

形態素解析による
品詞の傾向と文が与える印象との
関連性についての研究

神奈川県立厚木高校 77期生 E組 4班 α

—目次—

- | | |
|---------|---------|
| 1.背景 | 6.実験結果 |
| 2.目的 | 7.考察 |
| 3.社会的意義 | 8.結論 |
| 4.仮説 | 9.展望 |
| 5.実験方法 | 10.参考文献 |

1. 背景

「僕は、お母さんの『いったよね？』が大っ嫌い。いつだって、間違ってるのは、僕だ。そして正しいお母さんの、憎らしいことしたら。」

(井上よう子『どこかいきのバス』より)

「現在の我々—それもまたグローバルの視点から見れば、極めて小さな点存在であるにもかかわらず、人間の常としてほとんど論理的に、現在の我々がそこから決して離脱できない性格のものである」

(村上陽一郎『科学史・科学哲学入門』より)

柔らかい・硬いなどの
⇒ 印象を左右する要因は？

2. 目的

~~文章の柔らかさ・硬さの原因を明らかにする~~

⇒今回は「~~品詞の傾向~~」に着眼

3. 社会的意義

- 「柔らかい」「硬い」といった文の傾向を文章生成AIに学習させ、文の硬さを調整する

⇒より作りたい文章に相応しい印象を持った文章を生成することが可能に

例)

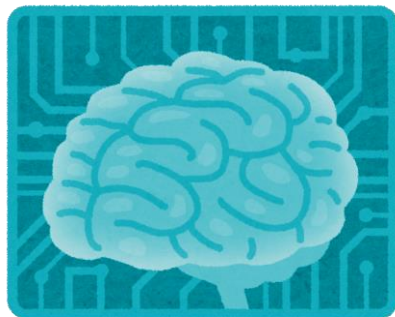
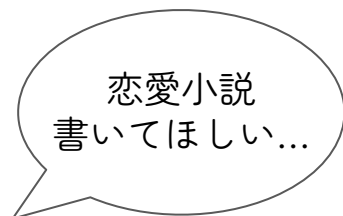


図1 AIのイラスト

助動詞が○%
連体詞が△%
形状詞が□%…



文章を生成

従来より恋愛小説らしい文体に!

印象を左右する要因:

品詞の傾向

例) ミステリー小説 ⇒ 説明口調…接続詞や助詞が多くなる

恋愛小説 ⇒ 会話口調…感動詞が多くなる

5. 実験方法

1. 各ジャンルの電子書籍を形態素解析ツール「MeCab」を

利用して形態素解析する

2. ジャンルごとに含まれる品詞の出現頻度に有意差があるか
検証し,各品詞の傾向を考察する

特定の品詞の出現頻度に
有意差が見られる



その品詞が文の印象を
左右していると考えられる

5. 実験方法

1. 形態素解析ツール「MeCab」を用いて,

以下の書籍の形態素解析を行う

〈ミステリー小説〉

・夏と花火と私の死体
ート

・銃とチョコレート

〈恋愛小説〉

・吉祥寺の朝日奈くん
を

・くちびるに歌

5. 実験方法

1-1 電子書籍の本文を
Google Docsに起こす

1-2 自動読み取りに誤りが
無いか全文確認する

1-3 MeCabを利用し、
形態素解析を行う

表1 MeCabによる形態素解析

Requirement already satisfied: mecab-python3 in /usr/local/lib/python3.10/dist-packages (1.0.8)		Requirement already satisfied: unidic-lite in /usr/local/lib/python3.10/dist-packages (1.0.8)	
「	「	補助記号-括弧開	
まさか	まさか	まさか	副詞 1
我々	ワレワレ	ワレワレ	我々 代名詞 0
の	ノ	の	助詞-格助詞
い	イル	居る	動詞-非自立可能 上-一段-ア行 未然形-一般 0
ない	ナイ	ない	助動詞 助動詞-ナイ 連体形-一般
とき	トキ	時	名詞-普通名詞-副詞可能 2
に	ニ	に	助詞-格助詞
色々	イロイロ	イロイロ	色々 形状詞-一般 0
な	ナ	だ	助動詞 助動詞-ダ 連体形-一般
こと	コト	コト	名詞-普通名詞-一般 2
が	ガ	ガ	助詞-格助詞
はじまっ	ハジマッ	ハジマル	始まる 動詞-一般 五段-ラ行 連用形-促音便
て	テ	て	助詞-接続助詞
しまう	シマウ	仕舞う	動詞-非自立可能 五段-ワア行 終止形-一般 0
なんて	ナンテ	なんて	助詞-副助詞
。	。	。	補助記号-句点
しかも	シカモ	シカモ	然も 接続詞 2
そこ	ソコ	ソコ	其処 代名詞 0
に	ニ	に	助詞-格助詞
悪人	アクニン	アクニン	悪人 名詞-普通名詞-一般 0
たち	タチ	達	接尾辞-名詞的-一般
が	ガ	ガ	助詞-格助詞
関わっ	カカワッ	カカワル	関わる 動詞-一般 五段-ラ行 連用形-促音便 3.0

5. 実験方法

1-4 出力されたデータを Google Sheetsに起こす

表2 MeCabで出力されたデータ

```

Requirement already satisfied: mecab-python3 in /usr/local/lib/python3.8/site-packages (from mecab-python3==1.0.2)
Requirement already satisfied: unidic-lite in /usr/local/lib/python3.8/site-packages (from mecab-python3==1.0.2)
Requirement already satisfied: mecab-python3 in /usr/local/lib/python3.8/site-packages (from mecab-python3==1.0.2)
Requirement already satisfied: unidic-lite in /usr/local/lib/python3.8/site-packages (from mecab-python3==1.0.2)

```

「	マサカ	マサカ	「	補助記号-括弧開	
まさか	マサカ	マサカ	まさか	副詞	1
我々	ワレワレ	ワレワレ	我々	代名詞	
の	ノ	ノ	の	助詞-格助詞	
い	イ	イル	居る	助詞-非自立可能 上二段-ア行	未然形一般
ない	ナイ	ナイ	ない	助動詞	助動詞-ナイ
とき	トキ	トキ	時	名詞-普通名詞-副詞可能	連体形一般
に	ニ	ニ	に	助詞-格助詞	
色々	イロイロ	イロイロ	色々	形状詞-一般	
な	ナ	ダ	だ	助動詞	助動詞-ダ
こと	コト	コト	事	名詞-普通名詞-一般	連体形一般
が	ガ	ガ	が	助詞-格助詞	
はじまっ	ハジマツ	ハジマツ	始まる	動詞-一般	
て	テ	テ	て	助詞-接続助詞	
しまう	シマウ	シマウ	仕舞う	動詞-非自立可能 五段-ワア行	終止形一般
なんて	ナンテ	ナンテ	なんて	助詞-副助詞	
。			。	補助記号-句点	



表3 スプレッドシートに起こしたデータ

「	マサカ	マサカ	「	補助記号-括弧開	
まさか	マサカ	マサカ	まさか	副詞	
我々	ワレワレ	ワレワレ	我々	代名詞	
の	ノ	ノ	の	助詞-格助詞	
い	イ	イル	居る	助詞-非自立可能 上二段-ア行	未然形一般
ない	ナイ	ナイ	ない	助動詞	助動詞-ナイ
とき	トキ	トキ	時	名詞-普通名詞-副詞可能	連体形一般
に	ニ	ニ	に	助詞-格助詞	
色々	イロイロ	イロイロ	色々	形状詞-一般	
な	ナ	ダ	だ	助動詞	助動詞-ダ
こと	コト	コト	事	名詞-普通名詞-一般	連体形一般
が	ガ	ガ	が	助詞-格助詞	
はじまっ	ハジマツ	ハジマル	始まる	動詞-一般	五段-ラ行
て	テ	テ	て	助詞-接続助詞	連用形-促音便
しまう	シマウ	シマウ	仕舞う	動詞-非自立可能 五段-ワア行	終止形一般
なんて	ナンテ	ナンテ	なんて	助詞-副助詞	
。			。	補助記号-句点	

色々	イロイロ	イロイロ	色々	
な	ナ	ダ	だ	助動詞 助動
こと	コト	コト	事	名詞-普通名詞-一
が	ガ	ガ	が	助詞-格助詞
はじまっ	ハジマツ	ハジマル	始まる	動詞-一般
て	テ	テ	て	助詞-接続助詞
しまう	シマウ	シマウ	仕舞う	動詞-非自立可能
なんて	ナンテ	ナンテ	なんて	助詞-副助詞
。			。	補助記号-句点

色々	色々	形状詞-一般
な	だ	助動詞
こと	事	名詞-普通名詞-一
が	が	助詞-格助詞
はじまっ	始まる	動詞-一般
て	て	助詞-接続助詞
しまう	仕舞う	動詞-非自立可能
なんて	なんて	助詞-副助詞
。	。	補助記号-句点

5. 実験方法

1-5 品詞順に並び替え， 品詞の数を集計する

表3 スプレッドシートに起こしたデータ

「	マサカ	マサカ	「	補助記号-括弧前		
まさか	マサカ	マサカ	まさか	助詞		
我々	ワレワレ	ワレワレ	我々	代名詞		
の	ノ	ノ	の	助詞-格助詞		
い	イル	居る	い	動詞-非自立可能 上一段-ア行	未然形-一般	
ない	ナイ	ない	ない	助動詞	助動詞-ナイ	連体形-一般
とき	トキ	時	とき	名詞-普通名詞-副詞可能		
に	ニ	に	に	助詞-格助詞		
色々	イロイロ	色々	色々	形状詞-一般		
な	ナ	だ	な	助動詞	助動詞-ダ	連体形-一般
こと	コト	事	こと	名詞-普通名詞-一般		
が	ガ	が	が	助詞-格助詞		
はじまっ	ハジマツ	ハジマル	はじまっ	動詞-一般	五段-ウ行	連用形-促音便
て	テ	て	て	助詞-接続助詞		
しまっ	シマウ	シマウ	しまっ	動詞-非自立可能 五段-ウア行	終止形-一般	
なんて	ナンテ	なんて	なんて	助詞-副助詞		
				補助記号-句点		



表4 品詞順に並び替えられたデータ

色々	イロイロ	イロイロ	色々	形状詞-一般		
の	ノ	ノ	の	助詞-格助詞		
に	ニ	ニ	に	助詞-格助詞		
が	ガ	ガ	が	助詞-格助詞		
て	テ	テ	て	助詞-接続助詞		
なんて	ナンテ	ナンテ	なんて	助詞-副助詞		
ない	ナイ	ナイ	ない	助動詞	助動詞-ナイ	連体形-一般
な	ナ	ダ	だ	助動詞	助動詞-ダ	連体形-一般
我々	ワレワレ	ワレワレ	我々	代名詞		
はじまっ	ハジマツ	ハジマル	はじまっ	動詞-一般	五段-ウ行	連用形-促音便
い	イル	イル	い	動詞-非自立可能 上一段-ア行	未然形-一般	
しまっ	シマウ	シマウ	しまっ	動詞-非自立可能 五段-ウア行	終止形-一般	
まさか	マサカ	マサカ	まさか	助詞		
「			「	補助記号-括弧前		
。			。	補助記号-句点		
と	コト	コト	と	名詞-普通名詞-一般		
とき	トキ	トキ	とき	名詞-普通名詞-副詞可能		

色々	色々	形状詞-一般
な	だ	助動詞
こと	事	名詞-普通名詞-一般
が	が	助詞-格助詞
はじまっ	はじまっ	動詞-一般
て	て	助詞-接続助詞
しまっ	しまっ	動詞-非自立可能
なんて	なんて	助詞-副助詞
。	。	補助記号-句点

が	が	助詞-格助詞
て	て	助詞-接続助詞
なんて	なんて	助詞-副助詞
ない	ない	助動詞
な	だ	助動詞
我々	我々	代名詞
はじまっ	はじまっ	動詞-一般
い	い	動詞-非自立可能
しまっ	しまっ	動詞-非自立可能

5. 実験方法

表5 集計されたデータ

題名	章	単語数	連体詞	名詞	固有名詞	補助記号	副詞	動詞	代名詞	接尾辞	接頭辞	接続詞	助動詞	助詞	形容詞	形状詞	感動詞	
銃とチョコレート	prologue	2101	15	464	12	279	35	273	44	41	5	2	248	606	29	19	2	
	全体	13582	94	2652	239	1658	169	1944	406	237	55	18	1573	4043	182	101	18	
	一章	1	3414	17	678	44	430	55	497	89	63	18	8	374	996	53	29	3
		2	2337	14	442	62	277	24	336	75	40	13	5	281	696	19	10	4
		3	3544	28	747	43	406	44	492	83	50	11	2	429	1103	46	20	5
		4	4287	35	785	90	545	46	619	159	84	13	3	489	1248	64	42	6
		全体	17157	130	3188	368	2274	196	2426	549	279	51	28	2003	5086	216	139	27
	二章	1	4411	32	806	110	610	54	599	143	66	15	9	523	1298	55	34	12
		2	4018	24	716	90	538	50	572	135	80	17	3	468	1206	42	27	6
		3	4958	40	1025	107	628	47	707	117	91	9	11	560	1468	57	34	5
		4	3770	34	641	61	498	45	548	154	42	10	5	452	1114	62	44	4
		全体	19066	173	3454	354	2592	233	2742	619	260	44	38	2240	5587	275	147	30
	三章	1	3702	40	600	113	472	44	526	144	56	3	12	454	1097	49	33	4
		2	4150	28	746	76	562	53	606	136	47	10	9	493	1231	65	33	7
		3	5646	63	1117	68	779	64	795	169	67	14	10	646	1648	73	42	6
		4	5568	42	991	97	779	72	815	170	90	17	7	647	1611	88	39	13
		全体	16967	120	3146	405	2305	229	2444	483	231	33	29	1957	4892	254	186	38
	四章	1	3456	33	661	66	485	36	476	101	68	9	8	376	998	58	34	8
		2	3689	27	677	91	484	57	514	117	58	7	6	432	1046	62	47	12
		3	4685	29	826	168	612	61	666	125	54	7	5	559	1387	62	51	10
		4	5137	31	982	80	724	75	788	140	51	10	10	590	1461	72	54	8
		全体	16967	120	3146	405	2305	229	2444	483	231	33	29	1957	4892	254	186	38
	epilogue	6834	63	1294	127	804	120	913	247	155	24	15	819	2022	101	56	5	
	合計	75707	595	14198	1505	9912	982	10742	2348	1203	212	130	8840	22236	1057	648	120	

5. 実験方法

表5 集計されたデータ

題名	章	単語数	連体詞	名詞	固有名詞	補助記号	副詞	動詞	代名詞	接尾辞	接頭辞	接続詞	助動詞	助詞	形容詞	形状詞	感動詞	
銃とチョコレート	prologue	2101	15	464	12	279	35	273	44	41	5	2	248	606	29	19	2	
	全体	13582	94	2652	239	1658	169	1944	406	237	55	18	1573	4043	182	101	18	
	一章	1	3414	17	678	44	430	55	497	89	63	18	8	374	996	53	29	3
	2	2337	14	442	62	277	24	336	75	40	13	5	281	696	19	10	4	
	3	3544	28	747	43	406	44	492	83	50	11	2	429	1103	46	20	5	
	4	4287	35	785	90	545	46	619	159	84	13	3	489	1248	64	42	6	
	全体	17157	130	3188	368	2274	196	2426	549	279	51	28	2003	5086	216	139	27	
	二章	1	4411	32	806	110	610	54	599	143	66	15	9	523	1298	55	34	12
	2	4018	24	716	90	538	50	572	135	80	17	3	468	1206	42	27	6	
	3	4958	40	1025	107	628	47	707	117	91	9	11	560	1468	57	34	5	
	4	3770	34	641	61	498	45	548	154	42	10	5	452	1114	62	44	4	
	全体	19066	173	3454	354	2592	233	2742	619	260	44	38	2240	5587	275	147	30	
	三章	1	3702	40	600	113	472	44	526	144	56	3	12	454	1097	49	33	4
	2	4150	28	746	76	562	53	606	136	47	10	9	493	1231	65	33	7	
3	5646	63	1117	68	779	64	795	169	67	14	10	646	1648	73	42	6		
4	5568	42	991	97	779	72	815	170	90	17	7	647	1611	88	39	13		
全体	16967	120	3146	405	2305	229	2444	483	231	33	29	1957	4892	254	186	38		
四章	1	3456	33	661	66	485	36	476	101	68	9	8	376	998	58	34	8	
2	3689	27	677	91	484	57	514	117	58	7	6	432	1046	62	47	12		
3	4685	29	826	168	612	61	666	125	54	7	5	559	1387	62	51	10		
4	5137	31	982	80	724	75	788	140	51	10	10	590	1461	72	54	8		
epilogue	6834	63	1294	127	804	120	913	247	155	24	15	819	2022	101	56	5		
合計	75707	595	14198	1505	9912	982	10742	2348	1203	212	130	8840	22236	1057	648	120		

5. 実験方法

表5 集計されたデータ

題名	章	単語数	連体詞	名詞	固有名詞	補助記号	副詞	動詞	
	prologue	2101	15	464	12	279	35	273	
	一章	全体	13582	94	2652	239	1658	169	1944
		1	3414	17	678	44	430	55	497
		2	2337	14	442	62	277	24	336
		3	3544	28	747	43	406	44	492
		4	4287	35	785	90	545	46	619
	二章	全体	17157	130	3188	368	2274	196	2426
		1	4411	32	806	110	610	54	599
		2	4018	24	716	90	538	50	572
		3	4958	40	1025	107	628	47	707
		4	3770	34	641	61	498	45	548

5. 実験方法

2. 有意差があるか χ 二乗検定と残差分析を用いて検定する

2-1 作品ごとに全単語あたりの

各品詞の割合を計算
↓

同ジャンル内の作品同士で
各品詞の割合を平均する

< χ 二乗検定 >

ある不規則な事象の出現頻度が偶然なのかどうか検定するもの

< 残差分析 >

単純な差ではなく、母数を考慮して相対的に出現頻度の差を分析したもの

5. 実験方法

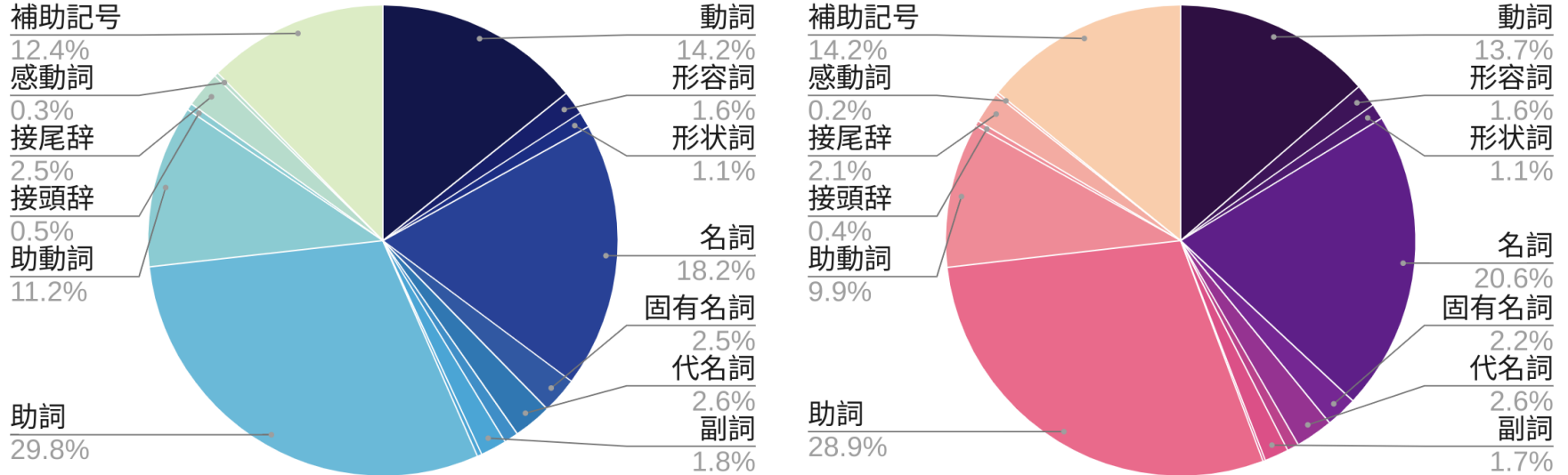


図2 各品詞の割合の平均(左:ミステリー小説 右:恋愛小説)

5. 実験方法

2-2 js-STAR_XRを用いて

χ^2 乗検定及び

残差分析を行う

①前項で算出した割合を

1万倍にし,1万語あたりに
出現する数に換算する

②小数第1位を四捨五入し,

これを「実測値」として検定を行う。

表7 割合の平均から実測値の算出

種類	ジャンル	動詞	形容詞	形状詞	名詞	固有名詞
割合の平均 (円グラフの値)	ミステリー	0.1423768	0.0164264	0.0107684	0.1823152	0.0249146
	恋愛	0.1366978	0.0156500	0.0106984	0.2060481	0.0217346

①

種類	ジャンル	動詞	形容詞	形状詞	名詞	固有名詞
割合の平均 ×10000	ミステリー	1423.76813	164.26381	107.68428	1823.15155	249.14555
	恋愛	1366.97849	156.49988	106.98353	2060.48082	217.34589

②

種類	ジャンル	動詞	形容詞	形状詞	名詞	固有名詞
上の値を 四捨五入	ミステリー	1424	164	108	1823	249
	恋愛	1367	156	107	2060	217

6. 実験結果 - χ 二乗検定 -

表8 χ 二乗検定を行った結果

	ミステリー小説		恋愛小説	
	実測値	残差分析の結果	実測値	残差分析の結果
動詞	1424		1367	
形容詞	164		156	
形状詞	108		107	
名詞	1823	▽▽	2060	▲▲
固有名詞	249		217	
代名詞	263		259	
連体詞	97		81	
副詞	181		166	
接続詞	31	▲	16	▽
助詞	2981		2888	
助動詞	1118	▲▲	993	▽▽
接頭辞	46		37	
接尾辞	247		214	
感動詞	27		20	
補助記号	1240	▽▽	1418	▲▲

(▲有意に多い、▽有意に少ない)p<.05)

(▲▲有意に多い、▽▽有意に少ない)p<.01)

ミステリー小説

接続詞,助動詞が有意に多い

恋愛小説

名詞,補助記号が有意に多い



文の印象を左右する

品詞である

6. 実験結果 -残差分析-

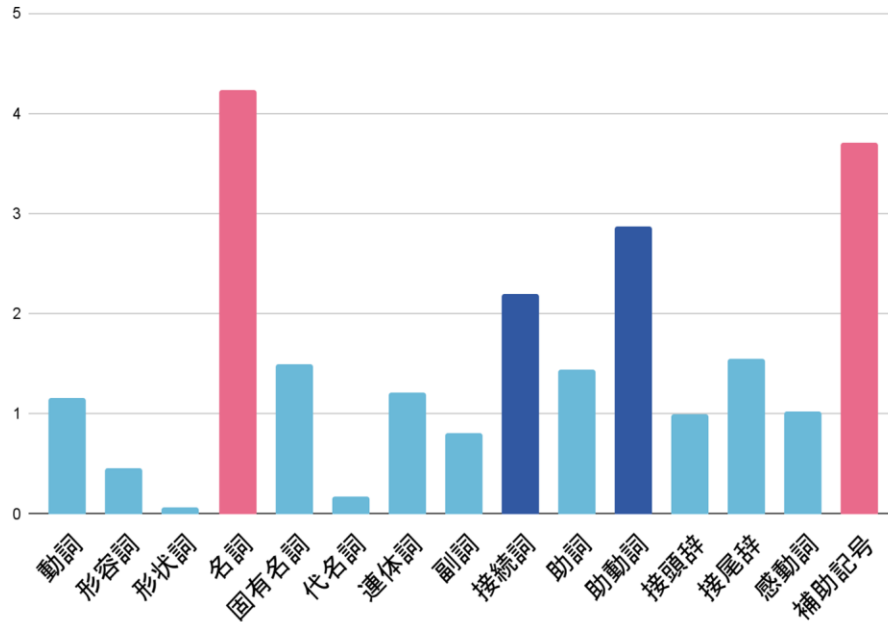


図3 残差分析を行った結果
(青:ミステリー小説 赤:恋愛小説)

○恋愛小説により多く出現する品詞が「名詞」「補助記号」のみだった

○修飾語の役割を持つ「形容詞」「形状詞」「副詞」は出現頻度の差が小さかった

7. 考察 - χ^2 二乗検定-

○ ミステリー小説 … 接続詞・助動詞が有意に多い

⇒ 文の順序や文法に重点を置いた構成である

○ 恋愛小説 … 名詞・補助記号(!,?,…など)が有意に多い

⇒ 意味を伝えることに重点を置いた構成である

7. 考察 - χ 二乗検定-

○小説内の会話には,口頭での会話と比べて間が少ない

⇒特に少なかった「感動詞-フィラー」(言い淀み)が影響し,感動詞には有意な差が出なかった。

○小説内の文章では,主語や目的語などの助詞を伴う表現を省略することが少ない

⇒助詞には有意な差が出なかった。

7. 考察 -残差分析-

○ミステリー小説は文法を意識した文章のため,様々な品詞が満遍なく使われている

⇒恋愛小説に多かった品詞がより強調された

○修飾語はジャンルを問わず文を構成する際に必要になる

⇒形容詞,形状詞,副詞は出現頻度に差が出なかった

8. 結論

○品詞の傾向は,文の印象を左右している

○名詞などを多用し,意味を伝えることを重視した場合

⇒ ~~「柔らかい文章」~~ になる

○接続詞などを多用し,文の構成を重視した場合

⇒ ~~「硬い文章」~~ になる

9. 展望

- サンプル数や文章の種類を増やし,分析の精度を高める
- 品詞以外の要素にも注目した研究を行う
- 生成AIを用いて品詞の傾向を調整した文章を作成し,
実際に印象が変化するか確かめる

10. 参考文献

- [1] 「Quiita:mecab + NEologd + python3 で形態素解析」
<https://qiita.com/sudo5in5k/items/f89d9dc1bec1ed221ede>
- [2] 「品詞分析から見る夏目漱石の前期作品の文体の特異性」
杉浦清人,情報処理学会研究報告(2018)Vol.2018-CH-117,No.12,1-7
- [3] 「ボーカロイド楽曲の人気要因に関する研究—歌詞の品詞分析によるJ-POP楽曲との比較—」
中井悠加・上村愛結,島根県立大学収録誌『人間と文化』(2022),巻5,pp.77-86
- [4] 「RNN言語モデルを用いた日本語形態素解析の実用化」
森田一・黒橋禎夫,情報処理学会第78回全国大会講演論文集(2016),pp.13-14
- [5] 「『あ』系感動詞における語の認定について」
姚瑶,早稲田大学大学院文学研究科紀要(2021),pp.209-220
- [6] 「日本語学習者の習熟度別に見たフィラーの分析」
小西円,国立国語研究所論集(2020),pp.91-105
- [7] 「福岡市方言における形容詞型活用の諸相について：10代・20代でのヒマ(暇)イ・スルカローの使用実態」
富田あかね,首都大学東京言語研究会『言語の研究』(2017),pp.1-30
- [8]js-STAR_XR
<https://www.kisnet.or.jp/nappa/software/star/index.htm>