

文学作品の線形空間論による解析

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

宮沢 賢治、堀井 清之、金田 知子

白百合女子大学文学部

文学作品における文体解析は、作品を解析する人の知識・経験・考え方などによるところが大きい。ここでは、数学的手法を用いて文学作品を評価することを試みた。作品をイメージとしてとらえ、作品に使用されている話法について分類し、頻度を数値データに置換する。その数値データに線形空間論を適用して、その作品の構成、性質などを考察する。例として夏目漱石の作品「草枕」を取り上げた。

1. はじめに

文学作品を構成するには視覚的表象である文字は最小構成部品である。特定の意味を持つ文字を組み合わせた単語が文学作品を構成する2番目の部品である。文学作品のフレームを構成するのが単語を組み合わせることで構成される文章である。文章を組み合わせることで物語性(story)を創造し文学作品が形作られる。文学作品を構成する部品の中で、文字は変えようがない。しかし、文字を組み合わせることで単語になると、使われ方によって名詞、動詞、副詞等のパリエーションが存在する。また、同じものを示す名詞でも状況によって使い分けられているし、別の単語でも同じものを指すときもある。単語を組み合わせることで構成される文章は、同一の内容を表現するのに無数のパリエーションが存在し、この文章の構成如何で作品の文学性が評価される程重要な文学作品のフレームである。文章そのものを分類するのに文型や文体等がある。文体の構成を話法で分類し、解析手法として線形空間論(linear-space-theory)を適用する。

2. 文学作品の数値化

文学作品の評価は、作品の持つ物語性、文章の構成さらに用語の使われ方等、小さな部品から大きな構成部品、ひいては作品全体に渡る観点からなされる。但し、この評価も数値的になされるわけではなく、用語が使われる周辺の文章や雰囲気など数値的な取り扱いが困難なパラメータを前提としてなされる。このため、評価を下す文学者の知識・経験・考え方異なる評価が可能であり、それぞれの側面から見た妥当な評価である。しかし、数学的手法を導入した計算機を前提とする文学作品の解析を行うには何らかの数値化が必要である。

文学作品を構成する部品は、文字、単語、文章そして物語性に分類される。文学作品を解析するに当たり、可能な限り微小な構成部品から数値化することが望ましい。これは作品の細部まで解析対象となるためである。しかし、文字を数値化しても作品の長短が評価できるだけで

あり、あまりにも原始的な微小部品である。但し、母音や子音の頻度を数値化してみることは言語学的観点からみれば有益な解析結果が得られるかも知れない。文字を組み合わせることで得られる単語は作者によつて異なると考えられる。作品を解析する上で、数値化可能な部品は単語であろう。しかし、単語には多くの種類がありその総てを数値化の対象とするには多大な労力を必要とするであろう。

文体の構成を考察するのに、話法を取り上げた。話法を次の3つに分類する。一つは、客観型話法、二つ目は会話法、そして最後にその他として中間型話法とにである。分類にしたがって作品の章毎に頻度を数えることにより数値化を行った。また、作品の話は時間的に進んでいく。章毎のデータは時系列データであると言える。この目標は、数値化されたデータがどのように使われているかを分析し、その結果から普遍的な作者の意図や作品の構成方法などを見出そうとする点にある。このため、数値化されたデータの客観性、データの重複度、作品を構成するデータの独立性等を確保する手段が必要となる。

3. 数学的手法

ベクトルは大きさや方向を持つ量を意味するが、作品の持つ志向性をベクトルで表現するために用いる。数値化した話法データを3次元のベクトルと考える。つまり、客観型話法、中間型話法、会話法で構成されるベクトルである。次に、データの積和を計算するために内積(inner-product)の概念を用いる。内積があればベクトル間の角度の概念が必要となる。角度でお互いの話法の親密度がわかる。話法は、お互いに独立していると考え、データの直交化を行う。そして、ベクトルデータの大きさを揃えるために正規化を導入する。これは、データ抽出者の主観の相違を削減するために用いる。最後に線型変換である。これはデータの平均や変化率を系統的に計算する。最近データ圧縮や多重解像度解析に有効なウェ

ープレート変換による多重解像度解析を活用する。ウェーブレット変換には八 - ル基底であるドビッシー 2 次基底を使用する。

(1) 例題

ここでは、夏目漱石の作品「草枕」の話法について取り上げる。各章で登場する客観型話法、中間型話法および会話法の頻出度を表 1 [1] に示す。

表 1 話法による分類

章	客観型話法	中間型話法	会話法
1	2	31	67
2	36	25	39
3	62	36	2
4	44	21	35
5	23	12	65
6	57	42	1
7	50	49	1
8	47	16	37
9	25	9	66
10	39	41	20
11	31	31	38
12	46	41	13
13	60	17	23

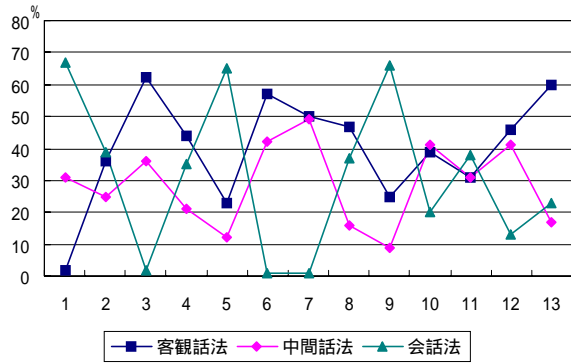


図 1 章毎の客観型話法、中間型話法、会話法の頻度グラフ

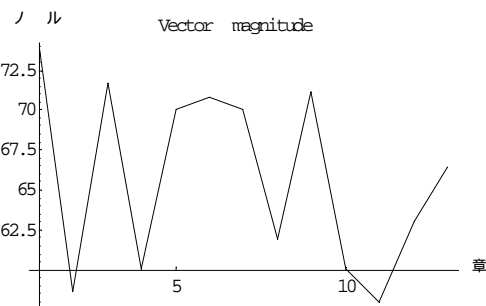


図 2 章における話法ベクトルの大きさ

次に、表 1 をグラフ化した結果を図 1 に示す。

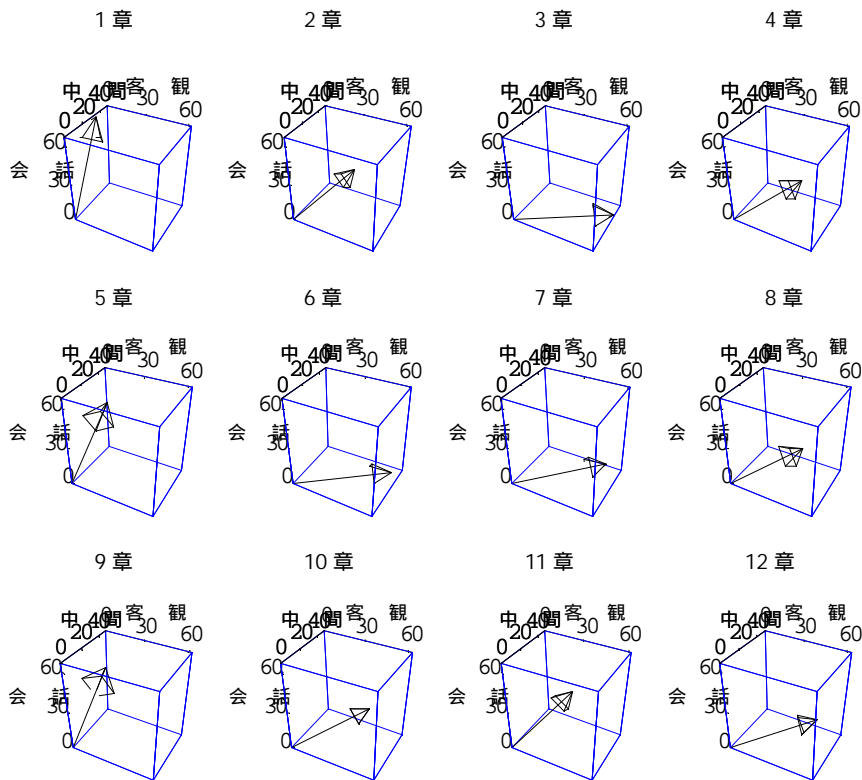


図 3 各章毎の話法頻出度の 3 次元表示 (x: 客観型話法、y: 中間型話法、z: 会話法) 続く

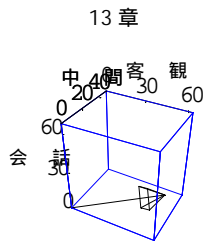


図3 続き 13章の話法頻出度の3次元表示
(x:客観型話法、y:中間型話法、z:会話法)

それぞれの話法を3次元ベクトルの独立な要素として考える。章におけるベクトルの大きさを図2に、3次元の互いに直交するx,y,z軸上にそれぞれ客観話法、中間話法そして会話法の頻出度を図3に示す。上段から左に1章、2章、3章、4章で、次が5章から8章、9章から12章で、次のページに13章のベクトル図となっている。この図2から章を追う毎に矢印は原点を軸としてスパイラル状に回転していることがわかる。

(2) 話法ベクトル間の角度

ベクトル間の角度が小さいと同一話法ベクトルであり、角度が90度に近いと互いに独立性の高い話法ベクトルであることを意味する。「草枕」は13章からなり、各章の客観型話法の頻出度を要素とする13次の客観型話法データベクトル、中間型話法の頻出度を要素とする13次の中間型話法ベクトル、さらに会話型話法の頻出度を要素とする13次の会話型ベクトル間の角度を調べる。

中間話法と客観話法ベクトル間の角度は、26.47度、
中間話法と会話法ベクトル間の角度は、55.88度で、
客観話法と会話法ベクトル間の角度は、56.44度である。

図1から客観話法と会話法は互いに相反する傾向があるが、線型空間論的にも異なるベクトルであるといえる。

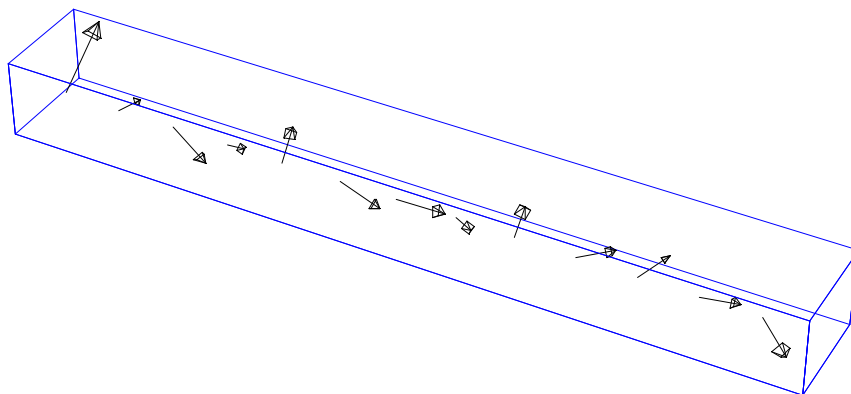


図6 正規直交化した話法ベクトルを共通軸上に3次元表示(左から1章、2章で13章まで)

(3) 話法ベクトルの正規直交化

各話法ベクトル間で重複する要素を取り除き、それぞれのベクトルが重複する要素を含まない互いに独立な直交したベクトル成分だけを抽出し、ベクトルの大きさを統一するためにベクトルの大きさを正規化する。

グラム・シュミット(Gram Schmidt)の方法で直交化し、それぞれのベクトルのノルムが1となるように正規化した計算結果を図4、5、6に示す。図4は正規化直交化された結果をグラフ化したもので、図5は章毎のベクトルの大きさである。図6は直交化した話法ベクトルを共通軸上に3次元表示した結果である。共通軸は客観型話法をx軸方向としている。奥から順に第1,2,...,13章に並んでいる。図2と比較して、ベクトルの方向だけでなく大きさに顕著な相違が存在することがわかる。

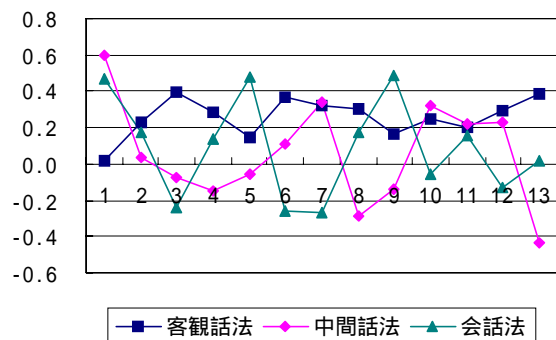


図4 章毎の正規直交化した客観、中間、会話法のグラフ

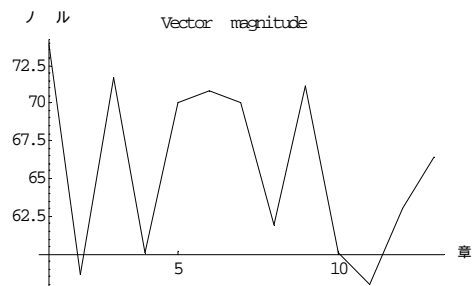


図5 章におけるベクトルの大きさ

(4) ウェーブレット変換による多重解像度解析

話法ベクトルの正規直交化 3 次元ベクトルにウェーブレット変換を適用し、多重解像度解析を用いて草枕を構成する話法の頻出度合いを作品全体の平均から各章毎の変化まで調べる。ウェーブレット変換は、データの個数は、2 のべき乗でなければならない。データにゼロを追加して、2 のべき乗の 16 個のベクトルを構成し、ドブシ-2 次基底でウェーブレット変換によるベクトルデータの多重解像度解析を行う。その計算結果を図 7a,b,c,d,e に示す。全体で 5 レベルからなり、各レベルの上段の図は 3 次元ベクトル表示であり、客観型話法、中間型話法および会話法ベクトルをそれぞれ軸方向 x 、幅方向 y そして高さ方向 z 軸へ割り振ってある。左側の原点が 1 章で、一番右側が第 13 章のベクトルである。

レベルは値が小さいほど平均的な情報を与える。

下段の図は、横軸が序章を原点とする章、縦軸が 3 次元ベクトル図の各成分値である。

図 7a のレベル 1 は作品全体の平均的な話法の頻出度を表す。「草枕」では、作品全体が平均的に客観型話法で構成されていることがわかる。

図 7b のレベル 2 は全章を 2 分割した場合の話法の頻出度を表す。この場合でも「草枕」の大部分の章は客観型話法で構成されている。

図 7c のレベル 3 は全体を 4 分割した場合の話法の頻出度を表す。始めの章は中間型話法と会話法が支配しているが中頃の章では減少して、また増加している。客観型話法は、序章に少なくだんだんと増加し、中頃から支配的になりつつある。

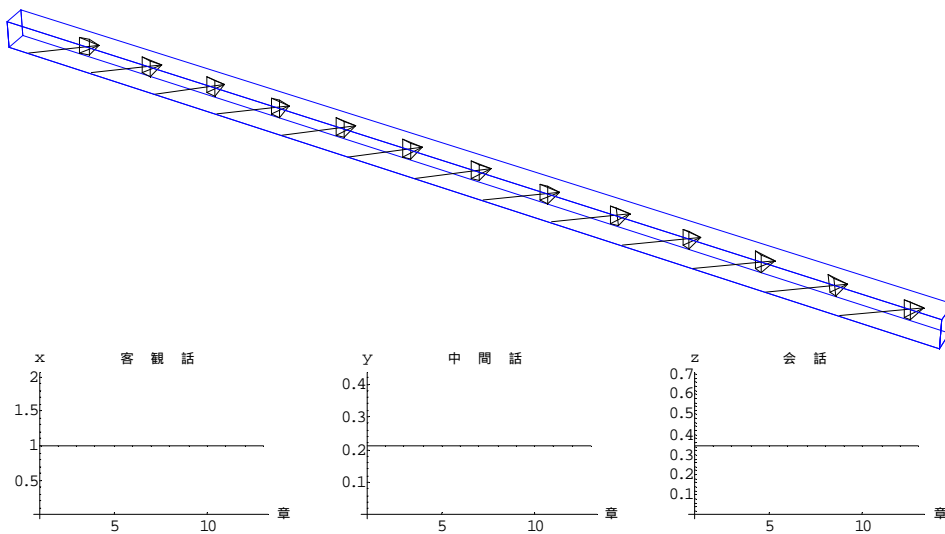


図 7a 多重解像解析度レベル 1

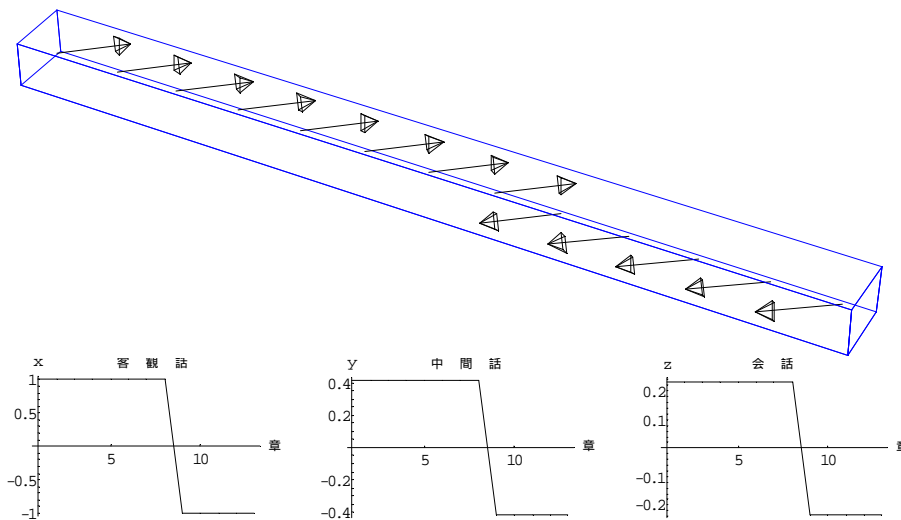


図 7b 多重解像解析度レベル 2

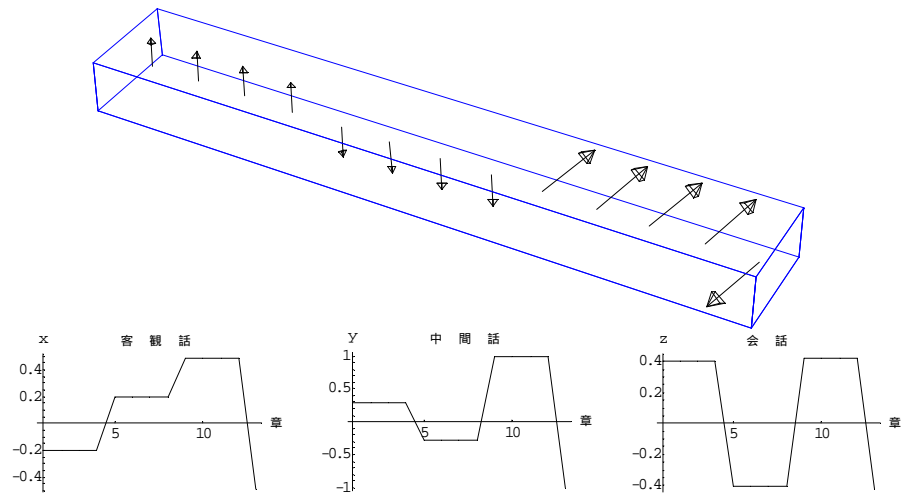


図 7c 多重画像解析度レベル 3

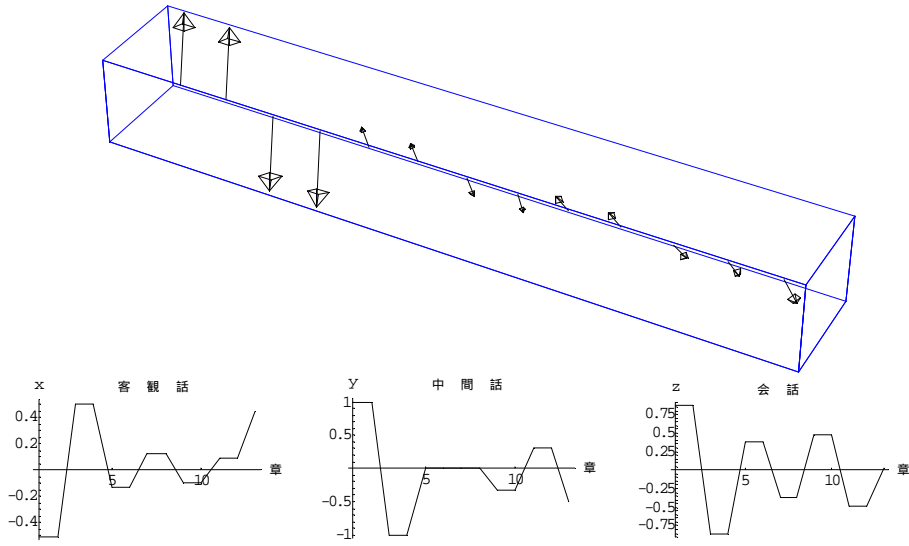


図 7d 多重画像解析度レベル 4

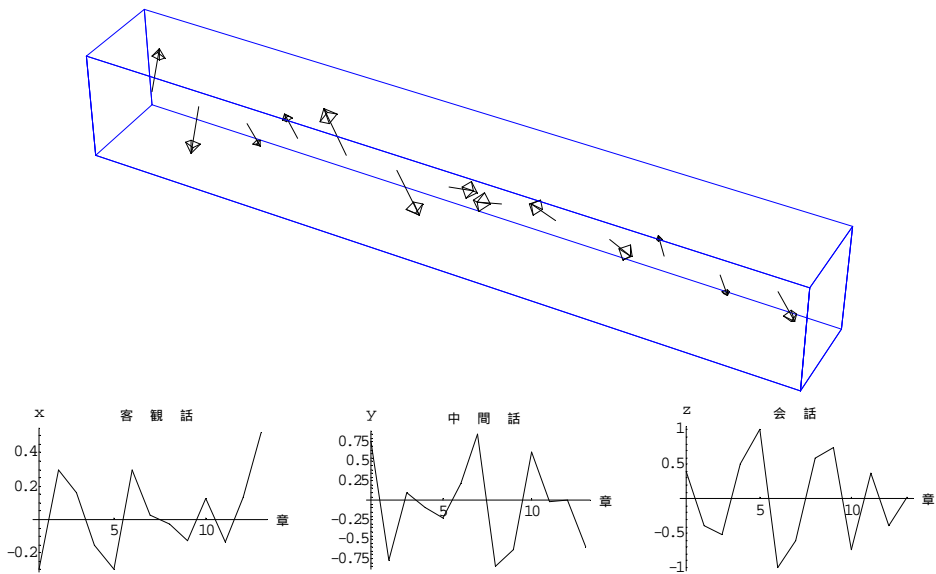


図 7e 多重画像解析度レベル 5

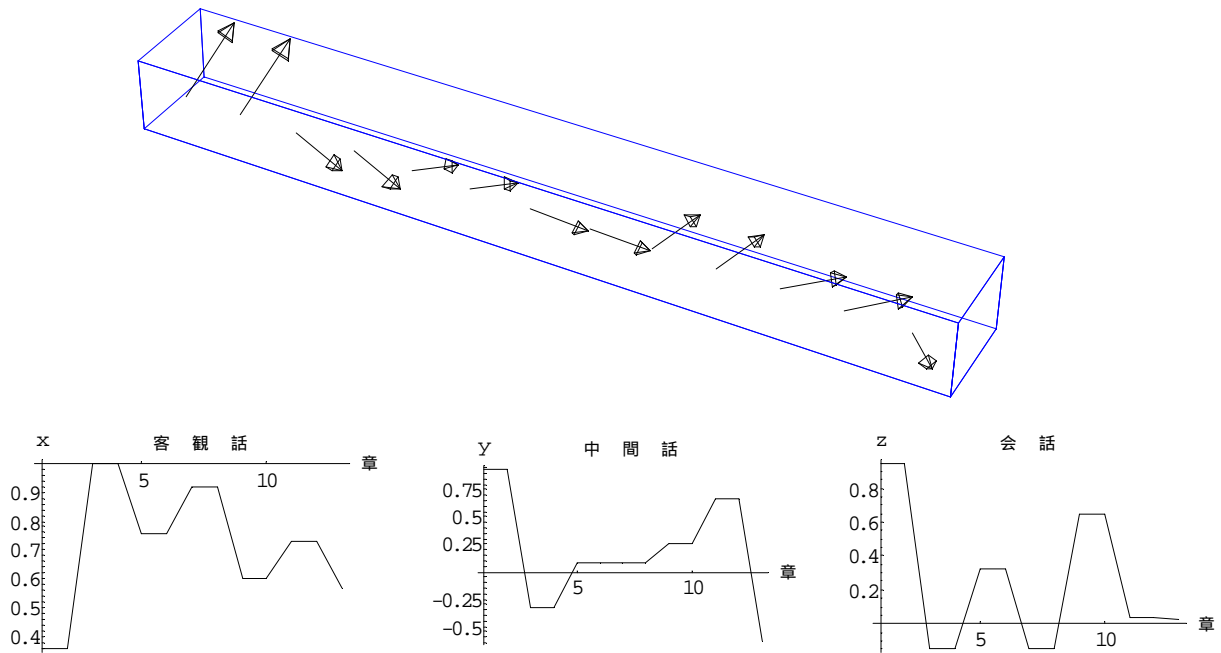


図8 レベル1からレベル4までの和

図7dのレベル4は全体を8分割した場合の話法の頻出度を表す。作品は13章で、ゼロを追加して16個の要素から各ベクトルが構成されているから、2章毎に平均化された情報である。始めの1,2章は中間型話法と会話法で構成されていて、3,4章は客観型話法となる。続いて5,6章は中間型話法と会話法が、7,8章は再び客観話法が増加し支配的になる。9,10章は客観型話法が支配し、11章以降は、中間型話法が11,12章でやや増加するが客観型話法が最後には客観話法が支配的になっている。会話法についていえば始めは支配しているが中間の章以降で急激な増減振動している。

図7eのレベル5は各章毎の話法の頻出度であり、どの話法も大きく振動的変化をしている。レベル5はデータの平均化がなされていないため隣接するデータ間のバラツキが強調された結果を含んでいると考えられる。つまり、レベル5は雑音である。この意味でレベル1-4までの結果が信頼できる。レベル1から4までを加算した結果を図8に示す。

4. まとめ

話法の頻出度を解析した結果、「草枕」は中間型話法と会話法で始まっているが、全体平均では客観型話法が全体を構成していること、話法が章により振動的に変化していくことがわかった。基礎データは作品を通してデータ抽出者の印象から平均的な意味でなされた結果で、分類する人間の主観が入っている。しかし、このようなデータをグラム・シュミットの方法で直交化することによりデータ間の重複要素が取り除かれ、正規化する事に

よりデータ抽出間の隔たりが除かれ、また、ウェーブレット変換の多重解像度解析によりデータのばらつきが低減される。このような手法を導入することにより、解析結果に客観性がえられる。このことにより、作品の文体の特徴を掴めると考える。また、文体だけでなく、登場人物や、使われている色や、表現などに適用することにより、作品の特徴や作品の山場がどこにあるかなど解析可能と考える。

参考文献

- [1] 宮澤健太郎著：「草枕」の文体論的考察、白百合女子大学研究紀要第33号（平成9年12月）
- [2] 宮沢賢治著：「宮沢賢治 近代と反近代」、洋々社（1991年9月10日）
- [3] 宮澤健太郎著：「漱石の文体」、洋々社（1997年9月30日）
- [4] 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
- [5] 斎藤兆古著：Mathematicaで学ぶウェーブレット変換、朝倉書店（1998）

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 Literature

Harumi Iwasaki
Computational Science Research Center, Hosei University

Yoshifuru Saito
College of Engineering, Hosei University

Kenji Miyazawa , Kiyoshi Horii and Noriko Kaneda
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 "Kusamakura" 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 "Kusamakura" is spirally rotating with sinusoidally modulated magnitude. Thus, we have succeeded in analyzing the structure of "Kusamakura" using linear space theory.

Keywords.

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