



Title	Random Forestsを使用したAgatha Christie作品の文体分析 : Dorothy Sayersと比較して
Author(s)	土村, 成美
Citation	言語文化共同研究プロジェクト. 2017, 2016, p. 43-54
Version Type	VoR
URL	https://doi.org/10.18910/62018
rights	
Note	

The University of Osaka Institutional Knowledge Archive : OUKA

<https://ir.library.osaka-u.ac.jp/>

The University of Osaka

Random Forests を使用した Agatha Christie 作品の文体分析 * Dorothy Sayers と比較して

土村 成美

大阪大学大学院言語文化研究科

〒 560-0043 豊中市待兼山町 1-8

E-mail: t.naru.425@gmail.com

あらまし 本研究ではミステリー作家 Agatha Christie 作品における特徴語に関して、同時代の女性ミステリー作家 Dorothy Sayers との比較を通して計量的な分析を行うことを目的とする。分析手法として、アンサンブル学習の一種である Random Forests を使用し、高頻度語を変数とした 2 作家の作品の分類を行い、分類に寄与した特徴語について検討を行う。Random Forests を実行した結果、Christie と Sayers の作品を 90% を超える正解率で分類することが出来た。抽出された特徴語として、Sayers が *everybody, anybody* のような *-body* 形名詞や *till* が挙がるのに対して、Christie は *everyone, anyone* のような *-one* 形名詞や *until* を多用していることが明らかになった。また、Christie 側には女性代名詞や動詞の過去形が多く特徴語として抽出され、これらの語を Christie が特徴的に用いることが明らかになった。動詞の過去形の中でも動作にまつわる動詞に注目すると、発話を導く際に発話動詞に加えて登場人物の動きを表現したり、頷くという行為一つを取っても様々な *-ly* 副詞と共に共起させて表現したりと、登場人物の行動の描写が工夫されており、Christie 作品が読者にとって情景描写を行いやすく、現在でも広く普及している一要因となっているのではないかと考えられる。

キーワード Agatha Christie, Random Forests, 特徴語

A Stylometric Approach to Agatha Christie's Works Using Random Forests

Comparing with Dorothy Sayers

Narumi TSUCHIMURA

Graduate School of Language and Culture, University of Osaka

1-8 Machikaneyama-cho, Toyonaka, Osaka, 560-0043 Japan

Abstract This study tries to make a quantitative analysis on characteristic words in Agatha Christie works, comparing with her contemporary Dorothy Sayers. This study applies Random Forests, one of ensemble learning methods, for classifying the two writers' works. The variables used in Random Forests are the most frequent words, and we discuss the 100 the most characteristic words. The results of this analysis revealed that Christie's use of characteristic words contrasted with Sayers', especially between synonyms: Christie tends to use *someone, anyone* and *until* while Sayers tends to use *somebody, anybody* and *till*. Moreover, Christie tends to use words

related to the female gender and a number of past tense verbs. Among the past tense verbs, we focus on the words related to movement. It revealed that Christie described characters' behaviour in her novels by using a number of co-occurring words like *-ly* adverbs, and I assume that this makes readers to imagine scenes in her novels easily. It may be one of reasons why her novels are read and loved all over the world now.

Keywords Agatha Christie, Random Forests, characteristic words

1. はじめに

本研究はミステリー作家 Agatha Christie 作品の語彙的特徴に関して、統計的手法を用いて分析を行うことを目的とする。Christie 作品に関する計量的手法を用いた研究としては、語彙多様性に関して分析を行った Lancashire & Hirst (2009) や、語彙面と統語面からのアプローチを行った Le et al. (2011) などが存在する。また、稲木 (2009; 2013) のように、Christie 作品のスマールコーパスを用い、WordSmith Tools を用いて抽出された特徴語を元に、Christie の作品中での読者を誤誘導するテクニックを考察した研究もある。しかし、これらはいずれも分析に用いる作品が限定的であり、Christie のデビュー作品から晩年の作品まで網羅的に分析を行った研究は見当たらない。本研究では短編を含めた Christie のミステリー作品を網羅的に用いた分析を行う。

Christie との比較対象として、Christie と同時代に活躍したイギリスの女性作家 Dorothy Sayers の作品を用いる。Sayers は Peter Wimsey 卿作品が代表的なミステリー作家である。日本では紹介が遅れたために、他の作家と比べると知名度はあまり高くないが、ミステリー黄金時代の 1930 年代には作品の初版部数が Christie と拮抗するほど (Malling and Peters, 1998) 人気の高い作家である。また P. D. James や Minette Walters など多くの女性ミステリー作家が理想の作家として挙げるのも Sayers であり (森, 1998), 「ミステリーの 2 大女王」として Christie と並んで語られることが多いため、本研究で Christie と比較を行う作家として選定した。

本研究では Sayers 作品を比較対象とし、機械学習の手法を用いた Christie 作品の分析を行う。機械学習の一手法である Random Forests を用いた 2 作家の作品の分類を行い、Christie 作品における特徴語の検討を行う。この分析を通して Christie の語彙使用の特徴の一端を示すことを試みるのが本研究の目的である。

2. 使用データ・分析手法

2.1. リサーチクエスション

本研究では以下の 2 つのリサーチクエスションを設定した。

1. 機械学習の手法を用いて Christie の作品と Sayers の作品とを分類することは可能か

*本論文は、JADH2016(September 12th to 14th, University of Tokyo) において行った研究発表 “Stylistic Analysis of Agatha Christie’s Works: Comparing with Dorothy Sayers” に加筆・修正を行ったものである。本研究を進めるにあたり、田畑智司先生 (大阪大学)、ホドシチェク・ボル先生 (大阪大学) をはじめ、研究プロジェクトの先生方及びメンバーの方々に研究会において様々な助言を頂いた。ここに記して深く感謝申し上げたい。なお、本稿における誤りは全て筆者の責任である。

2. 分類に寄与した特徴語のうち、Christie と Sayers とで語の用い方にどのような違いがあるのか

2.2. 使用データ

分析に使用するデータについて、それぞれの作家の作品数、異なり語数、総語数を以下の表 1 に示した。両作家とも長編・短編を含む。両作家とも複数ジャンルでの作品執筆を行っているが、ジャンルの違いによる使用語彙の違いを考慮して、両作家ともミステリー作品に限定しての分析を行う。

表 1: それぞれの作家の使用データの構成

Author	Number of texts	Types	Tokens
Christie	221	40,885	5,230,256
Sayers	55	31,967	1,430,257

2.3. Random Forests とは何か

本研究では分析手法として Random Forests を用いる。Random Forests とは、Breiman (2001) が提唱した、分類・回帰を行うためのアンサンブル学習の一種である。Random Forests は、元データからブートストラップサンプリングされたデータを用い、多数の決定木の作成を行う。その多数の決定木の予測結果を組み合わせ、多数決により結果を得る。Random Forests はデータの 3 分の 2 を用いてモデルの作成を行い、残りの 3 分の 1 を用いて作成したモデルのテストを行う。

Random Forests はテキストや著者の識別においても利用されている。金・村上 (2007) では、小説、作文、日記の 3 種類の文章を用い、Random Forests を利用した書き手の同定を行い、他の分類法 (k 最近傍法、学習ベクトル量子化法、サポートベクターマシン、バギング、ブースティング) と分類精度を比較した。その結果、Random Forests は他の分類法より正解率が高いことが報告されている。小林・田中・富浦 (2011) は、英語母語話者と非英語母語話者が執筆した英語学術論文について、メタ談話標識を変数としてランダムフォレストを用いた分類を行い、母語話者と非母語話者それぞれに特徴的な表現の抽出を行なった。その結果、母語話者の英語は -ly 副詞や従属接続詞、hedges を多用するのに対し、非母語話者は *on the other hand* や *moreover* のような接続表現や *show* や *think* のような同一の基本動詞を多用していることが明らかになった。また、田畑 (2012) では、Charles Dickens と Wilkie Collins の作品を識別する特徴語の抽出を Random Forests を用いて行った。対数尤度比やカイ二乗値など従来の特徴語抽出の指標は、固有名詞のようにある特定のテキストのみに出現する語を特徴語として抽出してしまうという問題点があるが、Random Forests を用いるとその問題を克服した特徴語抽出が可能であることを示した。

2.4. 分析手法

Random Forests を行うに際し、変数として高頻度語の出現頻度を用いた。変数の選定理由として、高頻度語は Jane Austen 作品の計量的な文体分析を行った Burrows(1987, 1992) などにおいて、著者の文体を識別する指標としての有効性が示されているためである。作品ごとの総語数は異なるため、頻度は調整頻度を用いた。分析には統計解析ソフトウェア R を用い、randomForests パッケージを使用して Random Forests を実行した。分析に用いる変数を 100 語から 1000 語まで 100 語ずつ変化させ分析を行った。

3. 分析結果

3.1. Random Forests の実行と overfitting の問題

上記のように変数を 100 語から 1000 語まで 100 語ずつ変化させて Random Forests を実行したところ、分類精度は 92.7%–95.3%であった。以下の表 2 に Random Forests の実行例を示す。

表 2: Random Forests 実行結果の一例

Call:			
randomForest(formula = author ~.,			
data=cs.all, proximity=T, importance=T, ntree=500)			
Type of random forest: classification			
Number of trees:		500	
No. of variables tried at each split:		24	
OOB estimate of error rate:		4.71%	
Confusion matrix:			
	Christie	Sayers	class.error
Christie	221	0	0.0000000
Sayers	13	42	0.2363636

ところが、Christie と Sayers で作品数・総語数共に 4 倍ほどの大きな差があるために overfitting してしまい、Sayers 作品が十数作品ほど常に Christie 作品として誤判別されてしまうという問題が発生した。そのため、分析を実行する都度両作家から 50 作品ずつランダムサンプリングを行い、先ほどと同様に Random Forests を実行した。ランダムサンプリングを行った際の分類精度は 94%–100%となり、分類精度が向上した。これより先はランダムサンプリングを行った際の Random Forests の結果を用いて考察を行う。変数が 600 語の時に最も高い平均分類精度が得られたため、変数が 600 語の際の実行結果を元に考察を行う。

表 3: Random Forests 実行結果の一例 (データのランダムサンプリング後)

Call:			
randomForest(formula = author ~.,			
data=cs50, proximity=T, importance=T, ntree=500)			
Type of random forest: classification			
Number of trees:		500	
No. of variables tried at each split:		24	
OOB estimate of error rate:		0%	
Confusion matrix:			
	Christie	Sayers	class.error
Christie	50	0	0
Sayers	0	50	0

3.2. Christie と Sayers の特徴語の比較

randomForests パッケージには varImpPlot() 関数が実装されており, グループの判別に大きく寄与した変数を求めることが可能である。Christie 作品と Sayers 作品の判別において大きく寄与した語上位 30 語を図示したものが以下の図 1 である。

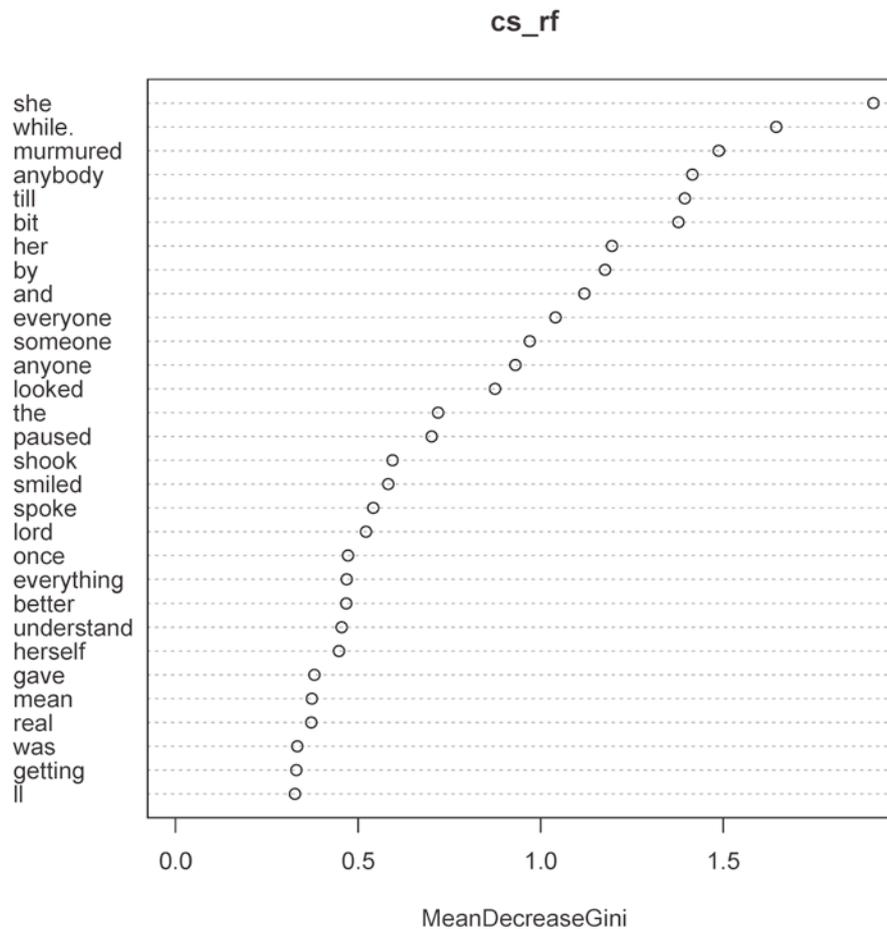


図 1: ジニ係数平均減少値に基づく, 分類に大きく寄与した語

Christie 作品と Sayers 作品の判別において大きく寄与した上位 100 語について、作家ごとにより高頻度で出現する語を特徴語として分類したものを表 4 に示した。この特徴語のリストに基づいてこの先考察を行う。

表 4: Christie と Sayers の特徴語

Christie
she, murmured, her, everyone, someone, anyone, looked, paused, shook, smiled, spoke once, everything, understand, herself, gave, mean, real, was, am, slowly, head, nodded eyes, really, looking, is, were, possible, are, smile, girl, voice, thoughtfully, world, stared yet, sometimes, going, suddenly, face, slightly, do, sharply, most, least, different perhaps, minute, will, told, not, sure, one, shall, there, always, can, what, talking, you
Sayers
while, anybody, till, bit, by, and, the, lord, better, getting, ll, upon, o, somebody, if, turn when, this, after, gentleman, replied, into, time, book, set, get, nobody, however glass, clock, only, found, look, make, over, take, in, put, god

4. 考察

4.1. *-one vs -body, until vs till*

表 4 において Christie と Sayers の特徴語を比較すると、まず両者で対照的であるのは類義語であり、Christie は *everyone, someone, anyone* といった *-one* 形の語が特徴語となっているのに対し、Sayers は *anybody, somebody, nobody* といった *-body* 形の語が特徴語として抽出されている。抽出された特徴語の中で対となっている *someone* と *somebody, anyone* と *anybody* とについて、それぞれの作家の使用頻度を表 5 に示した。

表 5: Christie と Sayers の *-one* 形, *-body* 形の使用頻度 (100 万語あたり)

Author	<i>everyone</i>	<i>everybody</i>	<i>anyone</i>	<i>anybody</i>
Christie	207.64	114.91	416.81	97.51
Sayers	20.28	125.85	72.02	461.46

表 5 から分かるように、Christie は *-one* 形の名詞を圧倒的に多用しており、それに対し Sayers は *-body* 形の名詞を多用していることが分かる。

また、これと同様に対照的であるのが、Christie は *until* が特徴語として挙がっているのに対し、Sayers は *till* が特徴語となっている。表 6 に示すように、両者は *until* と *till* について対照的な使用をしている。Sayers は *till* を多用しているのに対し、Christie は *until* の使用頻度が圧倒的に高い。

表 6: Christie と Sayers の *until* と *till* の使用頻度 (100 万語あたり)

Author	<i>until</i>	<i>till</i>
Christie	257.54	107.83
Sayers	84.60	491.52

Christie, Sayers 共に *-one* 形名詞と *-body* 形名詞, *until* と *till* を使用しているため, 何らかの使い分けをしている可能性があると考えた。 *-one* 形の名詞と *-body* 形の名詞の差異について, また *until* と *till* の差異について, Quirk et al. (1985) は次のように述べている。

The pairs of pronouns with personal reference (eg: *everybody*, *everyone*) are equivalent in function and meaning, but the pronouns in *-one* are regarded as more elegant than those in *-body*. (Quirk et al. 1985: 376, 378)

In the Survey of English Usage corpus of (largely) British English, *until* occurs twice as often as *till*, and is used approximately equally as conjunction and as preposition, and equally too in spoken and written materials. On the other hand, *till* occurs chiefly in spoken texts, and is predominantly a preposition. (Quirk et al. 1985: 533, 534)

till は *until* に比べて口語的であるとの記述があり, *-one* 形の名詞は *-body* 形の名詞より格調高いと説明されているため, ここで会話文では *-body* 形の名詞と *till* がより多く使用され, 地の文では *-one* 形の名詞と *until* がより多く出現するとの仮説を立て, それぞれの作家のこれらの語の使用頻度を調査することとした。以下の表 7, 表 8, 表 9 にそれぞれ *everyone* と *everybody*, *anyone* と *anybody*, *until* と *till* の地の文と会話文における出現頻度をまとめた。頻度は全て 100 万語あたりに調整している。

表 7: Christie, Sayers 作品の *everyone*, *everybody* の出現頻度内訳

	Christie		Sayers	
	<i>everyone</i>	<i>everybody</i>	<i>everyone</i>	<i>everybody</i>
conversation	151.62	89.10	13.98	125.85
narrative	56.02	25.81	6.29	51.04

表 8: Christie, Sayers 作品の *anyone*, *anybody* の出現頻度内訳

	Christie		Sayers	
	<i>anyone</i>	<i>anybody</i>	<i>anyone</i>	<i>anybody</i>
conversation	360.02	86.61	51.74	394.34
narrative	56.78	10.90	20.28	67.12

表 9: Christie, Sayers 作品の *until*, *till* の出現頻度内訳

	Christie		Sayers	
	<i>until</i>	<i>till</i>	<i>until</i>	<i>till</i>
conversation	177.24	77.82	40.55	332.11
narrative	80.30	30.02	44.05	158.01

以上の表 7, 8 より, *-one* 形名詞と *-body* 形名詞については, 両作家共に地の文よりも会話文において *-one* 形名詞も *-body* 形名詞も多用していることがうかがえる。地の文と会話文との分類では, 両作家ともこれらの語の使い分けは行っていないようである。表 9 の *until* と *till* については, Christie はどちらの語も会話文において多用していることが分かる。しかし Sayers は, *till* は会話文での頻度が高く, *until* は逆に地の文での頻度の方が高くなっている。

4.2. 女性代名詞

Christie 側の特徴語として, 女性代名詞 *she*, *her* が挙がっている。特に *she* については, 図 1 から分かるように, ジニ係数平均減少値が最も大きく, Christie と Sayers の作品を判別する上で最も寄与率が高くなっている。*she* と *her* の両作家の使用頻度を表 10 に示す。

表 10: Christie, Sayers 作品における *she*, *her* の使用頻度 (100 万語あたり)

Author	<i>she</i>	<i>her</i>
Christie	12,238	9,215
Sayers	5,515	4,387

Christie 作品では Sayers 作品に比べると *she*, *her* 共に圧倒的に高頻度で使用されている。Sayers の作品では探偵は男性の Lord Peter Wimsey のみであるが, Christie 作品では Miss Marple や Tommy Beresford の相方である Tuppence Beresford といった女性探偵が主人公である作品も存在する。このような登場人物が *she* の生起頻度に影響を与えているのかどうかについては, 作品の主人公を考慮した分析が必要となるため, 今後の課題としたい。

4.3. 動詞の過去形

他に Christie の特徴語で特徴的なのは, Sayers に比べると動詞の過去形が多いことである。動詞の過去形は地の文での narration に多用されるため, これは小説テキストならではの特徴であると言える。*murmured*, *told* のような発話に関する語, *looked* や *stared* のような視覚に関する語などが挙がっているが, ここでは動作に関する動詞 *paused* と *nodded* に注目する。

はじめに *paused* について, *paused* を含む 2 語の cluster を検索すると, 出現頻度が上位のものは以下ようになった。cluster, コロケーションの検索及びコンコーダンスラインの表示には, コンコーダンサー CasualConc を使用した。

表 11: *paused* を含む 2 語 cluster

Rank	Cluster	Freq	Rank	Cluster	Freq
1	he paused	681	6	paused then	78
2	paused and	403	7	poirot paused	62
3	she paused	391	8	paused to	35
4	paused a	174	9	i paused	33
5	paused for	97	10	paused in	33

ここで2位に挙がっている *paused and* について、*paused* とどのような語を等位接続詞 *and* を用いて繋いでいるのか確かめるため、*paused and* を node とした際に右側 1 語目に共起する語を検索した結果を以下表 12 に示す。最低頻度を 5 に設定したところ、5 語が抽出された。

表 12: *paused and* の R1 における共起語

Rank	Context word	Freq
1	then	196
2	said	28
3	added	25
4	went	21
5	looked	10

Christie は *paused and then* という表現を他の表現に比べて多用していることが分かる。*paused and then* のコンコーダンスラインの一部を以下に示す。

	Kwic - 196 found in 69 files	ファイル名
38	. Countess Anna Radzky. Mrs Macatta. Mr James Thesiger ' He paused and then asked sharply:	The Seven Dials Mystery.txt
39	'No, no, it's nothing to do with that. It's just ' He paused and then asked, 'What did it say? At	The Moving Finger.txt
40	He paused and then asked: 'What made Miss Megan	The Moving Finger.txt
41	'I came straight back here.' She paused and then asked, 'Why did you think	The Clocks.txt
42	Madame Joilet paused and then asked with a sudden gleam of	4.50 from Paddington.txt
43	dangerous to those who do not know how to handle them!' She paused and then asked, 'You have brought what	The Pale Horse.txt
44	ed a few words with her on the day she left Jerusalem.' She paused and then blurted out: 'As a matter of	Appointment with Death.txt
45	'I wanted to see for myself. I've been worried.' He paused and then broke out, 'Why, Lucy	The Hollow.txt
46	'He was this summer.' She paused and then burst out: 'Mark will have it	The Body in the Library.txt
47	At the doorway Jane paused and then came back.	Death in the Clouds.txt
48	Half reluctantly, Rowley followed her. At the door, Lynn paused and then came swiftly back. She spoke	Taken At The Flood.txt
49	He paused and then , clearing his throat, he	Lord Edgware Dies.txt
50	Miss Lemon paused and then continued:	Hickory Dickory Dock.txt
51	a week ago that she would like nothing better than to ' he paused and then deliberately quoted, "'to put	Death on the Nile.txt
52	Frankie paused and then explained her plans for	Why Didn't They Ask Evan...
53	ght. A man's got to stick up for his wife. Dick would.' She paused and then explained: 'You see, I've	The Moving Finger.txt
54	'It is indeed.' Poirot paused and then hazarded a guess. 'Are you,	Dead Man's Folly.txt
55	He paused and then he said slowly:	Evil Under the Sun.txt
56	'Ah, you think that?' He paused and then , his voice changing, he said:	The Arcadian Deer.txt
57	He paused and then leaned a little farther	The Pale Horse.txt
58	He paused and then murmured.	The Sittaford Mystery.txt
59	less. I writhed in impotent agony. They came down the path, paused and then Poirot himself appeared, his	The Big Four.txt
60	He paused and then quietly, circumspectly, he	One, Two, Buckle My Sho...
61	r the sun. It was almost a quotation from Ecclesiastes.' He paused and then quoted himself: 'Yes, also	Evil Under the Sun.txt
62	He paused and then remarked abruptly:	Death in the Clouds.txt
63	She paused and then remarked in a tone of	They Do It With Mirrors.txt
64	He paused and then resumed:	Five Little Pigs.txt
65	He paused and then resumed: 'Discounting the	The Mirror Crack'd From...
66	ix running once and I wasn't. I'm not the bilious type.' He paused and then said:	4.50 from Paddington.txt
67	He paused and then said:	After the Funeral.txt

図 2: *paused and then* のコンコーダンスラインの一部

図2から明らかであるように、Christieは*paused and then*の後に発話を導く動詞を置き、発話内容を続ける、という形を多用している。Christieは登場人物に発話をさせる際、単に*said*や*asked*のみで発話内容を導くのではなく、発話の際の登場人物の動きも表現していることが分かる。

次に*nodded*に関して、*nodded*を含む2語のclusterの出現頻度上位50位を以下に示す。

表 13: *nodded* を含む 2 語 cluster

Rank	Cluster	Freq	Rank	Cluster	Freq
1	poirotnodded	480	26	nodded gently	20
2	nodded his	256	26	nodded then	20
3	he nodded	211	26	nodded vigorously	20
4	she nodded	184	26	other nodded	20
5	nodded he	125	30	nodded at	19
6	nodded her	121	31	doctor nodded	18
7	nodded and	120	32	nodded sympathetically	17
8	nodded thoughtfully	93	32	race nodded	17
9	i nodded	81	32	who nodded	17
10	and nodded	64	35	man nodded	16
11	nodded to	52	35	neele nodded	16
12	marple nodded	51	35	nodded a	16
13	nodded in	40	35	nodded the	16
14	tuppence nodded	34	39	girl nodded	15
15	battle nodded	30	39	latter nodded	15
16	nodded approval	28	39	luke nodded	15
16	tommy nodded	28	39	nodded approvingly	15
18	inspector nodded	27	39	nodded slowly	15
18	nodded yes	27	39	then nodded	15
20	satterthwaite nodded	26	45	nodded towards	14
21	nodded she	24	46	nodded as	13
22	nodded comprehendingly	23	46	nodded it	13
22	nodded gravely	23	48	quin nodded	12
22	nodded with	23	48	weston nodded	12
25	nodded i	22	50	frankie nodded	11

表13より、*nodded*と様々な-ly副詞が共起する形がChristie作品において多用されていることを読み取ることが出来る。Christieは頷くという行為一つをとっても、様々な副詞を共に用いて、どのような頷き方をしているかを描写しているのである。Sayers作品については、*nodded*の総出現頻度が119回と、Christieの2,322回と比較して圧倒的に少なく、*nodded and*が20回、*wimsey nodded*が19回出現する以外は全て出現頻度が10回未満で極端に少なく、-ly副詞との共起も殆ど見られなかった。

以上のような *paused, nodded* の使用のされ方から、Christie は登場人物の行動を表現する際に様々な語を併用して描写することで、登場人物がどのように動いているのか、小説の場面の情景描写をより読者にさせやすいようにしているのではないかと考えられる。森(1998)では、Christie 作品は Conan Doyle の Sherlock Holmes シリーズと並んで映像向きであることが述べられているが、このような登場人物の行動の描写が映像向きであることに繋がっており、現在でも広く Christie の作品が親しまれている一つの要因となっているのではないだろうか。

5. おわりに

本研究では、Random Forests による Christie 作品と Sayers 作品の分類、及び分類に寄与した特徴語の検討を行った。Christie 作品と Sayers 作品の分類は、90%を超える高い精度で行うことが可能であった。全作品を Random Forests にかけると *overfitting* の影響で分類精度がやや落ちたため、最終的にはランダムサンプリングを行なって両作家の作品数を揃えて Random Forests を実行したところ、分類精度が向上した。

そして各作家の特徴語を見てみると、Sayers は *-body* 形名詞や *till* を多用するのに対し、Christie は *-one* 形名詞や *until* を好んで用いることが明らかになった。両作家ともこれらの対になる語を両方とも作品中で使用しており、地の文と会話文とで語の使い分けをしているのか調査したが、目立った使い分けはしていないことが見受けられた。

また、Christie は女性代名詞 *she, her* を Sayers と比較して圧倒的に多用していることが明らかになった。他には動詞の過去形が特徴語として多く挙がっていた。動作を表す動詞に関して *cluster* とコロケーションで使用のされ方を確認したところ、Christie は登場人物の行動の表現について様々な語を並置して描写することで、場面の情景描写を読者にさせやすいようにしているのではないかと推察された。

今後の課題としては、第一に先にも述べたように、Christie 作品に女性代名詞 *she* と *her* が頻出するのは、女性探偵を主人公とする作品が一定数含まれるためという可能性も考えられる。そのため、作品の主人公と女性代名詞の使用との関係性を調査したい。第二に、Quirk et al. (1985)における *until* と *till* の違いに関する記述では、品詞に関する記述も見られるが、本研究ではタグ付けを行っていないテキストについて分析を行ったため、Quirk et al. (1985)にあるような品詞の違いを考慮した考察は出来ていない。今後はアノテーションを施したテキストの分析も行いたいと考えている。第三に、*-one* 形名詞と *-body* 形名詞の使い分けがされているかどうか調べる際に会話文と地の文の区別を行なったが、Christie が話者ごとにこれらの名詞の使い分けを行っている可能性も考えられる。話者タグ付けを行っての分析も今後行いたいと考えている。

文 献

- [1] Breiman, L. (2001). Random Forests. *Machine Learning*, 45: pp.5-23
- [2] Burrows, J. (1987). *Computation into Criticism: A Study of Jane Austen's Novels and an Experiment in Method*. London: Clarendon Press.

- [3] Burrows, J. (1992). Not Unless You Ask Nicely: The Interpretative Nexus Between Analysis and Information. *Literary and Linguistic Computing*, 7(2): pp.91–109.
- [4] 稲木昭子 (2009) 「発話動詞と共起する-ly 副詞：誤誘導のテクニック」『英語文化学会論集』18: 1–14.
- [5] 稲木昭子 (2013) 『謎解きのことば学：アガサ・クリスティの英語を楽しむ』東京：英宝社.
- [6] 金明哲・村上征勝 (2007) 「Random Forests 法による文章の書き手の同定」『統計数理』55(2): pp.255–268.
- [7] 小林雄一郎・田中省作・富浦洋一 (2011) 「ランダムフォレストを用いた英語科学論文の分類と評価」『情報処理学会研究報告 IPSJ SIG Technical Report』第 90 号 (2011-CH-90): 53–68.
- [8] Lancashire, I. and Hirst, G. (2009, March). *Vocabulary Changes in Agatha Christie's Mysteries as an Indication of Dementia: A Case Study*. Paper presented at the 19th Annual Rotman Research Institute Conference, Cognitive Aging: Research and Practice, Toronto.
- [9] Le, X., Lancashire, I., Hirst, G. and Jokel, R. (2011). Longitudinal detection of dementia through lexical and syntactic changes in writing: a case study of three British novelists. *Literary and Linguistic Computing*, 26(4): pp.435–461.
- [10] Malling, S. and Peters, B. (ed.) (1997). *AZ Murder Goes: Classic*. Scottsdale: Poisoned Pen Press.
- [11] 森英俊編著 (1998) 『世界ミステリ作家事典：本格派篇』東京：図書刊行会.
- [12] Quirk, R., Greenbaum, S., Leech, G. and Svartvik, J. (1985). *A Comprehensive Grammar of the English Language*. London: Longman.