

## 発表要旨

### 『細雪』の計量テキスト分析

#### A Quantitative Textual Analysis of “The Makioka Sisters”

小谷内 郁宏

KOYAUCHI Ikuhiro

(2021年11月4日受理)

#### 要約

『細雪』(1948年)は、谷崎潤一郎にとって最長の長編であり、上巻(29章)、中巻(35章)、下巻(37章)、文庫本三分冊になる。そのテーマについて、CiNii(学術情報データベース)サイトにて、「細雪」をキーワードに論文検索すると、姉妹の関係性について、あるいは四姉妹の内のひとりに焦点を当てたテーマ設定が多く見受けられる。

この長編小説の冒頭は、次女幸子が背後から来た四女妙子に「こいさん、頼むわー」と化粧の手伝いを頼む場面に始まり、最終場面では三女雪子の縁談もまとまり、その準備のため東京に向かう車中、体調をくずし「下痢はとうとうその日も止まらず、汽車に乗ってからまだ続いていた」と語られ、結局幸子と悦子に始まり、雪子で終わる。そこには長女鶴子はおらず、この物語が次女幸子を惣領とし、三女雪子、四女妙子をめぐり、チェーホフ『三姉妹』(1900年)に通じる、実質三姉妹的な関係性に基いている。

この発表は、その仮定を裏付けるため、計量テキスト分析ソフトHD Coderを用い、『細雪』のテキストを分析、解釈する。

- I. 『細雪』について
- II. 計量テキスト分析
- III. まとめ

## I. 『細雪』について

谷崎は第二次世界大戦中の1941年に谷崎潤一郎訳源氏物語を完成させた。

そして1943年、月刊誌『中央公論』1月号と3月号に『細雪』の第1回と第2回が掲載された。

東京日本橋生まれの谷崎だが関西に憧れを抱き、関西出身の夫人松子、義姉、義妹たち四姉妹の生活を題材にして長編小説化した。

当時軍部からは「内容が戦時にそぐわない」として6月号の掲載は止められた。谷崎はそれでも執筆を続け、1944年(昭和19年)7月には私家版の上巻を作り、友人知人に配ったが、それも軍により印刷・配布を禁止された。中巻も完成したが出版できなかった。疎開を経て、終戦後は京都の鴨川べりに住まいを移し、1948年に作品を完成させた。

作中には年号の表記がなく、正確な年代は不明である。しかし、作中において四季の移り変わり、阪神間を襲った大きな気象災害(土砂災害)が克明に描かれているため、この作品は日中戦争勃発の前年1936年秋から日米開戦の1941年春までの設定とされている。

### 1.1 登場人物

大阪船場の商家である蒔岡家の四姉妹が中心である。

- 長女鶴子、本家の奥様であり、夫は婿養子の銀行員辰雄。
- 次女幸子、谷崎の妻松子がモデル。分家の奥様、「御寮人さん」と呼ばれる。夫は、婿養子で計理士の貞之助。そして娘悦子がいる。
- 三女雪子、優柔不断で「雪姉ちゃん」と呼ばれる。小説中、五人と見合いをする。
- 四女妙子、自由奔放で、「こいさん」(末娘の意)と呼ばれる。三人と恋愛をする。

その他、雪子に見合いを紹介する重要人物として行きつけの美容院の女主人に井谷がいる。

男性陣としては、妙子の恋人、貴金属商の三男坊の奥畑、妙子の恋人で写真師の板倉、バーテンダーの三好の三人がいる。

そして雪子の縁談相手として、順に瀬越、野村、沢崎、橋寺、御牧の五人である。

### 1.2 あらすじ

#### <上巻>

蒔岡家は大阪の中流上層階級の商家であり、その本家は長女の鶴子、その夫で婿養子となり蒔岡の名を継いだ辰雄、6人の子供の一家である。

分家は阪神間の閑静な郊外、芦屋にあり、次女の幸子、同じく婿養子となり蒔岡の名を継いだ夫の貞之助、幼い娘の悦子の3人家族である。鶴子と幸子には雪子と妙子という2人の未婚の妹がおり、本家と分家を行き来している。

冒頭、雪子は30歳にして未婚であることが語られる。以前、妙子が奥畑と駆け落ちをし、妙子と間違えられて雪子の名が地元の新聞に載ってしまったことがある。

記事は蒔岡家の恥となり、雪子のみならず妙子の名をも汚すことになった。妙子は人形作りに慰めを見出すようになる。

井谷が雪子に瀬越という男性との見合い話しを持ってくる。一家はこの縁談に乗り気になるが、調べたところ、瀬越の母親の遺伝性精神病が判明したため、断わる。

数か月後、幸子に友人からの雪子の縁談があったが、相手は中年やもめの野村であった。幸子は、外見が老けていることもあり、乗り気になれなかった。

一方で長女鶴子の夫、辰雄が支店長として東京に転勤になり、本家は東京渋谷に引越すことになる。妙子は人形作りの忙しさを口実に芦屋に留まり、雪子は行くことになった。

雪子は野村との縁談に乗り気ではないものの、芦屋に帰る口実として見合いに同意する。

雪子は野村の外見や身の上に対して何も不満を感じることはなかったが、無神経さに嫌気がさし、野村の求婚を断わり、雪子はまた東京に送り返される。

#### <中巻>

写真師の板倉は妙子の人形の写真を撮っており、妙子と知り合いである。

関西地方が大洪水に襲われた。妙子は洋裁学院にいたが、板倉が妙子を救出した。そして妙子は好意を持つようになる。やがて妙子と板倉の関係は幸子に知られるが、幸子は板倉の低い出自のため反対するが、妙子は彼と結婚するつもりでいる。

妙子は洋服店を開くことに決め、本家に資金援助を依頼するため東京に行くが、板倉が病気になり大阪に呼び戻される。板倉は入院していたが、手術の合併症からくる壊疽のために亡くなる。板倉の死で、妙子が低い出自の男と結婚するのではという幸子の憂いはひとまず消えた。

### <下巻>

辰雄の長姉が幸子に雪子の縁談を紹介する。名古屋の名家の出の沢崎である。幸子、雪子、妙子と悦子は辰雄の姉を訪ねて大垣に行き、そこで雪子は見合いをするが、結局沢崎の方から断ってきた。

幸子は妙子がまた奥畑と縫いを戻したと聞き、鶴子は妙子に東京に来よう申し渡すが、妙子は拒否し絶縁される。

井谷が雪子に縁談を持ち込む。この橋寺は魅力的な相手だったが、雪子の引っ込み思案のため橋寺は縁談を打ち切ってしまう。

その後、幸子は妙子が奥畑の家で重い病気になっているとの知らせを受ける。妙子は赤痢と診断されるが、結局妙子は一家の友人の病院に移され、緩やかに回復する。

一方で、幸子は妙子が三好というバーテンと関係があるらしいとも聞く。幸子は驚くが、妙子と奥畑は結婚するしかないと考える。

井谷は幸子に、雪子の縁談があると告げる。

公家華族の御牧子爵の庶子45歳で、東京で御牧に会ってみると、その気さくな人柄にたちまち魅了される。

妙子は、三好の子を妊娠しており4か月だということを幸子に告げる。幸子と貞之助は妙子が有馬で秘密裏に出産するよう手配するが、結局妙子は死産し、三好と所帯を持つ。そして最後に、蒔岡家は御牧家から求婚への返答を求められ、雪子はそれを受け入れる。

## II. 計量テキスト分析

### 2.1 総抽出語数と異なり語数

下記表1の縦軸見出しの「異なり語数(使用)」とは総抽出語数に対し異なる語彙の数であり、すなわち異なり語数では「幸子」という語彙が何度出てもそれは1語とカウントされる。(使用)とは、機能語(助詞、助動詞など)を除いた語数を示す。計算では、(使用)語数が用いられる。

「文」とは、句点「。」で区切られた文がいくつあるかということである。

「段落」は、ファイル内の改行マークごとの部分を段落として数える。

今回の手順では、基本テキストとして青空文庫より『細雪』のテキストファイルをダウンロードした。比較例として、太宰治の中編小説『人間失格』を挙げたが、文数で見ると『細雪』は『人間失格』の5倍弱であり、異なり語数(使用)は3倍になる。

前処理を終えた後、各章ごとにKH coderで文字数を計量し、最後に各章のファイルを合体し、全体の語数を数えた。

ちなみに、全体で「異なり語数(使用) / 総抽出語数(使用)」 $12,876/13,454 \approx 0.095$

表1

	例： 人間失格	細雪 上巻	細雪 中巻	細雪 下巻	細雪 全体
総抽出語数(使用)	46,655 (16,359)	98,614 (35,331)	128,719 (46,226)	148,343 (52,539)	376,211 (134,454)
異なり語数(使用)	4999 (4442)	7,115 (6,407)	8,232 (7,487)	8,807 (8,042)	13,907 (12,876)
文	1,600	2,537	3,023	3,206	8,806
段落	811	1,491	1,431	1,471	4,431

(9.5%) という数値が書き手の語彙率となり、語彙力の豊かさのひとつの指標ともなる。各章のボリュームを総抽出語数から算出すると、「上巻」26%、「中巻」34%、「下巻」40%となる。「上巻」に比較して、「下巻」の総量が大きいのは、雪子、妙子の縁談、恋愛のエピソードが多いこと、「中巻」は大水害に関する記述が多いことが考えられる。

## 2.2 クレンジング

前処理の第一段階として、ファイルの前処理が必要となる。青空文庫からテキストファイルを落とし込む場合、漢字のルビも本文に入り、それらを一つひとつ除去しなければならず、難読漢字を使用する作家であればあるほどルビも多く、その除去作業が余計に必要となる。

## 2.3 コーディング

テキスト解析をする際、コーディング作業が必要となる。例えば、文中の人名で「幸子」と「御寮人」と「中姉ちゃん」は同一人物、「妙子」と「こいさん」も同一人物であり、予備作業として両単語を関連づけることで解析作業の精度を上げる。本発表の場合も、人名を中心にその作業を行った。

下記例のように、テキスト上でその人物に関連する単語がある場合、or で繋ぎ、そしてできあがった一覧表をテキストファイル化し、KH Coder に読み込ませる。

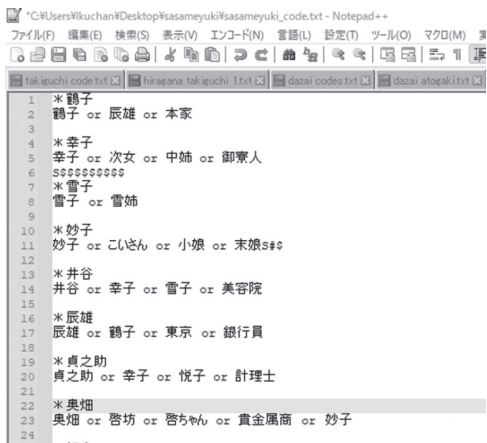


図1 コーディング

## 2.4 Jaccard係数と特徴語

次に、『細雪』の構成を調べるため、上中下の3巻において、特に多く出現している言葉、すなわちそれぞれの章を特徴づけていることばを上位10ずつリストアップする。「ツール」→「外部変数と見出し」→「特徴語」の項を選択しリストアップする。下記図2の数値はそれぞれの語と3巻との関連をあらわすJaccardの類似性測度をもって、値が大きい順に10語が選択されている。

末吉 (『テキストマイニング入門』, p.214) によると、Jaccard係数はKH Coderに主たる分析手法として採用されている。それは、「ある語A」と「ある語B」の関連性（類似性・共起性）の程度を表わす係数で、数式でいうと(Jaccard係数 = 「A」と「B」の両方が同時に出現した回数 / 「A」と「B」のどちらか片方だけでも出現した回数) であり、0から1の値を取り、その値を比較することで共起性や類似性の大きさを相対的に比較するための1つの指標となる。

目安として、「0.1」以上で「関連がある」、「0.2」以上で「強い関連がある」そして「0.3」以上で「とても強い関連がある」とされる。これらの語は単なる頻出語ではなく、各章をそれぞれに特徴付けている語と言える。

	上巻		下巻		中巻
雪子	.115	幸子	.142	妙子	.109
貞之助	.072	雪子	.135	こいさん	.082
人	.068	妙子	.091	自分	.079
姉	.064	こいさん	.082	悦子	.078
悪い	.045	人	.079	板倉	.068
本家	.035	自分	.077	家	.059
子供	.035	貞之助	.070	風	.052
一つ	.032	又	.069	東京	.044
井谷	.030	家	.060	奥畑	.040
夫人	.025	子	.055	外	.039

図2 各巻における10の特徴語

上巻においては、「雪子」が0.115で最も関連が強い。

中巻においては、「妙子」と「こいさん」が同一人物、0.185で高い。下巻においては、「幸子」が0.142、「雪子」が0.135、「妙子」と「こいさん」は同一人物で0.173である。あと、男性では上巻、下巻で妻幸子とともに活動的な夫、計理士「貞之助」の係数は高い。

## 2.5 四姉妹の登場回数

四姉妹名を比較検討すると、長女鶴子の回数は非常に少ない。最も多い雪子の60分1程度（2434対43）である。このことから、鶴子が本家の別格な人間であること、鶴子家族が転勤で東京に引っ越してしまうことで、上巻から下巻にかけて、鶴子が三人とは物理的、心理的関係性の薄い人物設定となっている。

上巻を見ると、幸子と雪子の回数が多く、妙子は半分程度である

しかしながら、中巻では幸子の1.5倍程度である。この点からも中巻は妙子の巻といって差し支えないだろう。

下巻では、幸子と雪子が500回程度で拮抗している。

三巻を通しての回数は、幸子と妙子が1400回程度、雪子は1100回程度であり、幸子の作品における重要性は想定内であるが、妙子が雪子を凌駕している点は作品解釈の新しい視点ともなり得る。

総じて、上巻は幸子と雪子の巻、中巻は妙子の巻、下巻は三人の巻と言える。

全体的に見ると、鶴子を除いた三人は同等

であるが、活動的に生きる幸子と妙子、一方受身的に人生を進めている雪子は、三人の中では登場回数が少ない。

		上巻	中巻	下巻	上中下巻	個人別
鶴子	長女	19	10	14	43	43
幸子	次女	376	417	517	1310	
御家人		4	16	14	34	
中巻		20 (400)	30 (483)	23 (554)	73 (1417)	1417
雪子	三女	393	210	476	1079	
雪飾		23(416)	13 (223)	8 (485)	45 (1124)	1124
妙子	四女	158	412	327	897	
こいさん		64 (222)	251(663)	250 (577)	565 (1482)	1482
悦子	幸子の娘	150	200	118	468	468

図3 四姉妹の登場回数

## 2.6 クロス集計マップ

次に、各巻と登場人物の相互関連性のクロス集計マップを挙げる。横軸に各巻の並びがあり、縦軸には登場する人物がある。

見方としては、中央の上巻では、「幸子」と「雪子」の四角の面積が大きく、右の中巻では「幸子」と「妙子」の面積が大きい。そして左の下巻では鶴子以外の三人の面積がほぼ同等に大きい。四角の面積が大きいことは、ある語Aとある語Bとの関連性（類似性・共起性）を示すJaccard係数が各巻において大きいことを示す。

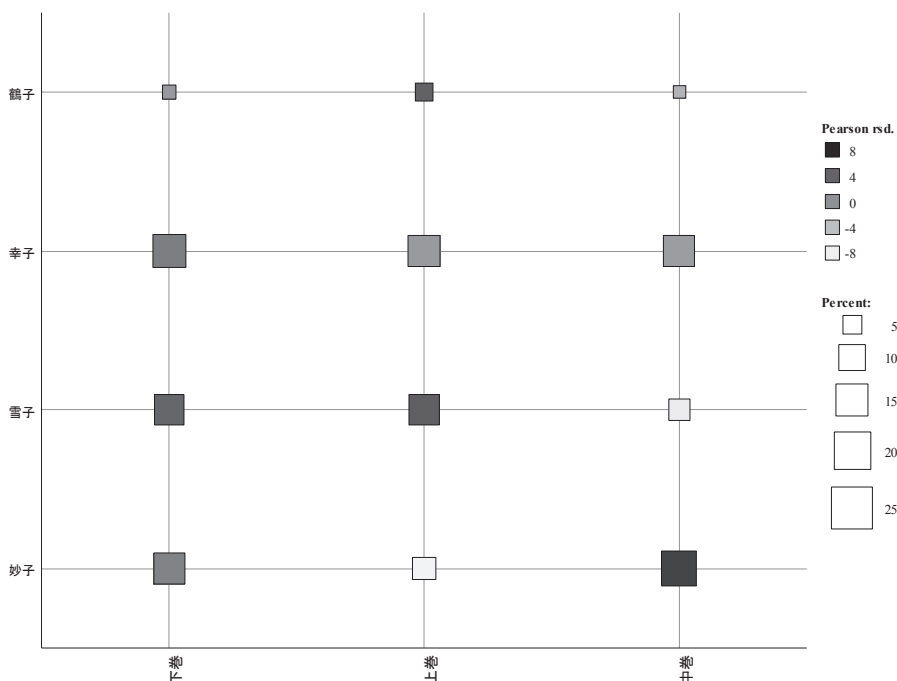


図4 人物と巻のクロス集計マップ

## 2.7 対応分析

次に、空間的に各巻を内容的な隣接関係で配置し、登場人物も各章での影響力の大きさ、すなわちそのJaccard係数の大きさと関係の隣接関係で配置する「対応分析」を提示する。

縦横二成分が交差する原点は、時系列、空間的、意味的な出発点と言える。原点より離れれば離れるほど、親密度、関係性は薄くなっていく。

原点に近い場所（空間）は、近いところから「蘆屋」そして「船場」であり、遠くに「東京」がある。時系列は、Y軸下から上巻、中巻、下巻となり、上に行けば行くほど時は下る。

原点に近い人物、それは鶴子、幸子・貞之助夫婦そして蔦岡家である。

右上、中巻には、妙子を中心に奥畑、板倉、三好の三人の恋人たちがいる。

左上、下巻周囲には、雪子を中心に瀬越、野村、沢崎、橋寺、そして遠くに夫となる御

牧がいる。そして、上巻と下巻の間に雪子に瀬越、橋寺そして御牧と三人の縁談相手を紹介する美容院の女主人井谷が重要人物として、大きな円として示される。

## Ⅲ. まとめ

『細雪』は、大阪船場、兵庫蘆屋、東京渋谷を舞台とし、三女雪子の五つの縁談、四女妙子の三人との恋愛、次女幸子の妹たちに対する献身を描いた作品である。

長女鶴子は、船場、薪岡家を体現する人物だが、あくまで「天の声」的に登場し、やはり幸子、雪子、妙子の三人こそ「三姉妹」的の分身である。

たとえ事の顛末が時に揺れても、経過ともに大団円的、予定調和的に収束していく。

『細雪』は、松子夫人の四姉妹をモデルとしているが、その原型は明治、大正期のロシア文学の流行時、谷崎自身も読んだに相違な

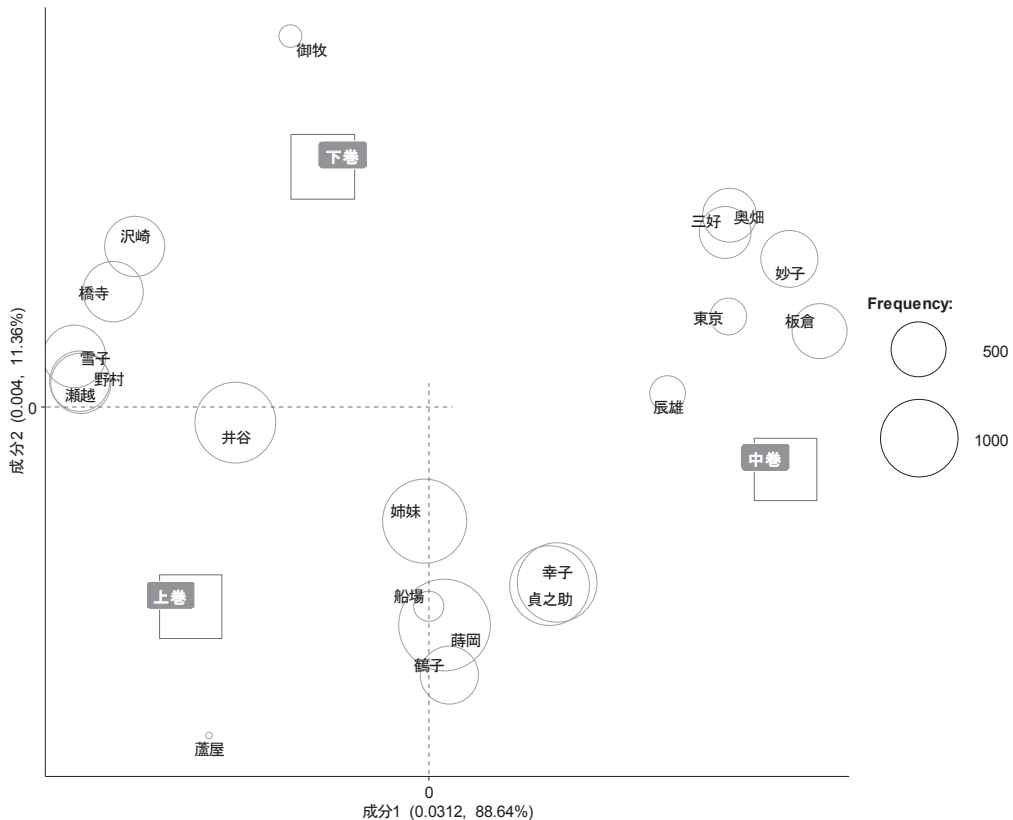


図5 巻と登場人物間の相互関係を示す対応分析



いチャーホフの『三姉妹』的世界に通じる。

長女は「保護者」、次女は「放蕩者」、三女は「冒険者」というアーキタイプ（原型）である。当て嵌めると、幸子が「保護者」、雪子が「放蕩者」、妙子は「冒険者」であり、その三者関係は時に揺らぎはするが、ストーリーが進むにつれて最終的には予定調和に向かう関係性である。

今回、谷崎潤一郎の『細雪』（1948年）を素材に、KH Coderを利用した計量テキスト分析方法を応用して分析、解釈した。

文学における計量テキスト分析として、全作品のデータを入力し、その作品群に潜む作家の傾向を抽出分析する方法がある。

もうひとつの方法として、発表者の場合、一つの作品に集中し、その中での語彙の選択と配置、登場人物の関係性の接点を詳細に探る方法を選択した。

今回の発表をすることで、従来の伝統的な文学研究に、さらに計量分析を加味することで、テキストへの第三者的俯瞰という手段が上乗せできる可能性を感じることができた。換言すれば、定性的な情報分析に定量的な分析の視点を合体させたことである。

今後、今回試用した方法論をさらに深化させていきたいと考えている。

#### 使用ソフト

KH Coder Ver.3.Beta.03i (<https://www.khcoder.net> よりダウンロード)

#### 使用デジタル・テキスト

青空文庫 <https://www.aozora.gr.jp/>よりテキスト・データをダウンロード

『細雪 上巻・中巻・下巻』

底本：「谷崎潤一郎全集 第十五巻」中央公論社 1968（昭和43）年1月25日発行

初出：一～八「中央公論 第五十八巻第一号」1943（昭和18）年3月1日発行

九～十三「中央公論 第五十八巻第三号」1943（昭和18）年1月1日発行

入力：砂場清隆 校正：小島大樹 2016年6月18日作成 2018年5月23日修正

#### 引用文献・参考文献

樋口耕一『社会調査のための計量テキスト分析 第2版』ナカニシヤ出版, 2020年

小林雄一郎『ことばのデータサイエンス』朝倉書店, 2020年

末吉美喜『テキストマイニング入門』オーム社, 2020年

金明哲『テキストアナリティクスの基礎と実践』テキストアナリティクス1, 岩波書店, 2021年

金明哲・中村靖子編『文学と言語コーパスのマイニング』テキストアナリティクス7, 岩波書店, 2021年

黒橋禎夫『改訂版 自然言語処理』NHK出版, 2020年

伊藤公一朗『データ分析の力 因果関係に迫る思考法』光文社, 2017年

エレツ・エイデン, ジャン＝バティ・ミシェル『カルチャロミクス 文化をビッグデータで計測する』阪本芳久訳, 草思社, 2019年

江崎貴裕『分析者のためのデータ解釈学』ソシム, 2021年

アンソニー・ケイ『文章の計量』吉岡健一訳, 南雲堂, 1996年

李在鎬編『文章を科学する』ひつじ書房, 2017年

有馬明恵『内容分析の方法 第二版』ナカニシヤ出版, 2021年

安藤宏『日本近代小説史』中央公論新社, 2019年

高橋和子「コロナ渦における遠隔ダンス授業の成果と課題」『環境と経営』第27巻第1号, 静岡産業大学論集, 2021年

樋口耕一「コンピュータ・コーディングの実践:漱石「こころ」を用いたチュートリアル」『年報人間科学24-2』pp.193-214, 大阪大学人間科学部社会学・人間学・人類学研究室, 2003年

小谷内郁宏「記号としての三姉妹」『静岡学園短期大学研究報告』第9号, pp.11-20, 1996年

