

語彙指標を用いた流行歌の歌詞の通時的分析

小林雄一郎
東洋大学

天笠美咲
東洋大学

鈴木崇史
東洋大学

本研究の目的は、計量文体論の技法を用いて、日本の流行歌の時系列変化を明らかにすることである。具体的には、1977年から2012年までに発表された773曲の歌詞における26種類の語彙指標（品詞、語種、文字種、語彙レベル）を比較した。その結果、1990年頃を境に、語種と文字種の頻度が大きく変化していることが分かった。特に、外来語とカタカナの頻度が著しく減少し、漢語と漢字の頻度が増加した。本研究は、計量文献学に新たな知見をもたらす、日本の現代文化を対象とする社会学研究に客観的な資料を与えるものである。

A Chronological Variation of Lexical Indices in the Lyrics of Popular Songs

Yuichiro Kobayashi
Toyo University

Misaki Amagasa
Toyo University

Takafumi Suzuki
Toyo University

The purpose of this study is to investigate the chronological variation of Japanese popular songs using stylometric techniques. The researchers compared the lyrics of 773 songs released between 1977 and 2012 to examine the use of 26 different lexical indices, such as part-of-speech, word type, character type, and vocabulary level. The results suggest that the frequencies of word types and character types greatly changed around 1990. The frequencies of foreign vocabulary and *katakana* characters fell sharply and those of Sino-Japanese words and Chinese characters rose in the lyrics of popular songs. The findings of this study provide an important case study for computational stylistics and can also be useful for understanding cultural trends in Japan.

1. はじめに

流行歌は、その楽曲を生み出した時代や社会の表象である。例えば、オリコン年間ランキング上位にランクインする曲は、多くの人々に聞かれている曲であり、数多くの人々の共感を得ている曲であると言える。無論、個々の歌手や楽曲によって、音楽のテーマやスタイルは異なるであろう。しかしながら、ポピュラー作品の表現は、個別的な関心によって一見異なる形式を取りながらも、その時代を生きた人々の価値観を反映し、何らかの時代性を帯びているものである(南田, 2010)。

また、楽曲において、歌詞は重要な意味を持つ。歌詞は、書き手の内面を窺い知る手がかりとなり、聞き手が自分を映す鏡としても機能している(見崎, 2002)。従って、歌詞を分析することは、その時代の背景や人々の心理を理解する上で重要な意義を持つ。

本研究の目的は、計量文体論の技法を用いて、1977年から2012年までの36年間における流行歌を通時的に分析し、歌詞における言語使用の時系列変化を明らかにすることである。

2. 関連研究

歌謡曲や流行歌の歌詞を題材として、特定の時代における文化や社会思想を論じた研究は多い。例えば、見田(1968)は、明治元年から昭和38年までの451曲を対象とし、「怒り」、「喜び」、「孤独」、「あこがれ」などのモチーフ因子を分析し、戦前から戦後にかけての日本人の心情の変化を追った。また、久保(1995)は、1965年から1989年までの間に発表された302曲の流行歌の歌詞をデータベース化し、1960年代に見られた「若者モノ」が1975年以降に減少していく一方で、「恋愛モノ」が台頭していく過程などを示した。

そして、計量文体論的な手法による歌詞分析の古典的研究としては、水谷(1982)が挙げられる。この研究は、数量化 III 類などの統計的手法を援用し、昭和初期の歌謡曲における語彙使用の類似度を明らかにした。また、伊藤(1997-2000)による一連の研究は、松任谷由美という特定の女性歌手の歌詞を言語学的に分析したものである。さらに、細谷・鈴木(2010)や小林・狩野・鈴木(2013)

は、主成分分析やランダムフォレストなどの手法を用いて、複数の女性歌手の歌詞を分析している。

3. 分析の手順

3.1. 分析データ

本研究の分析データは、1977年から2012年までのオリコン年間チャートのシングルトップ20位にランクインした楽曲である。¹ データの作成にあたり、両A面シングルの場合は、2曲とも分析対象とした。また、複数年にわたって同じ曲がランクインしている場合は、最初にランクインした年のデータとした。なお、2013年10月の時点において、「歌ネット」² および「うたまっぶ」³ で歌詞が入手できなかった14曲は、分析の対象外とした。その結果、分析対象は合計773曲となった。表1は、分析データの概要をまとめたものである。

表1. 分析データの概要

総文字数	564922
総語数	306033
総文数	6713

註: MeCab 0.996⁴ と UniDic 2.2.0⁵ の解析結果に基づく

3.2. 分析対象とする言語項目

これまでの言語学的な歌詞分析においては、使用語彙が主な関心の的となってきた。全ての出現語彙を対象とする場合(小林・狩野・鈴木, 2013)もあれば、高頻度語(金城, 2013; 細谷・鈴木, 2010)、名詞(茅根, 2002)、動詞(茅根, 2002)、人称代名詞(茅根, 2002; 藤掛・西村・菅沼・田賀・澤柳, 1994; 藤川, 1999)、文末表現(鈴木・山口, 2000)、コロケーション(塚本, 2014)、漢字とルビの組み合わせ(鈴木・山口, 2000)、英語表現(藤掛・西村・菅沼・田賀・澤柳, 1994)といった特定の表現を対象とする場合もある。しかし、いずれの場合においても、語彙の表層形のみが分析されることが多かった。

これに対して、本研究では、品詞、語種、文字種、語彙レベルから構成される26種類の語彙指標を用いた分析を行う(表2)。これらの指標の頻度は、日本語文章難易度判別システム

jReadability⁶によって算出される。本研究で用いる語彙指標は、日本語教育での活用を想定したものであるが(李・柴崎, 2012)、計量文体分析にも応用できる。日本語を対象とする計量分析において、これまでも品詞構成率や語種の有効性は古くから知られてきた(樺島・寿岳, 1965; 安本, 1958; 安本, 1965)。

歌詞は、書き言葉における他のジャンルと比べて、総語数が極めて少ないために、テキストの内容による影響を受けやすい。しかし、品詞、語種、文字種、語彙レベルなどの語彙指標を用いることで、比較的短いテキストからも安定した頻度情報を得ることができる。そして、文章の内容による影響を軽減し、文体の違いを明らかにすることが可能になる(Tabata, 2002)。さらに、英語圏では、様々な語彙指標に基づく計量文体分析の方法論も模索されている(McCarthy, Lewis, Dufty, and McNamara, 2006)。

表2. 本研究で用いられる語彙指標

品詞	感動詞 形状詞 形容詞 助詞 助動詞 接続詞 代名詞 動詞 副詞 固有名詞 普通名詞 連体詞 その他の品詞
語種	和語 漢語 外来語 混種語
文字種	ひらがな カタカナ 漢字
語彙レベル	初級前半語 初級後半語 中級前半語 中級後半語 上級前半語 上級後半語

3.3. 分析手法

本研究で主に用いる分析手法は、重回帰分析である。この手法の目的は、複数の説明変数を用いて、特定の変数(目的変数)を予測、もしくは説明することである(豊田, 2012)。本研究では、

¹ 本研究のデータは、天笠(2014)において、楽曲のジャンルによる歌詞の違いを分析するために構築されたものである。

² <http://www.uta-net.com/>

³ <http://www.utamap.com/>

⁴ <http://taku910.github.io/mecab/>

⁵ <http://osdn.jp/projects/unidic/>

⁶ <http://jreadability.net/>

品詞、語種、文字種、語彙レベルから成る26種類の語彙指標を説明変数とし、楽曲が流行した年を目的変数とする。そして、重回帰分析を行うにあたっては、赤池情報量基準(AIC)に基づくステップワイズ法による変数選択を行う。

分析の手順としては、予備的分析として、変数間の相関関係を初めに調べる。次に、重回帰分析を用いて、26種類の語彙指標と年代の関係をモデル化する。そして、最後に重回帰分析による楽曲が流行した年の予測実験を行う。

回帰分析を用いてテキストの執筆年代を分析した研究としては、Tsukamoto (2004)を挙げることができる。この研究は、16世紀から20世紀までの英語の散文の執筆年が *wh* 限定詞の頻度、*been* の頻度、形容詞の頻度から予測できることを明らかにした。また、金(2009)は、芥川龍之介の作品の執筆年が格助詞や接続助詞などの頻度と強く関連としていることを示した。

4. 結果と考察

4.1. 相関分析

まず、Pearsonの積率相関係数を用いて、目的変数を含む全ての変数間の関係を調査した。表3は、相関係数の高い2変数の組み合わせ(上位10位)を示している。この表によれば、語彙指標間で最も相関の高い組み合わせは、外来語とカタカナ($r=0.90$)であり、動詞と助詞($r=0.84$)、

代名詞と助詞($r=0.80$)と続く。また、楽曲が流行した年と高い相関を持つ語彙指標は、漢語($r=0.79$)や漢字($r=0.72$)であった。

表3. 変数間の相関関係(上位10位)

順位	変数1	変数2	相関係数
1	外来語	カタカナ	0.90
2	動詞	助詞	0.84
3	代名詞	助詞	0.80
4	漢語	年	0.79
5	助詞	形容詞	0.73
6	漢字	年	0.72
7	和語	ひらがな	0.70
8	動詞	助動詞	0.69
9	助動詞	助詞	0.68
10	カタカナ	固有名詞	0.65

4.2. 回帰分析によるデータの記述

次に、重回帰分析を用いて、歌詞における語彙指標の使用率の時系列変化をモデル化した。前節の相関分析の結果、極めて相関の高い説明変数の組み合わせが見られたため、多重共線性の可能性を考慮し、AICに基づくステップワイズ法による変数選択を行った。その結果、形状詞、助動詞、接続詞、代名詞、動詞、固有名詞、連体詞、初級前半語、初級後半語、中級後半語、カタカナ、和語、漢語、外来語の14項目が選択された(表4)。

表4. 変数選択を行った重回帰分析の結果(1)

	Estimate	Standard Error	t value	p (> t)	Significance
(Intercept)	1073.83	208.6	5.15	0.00	***
形状詞	-10.01	3.18	-3.14	0.00	**
助動詞	3.83	0.92	4.15	0.00	***
接続詞	-40.83	14.42	-2.83	0.01	**
代名詞	-2.77	1.5	-1.84	0.08	
動詞	-3.05	0.89	-3.43	0.00	**
固有名詞	10.82	5.32	2.04	0.05	
連体詞	21.81	4.37	4.99	0.00	***
初級前半語	-1.71	0.55	-3.09	0.01	**
初級後半語	1.84	0.62	2.98	0.01	**
中級後半語	1.02	0.79	1.29	0.21	
カタカナ	-1.68	0.63	-2.66	0.01	*
和語	8.77	2.1	4.19	0.00	***
漢語	14.27	2.39	5.98	0.00	***
外来語	13.12	2.65	4.95	0.00	***

註: 表中のEstimateは回帰係数, Standard Errorは係数の標準誤差, t valueは「回帰係数が0である」という帰無仮説に基づく検定の結果, $p(>|t|)$ はt値から求めた有意確率, Significanceは有意水準($***=p<0.001$, $**=p<0.01$, $*=p<0.05$)をそれぞれ表す。また, Interceptは回帰式の切片を表す。

決定係数は 0.94, 調整済み決定係数は 0.89 であり, 「全ての回帰係数が 0 である」という帰無仮説は, 0.1%水準で棄却された ($F=21.98$)。そして, 回帰係数の p 値を見ると, 助動詞, 連体詞, 和語, 漢語, 外来語の使用率が楽曲の流行した年代と強い関係を持っていることが分かった。

4.3. 回帰分析によるデータの予測

さらに, 語彙指標に基づく重回帰分析を用いて, 楽曲が流行した年の推定実験を行った。実験にあ

たっては, 無作為に抽出した分析データの 3 分の 2 を訓練データとし, 残りの 3 分の 1 を評価データとした。

ステップワイズ法に基づく変数選択の結果, 感動詞, 形状詞, 助詞, 助動詞, 接続詞, 代名詞, 副詞, 固有名詞, 連体詞, 初級前半語, 中級前半語, 上級前半語, ひらがな, 和語, 漢語, 外来語の 16 項目が選ばれた (表 5)。

表 5. 変数選択を行った重回帰分析の結果 (2)

	Estimate	Standard Error	t value	p ($> t $)	Significance
(Intercept)	1134.40	268.35	4.23	0.00	**
感動詞	5.03	1.91	2.63	0.03	*
形状詞	-6.87	3.50	-1.96	0.09	
助詞	-1.12	0.67	-1.66	0.14	
助動詞	3.17	1.48	2.15	0.07	
接続詞	-59.66	16.50	-3.62	0.01	**
代名詞	1.67	1.98	0.85	0.43	
動詞	-2.74	2.34	-1.17	0.28	
固有名詞	5.13	5.00	1.03	0.34	
連体詞	14.31	6.12	2.34	0.05	
初級前半語	-2.72	0.63	-4.35	0.00	**
初級後半語	-2.35	0.87	-2.72	0.03	*
中級後半語	-1.53	0.98	-1.56	0.16	
ひらがな	1.72	0.76	2.28	0.06	
和語	8.25	2.55	3.23	0.01	*
漢語	15.85	2.81	5.65	0.00	***
外来語	13.07	2.96	4.41	0.00	**

註: 表中の Estimate は回帰係数, Standard Error は係数の標準誤差, t value は「回帰係数が 0 である」という帰無仮説に基づく検定の結果, p ($>|t|$) は t 値から求めた有意確率, Significance は有意水準 ($***=p<0.001$, $**=p<0.01$, $*=p<0.05$) をそれぞれ表す。また, Intercept は回帰式の切片を表す。

決定係数は 0.98, 調整済み決定係数は 0.94 であり, F 値は 24.59 であった。また, 楽曲が実際に流行した年 (実測値) と回帰式によって予測された年 (推定値) の差の平均は, 6.25 であった。

4.4. 語彙指標の分析

重回帰分析の結果 (表 4~5), 和語, 漢語, 外来語という語種の使用率が楽曲の流行した年と強く関連していることが分かった。そこで本節では, 語彙指標の時系列変化を詳しく見ていく。

まず, 図 1 は, 1977 年から 2012 年までの漢語の使用率の変化を視覚化したものである。図の横軸は年を表し, 縦軸は使用率を表している。この

図を見ると, 年代が進むにつれて, 漢語の使用率が徐々に増加している傾向が分かる。

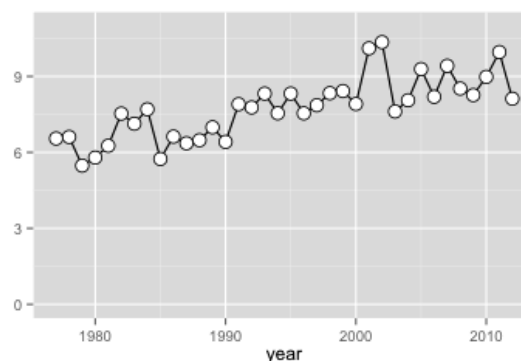


図 1. 漢語の使用率の変化

漢語の増加とともに使用率が減少している語種は、外来語である(図2)。この図を見ると、2010年代の使用率にわずかな増加が見られるものの、1990年あたりを境に、外来語の使用率が顕著に下がっていることが分かる。

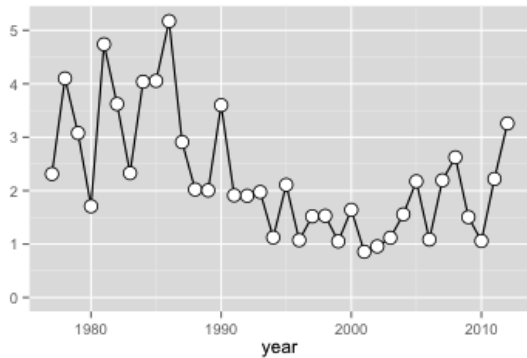


図2. 外来語の使用率の変化

次に、文字種に注目すると、漢字の使用率が増加し、カタカナの使用率が減少している(図3~4)。

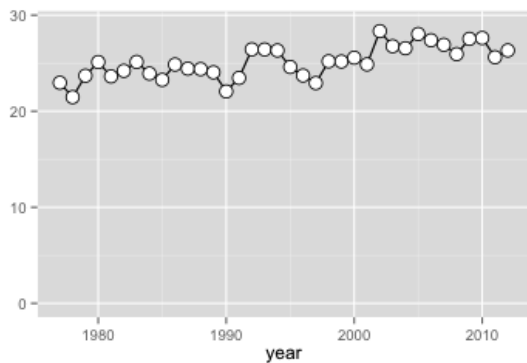


図3. 漢字の使用率の変化

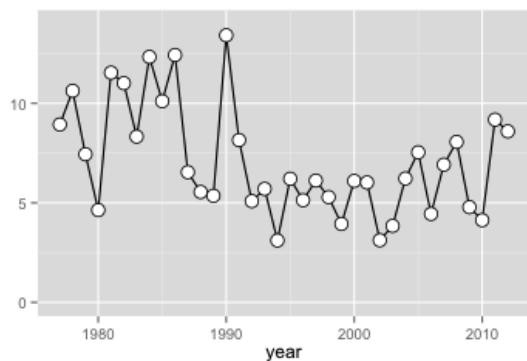


図4. カタカナの使用率の変化

これらの図からも分かるように、漢語の増加と漢字の増加、外来語の減少とカタカナの減少には、それぞれ密接な関係がある。そして、1980年代までと1990年代以降では、語種と文字種の使用率に明らかな違いが見られる。

伊藤(2014)は、使用言語モデルや歌風モデルの観点から、1970年、1980年、1990年、2000年、2010年という5つの年に流行した曲の歌詞を比較した。その結果、1970年代から1980年代は、洋風の雰囲気を出すために外国語や外来語の使用が多かったが、それが飽きられるにつれて、1990年代以降は和語と漢語を含む歌詞が増えたと指摘している。この「J-POP 日本語回帰説」は、本研究の分析結果ともおおむね一致する。

実際、1980年代前半から1990年頃までは、カタカナや外来語を含む歌詞が目立つ。例えば、1985年に年間1位となった「ジュリアに傷心」(チェッカーズ)における「キャンドル・ライトが／ガラスのピアスに反射けて滲む／お前彼の腕の中踊る／傷心(ハートブレイク) Saturday Night／悲しいキャロルがショーウィンドウで／銀の雪に変わったよ」や、1986年に年間1位となった「CHA-CHA-CHA」(石井明美)における「街で噂の 辛くち セクシー・ギャル／甘い誘いはねつける スパイシー・ギャル／花の金曜日(ウィークエンド) 匂いもファンキー・ナイト」などである。

一方、近年のヒット曲には、タイトルが英語表記であるにもかかわらず、歌詞に外来語や外国語を殆ど含まないものが多い。その例としては、2009年に年間1位となった「Believe」(嵐)や、2010年に年間1位となった「Beginner」(AKB48)などを挙げることができる。同様に、1999年から2005年までの間に、浜崎あゆみが年間チャートのトップ20位以内にランクインさせた17曲¹のタイトルは全て英語表記であるが、歌詞における外来語や外国語の比率は意外と小さく、古風な作風である(見崎, 2002)。

また、品詞に注目すると、年代が進むにつれて、連体詞や副詞が増加し、固有名詞や普通名詞が減少している。²そして、語彙レベルに注目すると、初級後半語が増加し、上級後半語が減少している。³紙面の都合上、これらの項目の分析に関しては、稿を改める。

¹ そのうちの1曲(2002年の「a song is born」)は、KEIKOとのデュエットである。また、両A面シングルは、1曲としてカウントした。

² 楽曲が流行した年と連体詞、副詞、固有名詞、普通名詞の相関係数は、それぞれ0.60, 0.40, -0.43, -0.63である。

³ 楽曲が流行した年と初級後半語、上級後半語の相関係数は、それぞれ0.40, -0.57である。

5. おわりに

本研究は、1977年から2012年までの36年間に発表された773曲の歌詞における26種類の語彙指標の使用率を比較し、言語使用の時系列変化を分析した。その結果、1990年頃を境に、語種と文字種の頻度が大きく変化していることが分かった。特に、外来語とカタカナの頻度の減少と、漢語と漢字の頻度の増加が明らかにされた。また、品詞や語彙レベルに関しては、連体詞、副詞、初級後半語の増加、固有名詞、普通名詞、上級後半語の減少が見られた。本研究の知見は、体系的な歌詞データと計量文献学の技法に基づくものであり、日本の現代文化を対象とする社会学研究に客観的な資料を与えるものである。

今後の展開としては、ランダムフォレストやLASSOなどの手法を用いることで、より精度の高いデータの記述と予測を実現し、より有益な知見を導き出すことが挙げられる。また、楽曲のジャンルや歌手のジェンダーによっても歌詞に違いが見られるため(北川, 1999)、各年代におけるジャンルやジェンダーの割合を調査することが求められる。さらに、個々の楽曲の歌詞に関する質的分析、そして、その楽曲が作られた背景などを仔細に検討することも必要である。

謝辞

本研究の成果の一部は、科学研究費補助金(若手研究B)「統計的データ解析と被験者実験を用いたテキスト・コミュニケーション研究」(代表:鈴木崇史)によるものである。

参考文献

- 天笠美咲 (2014). 『ヒットソングの歌詞のテキスト分析』 2013年度(平成25年度)東洋大学社会学部第1部メディアコミュニケーション学科卒業論文.
- 伊藤雅光 (1997-2000). 「ユーミンの言語学(1)-(46)」 『日本語学』 16(4)-20(8) 連載.
- 伊藤雅光 (2014). 「J-POPの歌詞に見られる日本語回帰現象について」 『日本語学』 33(15), 48-61.
- 樺島忠夫・寿岳章子 (1965). 『文体の科学』 京都: 綜芸社.
- 茅根滋 (2002). 「経済的観点から見る日米のヒット曲の歌詞構造とその法則性」 『金沢大学経済学会』 21, 83-96.
- 北川純子 (1999). 「『日本のポピュラー音楽とジェンダー』への展望」 北川純子(編) 『鳴り響く性—日本のポピュラー音楽とジェンダー』 (pp. 1-30). 東京: 勁草書房.
- 金明哲 (2009). 「文章の執筆時期の推定—芥川龍之介の作品を例として」 『行動計量学』 36(2), 89-103.
- 金城克哉 (2013). 「槇原敬之の歌詞の数量的分析—『君が笑うとき君の胸が痛まないように』から『Heart to Heart』まで」 『琉球大学欧米文化論集』 57, 23-42.
- 久保正敏 (1995). 「ニューミュージックに見る恋愛風景」 『情報処理学会研究報告』 CH-25, 49-57.
- 小林佳織・狩野恵里奈・鈴木崇史 (2013). 「女性グループの歌詞の計量テキスト分析」 『言語処理学会第19回年次大会発表論文集』 (pp. 338-341). 東京: 言語処理学会.
- 鈴木直枝・山口孝志 (2000). 「流行歌の歌詞にみる言語の変遷—過去34年間のヒット曲を通して」 『東北生活文化大学三島学園女子短期大学紀要』 31, 55-65.
- 塚本泰造 (2014). 「現代語『天使』のコロケーション小考—歌詞の中で『天使』はどのように振る舞うか」 『国語国文学研究』 49, 331-341.
- 豊田秀樹(編) (2012). 『回帰分析入門—Rで学ぶ最新データ解析』 東京: 東京図書.
- 藤掛和美・西村絵里・菅沼真琴・田賀小百合・澤柳千秋 (1994). 「現代流行歌歌詞の移り変わり—1970年代と1992年ヒット曲の比較から」 *CUWC gazette*, 49-60.
- 藤川大祐 (1999). 「ヒット曲の変化と子どもたちの状況」 『金城学院大学論集 人文科学編』 32, 119-132.
- 細谷舞・鈴木崇史 (2010). 「女性シンガーソングライターの歌詞の探索的分析」 『じんもんこん2010—人文科学とコンピュータシンポジウム論文集』 (pp. 195-202). 東京: 情報処理学会.
- 見崎鉄 (2002). 『Jポップの日本語—歌詞論』 東京: 彩流社.
- 水谷静夫 (1982). 『数理言語学』 東京: 培風館.
- 見田宗介 (1968). 『近代日本の心情の歴史』 東京: 講談社.
- 南田勝也 (2010). 「文化資料分析法—『歌詞』から社会をみわたせるか?」 工藤保則・寺岡伸悟・宮垣元(編) 『質的調査の方法—都市・文化・メディアの感じ方』 (pp. 86-96). 東京: 法律文化社.
- 安本美典 (1958). 「文体統計による筆者推定—宇治十帖の作者について」 『心理学評論』 2, 147-156.
- 安本美典 (1965). 『文章心理学入門』 東京: 誠信書房.
- 李在鎬・柴崎秀子 (2012). 「文章の難易度と語彙の関連性に関する考察—学年の違いを特徴づける語彙的要素とは何か」 石田基広・金明哲(編) 『コーパスとテキストマイニング』 (pp. 181-192). 東京: 共立出版.

- McCarthy, P. M., Lewis, G. A., Dufty, D. F., & McNamara, D. S. (2006). Analyzing writing styles with Coh-Metrix. *Proceedings of the Florida Artificial Intelligence Research Society International Conference*, 764-769.
- Tabata, T. (2002). Investigating stylistic variation in Dickens through correspondence analysis of word-class distribution. In Saito, T., Nakamura, J., & Yamazaki, S. (Eds.), *English corpus linguistics in Japan* (pp. 165-182). Amsterdam: Rodopi.
- Tsukamoto, S. (2004). Quantifying diachronic change: Part of speech analysis from early modern to modern English. *English Corpus Studies*, 11, 19-36.

Appendix 1: 品詞構成率 (百分率)

年	品詞												
	感動詞	形状詞	形容詞	助詞	助動詞	接続詞	代名詞	動詞	副詞	固有名詞	普通名詞	連体詞	その他
1977	0.88	1.21	2.46	27.51	8.76	0.02	4.58	9.43	1.94	0.76	21.19	0.88	20.36
1978	0.73	1.33	2.47	29.30	8.71	0.13	4.94	9.12	2.02	0.73	22.54	0.67	17.29
1979	0.30	0.70	2.08	24.50	5.56	0.12	3.60	9.26	1.71	0.61	22.05	0.70	28.81
1980	0.44	1.48	2.79	29.20	7.56	0.09	4.14	10.35	1.82	0.21	23.73	1.26	16.95
1981	0.22	1.36	3.02	25.13	6.14	0.13	3.19	8.75	1.81	0.79	21.92	0.55	27.00
1982	0.37	1.27	2.88	25.82	7.18	0.18	3.45	8.77	1.83	0.55	22.05	0.58	25.06
1983	0.15	1.39	3.08	25.90	8.57	0.08	3.38	9.10	2.21	0.83	21.69	0.64	22.98
1984	0.69	1.40	2.23	24.87	7.85	0.09	2.73	8.78	1.95	0.82	24.14	0.52	23.95
1985	0.31	0.60	1.81	19.03	6.28	0.20	2.48	6.95	0.98	0.37	19.08	0.51	41.40
1986	0.30	1.08	1.57	19.04	5.62	0.13	2.18	6.69	1.27	0.29	20.80	0.63	40.42
1987	0.12	1.33	2.29	24.47	8.20	0.15	2.64	9.91	1.34	0.28	21.15	0.60	27.52
1988	0.38	1.18	2.48	23.74	8.41	0.08	3.23	9.43	1.60	0.42	20.41	0.78	27.86
1989	0.18	0.82	2.44	23.64	7.31	0.09	2.66	9.00	1.43	0.19	21.56	0.78	29.90
1990	0.60	0.88	1.73	20.68	6.35	0.02	2.13	7.63	1.76	0.47	18.85	0.65	38.24
1991	0.22	1.18	1.77	21.38	7.13	0.11	3.08	8.33	2.12	0.20	19.34	1.01	34.13
1992	0.13	1.43	2.34	25.24	8.80	0.11	4.03	9.35	2.14	0.09	19.94	0.75	25.65
1993	0.18	1.69	2.69	25.66	9.35	0.15	3.84	10.71	2.39	0.27	19.76	1.01	22.29
1994	0.13	1.28	2.47	23.49	8.77	0.15	3.54	8.47	1.58	0.15	16.99	0.86	32.13
1995	0.38	1.18	2.64	25.93	9.25	0.17	3.80	9.18	2.27	0.18	21.62	0.80	22.59
1996	0.22	1.11	2.45	21.75	7.59	0.10	3.05	7.93	1.58	0.30	17.62	0.77	35.53
1997	0.14	1.44	2.81	24.60	8.70	0.06	3.32	8.55	1.58	0.20	17.89	0.63	30.08
1998	0.22	1.44	2.86	25.97	9.89	0.09	3.28	10.97	1.68	0.14	20.49	0.90	22.08
1999	0.27	1.27	3.23	23.34	8.65	0.07	2.78	8.38	2.15	0.19	18.45	0.77	30.44
2000	1.37	1.44	1.76	18.77	7.13	0.09	2.48	6.37	1.76	0.19	17.14	0.85	40.63
2001	0.41	0.96	2.53	23.60	8.62	0.17	3.57	8.36	3.36	0.12	20.29	0.86	27.16
2002	0.31	0.78	1.72	18.15	5.54	0.16	2.80	6.89	2.09	0.26	16.91	1.12	43.28
2003	0.64	1.28	2.95	31.18	10.18	0.30	4.89	10.73	3.05	0.23	21.15	1.30	12.13
2004	0.29	1.62	2.96	24.83	9.06	0.11	4.52	9.76	2.42	0.18	21.29	1.03	21.92
2005	0.54	0.77	1.93	21.91	7.42	0.14	3.16	8.90	1.61	0.36	19.46	1.00	32.80
2006	0.21	1.19	2.30	27.03	9.50	0.08	4.24	10.80	1.96	0.25	19.30	1.03	22.12
2007	1.32	1.15	2.49	25.75	8.63	0.05	3.63	10.75	3.06	0.72	18.98	1.00	22.47
2008	0.44	1.11	2.00	23.52	7.90	0.10	3.71	8.90	2.20	0.63	19.76	1.06	28.68
2009	0.42	1.29	2.22	24.75	8.96	0.11	3.70	10.09	2.10	0.17	19.47	1.07	25.64
2010	0.15	1.02	1.37	18.10	5.98	0.09	2.90	6.89	1.72	0.11	17.72	0.78	43.17
2011	0.23	1.33	2.01	24.00	8.30	0.09	3.79	8.04	2.47	0.60	18.98	1.02	29.15
2012	0.18	1.19	2.05	21.14	8.34	0.07	2.95	7.37	2.23	0.22	18.87	1.00	34.38

Appendix 2: 語種構成率, 文字種構成率, 語彙レベル構成率 (百分率)

年	語種				文字種			語彙レベル					
	和語	漢語	外来語	混種語	ひらがな	カタカナ	漢字	初級前半	初級後半	中級前半	中級後半	上級前半	上級後半
1977	89.77	6.55	2.31	1.37	68.10	8.94	22.96	33.28	22.20	22.37	16.02	4.89	1.25
1978	88.40	6.61	4.10	0.89	67.91	10.63	21.47	30.85	22.64	21.97	16.71	6.54	1.29
1979	90.51	5.48	3.08	0.93	68.86	7.44	23.70	29.66	25.47	21.89	15.46	6.78	0.73
1980	90.73	5.80	1.71	1.77	70.25	4.64	25.12	28.53	26.24	21.92	17.47	5.09	0.76
1981	88.18	6.26	4.74	0.82	64.84	11.54	23.62	28.41	23.16	22.29	18.58	6.85	0.72
1982	87.54	7.53	3.62	1.30	64.78	11.01	24.20	27.57	23.01	23.65	18.28	6.73	0.77
1983	89.13	7.13	2.33	1.41	66.54	8.33	25.13	30.36	24.48	21.08	16.86	6.02	1.19
1984	86.43	7.70	4.04	1.83	63.75	12.33	23.92	29.52	24.24	21.83	17.24	6.44	0.72
1985	89.01	5.74	4.05	1.19	66.63	10.11	23.26	27.22	23.12	24.67	18.44	5.85	0.71
1986	86.38	6.62	5.17	1.82	62.71	12.43	24.86	26.23	21.76	23.65	20.67	6.57	1.13
1987	89.45	6.36	2.91	1.28	69.03	6.54	24.43	25.17	22.53	24.14	20.57	6.51	1.07
1988	89.97	6.48	2.02	1.53	70.07	5.54	24.38	24.84	22.33	26.31	17.90	7.22	1.40
1989	89.43	6.99	2.01	1.57	70.60	5.36	24.05	29.41	22.51	21.30	17.66	8.30	0.81
1990	88.92	6.42	3.60	1.06	64.52	13.42	22.06	27.12	23.20	24.41	18.48	6.26	0.52
1991	88.42	7.90	1.91	1.76	68.38	8.15	23.46	25.64	26.04	24.59	16.87	6.64	0.22
1992	88.62	7.77	1.90	1.70	68.47	5.09	26.44	27.15	27.91	22.44	17.32	4.63	0.55
1993	88.37	8.32	1.97	1.34	67.87	5.70	26.43	28.10	23.46	23.57	17.60	6.55	0.72
1994	89.79	7.55	1.12	1.54	70.57	3.11	26.32	27.18	26.06	23.06	17.91	5.48	0.30
1995	88.35	8.32	2.11	1.22	69.19	6.21	24.60	29.00	25.52	22.82	15.77	6.21	0.68
1996	90.18	7.54	1.07	1.20	71.15	5.13	23.72	28.35	25.40	22.82	16.56	6.26	0.60
1997	89.15	7.86	1.52	1.46	70.95	6.12	22.93	28.31	26.58	22.07	16.63	5.64	0.78
1998	88.69	8.34	1.53	1.44	69.52	5.29	25.19	25.87	25.57	23.13	17.76	7.04	0.63
1999	89.00	8.42	1.05	1.52	70.89	3.94	25.17	28.46	26.08	23.10	16.35	5.36	0.66
2000	89.00	7.91	1.64	1.45	68.33	6.10	25.57	27.89	24.83	23.48	18.08	5.04	0.67
2001	86.80	10.11	0.86	2.23	69.10	6.02	24.87	26.73	24.84	23.47	19.16	5.54	0.26
2002	87.58	10.35	0.96	1.11	68.54	3.12	28.33	32.94	23.84	21.46	15.94	5.61	0.22
2003	90.40	7.61	1.12	0.87	69.39	3.84	26.77	28.65	27.40	22.39	15.19	5.56	0.82
2004	88.99	8.06	1.56	1.40	67.22	6.22	26.56	24.48	27.41	23.83	17.89	5.65	0.73
2005	87.17	9.29	2.17	1.37	64.43	7.54	28.04	27.69	24.34	22.42	18.72	5.69	1.14
2006	89.66	8.20	1.09	1.06	68.17	4.44	27.39	26.03	24.70	22.61	20.56	5.63	0.47
2007	86.34	9.42	2.19	2.05	66.17	6.92	26.91	28.21	24.85	22.73	16.65	6.84	0.72
2008	87.56	8.52	2.62	1.30	66.00	8.06	25.94	29.58	23.52	24.42	17.71	4.32	0.45
2009	89.18	8.26	1.50	1.06	67.68	4.78	27.53	25.59	26.11	23.09	19.06	5.63	0.53
2010	88.61	8.98	1.06	1.35	68.25	4.13	27.62	25.87	25.03	24.00	19.33	5.40	0.37
2011	86.41	9.96	2.22	1.41	65.20	9.18	25.62	27.85	23.99	24.47	18.22	5.04	0.42
2012	87.28	8.12	3.26	1.34	65.08	8.59	26.33	28.02	25.10	22.27	18.89	5.33	0.38