

The Mystery of Edwin Drood の原典および Thomas Power James による続編における語彙の豊かさの評価

後 藤 克 己*

1. はじめに

米国人 Thomas Power James (以下、T. P. James) は、1870 年に世を去った Charles Dickens の遺作となった *The Mystery of Edwin Drood* に続編を加えて完全版とし 1783 年に発表した。(Dickens, James, 1873) この続編部分 (part second) については、物語性、人物造形の一貫性、リテラシー (文法を含む) などの観点から多くの批評がなされている。その背景の一つには、T. P. James がこの続編を、“By the Spirit Pen of Charles Dickens, through a Medium.” (ibid. 中表紙) とアピールしたことがあると考えられる。

本稿は、これらの批評の中から、リテラシーに言及したものを概括するとともに、Dickens による原典 (以下、原典) と、T. P. James による続編 (以下、続編) の文体を、語彙の豊かさ (リテラシー) の観点で比較したものである。

2. リテラシー関連の先行批評

本章では、リテラシーに言及した 3 人の批評家による評価を概括する。まず、George F. Gadd は、続編について、“there is little enough in the second part of the book which, in any degree, serves to recall the genius of the first [Dickens’s original part]” (Gadd, 1905: 272) と批評し、また原典と続編での人物造形の不自然さを指摘するとともに、“presence of the awful grammatical vagaries” (ibid.) と、続編のリテラシーを批判しているが、その “grammatical vagaries” については具体例に言及していない。

次に、John Cuming Walters は続編を、“The work is self-condemned by its futility, illiteracy, and hideous American mannerisms;” (Walters, 2012: 106)¹⁾ と否定的に評している。ただし、その “futility, illiteracy, and hideous American mannerisms” について自らの評は述べず、他の批評家 (George F.

*国際人間学研究科言語文化専攻 博士前期課程

Gadd) による関連言説を引用するにとどめている。なお、Don Richard Cox は、Cox (1998) で、1870–1997 年の間に発表された、*The Mystery of Edwin Drood* の原作、続編、批評などに網羅的な評論を加えているが、そのなかで Walters (2012) を、“The book was a gold mine for Droodists in the early twentieth century; the copy in the British Library is literally falling to pieces.” (ibid. 380) と評している。

さらに、Dickens, Oscar Wilde, および Jack London の遺作の研究家で、また、自ら心霊現象の研究家でもある Arthur Conan Doyle は、Gadd と Walters による否定的な評価に対して、「文体などの特徴に原典との類似点がみられ、また米国風の語彙綴りも見られない」(Doyle, 1927: 344 趣意要約) と具体例を挙げて述べ、“By the Spirit Pen of Charles Dickens, through a Medium.” としたの続編に好意的な批評をしたうえで、全体として “I am content to say that it is far from final, but that it is worthy of respectful consideration.” (ibid. 349) と評している²⁾。このようにこれら 3 人の批評では、リテラシーについて一部具体例を挙げてはいるものの、概して抽象的・感覚的なものととどまっている。

3. リテラシーの観点からの定量的評価

作者のリテラシーは、創作されたテキストの文法(語法を含む)、識字(書き間違いを含む)、また、広くは文体全般に反映すると考えられる。本章では、これらのうち識字の観点から、原典と続編を、それぞれに用いられている語彙の難易度の高低、および語彙使用の多様さに着目して比較する。

3.1 評価に使用したツールとコーパス

本稿で評価に使用したツールは、JACET8000 分析プログラム(以下、V8AN)である。このツールは大学英語教育学会が刊行した『大学英語教育学会基本語リスト(以下、JACET8000)』³⁾を使用して、以下の評価機能を提供している。(清水、2003)

- ・平テキストの英文から、レマ化された頻度リストを作成する。(レマ化は JACET 8000 に基づいて行われる。)
- ・レマ化された語彙に、JACET 8000 のレベル指標(1～8、他)をつける。
- ・レベル指標ごとに、そこに属するレマ化された語の出現数と総語数を算出する。

本稿で評価に使用した原典と続編のコーパスは、筆者が、Dickens, James (1873) を OCR スキャンし、その出力を元の書籍と突き合わせて誤変換文字をチェック／修正したものである。また当コーパスを、テキスト全体での評価に加え、地の文あるいは発話部に限定した評価を行えるよう、地の文と発話部⁴⁾を切り出してサブコーパス化した。

3.2 評価結果

3.2.1 地の文での評価

作中の発話部は、登場人物の言葉づかい（リテラシー）を模倣していることから、地の文にこそ作者の言葉づかいが現れると考えられる。V8ANにて原典と続編の地の文を評価した結果は以下のとおりである。

まずToken語数については表1および図1からわかるように、原典の方が、難易度レベル8超の語数比率が高い。レベル3～8でも原典の方が高いが差は小さい。逆に、レベル1および2では続編の方が高い。つまり、原典の方に難易度の高い語彙がより多く用いられている。

表1 地の文でのToken語数の比較⁵⁾

難易度レベル		1	2	3	4	5	6	7	8	8超	計
原典	Token語数	39,670	3,063	1,709	1,038	813	728	572	451	2,911	50,955
	同 比率	77.9%	6.0%	3.4%	2.0%	1.6%	1.4%	1.1%	0.9%	5.7%	100%
続編	Token語数	55,542	4,532	2,053	1,250	914	775	680	531	1,970	68,247
	同 比率	81.4%	6.6%	3.0%	1.8%	1.3%	1.1%	1.0%	0.8%	2.9%	100%

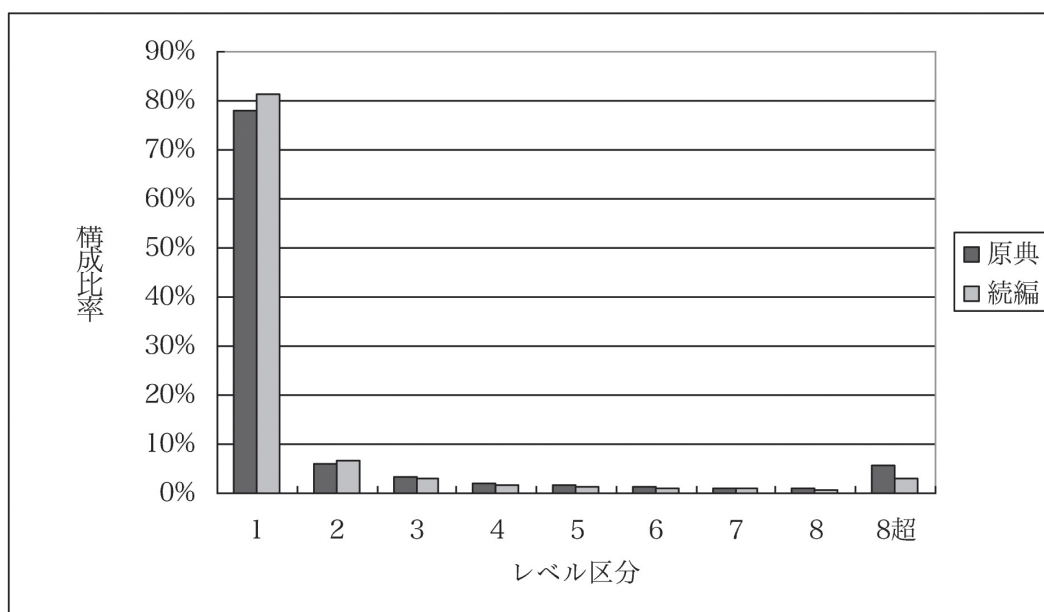


図1 地の文Token語数の難易度レベル別構成比率

つぎにIndex語数については表2および図2のように、原典で難易度レベル8超のIndex語数比率が高く、レベル1および2では続編の方が高い。つまり、原典の方に難易度の高い語彙がよ

り多く用いられている。なお、レベル 3～8 では、双方の比率は拮抗している。

表 2 地の文での Index 語数の比較⁶⁾

難易度レベル		1	2	3	4	5	6	7	8	8 超	計
原典	Index 語数	921	679	530	364	337	308	243	246	2,166	5,794
	同 比率	15.9%	11.7%	9.1%	6.3%	5.8%	5.3%	4.2%	4.2%	37.4%	100%
続編	Index 語数	868	621	399	287	261	213	181	184	1,158	4,172
	同 比率	20.8%	14.9%	9.6%	6.9%	6.3%	5.1%	4.3%	4.4%	27.8%	100%

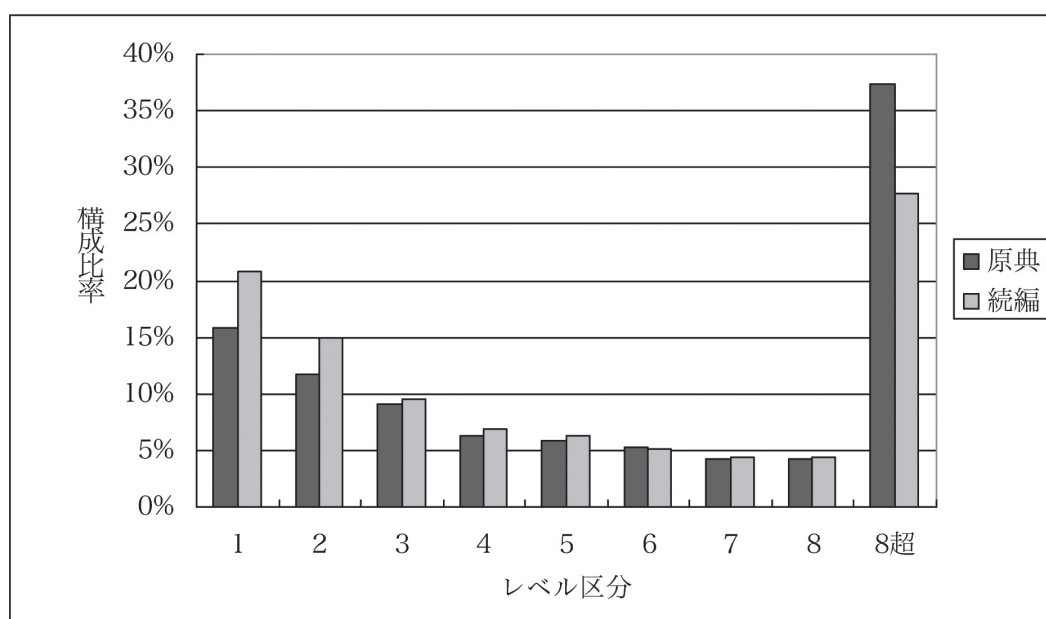


図 2 地の文 Index 語数の難易度レベル別構成比率

さらに Index 語数と Token 語数の比の比較では、表 3 と図 3 のように全てのレベルにて続編の方が Index/Token 比率が低く、その比率は全体で原典の約 1/2 になっている。これは、続編は原典より同じ語が用いられる傾向が高いことを表している。

表 3 地の文での Index 語数と Token 語数の比較⁷⁾

難易度レベル		1	2	3	4	5	6	7	8	8 超	全体
原典	Index/Token	2.3%	22.2%	31.0%	35.1%	41.5%	42.3%	42.5%	54.5%	74.4%	11.4%
続編	Index/Token	1.6%	13.7%	19.4%	23.0%	28.6%	27.5%	26.6%	34.7%	58.8%	6.1%

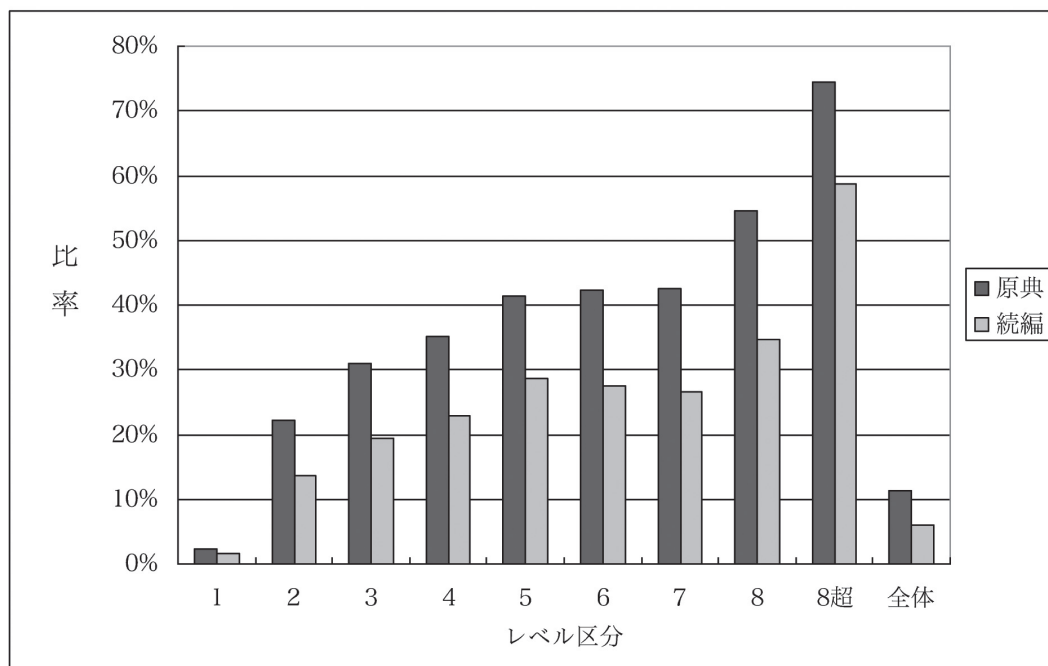


図3 地の文 Index/Token 比の難易度レベル別比率

3.2.2 発話部での評価

発話部の Token 語数については表4および図4のように、地の文の場合と同様に、難易度の高い語彙の比率は原典で高く、低い語彙のそれは続編で高い傾向がみられる。ただし、その差は地の文の場合と比較して小さい。

表4 発話部での Token 語数の比較

難易度レベル		1	2	3	4	5	6	7	8	8超	計
原典	Token 語数	34,305	1,923	863	628	382	396	195	223	1,154	40,069
	同 比率	85.6%	4.8%	2.2%	1.6%	1.0%	1.0%	0.5%	0.6%	2.9%	100%
続編	Token 語数	48,013	2,809	1,224	779	527	428	240	279	1,216	55,515
	同 比率	86.5%	5.1%	2.2%	1.4%	0.9%	0.8%	0.4%	0.5%	2.2%	100%

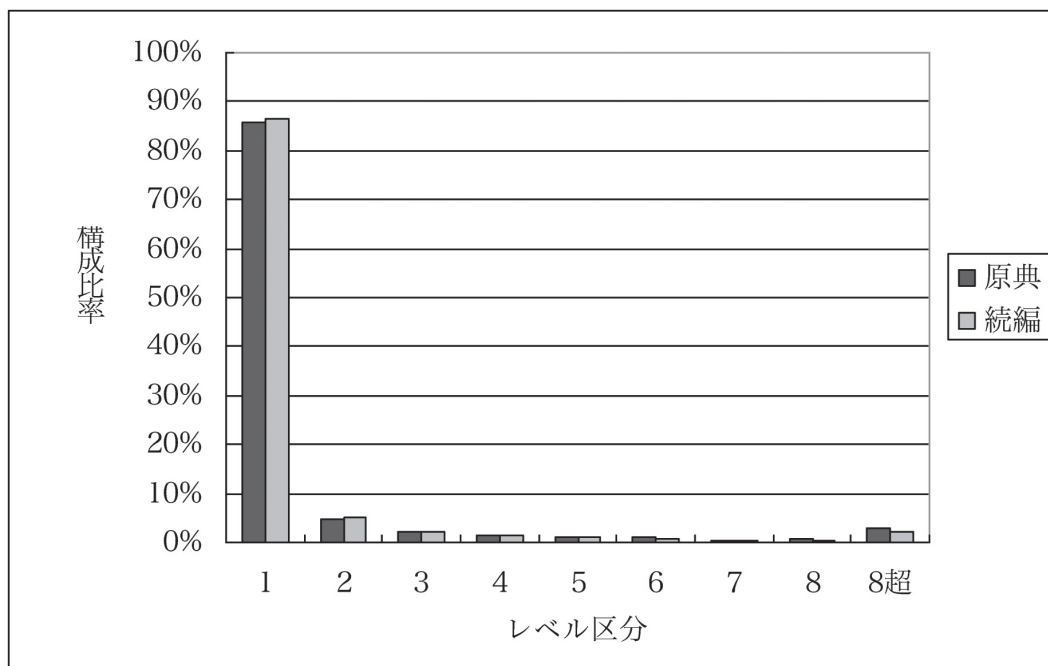


図 4 発話部 Token 語数の難易度レベル別構成比率

つぎに Index 語数については表 5 および図 5 のように、Token 語数の場合と同様に難易度の高い語彙の比率は原典で高く、低い語彙のそれは続編で高い傾向がみられるが、その差は、Token 語数の場合と同様に、地の文の場合より小さい。

表 5 発話部での Index 語数の比較

難易度レベル		1	2	3	4	5	6	7	8	8超	計
原典	Index 語数	852	549	343	255	197	181	122	147	857	3,503
	同 比率	24.3%	15.7%	9.8%	7.3%	5.6%	5.2%	3.5%	4.2%	24.5%	100%
続編	Index 語数	840	568	327	246	200	163	107	115	723	3,289
	同 比率	25.5%	17.3%	9.9%	7.5%	6.1%	5.0%	3.3%	3.5%	22.0%	100%

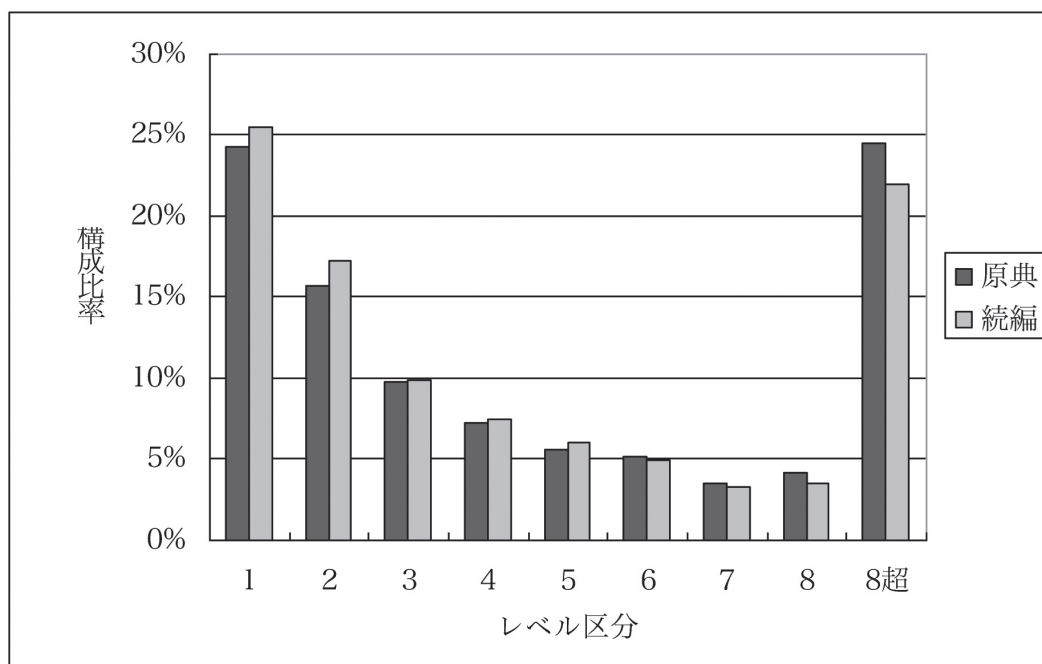


図5 発話部 Index 語数の難易度レベル別構成比率

さらに Index 語数と Token 語数の比較では表6と図6のように、地の文の場合と同様に、全てのレベルにて、続編の方が Index/Token 比率が低くなっている。なお、続編の語彙全体の Index/Token 比率は原典のその約 2/3 であり、地の文の場合より差は小さい。

表6 発話部での Index 語数と Token 語数の比較

難易度レベル		1	2	3	4	5	6	7	8	8超	全体
原典	Index/Token	2.5%	28.5%	39.7%	40.6%	51.6%	45.7%	62.6%	65.9%	74.3%	8.7%
続編	Index/Token	1.7%	20.2%	26.7%	31.6%	38.0%	38.1%	44.6%	41.2%	59.5%	5.9%

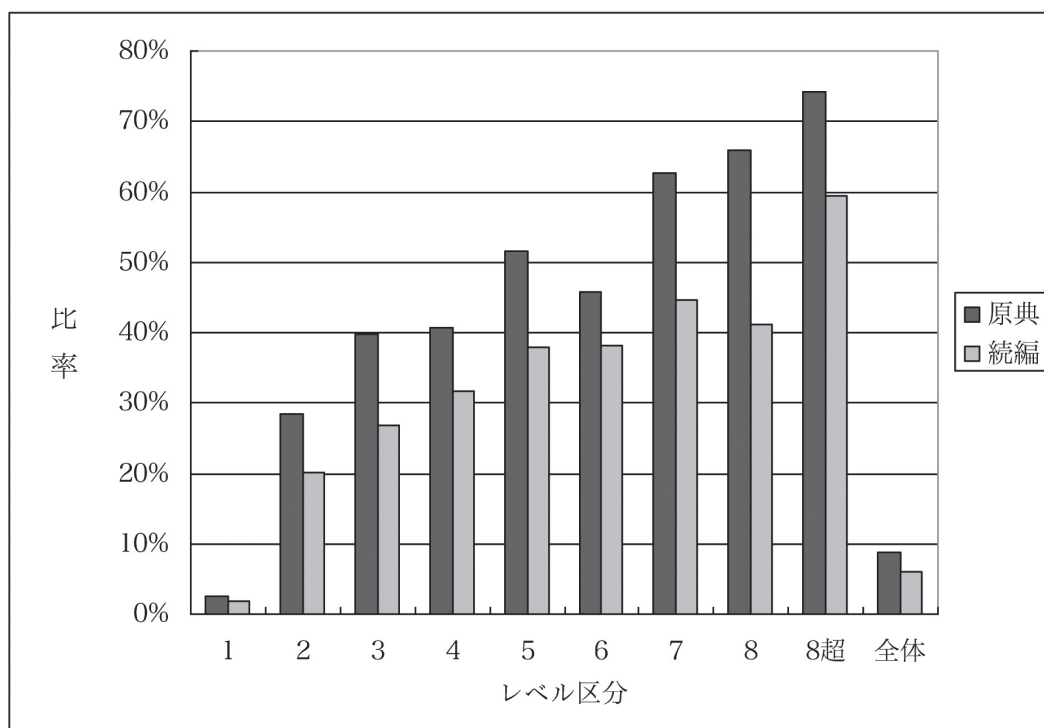


図6 発話部 Index/Token 比の難易度レベル別比率

3.2.3 地の文と発話部での Index/Token 比率の比較

Index/Token 比率は、語彙使用の多様さ（語彙の豊かさ）の尺度になると考えられる。テキスト全体⁸⁾、地の文、発話部の各コーパスにおいて、難易度レベル1から8超までの全語彙による Index/Token 比率をまとめて表7に示すが、ここから全体、地の文、発話部、のいずれの場合でも、原典の方が Index/Token 比率が高く、語彙がより豊かだといえる。また原典と続編のそれぞれについて、地の文と発話部の Index/Token 比率を比較すると、原典の地の文の比率は発話部のその約1.3倍だが、続編ではごく僅かな差（約1.03）となっている。

表7 Index/Token 比率一覧

	全体	地の文	発話部	地の文／発話部
原典	7.7%	11.4%	8.7%	1.30
続編	4.2%	6.1%	5.9%	1.03
原典／続編	1.8	1.9	1.5	

4. 結び

3章で行ったToken語数、Index語数およびIndex/Token比率の語彙難易度レベル別分析から以下のことが判明した。第1にToken語数およびIndex語数による識字レベル評価であるが、語彙のレベル別構成比率では、いずれも難易度の高いレベル8超では原典の方が続編より高く、一方、難易度の低いレベル1、2では続編の方が高くなっている。このことは、原典と続編の作者の識字レベルに差があることを示唆していると考えられる。

第2に地の文と発話部のIndex/Token比率の比較による識字レベル評価である。一般に、Index/Token比率については、日常語彙を多く含む発話部より地の文の方が高くなると考えられ、今回のコーパスでも原典にはその傾向が明らかに表れている。ところが続編では地の文での値が発話部より僅かに高くなっているにすぎず、語彙使用面で地の文の使用語彙と、登場人物のことばづかいを模倣する発話語彙との区別が、あまりなされていないものと推定される。このことを地の文が（含意された）作者のことばづかいを模倣することと考え合わせると、続編の作者の識字レベルが原典のそれより低いことを示唆していると考えられる。

本稿では原典と続編における語彙範疇での文体類似性を、原典と続編をそれぞれ1コーパスとして、語彙の難易度を、JACET8000の語彙レベルを基に比較することで評価した。今後、原典と続編のコーパスを、例えば章単位で分割するなどサブコーパス化し、各サブコーパスにおける語彙頻度データをもとに、テキストマイニング分野で多用されつつあるコレスポネンス分析など多変量解析ツールを用いて、より詳細な文体類似性評価研究をすすめる予定である。

注

- 1) この版は、同書のEstes, 1913版をOCRで再版したものである。
- 2) Doyleは続編全体ではなく、当時彼が入手していた続編の一部を基にこの批評をしている。そしてその点について、Waltersが指摘している“futility”について批評するには続編全体を精査する必要があるものの、“illiteracy and Americanisms”については一部による精査で足りると述べている。(Doyle, 1927: 344)
- 3) 『大学英語教育学会基本語リスト』は、BNC (British National Corpus, 1億語) を基準スケールとし、それに日本人の英語学習を反映した言語資料と、アメリカ英語を多く含んだ言語資料から作成したサブコーパスをからめて作製され (相澤他編, 2005: 2) であり、基本の8000語を難易度から以下の8つのレベルに分けている。(ibid. 5-6)
 - ・レベル1: 中学の英語教科書に頻出
 - ・レベル2: 高校初級、英字新聞の75%をカバー
 - ・レベル3: 高校の英語教科書、社会人の教養として身につけたいレベル
 - ・レベル4: 大学受験、大学一般教養の初級
 - ・レベル5: 難関大学受験、大学一般教養
 - ・レベル6: 英語を専門としない大学生、ビジネスマンの目標

- ・レベル 7: 英語専攻の大学生、ビジネスマンの目標
 - ・レベル 8: 日本人の英語学習者の最終到達目標レベル (除く、専門用語)
- 4) 本稿では、二重引用符で括られた部分のみを発話部としている。作品中に登場人物の発話／思考が自由間接表現された箇所がごく一部あり、該当部分はその人物のこぼれ言葉になっていると考えられるが、本稿では地の文として扱っている。
 - 5) V8AN ではコーパスの語彙は、難易度レベル 1～8、8 超に加えて、cont. forms, non-words, および proper nouns に区分される。ここで、cont. forms には、I've, you'll のようなアポストロフィによる省略形、方言の発話表記 'ope [hope] などが区分される。本稿では、レベル 1～8、および 8 超に区分された語を評価対象とし、それ以外は除外した。なお、表の Token 語数は通常の語数の数え方で表された語の出現数である。
 - 6) 表の Index 語数はレマ化された語の出現数である。例えば、「look, looked, looking」の 3 語は、Index 換算で 1、Token 換算で 3 となる。(清水、2003)
 - 7) 表の Index/Token は、Index 語数と Token 語数の割合を表す指標で、「Index 語数÷Token 語数×100」で算出される値である。(ibid.) なお、V8AN では i/t としてレベル全体での値が得られるが、本稿では、レベルごとに算出して表に加えた。
 - 8) テキスト全体の V8AN での処理結果は付録を参照

参考文献

- Cox, Don Richard (1998) *Charles Dickens's The Mystery of Edwin Drood: An Annotated Bibliograph*. New York: AMS Press, Inc.
- Dickens, Charles [& T. P. James] (1873) *Part Second of the Mystery of Edwin Drood. By the Spirit-Pen of Charles Dickens, Through a Medium. Embracing, Also That Part of the Work Which Was Published prior to the Termination of the Author's Earth-Life*. Brattleboro, VT: T. P. James.
- Doyle, Arthur Conan (1927) "The Alleged Posthumous Writings of Great Authors", *The Bookman* 66, New York. <https://www.unz.org/Pub/Bookman-1927dec-00342>
- Gadd, George F. (1905) "The History of a Mystery. A Review of the Solutions to 'Edwin Drood' (Chs. 3 & 4)", *The Dickensian* 1.
- Walters, John Cuming (2012) *The Complete Mystery of Edwin Drood: History, Continuations, and Solutions*. Memphis: General Books.
- (この版は、同書の Estes, 1913 版を OCR で再版したものである)
- 相澤一美、石川慎一郎、村田年 編集代表 (2005) 『「大学英語教育学会基本語リスト」に基づく JACET8000 英単語』、桐原書店
- 清水伸一 (2003) "JACET 8000 付属 CD-ROM のプログラムデータ使用法"、第 42 回 JACET 全国大会、<http://www.tcp-ip.or.jp/~shim/J8.htm>

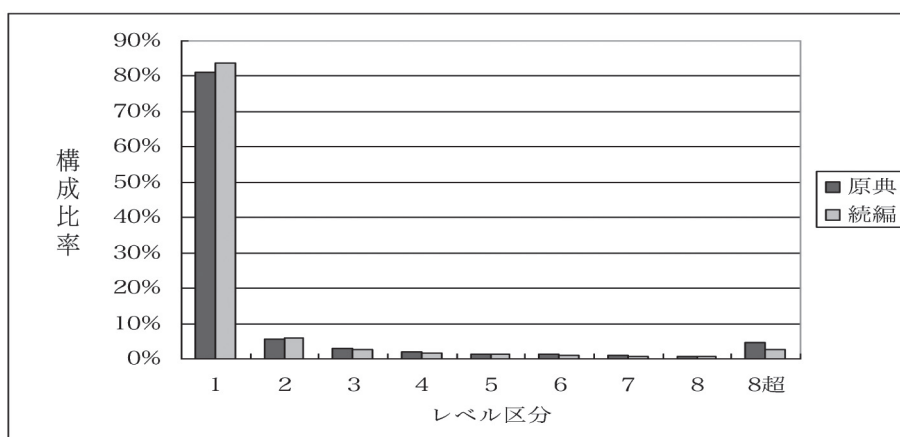
付録

テキスト全体（地の文＋発話部）での V8AN の処理結果は以下のとおりである。

Token 語数の比較を付表 1 および付図 1 に示す。

付表 1 全体（地の文＋発話部）での Token 語数の比較

難易度レベル		1	2	3	4	5	6	7	8	8超	計
原典	Token 語数	73,975	4,986	2,572	1,666	1,195	1,124	767	674	4,065	91,024
	同 比率	81.3%	5.5%	2.8%	1.8%	1.3%	1.2%	0.8%	0.7%	4.5%	100%
続編	Token 語数	103,555	7,341	3,277	2,029	1,441	1,203	920	810	3,186	123,762
	同 比率	83.7%	5.9%	2.6%	1.6%	1.2%	1.0%	0.7%	0.7%	2.6%	100%

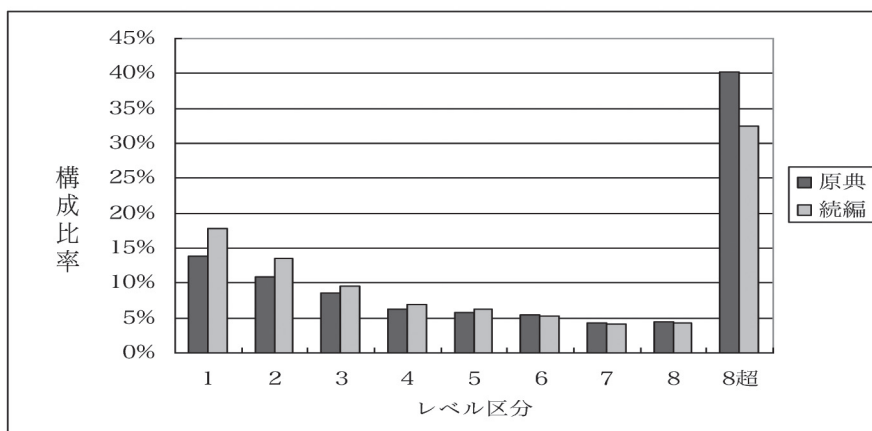


付図 1 全体（地の文＋発話部）での Token 語数の難易度レベル別構成比率

Index 語数の比較を付表 2 および付図 2 にしめす。

付表 2 全体（地の文＋発話部）での Index 語数の比較

難易度レベル		1	2	3	4	5	6	7	8	8超	計
原典	Index 語数	969	764	605	440	405	387	305	312	2,826	7,013
	同 比率	13.8%	10.9%	8.6%	6.3%	5.8%	5.5%	4.3%	4.4%	40.3%	100%
続編	Index 語数	924	704	496	356	322	273	216	222	1,683	5,196
	同 比率	17.8%	13.5%	9.5%	6.9%	6.2%	5.3%	4.2%	4.3%	32.4%	100%

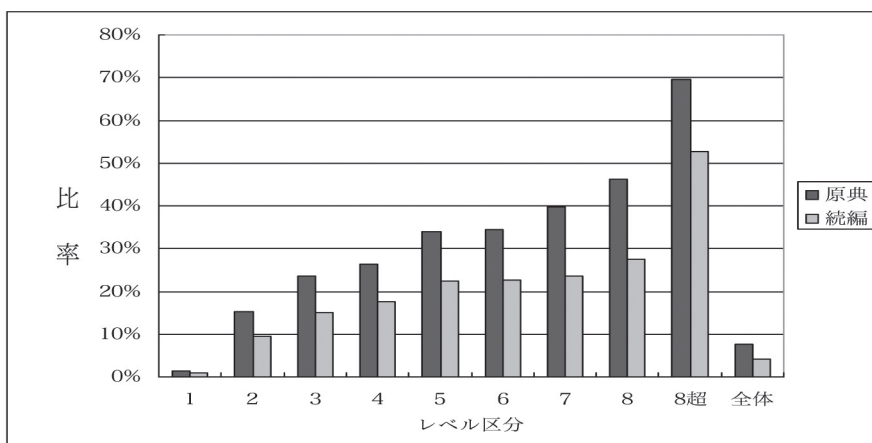


付図2 全体（地の文＋発話部）での Index 語数の難易度レベル別構成比率

Index/Token 比の比較を付表3および付図3に示す。

付表3 全体（地の文＋発話部）での Index / Token 比の比較

難易度レベル		1	2	3	4	5	6	7	8	8超	全体
原典	Index/Token	1.3%	15.3%	23.5%	26.4%	33.9%	34.4%	39.8%	46.3%	69.5%	7.7%
続編	Index/Token	0.9%	9.6%	15.1%	17.5%	22.3%	22.7%	23.5%	27.4%	52.8%	4.2%



付図3 全体（地の文＋発話部）での Index/Token 比率のレベル別比較