法政大学学術機関リポジトリ
HOSEI UNIVERSITY REPOSITORY

PDF issue: 2025-07-01

「坊ちゃん」と「三四郎」の線形空間論による文体解析

IWASAKI, Harumi / 堀井, 清之 / 宮澤, 賢治 / 斎藤, 兆古 / 岩崎, 晴美 / HORII, Kiyoshi / MIYAZAWA, Kenji / SAITO, Yoshifuru

(出版者 / Publisher)
法政大学計算科学研究センター
(雑誌名 / Journal or Publication Title)
法政大学計算科学研究センター研究報告 / Bulletin of Computational Science Research Center, Hosei University
(巻 / Volume)
12
(開始ページ / Start Page)
151
(終了ページ / End Page)
157
(発行年 / Year)
1999-03-31
(URL)
https://doi.org/10.15002/00024843

「坊ちゃん」と「三四郎」の線形空間論による文体解析

岩崎　晴美
法政大学計算科学研究センター

斎藤　兆古
法政大学工学部電気電子工学科

宮澤　賢治、堀井　清之
白百合大学文学部

文学作品の文体解析は、その作品を解析する人の感性と知識・経験・考え方などによるところが大きい。文学作品をある視点に基づき計数化を行い、数量化したデータに線形空間論を適用した文体解析の手法を提案している[1]。文学作品の文体解析に数学的手法を用いて評価することと可視化することを試みている。ここでは夏目漱石の「坊ちゃん」と「三四郎」の作品について、話法という視点でとらえて分類した頻出度[2],[3]を基に文体解析を行った結果を考察する。

1．はじめに

文学作品の計数化による解析は、単語・語句の長さや使われている詞によって頻度や因子分析を行っている例が多く、それらは主に言語学や心理学的見地からのものであると思われる。文学作品の文章をどう読み取るかは読者の感性によるものであり、それらは数量化し難いものである。作家が書くという行為には何らかの作家自体の深層心理が作品自体に表出されていると考えられる。人間の潜在意識には快・不快などのアンビバレンス（ambivalence）があると心理学的に考えている。その ambivalence が作品の中に表出していると考えられる。作品中に表出された ambivalence な語・言葉あるいは文章から作品の特徴や作者の内面のゆれが読み取れるのではないかと思われる。文学作品をある視点のもとに数量化し、作品の流れを時空間として解析する。数量を可視化することによって作品のイメージを掴む。ここでは夏目漱石の「坊ちゃん」と「三四郎」の２つの作品について、その作品の構成、性質などを考察する。

2．話法について

文体構成をどのような視点から捉えていくかである。作品を解析するキーに何をとるか、そのキーを ambivalence なもので分類する。ここでは話法に視点を置き ambivalence を考えて、次の３つに分類する。一つ目は作品のストーリーで筋を追う文章である「客観話法」、二つ目は登場人物の心理または心象が混入している主観的な文章である「中間話法」、三つ目は直接会話の文章である「会話法」に分けている。

3．解析

「客観話法」、「中間話法」と「会話法」の 3 つの話法で文章を分類し、頻出度で数量化する。線形空間論を適用する。作品の流れを時空間とし、3 つの話法はお互いに独立した３次元空間とする。数量化した話法データを3次元のベクトルデータとみなす。

数量化したデータにはデータ抽出者の主観、話法間の親密性が含まれているため客観化を行う。3 次元ベクトルデータである話法データを互いに独立な直交した成分に抽出する。直交化した成分のばらつきをなくすために正規化する。

正規直交化した話法ベクトルをイメージデータとみなし、それぞれの話法ベクトルを離散値系ウェーブレット変換による多重解像度解析を行う。

(1) 「坊ちゃん」と「三四郎」の数値化

章を時空間の区切りとし、夏目漱石の「坊ちゃん」と「三四郎」を話法で分類した頻出度を百分率で示したものを表 1 に示す[2],[3]。表１をグラフ化したものが図１である。

章における話法の大きさであるノルムを計算する。ノルムは話法の二乗和の平方根である。図 2 (a)の「坊ちゃん」では２章、３章と９章で特に大きい値となり、5 章と 8 章では小さい値を示している。図 2 (b)の「三四郎」では最後の 13 章が特に大きな値を示し、7 章と 11 章で小さな値を示している。ノルムの大きな値は特に少ない頻出度の話法があり、頻出度の高い話法と低い話法

表１．話法による分類（頻出度）
(a) 坊ちゃん

章	客観話法	中間話法	会話法
1	62	35	3
2	77	22	1
3	73.5	26	0.5
4	63	27	10
5	44	41	15
6	62	28	10
7	53	29	18
8	37	26	37
9	76	13	11
10	70	21	9
11	59	13	28

(b) 三四郎

章	客観話法	中間話法	会話法
1	69	19	12
2	71	18	11
3	68	22	10
4	71	10	19
5	66	25	11
6	67	12	21
7	54	20	26
8	63	11	26
9	65	12	23
10	67	14	19
11	62	9	29
12	68	12	20
13	76	1	23

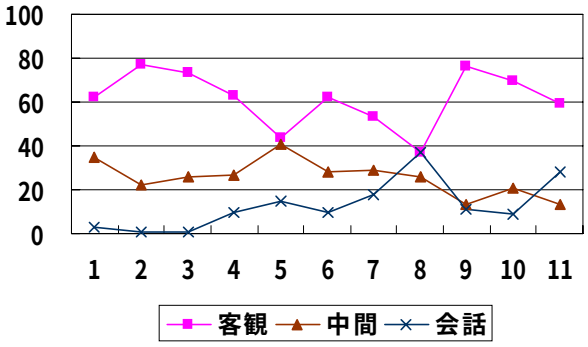
の差が特に大きいことを示している。ノルムの小さい値は頻出度が各話法にそれぞればらついていることを示している（表１、図１）。

夏目漱石の作品は一般に螺旋を描いているらしいと言われている。横軸に中間話法、幅軸に客観話法、縦軸に会話法として、縦軸を共通軸にして 3 次元空間に投影したものを図３に示す。

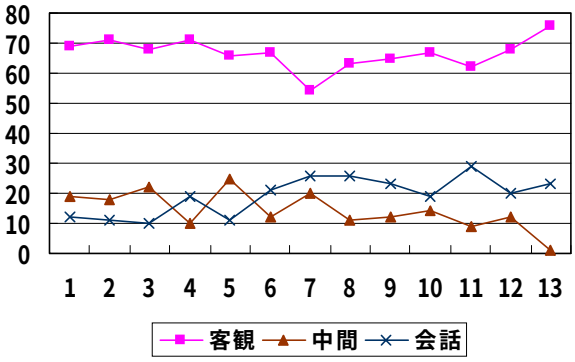
表１の頻度化したデータは抽出者の主観（親密性）が入っている。その親密性を各話法間の角度で示す。

	「坊ちゃん」	「三四郎」
客観話法と中間話法間の角度	24.9°	25.3°
中間話法と会話法間の角度	45.3°	37.4°
客観話法と会話法間の角度	48.8°	20.1°

「坊ちゃん」では客観話法と中間話法間が、「三四郎」では、３つの話法間ともに親密度は非常に高い。

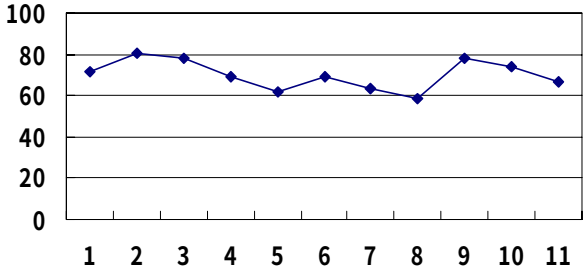


(a) 坊ちゃん

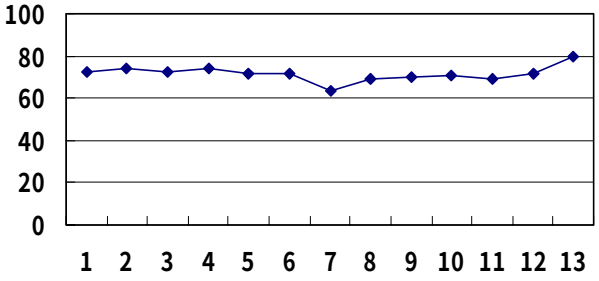


(b) 三四郎

図 1．話法の頻出度グラフ（横軸：章、縦軸：頻出度）

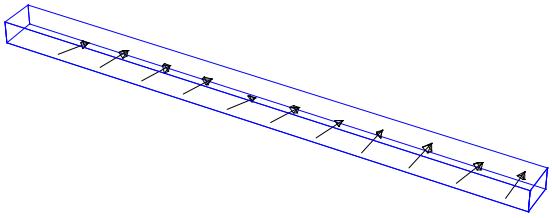


(a) 坊ちゃん

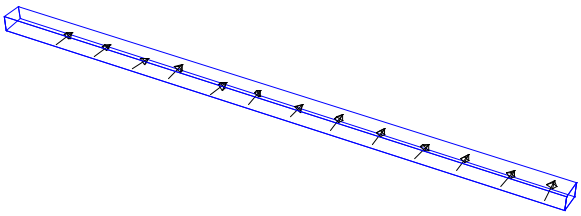


(b) 三四郎

図 2．話法のノルム（横軸：章、縦軸：ノルム）



(a) 坊ちゃん



(b) 三四郎

図 3．話法ベクトルを共通軸上に 3 次元表示（横軸:中間話法、幅軸:客観話法、縦軸:会話法）

(3) 正規直交化

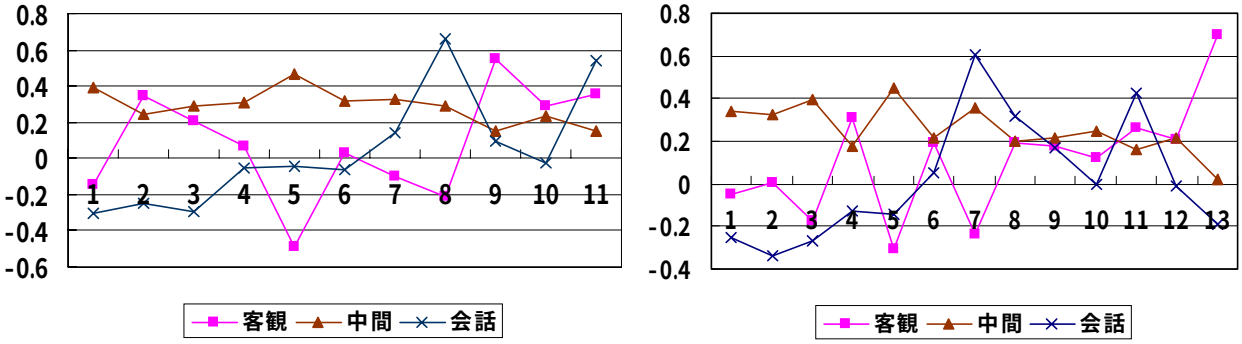
線形空間の直交化はグラムシュミットの方法で行う。第一成分に中間話法、第二成分に客観話法、第三成分に会話法とし、主観的な文章である「中間話法」を主体として解析を行った。第一成分の中間話法の値をそのまま直交成分とし、客観話法の値から第一成分の重なり部分を取り除いた直交成分を第二成分とする。会話法の値から第一成分と第二成分の重なる部分を取り除いた直交成分を第三成分とする。それぞれ直交化したベクトルデータを各話法のノルムが１となるように正規化する。

正規直交化したデータをグラフ化したものを図４に、章ごとのノルムを図５に示す。共通軸上に 3 次元空間に投影したものを図６に示す。正規直交化することによって章ごとの頻出度の変化が明確化する。

(4) ウェーブレット変換による多重解像度解析

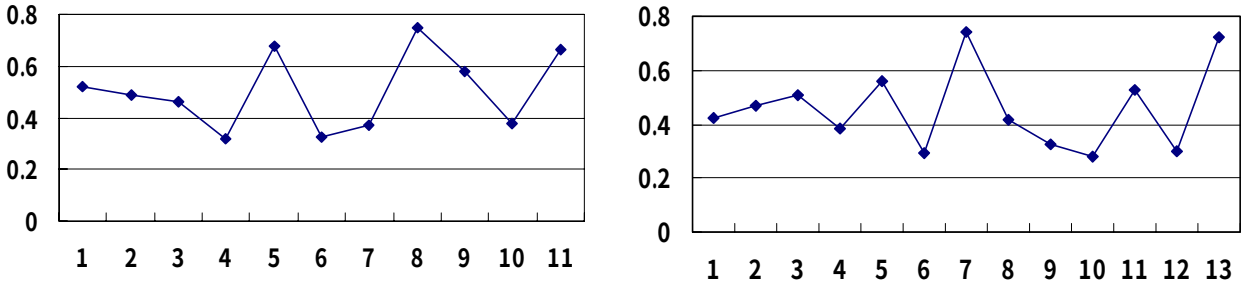
作品を構成する話法の頻出度を考察するために、正規直交化した話法ベクトルデータを離散値系ウェーブレット変換による多重解像度解析を行う。話法ベクトルデータはウェーブレット変換を行うため２のベキ乗個の 16 章にする。「坊ちゃん」には 11 章の話法ベクトルに 12 章から 16 章までゼロを追加し、「三四郎」も 13 章の話法ベクトルに 14 章から 16 章までゼロを追加し解析を行う。話法ごとにドビッシーの 2 次基底を用いてウェーブレット変換の一次元多重解像度解析を行う。解析結果のレベルは 5 段階で、レベル１から５まで頻出度のグラフと共通軸上に３次元ベクトル表示したものを図７に示す。

レベル１は作品全体の平均的な話法の頻出度を表す。第一成分の中間話法が高い値となる。「坊ちゃん」では



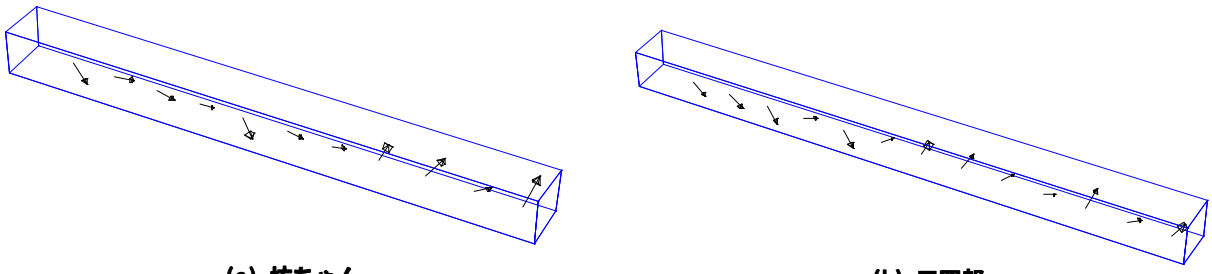
(a) 坊ちゃん　　(b) 三四郎

図４．正規直交化した話法データのグラフ（横軸：章、縦軸：頻出度）



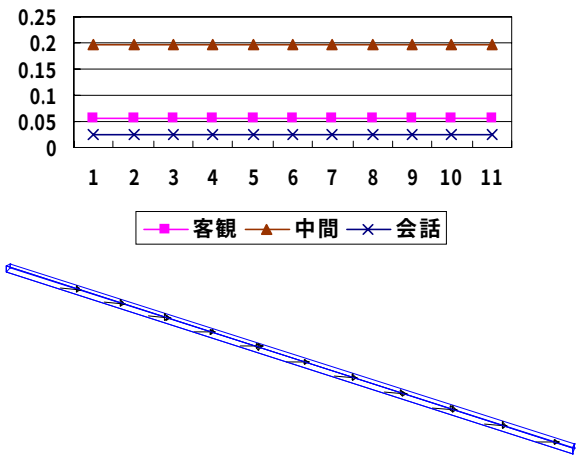
(a) 坊ちゃん　　(b) 三四郎

図５．正規直交化した話法ベクトルのノルム（横軸：章、縦軸：ノルム）

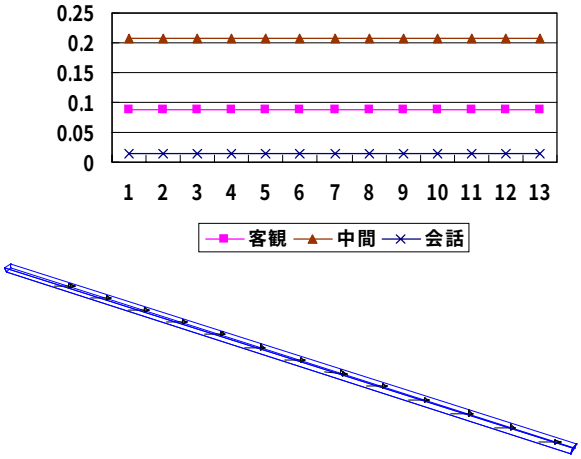


(a) 坊ちゃん　　(b) 三四郎

図６．共通軸上に３次元表示した正規直交化話法ベクトル（横軸:中間、幅軸:客観、縦軸:会話）

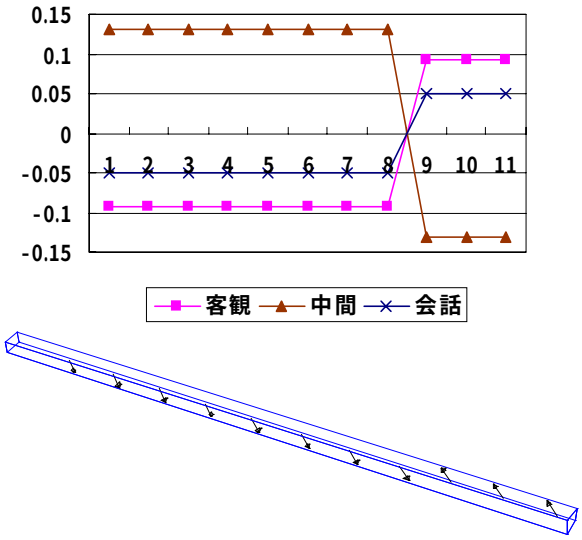


(a) 坊ちゃん

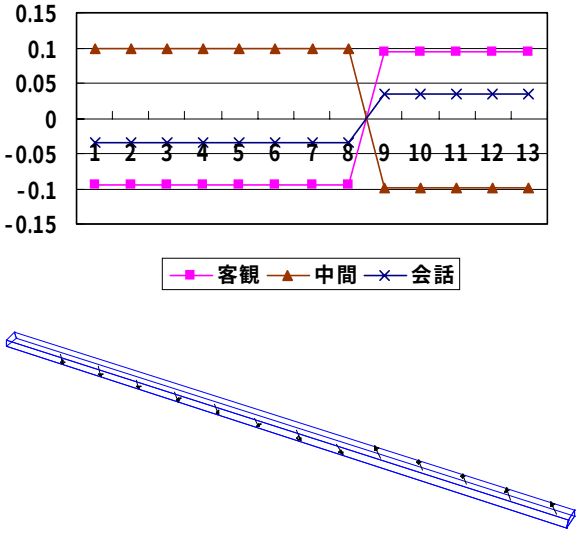


(b) 三四郎

多重解像度解析のレベル１

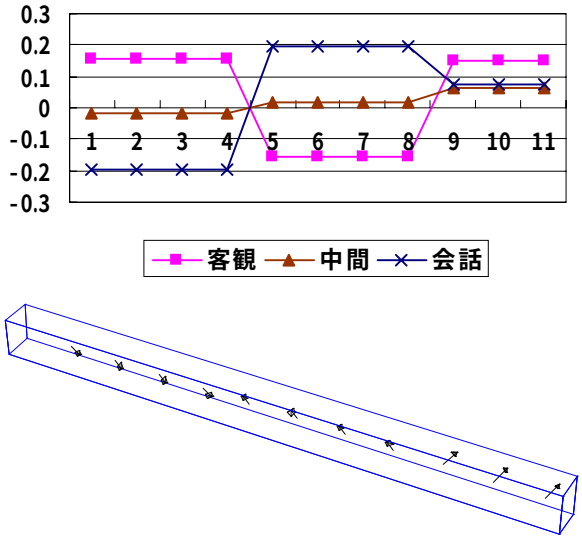


(a) 坊ちゃん

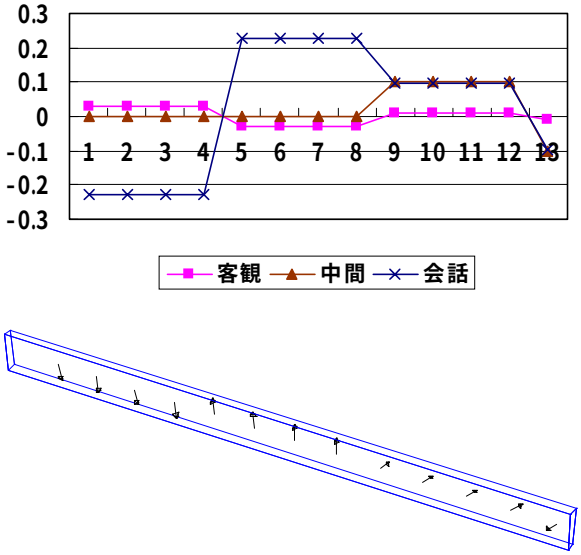


(b) 三四郎

多重解像度解析のレベル２

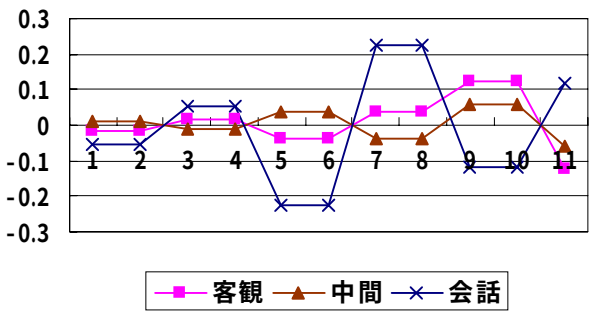


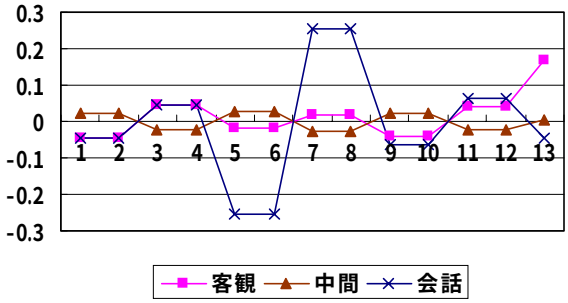
(a) 坊ちゃん

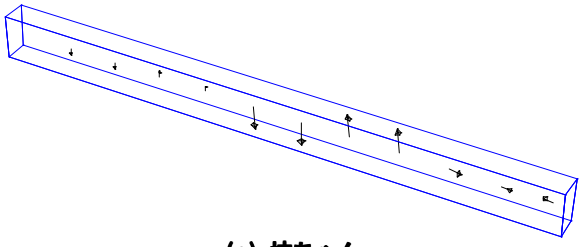


(b) 三四郎

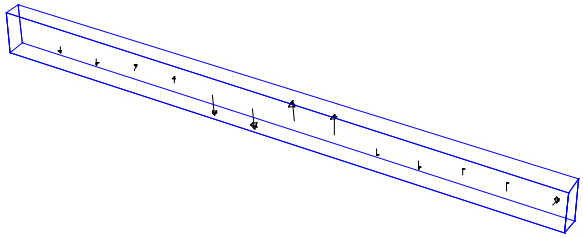
多重解像度解析レベルの３





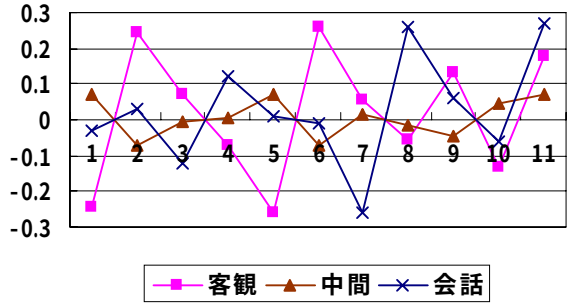


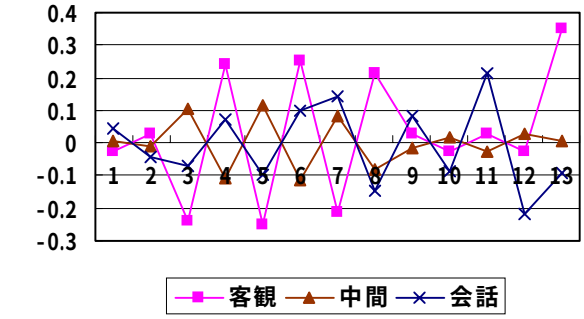
(a) 坊ちゃん

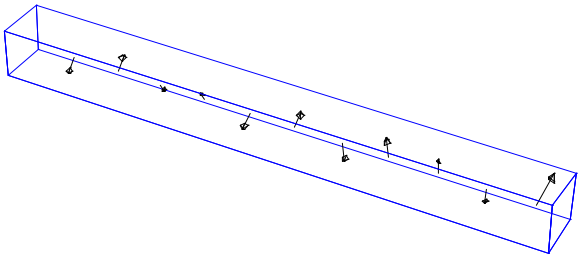


(b) 三四郎

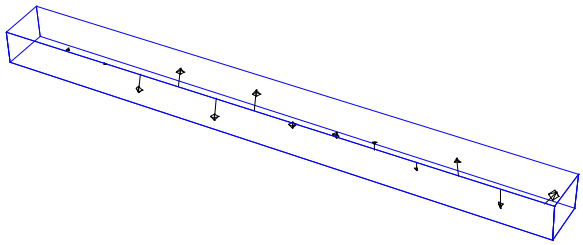
多重解像度解析のレベル 4







(a) 坊ちゃん

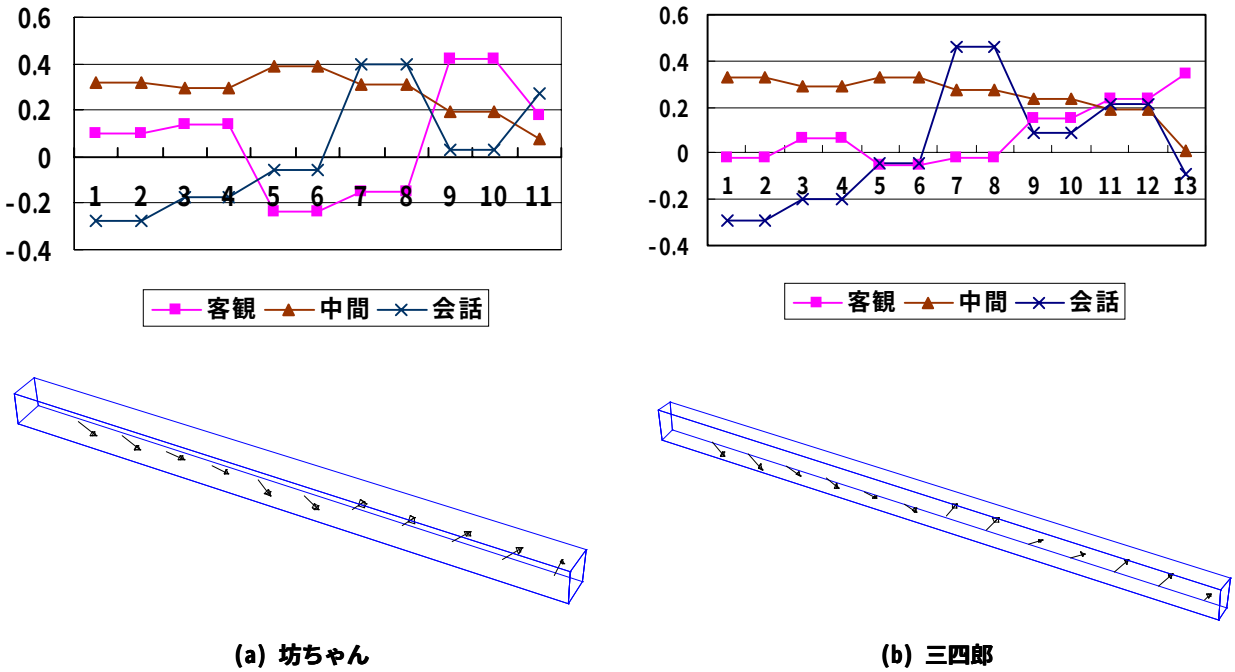


(b) 三四郎

多重解像度解析のレベル５

図 7．ウェーブレット変換による多重解像度解析結果（ドビッシーの 2 次基底関数）
上：話法ごとの頻出度グラフ（横軸:章、縦軸:頻出度）
下：共通軸上に３次元表示したもの（横軸:中間、幅軸:客観、縦軸:会話）

客観話法がやや会話法より多い。「三四郎」では客観話法が多い構成となっている。レベル２は 2 分割した平均的な話法の頻出度を表す。「坊ちゃん」では前半に中間話法が多く、後半に会話法が多くなっている。「三四郎」では前半部に中間話法がやや多く、後半部に会話法が多くなっている。どちらも同じ傾向の構成である。レベル３は４分割した平均的な話法の頻出度を表す。「坊ちゃん」では客観話法が 5 章から 8 章部で減少し、その後増



(a) 坊ちゃん　　(b) 三四郎

図 8．多重解像度解析結果のレベル１から４までを加算
上：話法ごとの頻出度（横軸:章、縦軸:頻出度）
下：共通軸上に３次元表示したもの（横軸:中間、幅軸:客観、縦軸:会話）

加している。会話法は章が進むにつれて増加している。「三四郎」も同じ傾向であるが客観話法の変化は小さく、会話法の増加が大きい。レベル４は８分割で２章づつに区切った平均的な話法の頻出度を表す。「坊ちゃん」では７章から８章部に会話法が特出している。９章以降は客観話法と会話法が頻出している。「三四郎」でも７章から８章部で会話法が特出している。９章以降でも客観話法と会話法が頻出している。このレベル４では８章までの前半部が同じ傾向の構成となっている。レベル５は章ごとの頻出度合いである。どちらも客観話法と会話法の変化が大きい。中間話法は他の話法と比較して少ない。「三四郎」では９章以降には中間話法はほとんど頻出していない。

線形空間論ではレベル５は平均化されていない値でノイズとしている。レベル５を除き、レベル１から４までの平均化データを加算したものを図８に示す。どちらも中間話法は章の流れとともに徐々に減少している。客観話法の変化は「坊ちゃん」では大きく、「三四郎」では小さい。客観話法は７章以降から大きく変化している。

4．まとめ

夏目漱石の「坊ちゃん」と「三四郎」は主人公の名前がタイトルとなっている作品である。この作品の発表年月日は

明治 39 年(1906) 4 月「ホトトギス」

明治 41 年(1908) 9 月 1 日から朝日新聞社連載

である。

「客観話法」、「中間話法」と「会話法」に３分類した話法の頻出度から解析した結果、「坊ちゃん」と「三四郎」は章の違いがあるが、話法についていえば前半部は同じ構成の作品である。１章から６章にかけては、会話法はやや増加しているといえる。７章から８章にかけて会話法が主体となっている。後半は、「坊ちゃん」では、９、10 章で客観話法が主体となり、最後の 11 章を会話法が主体で終っている。「三四郎」では、９章以降で客観話法を徐々に増加させ、最後の章を客観話法が主体で終っている。

参考文献

[1] 岩崎晴美,斎藤兆古,宮澤賢治,堀井清之,金田知子、“文学作品の線形空間論による解析”、法政大学計算科学研究センター研究報告、第 11 号、1998 年 3 月

[2] 宮澤健太郎、“漱石の文体”、洋々社、1997 年 9 月 30 日）

[3] 宮澤健太郎、“「三四郎」の文体論的考察”、白百合大学紀要 31、1995 年 12 月

[4] 斎藤兆古著：“Mathematica で学ぶウェーブレット変換”、朝倉書店（1998）

[5] 宮澤健太郎、“「文体」から賢治を読む －文体論からイメージ文学へ－”、日本語学 1997.September Vol.16,9、明治書院

キーワード.

計算文学、イメージ文学、正規直交化、ウェーブレット変換、多重解像度解析

Summary.

Linear Space Analysis of the “Bochan” and “Sanshiro” Styles

Harumi Iwasaki
Computational Science Research Center, Hosei University

Yoshifuru Saito
College of Engineering, Hosei University

Kenji Miyazawa and Kiyoshi Horii
Shirayuri Women's College

In the present paper, we propose a new methodology for analyzing the literature based on the linear space theory. At first, we classify the conversation sentences used in the story of “Bochan” & “Sanshiro” into three kinds of styles. Secondly, three dimensional normalized orthogonal vector space is constructed from three kinds of conversation styles by means of the Gram-Schmidt process. Third, discrete wavelets transform is applied to these three dimensional normalized orthogonal conversation vectors. As a result, it is revealed that a vector composed of the conversation vectors at each chapter in “Bochan” & “Sanshiro” is spirally rotating with sinusoidally modulated magnitude. Thus, we have succeeded in analyzing the structure of “Bochan” & “Sanshiro” using linear space theory.

Keywords.

Computational Literature, Image Literature, Normalized orthogonal system, Wavelet Transformation