元を終了後に「恐竜の復元」の方法の視点が増加したことを全体の傾向として確認することができた。さらに、それらのコード間の関連の強さがどのコード間においても高くなったことから、単元を通じて「恐竜の復元」の方法についての知識がネットワーク化され、それらの知識が構造化されたことを確認することができた。多くの言葉と関連を深めていたことが明らかとなり、学習した内容について知識が変容し、深くネットワーク化されたことが推察される。

このように単元のはじめと単元終了後のまとめの記述について KHCoder を用いてテキストマイニングによる分析を行い、全体の傾向として、①頻出語一覧の比較から生徒の知識が変容したこと、②共起ネットワーク図の比較から得られた知識がより深いネットワークとなったこと、③コーディング分析の結果の比較から具体的に「恐竜の復元」の方法の視点が増加したことを確認することができた。①～③より、テキストマイニングソフトウェアを用いて知識の構造の変化を捉えることができ、単元評価が可能となる。

5. 今後の展望

単元のはじめと終わりの記述をもとに KHCoder を用いて分析することで、単元を通じて学年全体の知識の変容や深いネットワーク化、系統的な知識としての知識の構造化を確認することができた。これは高橋(2019)で述べられていた「知識の段階として、知識が孤立している状態、つながりがある状態、深くネットワーク化されている状態、構造化されている状態」の中で、「構造化されている状態」であると考えられる。また、単元の目標であった「化石から産出地の情報を収集して骨の部位や恐竜の特徴を見だし、それらの情報をもとに恐竜の種類を推定すること」に対して、恐竜の復元に必要な内容や恐竜の復元の方法に関する知識の変容や構造化を確認することができた。これらのことから KHCoder を用いて学年全体の単元を通じた学習過程を確認することができ、単元の評価として扱うことが可能であると考えられる。

これからの授業実践においても生徒の学習過程について考えていくときに、このように KHCoder を用いたテキストマイニングを行うことで、知識がどのように構造化されたのかを見れるとともに、単元の評価とすることができ、その結果をもとに授業改善に向かうことが可能である。知識の構造化からの視点から単元評価が可能となり、評価の方法論として大きな可能性を秘めているといえる。

一方で、本研究における生徒の記述として、学習した内容を多く挙げて記述するという方式をとり、思いついた順で書いていた人も多かった。今回の記述としては特に問題はなかったが、言葉のまとまりを考えていくときには、記述の方法についても考えていく必要がある。

また、本研究では学年全体の学びを確認することができたが、個々の学習過程については追究できていない。個々の学習過程についてもどのように構造化されたのかを検討していく必要がある。個々の学びを分析して、個々の学びの評価についても考えていく必要がある。

最後に、これらの評価を踏まえて、恐竜の復元の単元で時代や産出地をもとに恐竜の種類を推定することができたといえるが、次回の恐竜の復元の単元を実践するときには、骨の大きさから恐竜の大きさを推定することや骨の個数から恐竜の個体数を推定することを通じて恐竜がいた時代の環境を数量的に復元することに取り組みたいと考えている。その評価は数量的に産出できたかどうかで行うことが可能であるが、他の理科の単元でも生徒が数量的に現象を扱うことが難しいことから、恐竜の復元においても難しいと考えられる。数量的に現象を扱うことについても知識の構造化の過程を探る必要があり、それをどの段階までできているのかを評価する必要がある。その評価を通じて、授業改善を行い、さらに生徒の学習が進ませられるようにしていきたい。

【謝辞】本研究を推進するにあたり、全面的な支援をいただいた袋井市教育委員会、袋井市立周南中学校の職員の皆様に、この場を借りて厚く御礼を申し上げます。

【附記】本研究は、令和元年～3年度科学研究補助金基盤研究(C)（課題番号 19K02728）研究代表者 石上靖芳）を受けての研究成果の一部です。

[註]

- *1 階層的クラスター分析とは、要素間の「距離」を計算して近い者同士を結び付ける分析方法を通じて、クラスター間に「より近い」や「比較的遠い」といった階層がある分析方法。
- *2 共起ネットワーク図とは、抽出した言葉の前後の回数を集計し、その集計が多いものから順位付けして数の多さを円の大きさで表し、サブグラフとして関係する言葉のグループごとに色分けして言葉のつながりを可視化するもの。
- *3 コンコーダンス分析とは、注目する言葉と言葉が文の中でどのような位置にあるのかを集計する分析方法。
- *4 コーディング分析とは、言葉のグループにコード名をつけ、グループに含まれる言葉が文に1つ以上あると1つカウントされ、調べたい文全体のうち何%そのグループに含まれる言葉が1つ以上含まれているのかを算出するもの。
- *5 ボーンベッドとは、大量の恐竜の骨の化石が含まれている地層。
- *6 類似度行列とはコーディングしたグループのうち、2つのグループに含まれる言葉が同一の文にあるかどうかを文全体に対する割合で産出するもの。
- *7 Jaccard 係数とは、ある言葉 a とある言葉 b があり、調べたい文に a と b の両方を含む文の数を a または b が含まれる文の数で割ることで算出する割合。式は $(a \cap b) / (a \cup b)$ で表す。
- *8 集計単位 H5 とは、これまでの集計では、A と B, B と C という文があれば A と B, B と C という関係のみ計測したが、そのように並んでいる場合、A と C という関係も見いだせるとし、A と C という関係もカウントする集計方法。

参考文献

青木成一郎(2019)「天文学講演におけるアンケートの自由記述欄に対する 計量テキスト分析」, 情報教育シンポジウム, 2019年8月, pp277-282

- 越中康治, 高田淑子, 木下英俊, 安藤明伸, 高橋潔, 田幡憲一, 岡正明, 石澤公明(2015)「テキストマイニングによる授業評価アンケートの分析ー共起ネットワークによる自由記述の可視化の試みー」, 宮城教育大学情報処理センター研究紀要, 第 22 号, pp67-74
- 福谷泰斗, 皆川直凡(2019)「テキストマイニングを用いた授業の理解状態把握に関する検討ー中学校社会科の授業実践をもとにした質的分析の試みー」, 鳴門教育大学情報教育ジャーナル, No. 16, pp7-16
- 樋口耕一(2014)『社会調査のための計量テキスト分析ー内容分析の継承と発展を目指してー』, ナカニシヤ出版
- 堀井大輔, 奥田援史(2018)「体育指導における模擬授業の効果ーテキストマイニングによる自己省察的分析ー」, 滋賀大学教育学部附属教育実践総合センター紀要, 第 26 巻, pp. 15-22
- 石川みどり, 小野塚知美, 良知祥吾, 奥井一機, 得丸定子(2016)「小学校での動物飼育授業における児童の心情変化:豚の飼育から出荷まで」, 上越教育大学研究紀要 35 巻, pp239-255
- 荻野正美(2015)「計量テキスト分析による若者の SNS 利用行動 およびリスク認知の検討」, プール学院大学研究紀要, 第 56 号, pp83-96
- 桑原智美, 藤田智子, 阿部睦子, 菊地英明, 佐藤安沙子, 西岡里奈, 倉持清美(2018)「調理実習における衛生に関する授業の効果」日本家庭科教育学会大会・例会・セミナー研究発表要旨集, pp102-103
- 文部科学省(2017)『平成 29 年中学校学習指導要領』, 東山書房, 26 ページ
- 文部科学省(2017)『平成 29 年中学校学習指導要領解説理科編』, 学校図書, 183 ページ
- 森健一郎, 高橋弾, 栢野彰秀(2017)「中学校理科「自然と人間」単元における特定外来生物の教材化とその評価ーテキストマイニングのコーディングを活用してー」, 日本科学教育学会研究会研究報告 Vol. 32 No. 1, pp25-28
- 小川博士, 高林厚志, 池野弘昭, 竹本石樹, 平田豊誠, 松本伸示(2019)「実社会・実生活との艦列を志向する真正の学習論に着目した中学校理科の単元開発とその実践ー生徒の科学や理科学習に対する態度に与える効果ー」, 理科教育学研究, vol. 59, No. 3, pp345-356
- 奥村仁一, 熊野善介(2016)「高等学校生物授業を事例とした生徒の言語活動に対する意識変化に関する実践的研究ービデオを活用した振り返り学習による効果の検証を通してー」, 理科教育学研究 56 巻 4 号, pp 421-434
- 大隅昇, 保田明夫(2004)「テキスト型データのマイニングー定性調査におけるテキスト・マイニングをどう考えるかー」, 理論と方法 vol. 19 No. 2, pp135-159
- 齋藤朗宏(2012)「日本におけるテキストマイニングの応用」, The Society for Economic Studies The University of Kitakyushu Working Paper Series No. 2011-12
- 齋藤めぐみ(2019)「保育者を志す学生を対象とした生活習慣改善の試みにおける「気づき」の探索的分析ー共起ネットワーク分析における一考察ー」千葉敬愛短期大学紀要, 41 号, pp1-12
- 高橋純編著(2019)『教育方法とカリキュラム・マネジメント』, 学文社, pp21-22
- 田村響太郎・石上靖芳(2019)「総合的な学習の時間における思考スキル活用に関する効果の検討ー課題解決学習における思考のはたらきが知識に与える影響の分析を通してー」, 静岡大学教育学部研究報告(人文・社会・自然科学篇), 第 70 号, pp103-116
- 田村響太郎・石上靖芳(2019)「思考スキルを用いた科学的思考力の育成に関する研究ー「恐竜の復元」を対象とした単元開発とその評価を通してー」, 静岡大学教育学部研究報告(教科

教育学篇), 第 51 号, pp111~126

吉村雅仁, 南美佐江(2018)「多言語を扱う英語授業の試み: 日本の中高等教育における言語意識教育と期待される効果」, 奈良教育大学教職大学院研究紀要「学校教育実践研究」, pp11-20