

子どもの哲学対話データサイエンス

— 中学1年生言葉データの定期的な形態素解析による事例研究 —

* 半沢 裕太, ** 田端 健人, *** 野坂 実央

Data Science of Children's Philosophical Dialogue

A Case Study by Regular Morphological Analysis of 1st. Grade Junior High School Language Data

HANZAWA Yuta, TABATA Taketo and NOZAKA Mio

要 旨

本研究の目的は、子どもの言語活動の成長を、数量的エビデンスと教師の実践感覚とを架橋しながら可視化することである。そこで本稿では、4月から8月にかけて毎月1回、計5回、中学1年生28名の教室で実施された「子どもの哲学p4c」の対話と振り返りシートの言語データを収集し、それを形態素解析システム「jReadability」にかけ、リーダビリティ・スコア、語彙レベル構成、延べ形態素数、異なり形態素数の4指標を計測し、解析結果を、対話のファシリテーターであった教師の実践感覚と照らし合わせながら、解釈した。成果として、「7月の対話が最も深まった」という教師の実践感覚が、対話とシートのリーダビリティ・スコアの差として実証された。また「単元にかかる時間が多いほど、生徒の理解や思考が深まる」という実践感覚が、対話の延べと異なりの形態素数の有意な増加として、さらに4月から8月にかけての生徒の言語能力の伸び、ならびに対話の場のセーフティの高まりが、形態素数の増加として実証された。加えて、本共同作業から、数値の意味が、対話非当事者の非実践者にはわからないが、対話当事者の実践者にはよくわかるという現象が新たに顕在化した。これによって、データサイエンスには、統計・コンピューター・人間の3構成要素が不可欠であることが具体化され、コンピューターにより統計的に処理された数値には、教師に蓄積された膨大な実践感覚から、深い意味を掘り起こす働き「センスマイニング」があることも明らかになった。

Key words : 定量的研究、定性的研究、数量的エビデンス、解析システム、実践感覚、子どもの哲学 p4c

1. リサーチ・クエスチョンと研究方法

子どもの言語能力の成長を、私たち教師や教育研究者は、どのように、またどこまで把握できるだろうか。私たちは、子どもの言語能力の成長を肌身で感知している。だが、経験的な感覚にとどまらず、その成長は、どこまで数値によって定量的に把握できるだろうか。

あるいは、その数値は、実践感覚とどれほど整合的だろうか。母語である日本語の言語能力については、漢字テストや国語テストによって、英語能力は英語テストによって、測定される。しかし、こうしたテスト以外に、言語能力を測定する方法は存在しないだろうか。これが本研究の問題意識である。

この問題意識を本研究ではさらに限定し、次のよう

* 宮城教育大学教職大学院 / 公立中学校
 ** 宮城教育大学教職大学院
 *** 宮城教育大学教務補佐

なりサーチ・クエスチョンを設定した。「子どもの哲学 (philosophy for children)」、通称「p4c」の対話を通して、子どもの言語能力はどれほど成長するのか。その成長を私たちは定量的かつ定性的に、どのように、またどこまで把握できるだろうか。私たちは、子どもの哲学 p4c の対話を通して、子どもや私たち自身が変化成長することを、経験的に直感している。この成長には、人間的成長や思考力の成長に加え、言語能力の成長も含まれる。この成長を、経験的実践感覚を超えて、数値によってどこまで可視化できるだろうか。

現在のところ、子どもの哲学 p4c の実践研究では、計量的研究は非常に少ない¹。本研究は p4c の研究領域に、「データサイエンス (data science)」という学術により、計量的研究の道を開拓する一つの試みである。

本研究は、上記リサーチ・クエスチョンを起点とし、次の方法によってこの問いの解明を試みる。第1に、子どもの哲学対話の言葉データを収集する。子どもの哲学対話の実践校である A 市立 B 中学校 1 年 C 組 28 名を調査対象とし、毎月 1 度行なわれた対話を、計 5 回、IC レコーダーで記録し、「話し言葉データ (対話データ)」を収集し、あわせて振り返りシートの「書き言葉データ (シートデータ)」を収集する。第2に、毎回の対話と振り返りシートを、文字起こしし電子化する。こうして得た「対話データ」と「シートデータ」が、本研究のローデータである。第3に、これを形態素解析し、延べ形態素数、異なり形態素数、リーダビリティ・スコア、語彙レベル構成、一文の平均文長などの値を測定する。形態素解析には、主に Web システムとして公開されている日本語難易度判定システム「jReadability」を活用する。第4に、この測定結果を、対話のファシリテーターであった教師 (第一執筆) の実践感覚と照らし合わせながら分析・解釈する。

2. 「データサイエンス」について

「データサイエンス」という日本語は、2008年刊の『広辞苑』第6版には記載がなく、2018年刊の『広辞苑』第7版になって登場する。それゆえ「データサイエンス」という言葉は、2008年から2017年の間に人口に膾炙したとみなしてよい。第7版によると、「統計学・計算機科学・情報科学などを応用し、各種のデータが持つ意味・法則性を探り出し、また、その分析手法を研究する学問分野」とある。しかし、この用語解説は不十分である。データサイエンスにとって決定的な構成要素が抜けているからである。

Blei & Smyth (2017) によると、「データサイエンスに決定的な構成要素」は3点あり、「統計的、コンピューターの、人間的」要素である。第3の人間の要素は、「データサイエンスが完全には自動化できないという現実」に対応する。「コンピューターを利用した現代の統計的ツールを、現代科学の問いに適用するには、人間による意味に満ちた判断と、訓練を積んだ深い知識とが必要である」(Blei & Smyth, 2017, p.8690)。子どもの言葉データサイエンスでも、この指摘通り、形態素解析の数値やその変化の「意味」は、人間的な知識、とりわけ現実の子どもを知る実践者の実践感覚による解釈を通さなければ、見出すことができない。それゆえ、データサイエンスは、理数系学問と人文系学問とを、また理論知と実践知とを融合する新たな学術領域である。加えて、現実とそのデータの「意味」を可視化する点で、現象学の領域でもある (cf. 田端2021b, §10)。尾身宜彦は、これを experience (経験; エクスペリエンス) と evidence (エビデンス) とを融合した「Ex-evidence (エクス・エビデンス; X-evidence)」と命名している (尾身, 2022, p. 未定)。

1 「子どもの哲学」を検索語として CiNii「本文あり」にて検索すると、32件がヒットする (2022.01.28アクセス)。検索対象外のノイズを除く31件中、数量的分析が含まれるものは、ワークシートのルーブリック・スコアをグラフ化した佐藤・中川 (2018) の1件しかない。同じく CiNii「本文あり」にて検索条件「p4c」で検索すると、先の検索結果と重複する論文も含め34件がヒットする (2022.01.28アクセス)。ノイズを除く33件のうち数量的分析が含まれるのは、「自己表現の方略のなさ」の1要因被験者内分散分析を行った植木・榊原 (2014)、振り返りシート・スコアをレーダーチャートにした梶形 (2018)、p4c 実践者へのインタビューから共起ネットワークを作成した久保 (2021) の3件である。海外の研究動向としては、子どもの哲学の実践と研究について、注目すべき論考を集めた最新のハンドブック (Gregory, M.R. et al., 2017) を見てみよう。同書には20ヶ国の著者による29本の論文が所収されているが、引用ではなく、著者たち自身が計量的分析を行なった論文は、1本 (Daniel, M.-F. et al., 2017) のみである。本ハンドブック外になるが、筆者が知る数少ない本格的な試みとしては、哲学対話が言語能力育成 (第二言語習得) に効果的であることを数量的に示した Gholamhossein & Mehdi (2010) の研究がある。

3. 「形態素解析システム」と「リーダビリティ・スコア」について

「形態素 (morpheme)」とは、言語学の用語で、意味をもつ言語要素の最小単位、それ以上分解すると意味をなさなくなるまで、言葉を分割した最小単位である。「形態素解析システム」とは、言葉の電子データをコンピューターで形態素へと自動的に分解し、その総数や異なり形態素数、平均文長、リーダビリティ・スコアなどを計算するプログラムである。

日本語の形態素解析システムには、幾つかの種類がある。テキストマイニングでよく知られる KH Coder²や、Web システム「User Local AI テキストマイニング」³も、形態素解析システムである。これらとややタイプを異にするシステムに、「リーダビリティ (Readability)」を数値化するものがある。「リーダビリティ」とは、文章の読みやすさ、読みにくさ、つまり文章の難易度を意味する。「リーダビリティ・スコア」とは、リーダビリティに与えられた指標のことである。

リーダビリティ研究が盛んになったのは、1940年代米国で、英語を母語としない移民児童の急増により、英語識字率を高める教育が重視されるようになったからとされる (cf., 柴崎, 2014, p.49)。リーダビリティ研究は、このように非母語話者の学習目的で出発したが、その後、母語話者か非母語話者かを問わず、研究と実用が進展する。米国では、医療や安全に関する情報は小学5年生レベルの英語で記載するよう求める法律が多くなったり (cf., 柴崎, 2014, p.49)、リーダビリティ・スコアが書籍を選択する指標として書店で利用されたりするようになる (cf., 柴崎・玉岡, 2010, p.449)。ちなみに Amazon 米国サイトには、書籍のリーダビリティ指標が複数掲載されている。一例として検索してみると、*Harry Potter and the Sorcerer's Stone* (『ハリーポッターと賢者の石』) は、「Reading age」が9～12

歳、「Grade level」で4～7学年、「Lexile measure (レクサイル指数)」で880と表記されている。

リーダビリティの研究と実用は、英語をはじめ、中国語、韓国語、デンマーク語、フランス語、スペイン語、ヘブライ語、ベトナム語など、様々な言語でなされており、難易度を判定する式は、現在までに200以上ある (cf., 柴崎, 2014, p.50)。「英語では、Gunning Fog index、Coleman Liau index、Flesch⁴ Kincaid Grade level、ARI (Automated Readability Index)、SMOG (Simple Measure of Gobbledygook)、Flesch Reading Ease が代表的な判定式」(cf., 柴崎, 2014, p.50) である。Fog 指数と Flesch 指数は、統計ソフト R の textFeatures 関数でも算出できる。

一方、日本語のリーダビリティ研究の歴史は浅く、開発され実用される公式もごく僅かである。

日本語の「リーダビリティ・スコア」を計算する実用的な Web システムは、筆者が知る限り現在のところ以下の3つである。

① 2007年公開「ことば不思議箱」⁵ (名古屋大学、佐藤・松崎研究室) の「日本語テキストの難易度を測る：第1の箱」⁶ (以下「不思議箱」)

② 2017年公開「リーダビリティ・リサーチ・ラボ」⁷ (長岡技術科学大学、柴崎秀子・原信一郎) の「日本語リーダビリティ測定 Ver.0.2」⁸ (以下「ラボ0.2」)

③ 2017年公開「jReadability PORTAL」⁹ (プロジェクト代表、早稲田大学、李在鎬) の「日本語文章難易度判別システム jReadability」¹⁰

開発目的やリーダビリティ公式などは、システムごとに異なる。

本研究では、③の「jReadability」を利用する。延べと異なりの形態素数、リーダビリティ・スコア、語彙レベル構成が一覧しやすいからである。補論では、①「不思議箱」と②「ラボ0.2」も利用し、スコアを比較する。

2 <https://kxcoder.net/>

3 <https://textmining.userlocal.jp/>

4 原著では Flesch とあるが、Flesch の誤植とみなし、引用者が改めた。

5 <http://kotoba.nuee.nagoya-u.ac.jp/>

6 <http://kotoba.nuee.nagoya-u.ac.jp/sc/obi3/>

7 <http://readability.nagaokaut.ac.jp/>

8 <http://readability.nagaokaut.ac.jp/readability/>

9 <https://jreadability.net/>

10 <https://jreadability.net/sys/ja>

jReadability の開発目的や特徴については、当該 Web ページの解説や李 (2016) や田端 (2021a) を参照されたい。不思議箱については、Sato et al. (2008) を、ラボ0.2については、柴崎・原 (2010) や柴崎・玉岡 (2010) を参照されたい。

ここで強調しておきたいのは、それぞれの開発目的とは異なる意図で、形態解析システムを、日本語母語話者である児童生徒の言語能力測定の代替として利用する挑戦的意義である。こうした意図で、児童生徒の言葉ローデータをシステムに挿入し解析・解釈する調査研究は、現在のところ稀有である (cf., 田端2021a, p.19)。それゆえ、上記①②③のシステムが、日本の児童生徒の言語能力をどれほど測定できるかは未知数である。児童生徒の実数値を挿入し、その計算結果を実践知と照らし合わせる作業を蓄積することで、これらシステムの解析能力についても明らかになるであろう。本稿はその一つの試行である。

4. 事例のデータサイエンス：解析結果と解釈

4-1. 事例となる言葉データの解説

まずは、ローデータについて解説しておきたい。

事例は、A 市立 B 中学校1年生 C 組の28名である。

4 月から8 月まで毎月1 回の p4c 対話の実施情報は、表1 の通りである。

問いは、p4c の慣例にならない、生徒たちが立て、生徒たちの投票で選出している。

整理番号は、実施日の月と日とを組み合わせている。整理番号417は、4 月17日データを意味する。

整理番号417は、いわゆる学級開きとして学級活動(以下「学活」)の時間に、それ以外は道徳科の時間に対話を実施した。道徳科の対話は、教科書『新訂 新しい道徳 1』(東京書籍)の該当単元のテキストを読んだ後に行なっている。テキストとの対応と、取扱い時数は、表2 の通りである。

1 時間扱いの整理番号417と517と712では、朝活動の時間に、テキストを黙読し、問いを立てた(15分)。その後学活の時間に、問いの投票、対話、振り返りシーターの記入を行なった(50分)。

3 時間扱いの612は、1 時間目に「いじめに当たるのはどれだろう」を扱い、グループに分かれ、Google Jamboard を利用し、その理由の意見交換を行なった。2 時間目に「傍観者でいいのか」を範読し、その後ノートのワークシートに各自の考えを記入した。3 時間目

表1：1 年 A 組の p4c 対話・実施情報

整理番号	実施日	問い	教科など
417	4 月 17 日	好き勝手にいい世界だったら 1 年 C 組はどうなる？	学活
517	5 月 17 日	短所は欠点か？	道徳
612	6 月 12 日	原因がいじめられる側にあったら、いじめてもよいか？	道徳
712	7 月 12 日	大変でも好きな仕事をする？楽で嫌いな仕事をする？	道徳
803	8 月 3 日	もし A 市が大都会だったら？	道徳

表2：p4c 対話と教科書の対応と時数

整理番号	対応する単元・テキスト	ページ数	時数
417	なし	---	1
517	3 自分との付き合い方を考えよう 「自分の性格が大嫌い！」(文：中村うさぎ)	16 - 19	1
612	4 いじめのない世界へ 「いじめに当たるのはどれだろう」(絵：テラカドヒトシ) 「傍観者でいいのか」(編：東京都教育委員会) 「ふたつの心」(文・絵：本山理咲)	21 - 32	3
712	7 働くってどんなこと？ 「新しいプライド」(文：遠藤功)	44 - 47	1
803	9 ふるさとのために 「ぼくのふるさと」(文：生徒作文)	52 - 54	2

に「ふたつの心」を黙読し、その後問いを立て、問いを投票、対話、最後に振り返りシートを記入した。

2時間扱いの803は、1時間目にテキストを範読し、その後グループに分かれ、ワークショップ形式で、A市の良さと課題を話し合った。2時間目に問いを立て投票、対話、振り返りシートを記入した。

振り返りシートの記入時間は、いずれも5分弱である。ただ、正確には同一時間ではない。

各回の実質的な対話時間は、整理番号417が約16分、517が約12分、612が約25分、712が約27分、803が約26分である。

ローデータは、各回のp4c対話のICレコーダー音源をもとに文字起こしした電子データ、各回の手書き振り返りシートの文字起こし電子データである。文字起こしでは、各回の対話の発言者を特定し、個人名で電子データを作成したが、個人情報保護の観点から、その後個人名を削除し、IDで管理した。振り返りシートには個人名が記載されており、個人ごとに整理した。これも個人情報保護の観点から、IDで管理した。電子化した文字記録の掲載は、紙幅の都合上省略する。

4-2. 各対話のポイント：KH Coderの共起ネットワークを利用して

次に、各回の対話のポイントを読者に示すため、対話の頻出語をつないだ文章を、表3に示す。この文章の作成にはKH Coder (ver.3.Beta.03a)の共起ネットワーク図を利用し、発言頻度の多かったキーワードと関連語を文章化した。KH Coderの共起ネットワーク図については、紙幅の都合から掲載を省略する。KH Coderの共起ネットワーク図を利用することで、主観的印象のバイアスを減らし、客観性を高めることができる。ただ、KH Coderの共起ネットワークでは、「抽出語の選択」設定で「否定助動詞」にチェックを入れても、「ない」という否定の文章がうまく表現されないため、文字起こしを担当した第3執筆者の記憶とスクリプトを照合して補った。

対話の言葉数(形態素数)が多いか少ないかは、問いの難易度の影響を受けるかもしれない。しかし、問いの難易度は原理的に主観に依存する。子どもにとって簡単な問いでも、大人には難しい場合やその逆もある。Aさんにとって簡単な問いも、Bさんには難しいこともある。それゆえ、問いの難易度を客観的に評価

し、各回の問いの難易度を比較することは、原理的に不可能である。個人的印象として、整理番号517の「短所は欠点か？」がいわゆる「ひねり」があり、応答が難しいとも感じられるが、それはあくまで個人の印象であり、生徒たちにとって本当にそうであるかは不明である。それゆえ、本稿では、対話の起点となる問いの難易度に特に違いがあったと断定はしないでおく。

また対話の形態素数の多寡は、問いが生徒たちの興味関心をどれだけ引くかにも影響されるであろう。各回の問いが、生徒の興味関心をどれほど引いたか、本研究では測定していない。ただ、p4cでは、あるテーマに対し、みんなで話し合ってみようことについて自分たちで問いを立て、投票で選んでいる。そのため、出された問いも、選ばれた問いも、生徒たちにとって比較的興味関心を引くものになっていると推察できる。少なくとも、生徒たちの興味関心を引かない問いが選出されるとは考え難い。そこで本稿では、選ばれた問いは、生徒たちにとって、ある程度興味関心を引くものであり、各回の程度に特段の差があったとはみなさない、という前提で分析を進めたい。

4-3. リーダビリティ・スコアの変化とその解釈

各回の対話とシートのリーダビリティ・スコア(RS)は、表4の通りである。スコアの算出では、各回のそれぞれのローデータをひとまとめにし、jReadabilityシステムに挿入した。

jReadabilityのリーダビリティ公式は、次の通りである(李, 2016, p.1)。Xがリーダビリティ・スコアである。

$$X = (\text{平均文長} \times -0.145) + (\text{漢語率} \times -0.126) + (\text{和語率} \times -0.042) + (\text{動詞率} \times -0.145) + (\text{助詞率} \times -0.044) + 11.724$$

このスコア(RS)は、値が小さいほど読むのが難しい(難易度が高い)ことを意味する。

表4のスコアを折れ線グラフで表したのが、図1である。

jReadabilityのRS(リーダビリティ・スコア)は、文章の難易度を絶対的に判定しているわけではない。システムの目的に最も当てはまりの良い計算式を開発し組み込んでいるため、目的が異なれば、別の計算式になり、RSも異なる。本稿では、jReadabilityのRSを、生徒たちの話し言葉と書き言葉の難易度の一つの目安

表3：各回のp4c対話のポイント

整理番号	対話略称	対話の問いとポイント
417	勝手対話	<p>問い：好き勝手にしていい世界だったら1年C組はどうなる？</p> <p>頻出語の文章化による対話のポイント： 「好き勝手にしていい世界だったら、学校に行かないと思う。」 「好き勝手にしていい世界だったら、ほかの人の意見を聞かなくなると思う。」 「好き勝手にしていい世界だったら、遊ぶと思う。」 「喧嘩してもすぐに仲直りできることが、仲が良いということ。」 「気軽に話せて助け合えるクラスが良いクラス。」</p>
517	短所対話	<p>問い：短所は欠点か？</p> <p>頻出語の文章化による対話のポイント： 「短所イコール欠点の関係だと思う。」 「短所イコール欠点ではないと思う。」 「短所は見方を変えれば長所になると思う。」 「自分の短所を長所として生かせるように少しずつ努力する。」 「短所を欠点と思うかどうかはひとそれぞれだと思う。」 「短所は長所になるかもしれないので、ポジティブに考える。」</p>
612	いじめ対話	<p>問い：原因がいじめられる側にあつたら、いじめてもよいか？</p> <p>頻出語の文章化による対話のポイント： 「どんな理由があつてもいじめてはダメ。」 「いじめられる側に原因があつても、いじめてはいけない。」 「嫌な人がいるのは悪くないけど、それを理由にいじめるのはダメ。」 「いじめや嫌がらせをしていた人が悪いから、いじめてもいい。」 「叩かれたりすると頭に来るから仕返ししてもいい。」 「喧嘩は両方だけど、いじめは片方が嫌な思いをすること。」</p>
712	仕事対話	<p>問い：大変でも好きな仕事をする？楽で嫌いな仕事をする？</p> <p>頻出語の文章化による対話のポイント： 「大変でも好きな仕事をすると思う。」 「楽で嫌いな仕事をすると思う。」 「大変な仕事は、難しいことに挑戦できて、達成感がある。」 「楽は楽しいという意味だから、嫌いな仕事はないと思う。」 「大変は大きく変わるという意味だから、大変な仕事をすると思う。」 「最初はできなくても、難しいことを達成すると自信が持てる。」</p>
803	町対話	<p>問い：もしA市が大都会だったら？</p> <p>頻出語の文章化による対話のポイント： 「A市が大都会になったら住まないと思う。」 「A市が大都会になつても住むと思う。」 「A市が東京みたいに大都会になったら、人口が増えてコロナの人も増えると思う。」 「A市が大都会になったら、今より土地の値段が高くなって住めなくなると思う。」 「A市が今のままだったら仕事がないから、大人になったら都会に行く。」 「田舎だけど自然があつて、ショッピングモールもあつたら住むと思う。」 「A市の良い所を知ってもらえば、A市に住む人が増えると思う。」</p>

とみなす。グラフからは次のことが読み取れる。

対話でもシートでも、整理番号612(いじめ対話)のRSが最小であり、文章の難易度が最も高い。

全ての回で、対話よりもシートの難易度が高い。これは、話し言葉よりも書き言葉の方が、難易度の高い

文章になることを示唆している。

712は、対話RSとシートRSとのギャップが大きい。授業の中で、生徒は、かなり簡単な文章で対話したが、振り返りシートは、難易度が高い言葉で書いている。

803は、jReadabilityの評価では、対話RSとシート

表 4 : 各回の対話とシートの RS

	対話 RS	シート RS
417	3.85	3.20
517	3.39	2.83
612	3.03	2.69
712	3.99	2.90
803	3.45	3.36

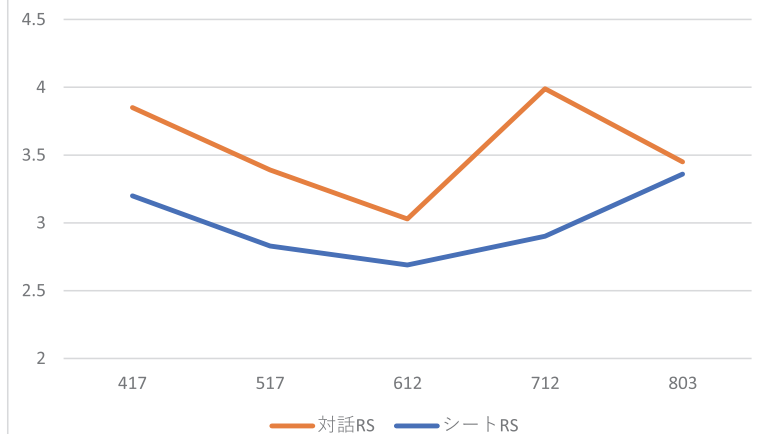


図 1 : 対話とシートの RS のグラフ化

RS とが近似している。生徒は、話し言葉と同程度の簡単な文章で、シートを書いている。

興味深いことに、712のスコアのギャップが何を意味するかの読み取り（解釈）で、非実践者と実践者として、際立った違いが生じた。非実践者の第2・第3執筆者は、対話の現場を参観せず、動画記録を視聴し、文字起こしし、電子スクリプトを形態素解析したが、このギャップが何を意味するか、見当がつかなかった。ところが、対話の実践者である第1執筆者には、このギャップの意味は即座に明らかであった。このスコアのギャップを第1執筆者に示したところ、「対話は簡単な言葉でしゃべっていたが、その間、生徒たちは深く考えており、対話の間に考えが深まったから、シー

トの言葉の難易度が上がった」と即答した。おそらく実践者は、現場でこの感覚を得ており、リーダビリティ・スコアはこの感覚を代弁していた可能性がある。

第1執筆者は、解析結果を知る前に、712の授業で学級の成長を最も感じたと言っていた。RSのギャップは、この実感を裏付ける一つのエビデンスの可能性はある。

逆に、第1執筆者にこうした実感があったことからすれば、jReadabilityの文章難易度指標RSは、実践者の実感を反映する数値になっているとも解釈できる。

4-3. 語彙レベル構成の変化とその解釈

語彙レベル構成について、対話を図2、シートを図

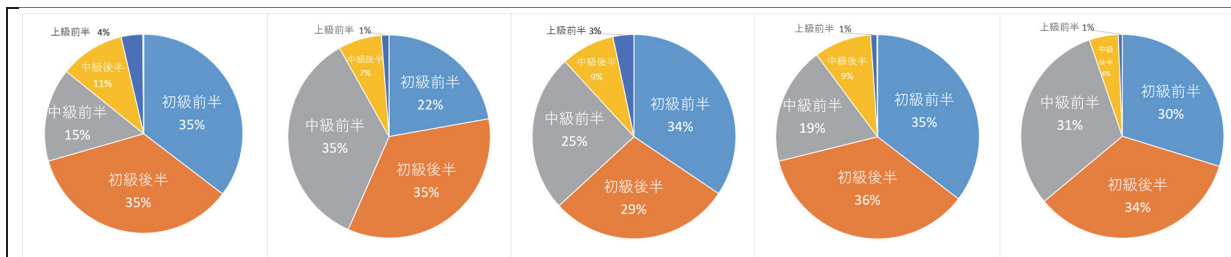


図 2 : 対話の語彙レベル構成

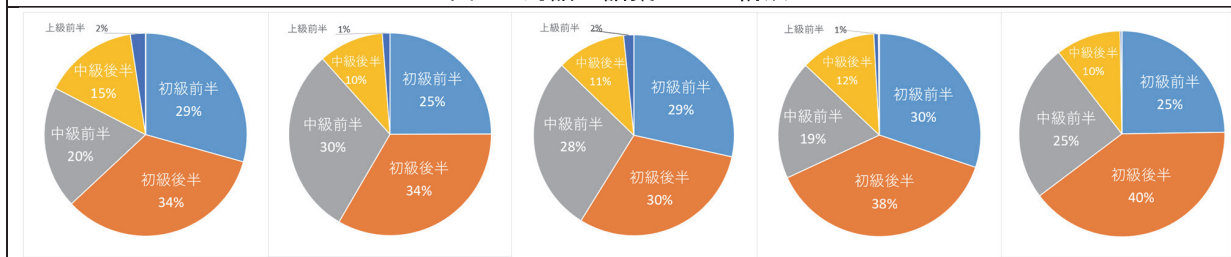


図 3 : シートの語彙レベル構成

3の一覧で示す。

jReadabilityの語彙レベル構成の円グラフは、パーセンテージの多い順に色分けされるため、システムの解析結果からExcelで作図し直した。右上から、初級前半(青)、初級後半(オレンジ)、中級前半(グレー)、中級後半(黄色)、上級前半(紺色)を表記している。円グラフは、左から、417、517、612、712、803の結果である。

初級前半・後半(青・オレンジ)の割合に注目し、対話とシートを比較してみよう。各回の初級の前半と後半を合計した割合を一覧にし(表5)、それを折れ線グラフで表現した(図4)。

全体に占める初級語彙の割合は、60～70%で推移している。ただ、割合のこの推移からは、特に何かを読み取ることは難しい。逆に、語彙レベル構成の指標は、少なくとも今回の対話の変化に対する識別力に乏

しい。

4-4. 形態素解析の基礎情報と解釈

次に、形態素解析の基礎情報を示す。シートデータでは、各回の変化や特徴が特段把握できなかったため、特徴的であった対話データにフォーカスする。

対話データは、2通りの仕方でjReadabilityに投入した。1つは、各回の授業での全生徒の発話をまるごとシステムに挿入する方法であり、もう1つは、発話ごとにシステムに投入する方法である。「発話ごと」とは、1人の生徒が話し始めてから、話し終わってコミュニティ・ボールをパスするまでの発言を、1回の発話としてカウントした。1回の発話は、1文で構成されることもあれば、複数の文で構成されることもある。文は、文頭から句点「。」までの単位である。「パ

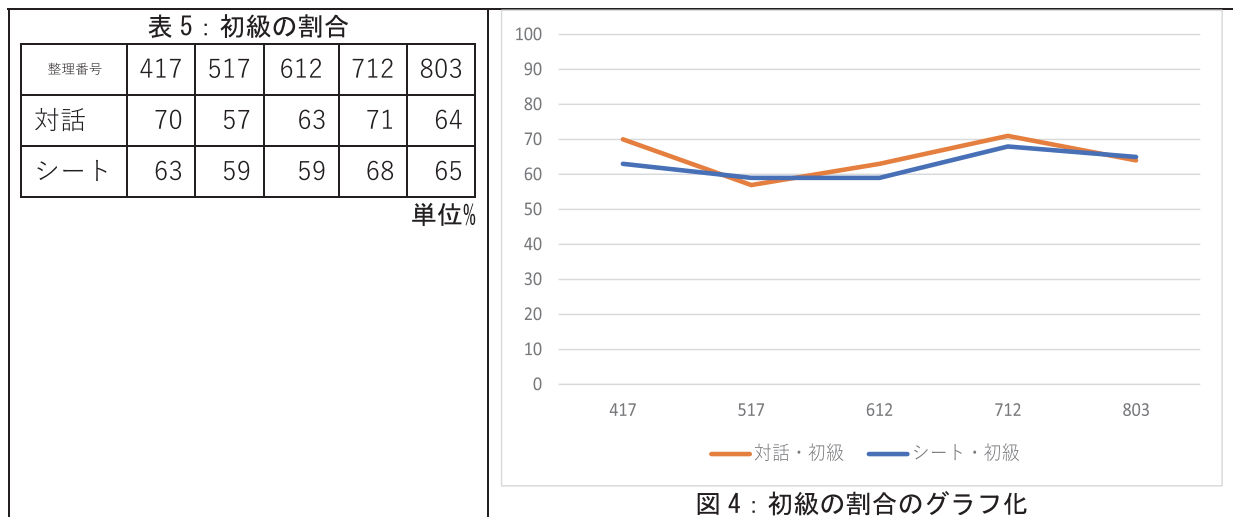


表6：各回の全生徒の発話情報

整理番号	417	517	612	712	803
発話者数(人)	17	16	16	18	22
個人を特定できなかった発話数	28	5	5	1	0
RS(リーダビリティ・スコア)	3.85	3.39	3.06	3.99	3.45
総文数	101	76	87	160	123
一文の平均語数	18.76	16.91	30.66	18.29	24.43
延べ形態素数	1895	1285	2667	2927	3005
異なり形態素数	256	189	299	293	345
対話時間	16分	12	25	27	26
1分あたりの延べ形態素数	118.44	107.08	106.68	108.41	115.58

スします。」とか、「A君。」なども文に含めた。つまり「A君。」も1文にカウントした。

まず、全生徒の発話をまとめて投入した発話情報の一覧を、表6に示す。

発話者数は、発話者を特定できた生徒の人数である。整理番号417では、個人を特定できなかった発話が28あったので、実際の発話者数は17名よりも多い。この回は、ほぼ全員の28名が発話している。

総文数、延べと異なりの形態素数は、対話時間の影響を受けるので、その多寡を単純に比較することはできない。延べ形態素数を時間(分)で割った「1分あ

たりの延べ形態素数」は、106.68～118.44で、およそ12語の幅である。

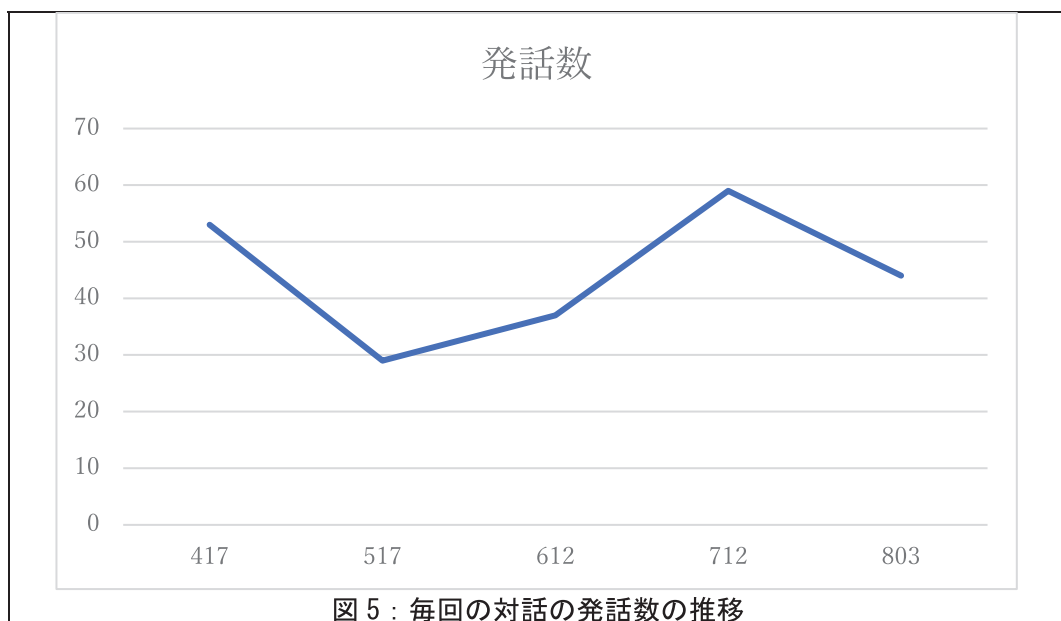
全生徒の発話をまとめて挿入する方法で参考になるのは、リーダビリティ・スコアであり、先に考察した通りである。

次に、発話ごとに挿入して解析した結果を一覧にしたのが、表7である。1発話ごとにシステムに挿入し、各項目ごとの数値を算出し、各対話ごとに平均(AVE)と標準偏差(SD)を計算した。

発話数をグラフ化したのが図5である。このグラフからは、第2回目の対話以降、発話数が増加する全体

表7：発話ごとの発話情報

整理番号	417	517	612	712	803
発話数	53	29	37	59	44
総文数の AVE	1.94	2.62	2.41	2.73	2.86
総文数の SD	1.17	0.93	0.88	0.82	1.15
一文の平均語数の AVE	18.29	17.03	30.13	18.49	27.99
一文の平均語数の SD	6.51	8.23	10.27	10.83	18
延べ形態素数の AVE	35.66	44.31	72	49.61	69.88
延べ形態素数の SD	25.48	21.89	34.33	27.85	34.06
異なり形態素数の AVE	25.45	27.79	40.41	30.92	42.02
異なり形態素数の SD	13.3	10.73	11.30	12.35	13.68



的な傾向（トレンド）がうかがわれる。

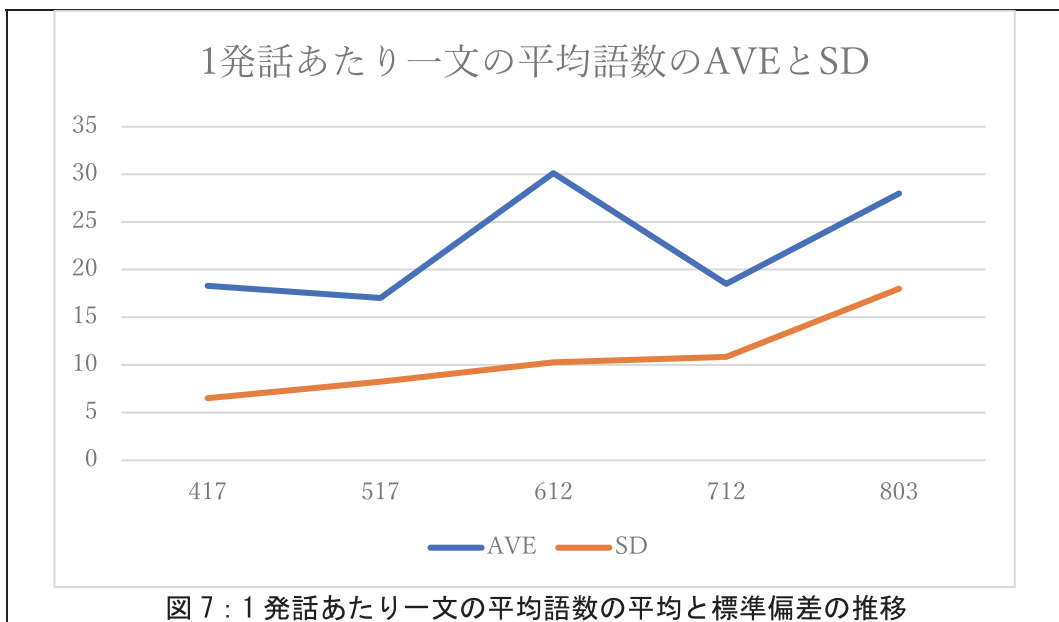
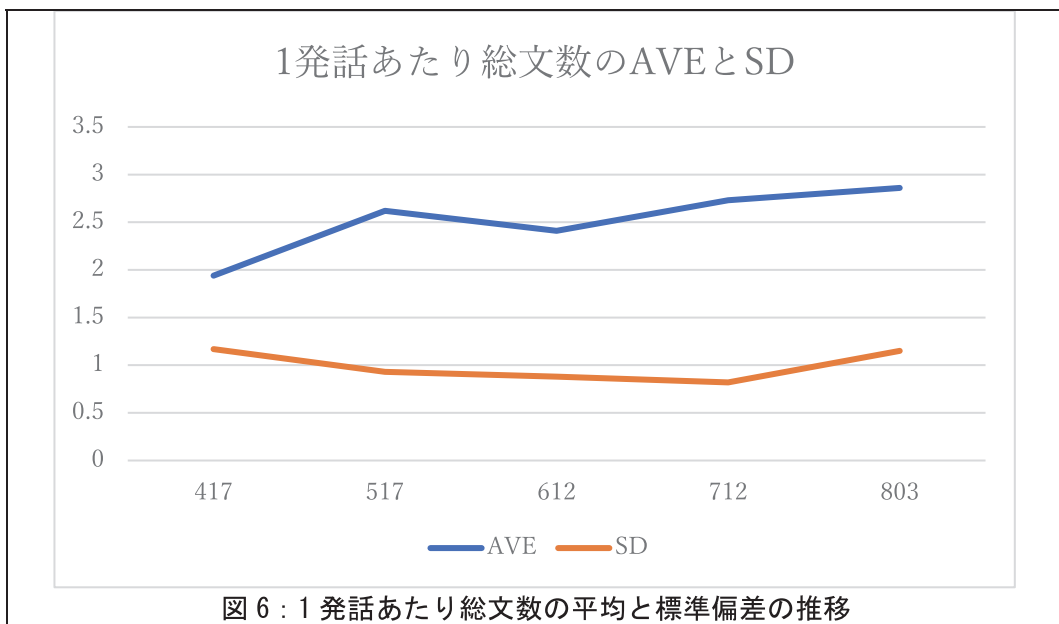
1 発話あたりの「総文数の AVE」は、1 発話を構成する文数の平均を示しており、整理番号417（4月17日の対話）は、1.94文で構成されていたことがわかる。対話の回数を重ねるごとに、1 発話を構成する文数が増加し、整理番号803では、2.86文になり、417よりも平均でおよそ1文増加している。グラフ化すると（図6）、総文数の増加傾向が明確になる。1回の発話で生徒は、より多くの文章を使うようになっている。

また標準偏差は803を別にすると、減少傾向にある。

つまり、ばらつきが小さくなっており、1 発話あたり多くの文章で発言する生徒が増えているといえる。

さらに、1 発話あたりの「一文の平均語数」の平均と標準偏差をグラフ化すると（図7）、ここからも増加傾向が見えてくる。この増加は、平均的に生徒たちは一文をいっそう長い言葉で表現するようになったことを意味する。総文数の傾向と合わせるならば、対話を重ねるごとに、生徒たちは、1回の発話で、より多くの文を使い、より多くの語で一文を構成するようになる、という傾向を示している。

図7では、標準偏差（SD）も増加傾向にあり、一文を



少ない語数で構成する生徒と、多くの語数で構成する生徒のばらつきが大きくなっていることを示している。

図8と図9は、延べと異なりの形態素数の平均と標準偏差である。

いずれも、1発話あたり一文の平均語数の平均と標準偏差の推移と、ほぼ同じトレンドを示している。1発話あたりの延べと異なりの形態素数も増加傾向にあり、標準偏差も増加傾向にある。このことは、対話を重ねるごとに、生徒たちは、1回の発話でより多くの語を活用するようになり、より多くの語彙を駆使するようになったことを意味する。

平均とばらつきの全体の推移を可視化するために、ノッチ付きの箱ひげ図でも描画してみた(図10)。ノッチに重なりがないケースでは、延べと異なりの形態素数平均の増加が、統計的に有意であることを示唆している。

以上の解析結果のトレンドは、次のようにまとめることができる。

- (1) 発話数の増加傾向
- (2) 1発話あたりの文数の増加傾向
- (3) 1発話あたりの一文の平均語数の増加傾向
- (4) 延べと異なりの形態素数の増加傾向

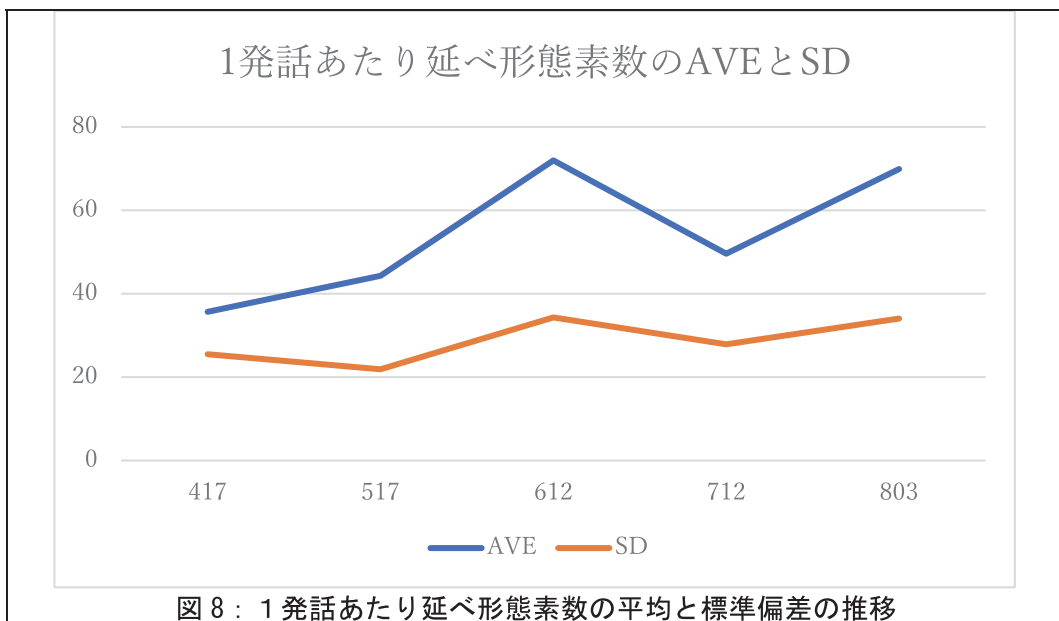


図8: 1発話あたり延べ形態素数の平均と標準偏差の推移

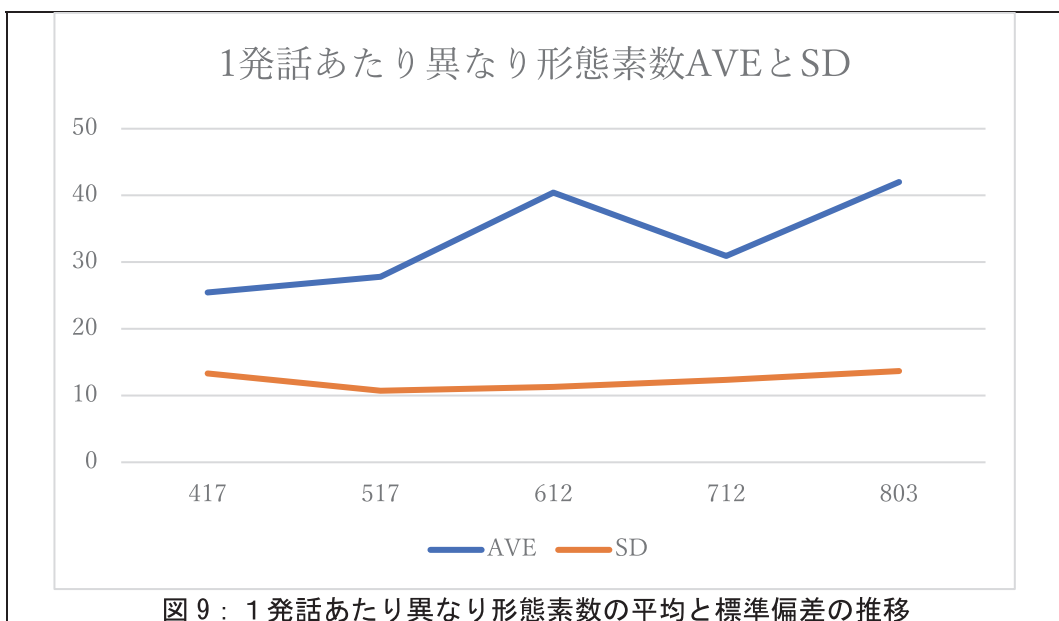


図9: 1発話あたり異なり形態素数の平均と標準偏差の推移

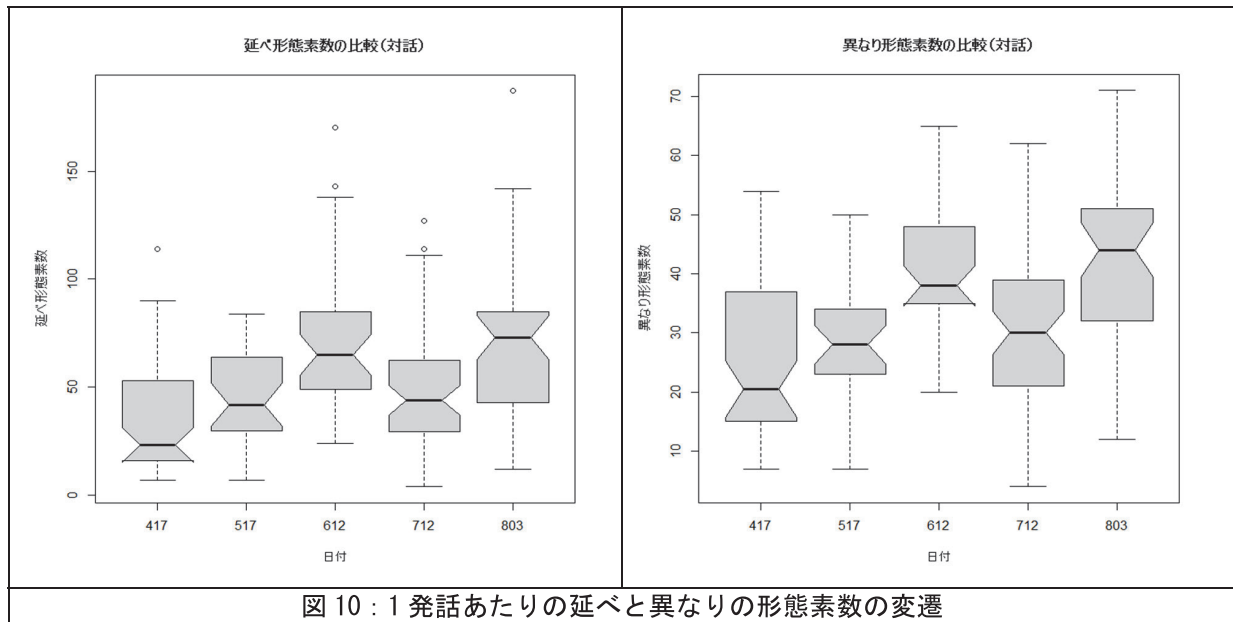


図 10 : 1 発話あたりの延べと異なりの形態素数の変遷

問いの難易度や興味関心の度合いに特段の差はないと仮定するならば、これら4点の増加傾向は、生徒たちの発話能力の向上ないし活性化を強く示唆している。

4-5. データ解析結果と実践知との往還

さらに図10の箱ひげ図には、あるパターンを読み取ることができる。

自由に問いを立てた4月を除き、教科書を刺激物として問いを立てた5月(517)と6月(612)、7月(712)と8月(803)のセットが、類似したパターン(それぞれ前の月より後の月の中央値が高い)を描いている。このパターンが何を意味するか、非実践者である第2・第3執筆には見当がつかなかった。ところが、実践者である第1執筆には、その意味が即座に明らかであった。第1執筆によると、5月<6月、7月<8月の差は、対話に入る前にかけた時間の差に対応している。5月と7月は1時間扱いであり、6月は2時間、8月は3時間扱いであった。実践者には、時数をかけるほど、生徒の学びが深まるという肌感覚があるが、形態素数のこのパターンは、その肌感覚をエビデンスで裏付けている可能性がある。

語彙数の中央値で明らかだが、同じ1時間扱いでも、5月<7月であり、複数時間扱いでも、6月<7月となっており、生徒たちは4月から8月にかけて、着実に語彙数や言葉数を伸ばしている。このエビデ

スは、先の(1)～(4)の傾向を補強するものである。それゆえ、このエビデンスにより、p4c対話の回を重ねるごとに、生徒たちは話し言葉の能力を伸ばしているとの推定に、状況証拠が一つ付加される。

さらに、言葉数と語彙数の増加傾向は、子どもの哲学p4cが重視する対話の場のセーフティ(安心安全)の高まりとも対応している可能性がある。p4cの経験的実践知からすると、セーフティが低い学級では、言葉数が少なく、みんなが同じことを発言する傾向がある。逆にセーフティが高まると、発言が多くなり、かつ多様な視点や発想の発言になる。この経験的実践知と、今回の結果は合致する。逆に、今回の結果から、言葉数と語彙数の増加や減少を、対話の場のセーフティの向上や低下の指数とする仮説が成立する。

5. 得られた知見と今後の課題・展望

以上、ポイントをまとめるなら、今回の形態素解析の結果は、7月対話に生徒の成長を感じたという実践者の肌感覚を裏付けるエビデンスとなった。7月対話は、簡単な言葉で対話しながら思考を深めたという教師の肌感覚と、対話とシートのリーダビリティ・スコアの差が一致した。7月シートには、「今回の対話が一番深まった」という生徒の記述も複数あり、実践者も生徒たちも、対話と思考の深まりを実感している。この実感が、リーダビリティ・スコアにも現れたこと

になる。

また各単元を扱う時数が増えると、学びが深まるという実践感覚についても、生徒の発言の言葉数や語彙数の増加によって裏付けられた。さらに、対話の言葉数と語彙数を、その対話の場のセーフティの高低、ひいてはそこで人間関係の深残の指標とする仮説が浮上した。

加えて、実践者と非実践者との共同作業により、非実践者にはわからないが、実践者にはその意味がわかる数値があることも判明した。対話とシートのRSのギャップの意味(4-3節)や、箱ひげ図のパターンの意味(4-5節)が、非実践者には見当がつかないのに対し、実践者には直ちに明らかであったからである。

また実践者と非実践者にとって、算出された数値がすべて意味をもつわけではないことも明らかである。しかし、現時点では意味をもたない数値も、今後の調査研究により、意味をもつようになるかもしれない。これは次のことを意味する。すなわち、形態素解析の計算結果は客観的であるが、その数値の解釈は主観的であることを免れない。これはデータサイエンス(「DS」)の第3要素に他ならず(2章)、統計的・コンピューターの・人間的であるDSのDSたるゆえんである。

今回の共同研究から、一步踏み込んで明らかになるのは、いくつかの数値が、実践者の中に眠る実践感覚を覚醒させたり、掘り起こしたりするという現象である。その事例として、本稿では、「生徒たちは簡単な言葉で対話しながらも、思考を深めており、深まった思考がふりかえりシートの言葉に現れた」という実践感覚(4-3節)や、「時数をかけるほど生徒の学びが深まる」という実践知(4-5節)が、数値の読み取

りの共同作業のなかで引き出された。

大量の言語データを解析し、データの背後に潜む有益な情報を掘り起こすことが「テキストマイニング(text mining)」と呼ばれるように、今回のデータサイエンスは、教師がもつ大量の実践感覚データの背後に潜む有益な意味を掘り起こし、明るみにもたらず「センスマイニング(sense mining)」という新たなコンセプトを導出した。

さらに、第2執筆者は、25年以上授業研究を専門とし、これまで数多くの授業を見学し、動画記録を視聴、授業の文字起こしの経験も多く、多数の文字記録を読んできた。今回も、動画を視聴し、文字記録を読み込んだが、7月対話が他よりも深まっているとか、生徒が考えながら聞いたり発言している、という確たる感触は得られなかった。ところが、システムによる形態素解析の結果は、生徒や教師の実感を微細に代弁し、実践者の実践感覚と一致していた(4-3節)。ここから、専門的な授業研究者でも見逃してしまう、微細ではあるが意味深い現実を、形態素解析システムが捕捉できることもある、という事例が示された。以上から、形態素解析システムは、実践者の高度で膨大な実践感覚を可視化し、相互主観化する有望なツールであると結論できる。

6. 補論

補論としてリーダビリティ・スコアについて、jReadability(以下「jR」)、不思議箱、ラボ2.0の3システムを比較しておきたい。同じローデータを挿入した結果は表8の通りである。

不思議箱、ラボ0.2はともに、学年レベルを示して

表8：3システムによる各回の対話とシートのリーダビリティ・スコア

	対話 (jR)	シート (jR)	対話 (不思議箱)	シート (不思議箱)	対話 (ラボ0.2)	シート (ラボ0.2)
417	3.85	3.20	7	7	5.16	6.10
517	3.39	2.83	6	6	7.16	6.86
612	3.03	2.69	7	7	5.62	5.38
712	3.99	2.90	6	6	6.55	7.08
803	3.45	3.36	6	6	6.41	6.56

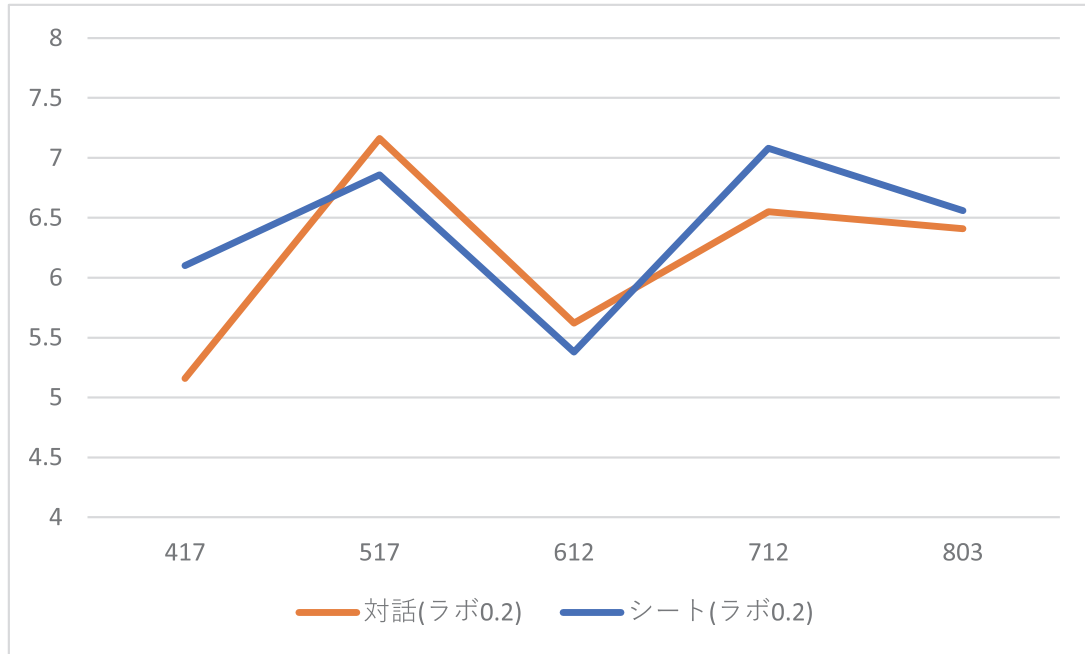


図11：対話とシートのラボ0.2スコアのグラフ化

いる。例えば、「6」は小学6年生レベルで、「7」は中学1年生レベルである。それゆえ、jRとは異なり、これら2つのシステムでは、値が大きいほど、難易度が高いことになる。

表8のように、今回のローデータに対して、不思議箱の識別力が弱いことがわかる。

ラボ0.2の評価は、生徒の学年に相応しい数値になっている。事例は、中学1年生、つまり7年生になったばかりの4月から8月までの時期であるため、ラボ0.2のスコアが、5～7であることは妥当である。ラボ0.2の設計からすると、生徒の学年とスコアとの相応は、事例の中学1年生が、中1国語の教科書の言葉と、ほぼ同じレベルで語り、書いていることを意味する。

ラボ0.2の結果を、図4と同じく、折れ線グラフに表すと、図11となる。

ラボ0.2のリーダビリティ公式は、jRとは異なるが、7月(712)のシートは、対話よりも難易度が高く、jRの結果と一致している。ラボ0.2の解析結果からも、7月の対話は比較的簡単な言葉で交わされ、シートは比較的難しい言葉で書かれていることになる。7月シートの難易度は、5回の中で最も高い。対話と思考が最も深まった、という教師と生徒の実感を裏付けるエビデンスである。

jRとラボ0.2の数値を比較すると、整理番号612の評

価が大きく食い違っている。612の対話とシートについて、jRが相対的に難易度が高いと評価し、ラボ0.2は相対的に難易度が低いと評価している。リーダビリティ公式が異なるため、評価に食い違いが現れるのは自然である。とはいえ、こうした食い違いを、どのように理解し受け止めるかは、今後の課題である。

付記1：学校や生徒の個人情報を保護した上で、データの公開をご快諾くださったA市立B中学校の校長先生に、心より感謝申し上げます。

付記2：本研究は、科学研究費助成事業、基盤研究B「グローバル世界を視野とする学力・非認知能力の効果的学校モデル」(研究代表：田端健人)(2020～2022年度、課題番号：20H01667)の研究成果の一部である。

付記3：第1執筆者は、①対話の授業実践、②授業のビデオ記録、③授業実施情報の提供、④解析結果の解釈を行なった。第2執筆者は、①ローデータの解析、②解析結果の解釈、③本稿本文の執筆を行なった。第3執筆者は、①対話とシートの文字起こし、②ローデータの解析、③解析結果の解釈、④本稿図表の作成を行なった。

付記4：2021年9月25日(日)開催の日本教育方法学会第57回大会(オンライン開催校：宮城教育大学)の「ラウンドテーブル①対話的な学びとコミュニティ形成：

討議倫理にフォーカスした事例提案」にて、第1執筆者が本研究の一部を発表した。

引用文献

- Blei, M. D. & Smyth, P. (2017) Science and data science, *Proceeding of Natural Academy of Science*, 114 (33) : 8689-8692.
- Daniel, M-F., M. Gagnon & E. A. Slusarczyk (2017) Dialogical critical thinking in kindergarten and elementary school. *The Routledge International Handbook of Philosophy for Children*, Routledge.
- Gholamhossein, S. & R. A. Mehdi (2010) A PBLT approach to teaching ESL speaking, writing, and thinking skills. *ELT Journal*, 65/2.
- Gregory, M.R. & J. Haynes & K. Murris (2017) *The Routledge International Handbook of Philosophy for Children*, Routledge.
- 久保順也 (2021) p4c を実践する教員らから見たその効果と課題. 宮城教育大学教職大学院紀要, 2: 21-28.
- 李在鎬 (2016) 日本語教育のための文章難易度に関する研究. 早稲田大学日本語教育学 21: 1-16.
- 榊形公也 (2018) 「考え、議論する」道徳のための新しいアプローチ. 桃山学院大学キリスト教論集, 53: 25-79.
- 楢木靖夫・榊原健太郎 (2014) 大学生の自己表現への方略のなさを低減する試み—話し合いに焦点をあてて—. 帝京科学大学紀要, 10: 177-182.
- 尾身宜彦・田端健人 (2022) 高等学校での定期考査とアンケート調査を有効活用する可視化システムの開発と活用. 宮城教育大学教職大学院紀要, 第3号.
- Sato, S., S. Matsuyoshi & Y. Kondoh (2008) Automatic Assessment of Japanese Text Readability Based on a Textbook Corpus. *LREC-08*: 654-660.
- 佐藤勇一・中川雅道 (2018) 福井高専における子どもの哲学 (P4C) の初実践について—附録 (対話文字起こし、ワークシート記述) 付—. 福井工業高等専門学校研究紀要 人文・社会科学, 50: 23-34.
- 柴崎秀子 (2014) リーダビリティ研究と『やさしい日本語』. 日本語教育, 158: 49-65.
- 柴崎秀子・原信一郎 (2010) 12学年を難易尺度とする日本語リーダビリティ判定式. 計量国語学, 27 (6) : 215-232.
- 柴崎秀子・玉岡賀津雄 (2010) 国語教科書を基にした小・中学校の文章難易学年判定式の構築. 日本教育工学会論文誌, 33 (4) : 449-458.
- 田端健人 (2021a) 子どもの言葉データサイエンス入門—形態素解析システム jReadability の活用と検証—. パイディア出版.
- 田端健人 (2021b) 全国学力・学習状況調査の平均正答率をどう受けとめるべきか?—「生きた数値」による<子ども教育データサイエンス>の構想—. 学ぶと教えるの現象学研究, 19: 1 - 20.