『源氏物語』の可視化*

井波 真弓**, 齊藤 兆古***, 堀井 清之**,
細井 尚子†, 山縣 貴幸†, 藤澤 延行†,
村井 祐一††, 山田 美幸†††, 熊谷 一郎††††

Visualization of “The Tale of Genji”

Mayumi INAMI, Yoshifuru SAITO, Kiyoshi HORII, Hisako HOSOI, Takayuki YAMAGATA,
Nobuyuki FUJISAWA, Yuichi MURAI, Miyuki YAMADA and Ichiro KUMAGAI

1. はじめに

近年、工学分野で発展してきた可視化手法が文学などの芸術の分野へも広がりを見せている。例えば、井波らによるウェーブレット解析を用いた『源氏物語』¹⁾や『銀河鉄道の夜』²⁾における感情の変化の可視化、山田・村井らによるシェイクスピア作品のストーリーの可視化^{3),4)}、細井による『源氏物語』の統計解析のためのテキストマイニング⁵⁾が行われている。これら過去の研究では、それぞれ異なる文学作品を対象としていたため、各可視化手法によりどのような情報が得られるかを比較することは困難であった。そこで、本稿では『源氏物語』を対象として各可視化手法を適用する。

『源氏物語』は、平安貴族社会の生活を背景として書かれており、その芸術性から日本文学の最高傑作とされている。『源氏物語』は、54帖からなる長編物語であり、作者は紫式部であるといわれている。各帖は、全体の構成の観点から3部に分けることができ、それぞれの登場人物、年紀、文章の違いなどから複数作者説も挙げられている。また、物語の成立時期は11世紀初頭とされているが、正確な年紀は不明である。そのため、従来国文学の立場から多くの研究がなされているが、一方でその文学的な価値からサイエンティフィックアート研究の対象として可視化分野でもしばしば取り上げられている。

本稿では、文学作品に対する可視化により、どのよう

な情報が得られるかを比較するため、『源氏物語』を対象として3つの可視化手法を解説する。2章では、ウェーブレット解析を用いた語りの構造の変化について可視化を行う。3章では、統計解析のためのテキストマイニング法の適用および数量化理論Ⅲ類を用いた特徴抽出を行う。4章では、物語のパターンを絵画的に可視化する方法について解説する。

2. ウェーブレット多重解像度解析

2.1 概要

文学作品は作家のメッセージがコード化されたものである。文学研究者や評論家はこのメッセージを作家固有の暗黙知としてとらえ、自己の暗黙知のコードを参照しながら文学作品を解析する。このように作品の分析は個人的経験に依存することが多い。そこで、線形空間論の正規直交系と離散値系ウェーブレット変換を導入することでより精緻な分析を行い、従来、文学者にしかできなかった文学作品の分析結果に客観的な評価を与えるとともに、文学作品の文体、文法構造、作品の暗黙知⁶⁾解析などに対して合理的で普遍性のある結果を導くことを目的に、これまで、『源氏物語』をはじめとする広範囲な文学作品に適用を試みてきた。

作品解析には、作者の表現した語、句、文章だけでなく、それを読み解く読み手の解釈も離散値系ウェーブレット変換のキーワードとして用いる。キーワードの選択には大きく分けると以下の二つの方法がある。まず、文章や文体の表現形式に注目⁷⁾してキーワードを選択した場合、解析者に依存せず一意的な結果が得られる処理を可能とし、従来指摘されていた文体論の問題点である客観性と全体性の提示を可能とした。次に、内容に主眼を置いてキーワードを選択^{8),9)}した解析から作品の暗黙知の可視化を可能とした。これは頻出度数の多い語彙が必ずしも作品において支配的になっているとは言えず、むしろ言葉の裏に秘められた言葉が抽出可能であることが判明し、統計学や確率論による手法では不可能であった暗黙知の可視化を可能にした。

* 原稿受付 2013 年 5 月 2 日

** 正会員 白百合女子大学 文学部 (〒181-8525 東京都調布市緑ヶ丘1-25, E-mail: minami@shirayuri.ac.jp)

*** 正会員 法政大学大学院 デザイン工学研究科

† 正会員 新潟大学 可視化情報研究センター (〒950-2181 新潟県新潟市西区五十嵐2の町8050, E-mail: yamagata@eng.niigata-u.ac.jp)

†† 正会員 北海道大学大学院 工学研究院 (〒060-8628 北海道札幌市北区北13条西8丁目, E-mail: murai@eng.hokudai.ac.jp)

††† 正会員 北海道大学大学院 文学研究科

†††† 非会員 明星大学理工学部 機械工学系

研究では離散値系ウェーブレット多重解像度解析を用いて『源氏物語』第一帖「桐壺」の語りの構造がどのように変化するかを検証する。

物語は「和歌」をどのように様式的に配置するかということにおいて構築される¹⁰⁾。「和歌」と「語り」は深く関わって展開してきたが、平安時代¹¹⁾になると表現上の異質性がきわだつようになった。

「語り」を担う物語の語り手は『源氏物語』の場合、複数存在しているといわれている。藍美喜子¹²⁾は「語り」の多重構造を時間軸で見ると、源氏に仕えた古女御、古女房の話を書きとめる筆録者の語り文、源氏の物語本を紹介する語りの三層にわたる語り手が存在し、それぞれの時間軸のズレによって、自由自在に思い出の物語の過去と現在を駆け巡る手法を獲得したと述べている。

従来「和歌」中心の助動詞の意味分析が行われ、「語り」特有の用法が説明されずにきたが、鈴木泰¹³⁾は「語り」分析に動詞に助動詞のつかない基本形を含めた研究が不可欠だとし、『源氏物語』から会話・和歌と語りの文章を区別し、和歌を除いた会話と語りの文章に注目しての過去の助動詞「キ・ケリ」と完了の助動詞「ツ・ヌ・タリ・リ」のついた形を対象にテンス・アスペクトについて論じている。

テンスとアスペクト¹⁴⁾とはどちらも時の表現についての意味のカテゴリーだが、テンスは当の出来事の発生時点と発話時点との前後関係をいい、過去・現在・未来に三分される。他方、アスペクトはテンスとは独立にその出来事自身の時間的なあり方による区別で、完了と未完了とに大別される。現代日本語について日本では用言が「た」をとまうか否かによって、過去および未来の意になるような時を表す文法範疇をテンス、時制という。また「出かけた」のように「た」をとまう形を「タ形」といい、「出かける」のようにとまわない形を「非タ形」または「ル形」（動詞の終止形は「る」でおわるから）とする。

現代小説の中では過去の出来事を述べる時、文末に「タ形」だけでなく、「ル形」も用いられる。「ル形」使用されることで、語り手の視点が自在に行き来し、作品を生き生きとさせる効果があること¹⁵⁾が指摘されている。石出靖男¹⁶⁾は、夏目漱石の『三四郎』と『草枕』の2作品の「タ形」文末・「非タ形」文末に注目し、ストーリー展開に速さや強弱の調節が行われることや作品の印象に与える影響について考察した。

現代日本語の作品は文末の「タ形」、「ル形」が視点の問題との関係を含めて多くの研究がなされているが、『源氏物語』においては「語り」における時間軸の研究に不可欠な助動詞のつかない基本形の表現が看過されてきた。また、日本語における過去と現在がいかに自由自在に作品を形成したかその構造を可視化したものがない。そこで、本稿では時間軸に注目し、離散値系ウェーブレット多重解像度解析を用い、語りの構造がどのように

変化するかを吟味する。本稿においては言語解析と形式を主眼とした解析を行うため、解析者へ依存しない一意的な可視化結果を与えている。

2.2 解析方法

2.2.1 解析対象

本稿では『新編日本古典文学全集』¹⁷⁾のテキストを用いる。段落の区切りは本テキストに従う。

2.2.2 キーワードの選択と方法

(1)作品の構成を継時的に見るために「タ形」、「ル形」、「引用」をキーワードとして選び、段落ごとの使用頻度を調べた。キーワードは著者が作品を読み、次のように選択した。「桐壺」に現れる語りの文末に注目し、過去の助動詞「キ・ケリ」と完了の助動詞「ツ・ヌ・タリ・リ」のついた形を「タ形」、「タ形」以外を「ル形」、和歌・会話を「引用」とした。Table 1は「桐壺」におけるキーワードの選択結果である。

(2)得られたデータに離散値系ウェーブレット変換の多重解像度解析を適用する。

2.2.3 分析

「タ形」、「ル形」、「引用」の要素にベクトルの概念を用いる。「タ形」を基準ベクトルとして、ベクトルをグラムシュミットの方法によって全て直交化し、重複要素を除く。さらに直交化されたベクトルを単位ノルムに正規化する。正規化されたデータにウェーブレット多重解像度解析を適用する。本研究では基底関数は演算処理の意味が把握できるドビッシーの2次^{18),19)}を採用する。

離散値系ウェーブレット変換で取り扱われるデータの個数は2のべき乗である。このため、 $n=32$ 個の要素からなるデータベクトルを考える。このベクトルの最初から17番目の要素まではTable 1の値で構成されており、残り15要素の値はゼロである。この様にして得られた

Table 1 Number of elements.

大段落	タ形	ル形	引用
1. 更衣を想う桐壺帝	1	6	0
2. 更衣に皇子誕生	4	8	0
3. 三歳の若宮袴着の儀	0	3	0
4. 更衣病み、命果てる	1	9	5
5. 若宮、更衣の里へ	0	3	0
6. 更衣の葬送、人々の愛惜	3	4	2
7. 帝、悲しみに沈む	1	3	1
8. 韃負命婦、母君訪問	3	22	23
9. 韃負命婦帰参、帝の哀傷	4	6	9
10. 若宮参内、祖母死去	4	3	1
11. 若宮の聡明さと美貌	1	5	2
12. 高麗人の観相、源姓	3	5	2
13. 先帝の四の宮（藤壺）入内	3	9	2
14. 源氏、藤壺の宮を慕う	0	5	4
15. 源氏の元服、左大臣家の婿	5	20	3
16. 左右大臣家の勢力の拮抗	1	2	0
17. 源氏、藤壺の宮を恋慕	1	10	0
合計	35	143	5

Fig. 1 Level 4 of the discrete wavelets multi-resolution analysis: Ta- form

Fig. 2 Level 4 of the discrete wavelets multi-resolution analysis: Ru-form

Fig. 3 Level 4 of the discrete wavelets multi-resolution analysis: Inyo

Fig. 4 Level 4 of the discrete wavelets multi-resolution analysis: patterns of 3 keywords. Solid: Ta- form, Dotted: Ru- form, Alternate long and short dash: Inyo.

行ベクトルを Y とする. さらに, W はウェーブレット変換行列である.

ウェーブレット変換行列を W とすればウェーブレットスペクトラム S は次式で与えられる.

$$S = WY \quad (1)$$

離散値系ウェーブレット変換で, 特にドビッシーの2次を採用した場合, ウェーブレットスペクトラムの第1レベル (ウェーブレットスペクトラムベクトル S の最初の1要素を除き, 他を全てゼロとする) は全要素の平均値に対応し, 第2レベル (ウェーブレットスペクトラムベクトル S の第2要素のみ残し, 残りの要素をゼロする) は全データを前半と後半に分けた場合の変化率, さらに第3レベル (ウェーブレットスペクトラムベクトル S の第3, 4要素のみ残し, 残りの要素をゼロする) は全データを4分割してグループしたデータ間の変化率を意味する. 従って, データの個数が $n=2^X$ であるならば, 第 $(X+1)$ レベルは, 隣接要素間の差を意味する.

2.3 結果および考察

「桐壺」の要素ベクトルに離散値系ウェーブレット多重解像度解析を適用した結果を Fig. 1, 2, 3, 4 に示す. 横軸は本作品の最初から最後まで大段落ごとに構成要素を時系列に並べた. 縦軸はキーワードの大段落ごとの頻度の変化率を表す. グラフは実線は「タ形, Ta-form」, 鎖線は「ル形, Ru-form」, 一点鎖線は「引用, Inyo」を表す.

実際のデータ数としては17段落までであるが, 解析には2のべき乗のデータが必要であるため最後の段落に18から32段落をゼロデータとした²⁰⁾. また, 結果はゼロを追加した段落を削除してある. 離散値系ウェーブレット多重解像度解析は, 全体, 半分, $1/4$, $1/8$... というように段階に分けて分析し, これをレベル1, レベル2, レベル3, レベル4... と呼ぶ. レベル1では作品全体の平均を示す. レベル2では半分に分けた1段落から8段落の平均と9段落から16段落の平均の変化を示す. レベル3では4等分, レベル4では8等分, レベル5では16等分, レベル6では32等分である.

2.3.1 解析結果

全体の流れを見るために分析対象を8等分したレベル4の結果を Fig. 1, 2, 3, 4 に示す. 本論文では1段落から4段落を第1パート, 5段落から8段落を第2パート, 9段落から12段落を第3パート, 13段落から16段落を第4パート, 17段落を第5パートと呼ぶ.

第1パートと第2パートにおいてはキーワードの「ル形」と「タ形」緩やかに増加している. 「引用」は第1パートから第2パートにおいて大きく変化し, 第2パートにおいて大きく増加している. 第3パートから第4パートにかけてキーワードの「引用」と「タ形」がほぼ同様の曲線を描き, キーワードの「ル形」が少なくなっている. 第4パートではキーワードの「タ形」と「引用」

が減少している。物語の終焉にあたる第5パートにおいては「タ形」のキーワードが増加し、「ル形」は緩やかに減少し、「引用」は緩やかな増加を見せている。

2.3.2 「タ形」と「ル形」の解析結果

キーワードの「タ形」とキーワードの「ル形」との関係を見ると、第1パートと第2パートで共に増加傾向にあるが、頻出度合いの変化が小さい。第3パートと第4パートにおいては頻出度合いの変化が多少大きくなり、相反する傾向が見られる。第5パートになると、キーワードの「タ形」は緩やかに増加し、「ル形」は緩やかに減少する。

第1パートは作品の冒頭の部分で、桐壺帝と更衣の関係や宮廷にまつわる人々の関係が語られる。皇子の誕生や袴儀などの行事がかたられるが、更衣は心労が重なり亡くなってしまう。第2パートでは更衣を取り巻くさまざまな人々の気持ちが和歌や会話で語られる。第3パートで「タ形」が「引用」と同様の傾向を示し始めると「ル形」は「タ形」とは相反する変化を示す。

2.3.3 「タ形」と「引用」の解析結果

キーワードの「タ形」とキーワードの「引用」は第1パートから第2パートにかけて増加しているが、「引用」の増加が顕著である。第3パートと第4パートではほぼ連動している。

「引用」は第3パート以降減少し、「タ形」は第1パートから第3パートにかけて増加する。更衣の死の悲しみが語られる部分から主人公光源氏の登場へと内容が変化すると、キーワードの「引用」は減少する。藤壺の登場と共にキーワードの「タ形」が減少し、「ル形」によって語られる。これから『源氏物語』が始まる部分がありと目の前に現れるように語られる。第4パートで一旦減少するが、第5パートで再び増加する。

2.3.4 「ル形」と「引用」の解析結果

キーワードの「ル形」とキーワードの「引用」は第1パートから第2パートにかけて増加しているが、「引用」の増加が顕著である。第3パートと第4パートでは相反する傾向を示している。「ル形」は第1パートから第5パートにかけて増減を繰り返し、「語り」が単調にならないよう文末に変化をもたせていると考えられる。「引用」は更衣を失った悲しみが表現される部分が非常に多い。

2.3.5 考察

(1) 第1から第2パートにかけて「タ形」と「ル形」は共に緩やかに増加しており、ほぼ同様の変化が見られた。しかし、第3から第4パートにかけて相反する傾向が見られた。第5パートにおいては「ル形」にはほぼ変化が見られなかったが、「タ形」は増加した。その理由として第3パートでは帝に起こった出来事が語られるため「タ形」が多い。さらに、第4パートでは藤壺の入内や源氏の結婚など新しい人間関係が構築される箇所であるが、「タ形」から「ル形」へと語り口に変化が見られ、

語り手が視点を自在に登場人物へ移動したり、臨場感を持たせたりするために効果的に使用された。

(2) 第2パートにおいて「引用」が大きく増加し、その後第3第4パートにかけて緩やかに減少していった。物語の語り手や女房の口を借りるのではなく「引用」を通して更衣の死という大きな感情の流れが表現されていることが考察された。

(3) 「桐壺」の導入部の「引用」に関しては文末表現に大きな特徴は見られないが、登場人物の心情は「引用」で表現され、「引用」が少なくなると、それに代わり「語り」の「ル形」と「タ形」で作品に変化をもたせていることが明らかとなった。

先行研究においては語りの構造に対して何らかの数学的方法を適用して客観的な解析が行われていなかった。しかし本研究では人手で選んだキーワードに対して離散値系ウェーブレット多重解像度解析を適用することで、語り手の表現方法の一端を可視化し、客観的考察を与えた。

3. 助動詞の自動検出と統計解析

3.1 概要

『源氏物語』については、現在でも複数作者説など不明な点が多く、国文学の立場からの研究だけではなく、統計的な手法を用いた研究などさまざまな視点から研究が行われている。村上ら²¹⁾は、長年の歳月をかけて『源氏物語』全巻の単語を品詞分類し、さらに助動詞などの詳細な出現頻度を算出し、数量化理論Ⅲ類による統計解析を行っている。しかしながら、一般に文学作品の品詞分類には多くの歳月が必要なこと、また、予期されるように解析結果には多くの不確定性を含むため、その理解には作品の内容の理解が必須であることは言うまでもない。この種の研究をさらに発展させるには、コンピュータを用いた文学作品のテキストマイニング法の発達と統計的可視化手法の導入が必要不可欠と思える。

本章では、『源氏物語』の助動詞の検索および特徴抽出のプログラムによる自動化²²⁾を試みる。また、数量化理論Ⅲ類による統計解析にクラスタリングを組み合わせた方法を用いて、『源氏物語』の第1部の33帖の類似性を評価することで、本手法の有効性を考察する。

3.2 解析方法

3.2.1 『源氏物語』の構成

『源氏物語』は、54帖からなる長編物語である。その各帖は、全体の構成の観点から3部に分けられており、第1部：1～33帖、第2部：34～41帖、第3部：42～54帖に分けられている。特に第1部では、2つの系列が存在し、それぞれTable 2に示すように紫の上系と玉鬘系に分類できる。本研究では、第1部の1～33帖を解析に用いた。

3.2.2 テキストマイニング法

助動詞に対するテキストマイニングを行うためには、

Table 2 Two groups in the part 1 of the tale of Genji

紫の上系	1. 桐壺, 5. 若紫, 7. 紅葉賀, 8. 花宴, 9. 葵, 10. 賢木, 11. 花散里, 12. 須磨, 13. 明石, 14. 滯標, 17. 絵合, 18. 松風, 19. 薄雲, 20. 朝顔, 21. 少女, 32. 梅枝, 33. 藤裏葉
玉鬘系	2. 帚木, 3. 空蟬, 4. 夕顔, 6. 末摘花, 15. 蓬生, 16. 関屋, 22. 玉鬘, 23. 初音, 24. 胡蝶, 25. 螢, 26. 常夏, 27. 篝火, 28. 野分, 29. 行幸, 30. 藤袴, 31. 真木柱,

注目する助動詞の出現頻度を明らかにする必要がある。本研究では、対象となるテキストからプログラムにより助動詞を検索し、その出現頻度を算出した。本研究では、2文字以上の助動詞である「たり」、「けり」、「なり」、「べし」、「めり」、「さす」、「らむ」、「らる」、「けむ」、「まじ」、「まし」、「まほし」の12種を対象とした。助動詞の検索では、通常の名詞の検索と異なり、活用形を考慮する必要がある。例えば、完了や断定の助動詞である「たり」の場合には、「たら」、「たり」、「たる」、「たれ」を含めて検索する必要がある。また、助動詞においては文末に現れることが多いため、対象となる助動詞に「。」や「,」を加えて検索することで検出精度の向上を図った。加えて、助動詞の「なり」については、動詞の「なる」と同様の活用形を持つため、判別が困難である。例えば、1帖の桐壺に出現する「出でおはすものなりけり」の場合は助動詞の「なり」であり、「この御子三つになりたもう年」の場合は動詞である。そこで、判別が困難な単語については、その前後の単語から助動詞を判別した。あらかじめ1帖と4帖に出現する「なり」をマニュアルで判別した後にリスト化し、残りの帖に対してリストと照合することでプログラムによる検索を行った。解析の対象とした『源氏物語』のテキストは、渋谷²³⁾により公開されているローマ字表記されたものを使用した。

3.2.3 数量化理論Ⅲ類とクラスターリング法

テキストマイニングにより得られた各帖における助動詞の出現頻度を基に各帖の関係を評価するため、統計解析手法である数量化理論Ⅲ類により検討した。数量化理論Ⅲ類は、ある二次元の度数分布 p_{ij} が与えられたときに以下の式で定義される相関係数 ρ_{XY} が最大となるようにスコア x_i, y_j を求める方法である。

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} \quad (2)$$

ただし、 $\sigma_{XY}, \sigma_X, \sigma_Y$ は以下の式で与えられる。

$$\sigma_{XY} = \sum_{i=1}^n \sum_{j=1}^m x_i p_{ij} y_j - \sum_{i=1}^n \left(x_i \sum_{j=1}^m p_{ij} \right) \cdot \sum_{j=1}^m \left(y_j \sum_{i=1}^n p_{ij} \right) \quad (3)$$

$$\sigma_X = \sqrt{\sum_{i=1}^n \left(x_i^2 \sum_{j=1}^m p_{ij} \right) - \left\{ \sum_{i=1}^n \left(x_i \sum_{j=1}^m p_{ij} \right) \right\}^2} \quad (4)$$

$$\sigma_Y = \sqrt{\sum_{j=1}^m \left(y_j^2 \sum_{i=1}^n p_{ij} \right) - \left\{ \sum_{j=1}^m \left(y_j \sum_{i=1}^n p_{ij} \right) \right\}^2} \quad (5)$$

本研究では、度数分布として助動詞の出現率を用いた。

本解析により助動詞の出現率が類似した帖で同様なスコアが得られる。

解析によって得られた各帖の関係から同一の系統の巻を自動的に分類するために、クラスターリング手法であるk-means法を用いた²⁴⁾。この方法は、あらかじめクラスター数がわかっており、各クラスターを構成する要素数が同程度になるようなデータに対して適用可能である。クラスターの分類では、初めに各要素をランダムなクラスターに分類し、各クラスターを構成する要素の重心を求める。次に、各要素を最も近い重心のクラスターに分類する。これらの操作をクラスターの構成要素および重心位置が変わらなくなるまで繰り返すことで要素を分類する。上記の数量化理論Ⅲ類では、複数の成分に関する関係が得られることから、各成分の寄与率を考慮して重心からの距離を算出した。テキストマイニングにより得られる助動詞の出現頻度やクラスターリングによる各帖の分類の自動化の有効性を検証するため、村上らにより算出されたデータに対して同様の解析を行った。

3.3 結果および考察

3.3.1 テキストマイニングによる助動詞の検索

Table 3はプログラムによる自動抽出により得られた助動詞の出現頻度である。各巻の中での助動詞の出現頻度を比較するために、代表的な帖として紫の上系から1.桐壺、玉鬘系から4.夕顔を例として示している。また、第1部の1-33帖の合計を合わせて示す。括弧内の数字は、村上²¹⁾により求められた助動詞の数である。各助動詞で示されている検出誤差は、解析対象の33巻に対して本手法により検出された助動詞数と村上らの結果とのRMS誤差である。1文字の助動詞については、前後の文字だけでは他の品詞との区別がつかず、検出精度が悪いため本研究では2文字の助動詞に限定した。表より、「たり」、「けり」、「なり」、「べし」の出現数が他の助動詞に比べて多いことがわかる。検出誤差を見ると、「な

Table 3 Frequencies of auxiliary verbs

	たり	けり	なり	べし
1 桐壺	40(36)	57(56)	49(33)	34(35)
4 夕顔	135(132)	98(97)	113(97)	77(76)
1-33 帖	2157 (2075)	1630 (1637)	1423 (1650)	1292 (1302)
誤差(%)	1.1	0.4	4.3	0.7
	めり	さす	らむ	らる
1 桐壺	3(6)	24(21)	4(4)	3(3)
4 夕顔	32(32)	20(18)	19(19)	20(20)
1-33 帖	411(416)	437(322)	282(299)	298(304)
誤差(%)	1.0	8.0	2.3	0.9
	けむ	まじ	まし	まほし
1 桐壺	6(6)	5(5)	3(3)	6(6)
4 夕顔	15(14)	18(17)	5(4)	4(4)
1-33 帖	202(206)	230(212)	192(184)	78(79)
誤差(%)	1.4	2.7	2.0	1.3

り」と「さす」のRMS誤差が大きい。これは、他の品詞にも同じ音の語があるため、その判別が困難だからである。しかしながら、大部分の助動詞は3%以下の誤差で検出できている。

3.3.2 数量化理論Ⅲ類およびクラスタリング解析

Fig. 5は村上らが行った助動詞の出現数を基にした数量化理論Ⅲ類による解析結果である²¹⁾。ただし、1文字の助動詞も含めて出現数の多い21種の助動詞を解析に用いている。横軸と縦軸は、式(3)–(5)におけるスコアで、寄与率の大きいものから第1成分と第2成分を示している。また、紫の上系と玉鬘系とをそれぞれグループで分けて示している。図より、紫の上系と玉鬘系とは16帖を除くと2つのグループに分けることができ、助動詞の出現頻度に明らかな差があることがわかる。このように数量化理論Ⅲ類では、同様な傾向を持つデータが近くに現われるため、データの特徴抽出が可能である。しかしながら、村上らの結果におけるグループ分けは物語の内容を考慮して行われており、紫の上系と玉鬘系が重なり合った領域では予備知識なしに判別することは困難である。一般にこのような分類ではその作品に対する深い知識と理解が必要である。そこで、数量化理論Ⅲ類による解析結果を基に、プログラムによる自動的なグループ分けの方法について検討する。

Fig. 6は村上らのデータを基に1~33巻の助動詞の出現頻度に数量化理論Ⅲ類とクラスタリングを適用した結果である。数量化理論Ⅲ類による解析では、村上らの結果と同様の分布が得られた。図中の記号は、四角が紫の上系、三角が玉鬘系の帖を示しており、塗りつぶした記号はクラス1、白抜きの記号はクラス2に分類された帖である。クラスタリングによる分類との対応をみると、紫の上系の群と玉鬘系の分類に有意な差が見られる。村上らの結果(Fig. 5参照)において、2つの系が重なった領域に位置する帖を除けば、誤分類された帖は16帖と32帖の2つだけである。したがって、クラスタリングによる自動的な分類でも助動詞の出現頻度による

Fig. 5 Configuration of 33 stories in the part 1 of the tale of Genji ²¹⁾

Fig. 6 Clustering of stories 1-33 (Murakami's data)

Fig. 7 Clustering of stories 1-33 (present data)

各系の特徴抽出が可能である。

次に、助動詞の出現頻度の検出誤差が、解析に及ぼす影響について検討した。Table 3に示す12種の助動詞について、プログラムによる自動検出の結果を用いて数量化理論Ⅲ類による解析およびクラスタリングを行った。これは、真の助動詞の出現頻度に対してRMS誤差で平均2%の誤差を含んだ場合に相当する。Fig. 7は、自動検出した助動詞の出現頻度に対して数量化理論Ⅲ類とクラスタリングを適用した結果である。助動詞の出現頻度に誤差を含む場合でも、特徴的な巻の位置関係はあまり変化していない。Fig. 6の場合と同様に、村上らのデータにおけるあいまいな領域のデータを除けば、誤分類は5帖である。しかし、紫の上系と玉鬘系の大部分はそれぞれ異なるクラスとして分類されている。今後、1文字の助動詞の自動検出が可能になれば、さらにクラスタリングによる誤分類が減少すると考えられる。

3.3.3 助動詞の出現頻度と文の長さ

最後に、自動検出により得られた助動詞の出現頻度と文章の長さに注目し、『源氏物語』の各帖の特徴抽出を行った。具体的には、紫の上系、玉鬘系の帖について、助動詞「けり」とそれを含む文の文字数を評価した。助動詞の「けり」は、伝聞の過去を表わすものであり、そ

(a)Murasakinoue

(b)Tamakatsura

Fig. 8 Number of characters in each sentence with the auxiliary verb "keri"

の出現数を見ることで物語が現在のことなのか過去のことなのかを特徴づける指標になると考えられる。Fig. 8は、助動詞「けり」が含まれている文のみに注目した場合の1文あたりの文字数とその出現頻度の関係である。紫の上系からは1.桐壺と5.若紫(a)を、玉鬘系からは2.帚木と4.夕顔(b)の結果を示している。図中の括弧で示した数字は、その帖の文の数である。Fig. 8 (a)と(b)の紫の上系と玉鬘系を比較すると、「けり」を含む文の出現頻度に違いがみられる。すなわち、玉鬘系では「けり」を含んだ文の出現頻度が紫の上系よりわずかに少ない。また、紫の上系では文字数が10~100文字の文において「けり」の使用が見られる。一方、玉鬘系では10~40文字程度の比較的短い文章での出現頻度が高いものの、100文字以上の文においても比較的良好に用いられている。これらの特徴は、同じ系の中の異なる巻においてもみられる。

本研究では、『源氏物語』の助動詞のテキストマイニングおよび特徴抽出の自動化について検討を行った。プログラムによる検索により、2文字の助動詞についてはRMS誤差で平均2%の誤差で助動詞の出現頻度が検出可能であり、わずかに誤差を含んだデータにおいてもクラスタリングにより各帖の関係を自動的に分類できるこ

とを示した。

4. 物語のパターンを絵画的に可視化する方法

4.1 研究の経緯

山田と村井はこれまでウィリアム・シェイクスピアの戯曲を中心に、英文テキストから物語のパターンを可視化する研究を行ってきた³⁾。シェイクスピアの数々の戯曲と『源氏物語』は、ともに世界中の言葉に翻訳され、愛されてきた古典である。それゆえ、我々の新しい試みは、専門性の垣根を越えて、たちまち多くの方々から、その有効性と問題点とをご教授頂くことができた。可視化情報学会で研究成果を講演するたび、文理双方の分野から多面的な質問と助言を受け、それが貴重な刺激となり、本研究がここまで展開してきた。この少し変わった研究が、可視化情報学会の場で、大切に育てて頂いたことに関して、最初に深く謝意を示しておきたい。開始当初、この分野の先駆者であられる白百合女子大学の井波先生と堀井先生に、相談をもちかけ、激励のお言葉を頂いた。そこから、互いの考え方とスキルを提示し続けることで、独自の展開に向かって進んでいった。

4.2 電子テキスト時代の文学研究

電子テキストの普及の社会的影響は、論文閲覧からスマートフォンのアプリまで、今更、説明するに及ばない。著書『ゲーテンベルクからグーグルへ』²⁵⁾は、言葉の電子化がもたらす全ての可能性を予測している。その中に我々が対象とする古典文学の新しい批評も示唆されている。しかし最大の問題は、研究の対象が、人が創作したアートという点である。この点において、自然現象を相手にする理工学の電子化革命との決定的な違いが生じる。文学・芸術では、作者のメッセージや感情世界を汲み取るという過程に研究の醍醐味がある。しかし、それは普遍的でも客観的でもなく、読者には個別の自由な解釈が許され、むしろその範囲が大きいほど作品の魅力が大きいと言えよう。同じ文学作品を、十代の頃に読むのと、いま読むのとでは感じ方や捉え方が違う。比して、理工学の論文は、百人が読んでも一通りしか解釈できないように書くことを指導される。法律や特許では、一見厳密に見えるが、実際はその逆で、恣意的に複数の解釈が許されるよう作文することがある。このように電子テキストは、デジタル数値と異なり、それ自体は人が創り上げた言葉であり、意図する曖昧さや比喩などが含まれ、気が遠くなるほど難しい研究対象である。電子化されたところで、有意義な結果を導くのは難しいだろうという謙虚な姿勢を前提としている。

そのような背景で、我々が最初に取り組んだのは、ウィリアム・シェイクスピアの『オセロ』、『ロミオとジュリエット』などの著名な悲劇であった³⁾。これらは主として愛と死の対照をテーマとしており、可視化結果は中盤から悲劇に展開する明瞭な模様を描いた。その後、『ハムレット』や『マクベス』そして『リア王』を代表

とする悲劇に始まり, シェイクスピア全 37 作品の可視化を試みた。さらに, それにチャールズ・ディケンズの長編小説の作品に応用し, 小説の可視化結果は乱雑で解釈が困難なものとなった。しかし, 物語を複数の視点で見る⁴⁾ことで, 幾多のサブテーマの同時進行が読者を楽しませている様子が立証された。解釈を拡大すれば, 読者の立場や心境によって見え方が異なるという, 文学の多面性に存在する価値を裏付けたことになった。これを応用した最新の研究では, あえて可視化対象を, 新聞記事や取扱説明書にも広げ, 文学作品と比較した。その結果, 文学作品だけに存在するいわば構造色のようなものがあることが示された^{26), 27)}。それは, 愛ゆえの死, 愛が引き起こす狂気や憎しみなど, 登場人物の人間心理や, 物語の異様な世界観を電子テキストから抽出するという試みに光を与えるものとなった。そして『源氏物語』はまさにそのような心理描写を豊富に含んでおり, 挑戦的な可視化ターゲットと言える。

4.3 『源氏物語』の電子テキスト解析

『源氏物語』は, これまで千年もの間, 日本で愛され, 多数の翻訳, 映画, 演劇, 漫画等が生まれ, 海外にも多くの翻訳本を生み出し, 多様な解釈も論じられている。その電子テキストは教育研究向けに一般公開されており, 我々はそれをコンピュータプログラムで読み込み, 物語の可視化に挑んだ。これまで英文文学作品で成果を出してきた我々のアプローチが『源氏物語』に通用するかという試験的な位置づけである。ここでは『源氏物語』全 54 帖から, 若かりし源氏の淡い恋心を描いた「空蟬」(第 3 帖) および, 源氏の父である帝と母である桐壺更衣の相思相愛の物語「桐壺」(第 1 帖) を取り上げる。

解析方法の概略を説明するならば, 可視化のテーマに関連したキーワード群をテキストから検索し, その頻度分布を正方形のキャンバス上に描くというものである。具体的には, まず作品のテーマを表す何組かのキーワードを検索し, それらの座標から物語の内容の経時変化を画像表象する。キーワードの検索にあたっては, 予め作品中の全単語を全自動で配列として記録する。このとき英文ではスペースや句読点を単語間の区切りとして認識可能である。例えば “We visualize a story.” だとスペースとピリオドの位置から判断して 4 単語から構成されるのがわかる。これに対して, 日本語では単語間でも文字が連続するため, 単語への分解は高度な自然言語処理を必要としてしまう。例えば「私達は物語を可視化する。」という一文は, 意味を理解せずに単語に分解することは不可能である。そこで本研究では, 単語分解ルーチンをスキップし, キーワードを連続文字列として直接検索するという暫定的方法を採用した。先の例の場合, 「物語」という単語を検索するとき, 「物」「語」の 2 文字が連続する部分を全文字走査で見つける。これにより構文解析をすることなくキーワード検索が可能となる。無論, 厳密な単語の抽出は, コンピュータ言語学や自然

言語処理において実現している技術である。しかし我々の可視化では, キーワードの分布さえ取得されれば, それ以上にテキストを分析する必要はない。このことにより, 英語と日本語の間でのテキスト解析の難易度はほとんど差異がなかった。換言すれば, 本可視化手法は対立概念をもつキーワード群の登場頻度分布のみで可視化するゆえ, 意味の解釈に立ち入ることなく文学作品を可視化することができるのである。なお, キーワード検索に関係する日英の言語間の差異として, 日本語における主語の省略だけは歴然としたものがあった。これは『源氏物語』の古文から現代文への翻訳や, 日本語から世界の言語への翻訳において大きな影響をもっている。翻訳にまつわる研究については末尾にも述べるが, 本研究の応用の一つになっている。

4.4 絵画的な表現へのプライオリティ

山田と村井のアプローチの最大の特徴は, 結果の表象の方法である。それは 2 つの, いわば個性的な手段から成る。ここで個性的と表現する理由は, それらが主観に基づく選択であるためである。まず一つは, 二次元のスクリーンにキーワード分布を展開することである。本来, 文学作品は, 読書の形式, 舞台の形式, ドラマや映画の形式のいずれでも, 始まりから終わりまで, 時間軸を 1 つだけ持っている。しかし人為的に時間軸を二次元に拡張するという方法を採用。この点は毎回, 議論を呼んでいるため, 本紙面を借りて説明させて頂きたい。時間というものは情報としては一次元を越えるものではない(物理学の紐理論や相対論を除く)。しかし我々は, 可視化の結果を絵画的に示したいという最終形を先に定義する。そのため時間を疑似二次元に展開する。時間の二次元化の理由付けは, もはや自然科学も客観性もなく, 我々の主張でしかない。これを以下のような様々な例えで, その主張を受けいれて頂いているのが現状である。例えばモノクロの図を輝度にあわせて疑似カラー表示することは, CT や MRI による人体の組織の可視化で常套手段になっている。これは見た人が視覚的に把握しやすくなるという利点の優先にほかならない。この場合, 輝度という 1 成分の情報を RGB の 3 成分に人工的に増殖している。対象がこと時間軸であるとき, 同じような拡張でも非日常的と解釈され理解が得られにくい。しかし以下の例はどうだろうか。1 本の毛糸でセーターを編むのは, 一次元の材料から三次元の品物をつくっているではないか。毎日見る壁時計でも, 長針と短針という 2 つの組み合わせで解りやすく表現されている。それらはいずれも, 本来は一次元である量を, 人の手により二次元に拡張し, その結果, 人に分かりやすさを提供している。数学の世界でさえ, 実数から複素数に拡張することで, 諸現象の豊かな記述能力を獲得している。このような次元の拡張は, 人による人のための操作であり, その定義の客観性や普遍性を論ずる領域にない。本研究の場合, 可視化の対象が, 人によって作られた文学であり,

その結果の表象でも、我々が主観的に決めるルールを適用しても違和感を唱えるべきではない。結果が有効であれば、そのルールもまた機能していると言えるのである。

もう一つの個性的な手段は、絵画や写真から抽出したカラーバーを採用して数値データを色彩化することである。標準的なレインボー（色相角が一定速度で一巡するカラー表示方法）では、カラフルであっても無機質な印象しか残さない。試行錯誤を繰り返すうちに我々が経験したことは、絵画からの色コードのサンプリングにより、その絵のもつ情景が、結果に上乘せされることである。また、風景画などの写真からの色コードのサンプリングは、我々の日常視覚認識で自然に機能する知覚作用、立体認識、陰性残像などの効果を発生させた。いわば錯覚を誘導し、結果の視覚認識で趣が生ずるのである。これらの事は、美術や彫刻における色彩心理と立体感の認知科学として知られる。CGの世界でも、透明性の表現は高度の技術が必要とされるが、それもまた、単なる幾何光学だけでなく色彩認知の脳内処理と連動しているのである。要するに、人は色そのものだけでなく色の縁辺対比、つまり少なくとも色の空間勾配情報を常に処理しているのである。このような理由から、絵画や写真から抽出したカラーバーの導入は、創作としての可視化という付加価値を与える。これまで多くの方から、可視化結果がとても綺麗で印象的だと感想を頂いている。

以上の2つのアプローチを組み合わせることで、作品がもつ物語のイメージを表現する。その先にある最終的な狙いは、文学作品の中に入り込んで、物語を追体験する読者が抱く風景そのものを、豊かな色彩とともに絵画的に再現することである。我々はこれを、作品に潜む印象世界の顕在的な再可視化(art-to-art)と位置づけ、なおも発展性が大きい研究であると認識している。これは客観と主観を対立視せず、融合させるという文理横断・学際研究の特徴と言えるだろう。

4.5 「空蟬」および「桐壺」の可視化

『源氏物語』から、「空蟬」および「桐壺」の可視化の結果の例を紹介する。我々のアプローチは、まず各作品における対立概念を構成するキーワード群を選考し、その後、電子テキストからその二次元化された分布を得て、指定するカラーバーをもちいて絵画的に可視化する。この間の処理時間は、数万の単語から成る作品でも10秒に満たない。そのためキーワード群の選定が適切かどうかを試行錯誤判定するのもまた、数分で完了する。ごく最近の研究では、この試行錯誤の過程をもコンピュータに任せ、我々が知恵を絞ってキーワード群を選ぶという作業を省略させることもできるようになりつつある。

ここでは「空蟬」について、2つのテーマで可視化を試みた。1つは、空蟬に思いを寄せる源氏、源氏の恋を応援する空蟬の弟（小君）、そして抵抗する空蟬による主要人物登場の可視化である。そしてもう1つは、空蟬への恋心と冷たい態度への恨めしい気持ちを表す源氏の

Fig. 9 Visual imagery from “Utsusemi” painted by Yoshimitsu Tosa, regarding contrast of Men and Women

Fig. 10 Visual imagery from “Kiritsubo” painted by Mitsunori Tosa, regarding contrast of Men and Women

Fig.11 Imagery from (a) “Utsusemi” and (b) “Kiritsubo”

相反する感情を抽出した解析である。

Fig. 9 では、源氏や小君による「男性」と空蟬による「女性」の登場頻度をテーマとして解析した。最初、男たちが女に近づこうと働きかけ（ピンク色）、それに女が反応し（緑グレー）、中盤から後半にかけて男女が呼応する可視化結果が得られた。この可視化結果は、空蟬に思いを寄せる源氏が弟の小君の協力を得て、空蟬を垣間見、さらに近づこうとするが、空蟬の抵抗に合い、逃げおおせられ（空蟬は蟬が脱皮するかのごとく薄衣を脱ぎ、源氏から逃げる）、口惜しい気持ちを露わにする物語内容と合致している。もう1つのテーマは、源氏の空蟬への思いをより端的に示したもので(Fig.11(a))、「恋」「愛」などのキーワードが「冷淡」「恨めしい」などの相反する気持ちに、最終的に混在しているのを見ることができる。つまり、空蟬の冷淡な行動がさらに源氏の恋心をつのらせていることがわかる。

「桐壺」に関しては、源氏の父親である帝と母親である桐壺更衣、そして桐壺亡き後、桐壺の瓜二つという理由で帝の下に嫁ぐ藤壺の登場人物の可視化を試みた。さらに、帝が桐壺に対して抱く「愛」や桐壺の「美」をポジティブキーワードとし、対して、帝の周りにいる女性たちからの「嫉妬」や桐壺の「死」といったネガティブキーワードで可視化を行った。

Fig.10 において、最初、桐壺（緑グレー）と帝（朱色）が出会い、すぐに混在する。中盤で、桐壺が亡くなり、帝の気持ち描写が増え、終盤に新たな藤壺（緑グレー）が登場する。別テーマで帝の桐壺に対する思いを見てみる(Fig.11(b))。最初、帝の桐壺の相思相愛の朱色から桐壺の死（緑グレー）、そして帝の悲しみを経て、帝は桐壺への一層の愛の気持ちを抱くことがわかる。可視化結果と「桐壺」の物語内容が一致していることが証明できる。

4.6 今後の展開

ここでは、著者らが開発してきた英文学作品用の文学可視化解析プログラムを、日本語テキスト用に書き換え、『源氏物語』の物語世界を絵画的に可視化した例を示した。その際、日本語の原文は単語への分解が容易でないが、構文解析を経由せず物語のテーマを構成するキーワード頻度分布を高速検出できるように工夫を行った。この応用の結果、冷たさゆえの恋心、死ゆえの愛の高まり、など『源氏物語』の物語構想のパターン抽出に概ね成功したと判断できる。本記事で掲載した結果は、『源氏物語』を対象とする文学の研究としては初歩的なものであるが、今後は、その様々な翻訳によって生み出される差異の可視化に取り組む予定である。すなわち古典 (Teika Genji)、現代日本語訳 (Yosano Genji)、英語翻訳 (Seidensticker Genji) の比較解析を行い、各言語の特色による細かな文章構造の違いを可視化する。その一部の成果は、可視化情報学会全国講演会にて発表した²⁸⁾。この成果からは、特に、古文のあいまいさ（主語や形容詞などの欠落や多義語の存在）が読み手（翻訳者）に解釈の自由度を与えていることが明らかになり、このことが『源氏物語』の翻訳に多様性を与え、多くの人々を魅了し続けている要因の一つと見なすことができると推論された²⁹⁾。

古典文学の『源氏物語』が抱く世界観は時空を超えて今も広がり続けている。一方、電子テキスト時代に生まれた本研究の文学可視化スタイルは、これまでの通説を確認する客観的診断方法として機能するだけでなく、解釈の多様性の発見や、作品の個性を表象するグラフィカルなレジュメとして応用が広がるものと期待している。

5. おわりに

本稿では、文学作品の可視化により得られる情報について、『源氏物語』を対象として3つの可視化手法を紹介した。それぞれの手法において、その目的や得られる情報は異なるものの、作品の理解や解釈に従来とは異なる新しい視点を加えるという点では共通している。また、本稿で紹介した可視化手法は、『源氏物語』を例とする古典文学だけではなく、現代文学や英文学などさまざまな文学作品に適用可能であり、今後の文学作品の理解に有用なツールとなっていくことが期待される。

参考文献

- 1) 井波真弓, 齋藤兆古, 堀井清之: 『源氏物語』における源氏と醜女末摘花の恋, 可視化情報, Vol. 29, Suppl. No. 1 (2009) pp. 329-332.
- 2) 井波真弓, 齋藤兆古, 堀井清之: 『銀河鉄道の夜』におけるオノマトベ, 可視化情報, Vol. 31, Suppl. No. 1 (2011) pp. 207-212.
- 3) Yamada, M., Murai, Y.: Story Visualization of Literary Works: How a Computer Reads Shakespeare's Plays, J. Visualization, Vol. 12 (2009) pp. 181-188.
- 4) Yamada, M., Murai, Y.: Stereoscopic story visualization in literary works demonstrated by Shakespeare's plays, J. Visualization, Vol. 13 (2010) pp. 355-363.
- 5) 細井尚子: 源氏物語における助動詞に基づく特徴抽出, 可視化情報, Vol. 31, Suppl. No. 2 (2011) pp. 97-98.
- 6) 堀井清之: 線形空間論を用いた文学作品における暗黙知の可視化, 可視化情報, Vol. 21, Suppl. No. 1 (2001) pp. 774-782.
- 7) 井波真弓, 齋藤兆古, 堀井清之: 『源氏物語』第四帖「夕顔」における「あやし」と美的理念, 可視化情報, Vol. 28, Suppl. No. 1 (2008) pp. 461-464.
- 8) 井波真弓, 岩崎晴美, 宮沢賢治, 土屋宏之, 齋藤兆古, 堀井清之: 『源氏物語』における源氏と空蟬の恋ウェブレット多重解像度解析, 可視化情報, Vol. 24, Suppl. No. 1 (2004) pp. 211-214.
- 9) 井波真弓, 齋藤兆古, 堀井清之: 『源氏物語』における源氏と藤壺の涙, 可視化情報, Vol. 29, Suppl. No. 2 (2009) pp. 31-34.
- 10) 『和歌文学論集』編集委員会[編者]: 和歌と物語, 風間書房, (1993) p. 1.
- 11) 広田収: 桐壺更衣の物語と歌の位相, 人文学 (143), 同志社大学人文学会 (1986) pp. 104-117.
- 12) 藍美喜子: 『源氏物語』語り手の横顔, 甲子園短期大学紀要 20, 甲子園短期大学 (2002), pp. 101-108.
- 13) 鈴木泰: 改訂版古代日本語同士のテンス・アスペクトー源氏物語の分析ー, ひつじ書房 (1992).
- 14) 金田一春彦, 林大, 柴田武編: 日本語百科大事典, 大修館書店 (1988) pp. 200-201.
- 15) 守谷三千代: 小説の中の視点と文法ー時制と相を中心にー, 早稲田大学日本語研究教育センター紀要 4, 早稲田大学, (1992), pp. 98-120.
- 16) 石出靖男: 夏目漱石の小説におけるタ形文末・非タ形文末の表現効果, 表現研究 (93), 表現学会 (2011) pp.1-10.
- 17) 阿部秋生, 秋山虔, 今井源衛, 鈴木日出男: 新編日本古典文学全集 20 源氏物語①, 小学館 (1994).
- 18) 堀井清之, 齋藤兆古: 特許「文学作品解析方法および解析装置」, 特願 JP10-102673A.
- 19) 齋藤兆古: 画像処理入門, 朝倉書店 (1998) pp.40-42.
- 20) 齋藤兆古: ウェブレット変換の基礎と応用ーMathematicaで学ぶ, 朝倉書店 (1998) p.39, pp.93-95.
- 21) 村上征勝, 今西祐一郎: 源氏物語の助動詞の計量分析, 情報処理学会論文誌, Vol. 40, No. 3 (1999) pp. 774-782.
- 22) 細井尚子, 山縣貴幸, 五十嵐裕也, 藤澤延行: 統計解析に基づく源氏物語の特徴抽出と可視化, 可視化情報, Vol. 32, Suppl. No. 1 (2012) pp. 361-362.
- 23) 渋谷栄一: 源氏物語の世界, <http://www.sainet.or.jp/~eshi-buya>, (参照日 2013 年 4 月 1 日).
- 24) Hartigan, J. A.: クラスタ分析, マイクロソフトウェア (1983) pp. 109-145.
- 25) Shillingsburg, P. L.: From Gutenberg to Google: electronic representation of literary texts, The University of Cambridge Press, Cambridge (2006)
- 26) 山田美幸, 村井祐一, 熊谷一郎: 多様なジャンルにおけるテキストパターンの可視化と文学作品の特徴について, 可視化情報, Vol.31, Suppl. No. 1 (2011) pp. 213-214.
- 27) Yamada, M., Murai, Y., Kumagai, I.: Story visualization of novels with multi-theme keyword density analysis, J. Visualization (in press, 2013) doi:10.1007/s12650-013-0163-4.
- 28) 山田美幸, 村井祐一, 熊谷一郎: 「源氏物語」における日本語テキストのストーリー可視化, 可視化情報, Vol.32, Suppl. No. 1 (2012) pp. 359-360.
- 29) 山田美幸, 熊谷一郎, 村井祐一: 「源氏物語」と「The Tale of Genji」のストーリー可視化: 日英翻訳テキストの可視化比較について, 可視化情報, Vol.32, Suppl. No.2 (2012) pp. 149-150.