The University of Osaka
Institutional Knowledge Archive

Title	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
Author(s)	樋口, 耕一
Citation	年報人間科学. 2003, 24-2, p. 193-214
Version Type	VoR
URL	https://doi.org/10.18910/10477
rights	
Note	

The University of Osaka Institutional Knowledge Archive : OUKA

https://ir.library.osaka-u.ac.jp/

The University of Osaka

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――

樋口　耕一

〈要旨〉

社会調査によって得られる質的データには新聞・雑誌記事、質問紙調査における自由記述、インタビュー・データなど様々なものがある。コンピュータ・コーディングとは、それらの質的データを計量的に、また多くの場合は探索的に分析するための手法である。

本稿の目的は、独自に開発されたソフトウェア「KH CODER」を用いてコンピュータ・コーディングを実践するための手順を詳細に記述し、これを通じて方法論とプログラムを紹介・提案することにある。本稿の記述は、各自のパソコン上で手順を追うことができるチュートリアルとなっており、題材として用いるデータは夏目漱石「こころ」である。

チュートリアルの中では、作品全体を通して頻繁に出現している言葉や、上・中・下それぞれの部で特徴的な言葉から、作品の構成・特徴を探る。また、人の死やその原因となりうる事柄を表す言葉が、作品全体のどの部分で頻出しているのかという集計を行うことで、人の死が作品中でいかに描かれているかを探索する。この結果として、「先生」という登場人物の自殺が突然・不自然になされているという指摘は、必ずしも当てはまらないことが再確認された。

キーワード

コンピュータ・コーディング、内容分析、質的データ分析、CAQDA、KH CODER

1　はじめに

計量的分析のためのコーディング　コンピュータ・コーディング（川端[8]）の目的は、社会調査によって得られる様々な質的データを、計量的に、また多くの場合は探索的に分析することにある(1)。例えば、質問紙調査における自由回答項目の場合であれば、特定の事柄に言及した回答がいくつあったのか、また、どういった人々が特定の事柄に言及したのかといった形で、計量的な分析を行うことになろう。

同じ質的データの分析にあたって、分析者が典型的な回答を引用し、解釈するという方法では、当然、どうしてその回答が典型的と言えるのかという根拠、基準が曖昧になってしまう場合が多い。それに対して、計量的な手法を用いた場合は、数の多い事例を「典型」的だとして取り上げたり、数の少ない特異な事例を「例外」的事例として取り上げるといったことが可能になり、曖昧さが大幅に軽減され得る。そして、安田・原[20: p. 6]が述べるところの、データを「客観的な計数または計量によって把握」するということが、より容易になる。

さらに、男性の回答の中に比較的多く出現している言葉を洗い出したり、あるいは、出現パターンが似通った言葉の組み合わせを調べたりといった、各種の探索的な統計手法を用いることで、データの把握・理解が進むことも多いであろう。この点も、計量的な分析を行うことの重要な利点である。

ただし、質的データを計量的に分析するためには、データを数量化するためのコーディング作業が不可欠である。上述の自由回答の例では、各回答の中に含まれる言葉を調べ上げたり、あるいは、あらかじめ研究者が用意したいくつかのカテゴリに回答を割り振るといったコーディング作業を行わなければならない。

コンピュータを利用する利点　従来の多くの研究では、このコーディング作業を人手で行っていた。しかし、計量的な分析に堪えるような大量のデータを、人手でコーディングすることには問題点も多い。まず、Krippendorff[2: p. 188]はデータが膨大な量になる場合「コンピュータの助けを借りなければ、分析者は偏った、不完全な、そして非常に選択的な印象を形成しがちである」ことを挙げている。また、Seale[3]はコンピュータを利用することのメリットとして、その他にも、信頼性があること、共同研究が可能となること、サンプルの選択に役立つことなどを挙げている。

すなわち、人手でコーディングを行っていると何年もかかるような場合でも、コンピュータを用いればわずかな時間でコーディング可能である。また、大量のコーディング作業を人間が行う場合には、常に同じ基準のもとにコーディングを行うことが難しいという、信頼性の問題が生じる。その点コンピュータを用いればコーディングの基準が揺らぐことはなく、さらにその安定したコーディング基準を第三者に明示し得るので、検証可能であるばかりか、複数の分析

者による共同研究も可能となる。また、大量のデータの中から、ある条件を満たすデータを発見するといったこともコンピュータの得意とするところであり、これによって、データの典型例や特異例を抽出するなど、計量的方法の利点を活かすことが容易になる。

量的な分析と質的な分析　かつてコンピュータ・コーディングは、質的データの長所として挙げられる、読んで分かるおもしろさ、生々しさというもの、すなわち見田[19]のいう「追体験的な了解可能性」を切り落としていく方向で利用されることも多かった（川端[8]）。しかし現在では、質的データの「追体験的な了解可能性」をより効果的に生かすための助け、ないしは解釈の助けとしても、コンピュータ・コーディングは積極的に用いられている。すなわち、ライフヒストリーの記述やインタビュー・データの分析といった、必ずしも計量的な分析が最終的な目標とはならない局面においてもコンピュータ・コーディングは利用されている。

川端[9]は、ライフヒストリーの記述時にコンピュータ・コーディングを用いるメリットとして次のようことを挙げつつ、その具体的な手順を提案している。例えば、コンピュータを用いてデータ中から言葉を切りだし、多く用いられている言葉が、インタビューのどの部分で頻出しているのかをグラフやマップの形に表す。このように、計量的なやり方でデータ全体を見渡すことで、活かすべき細部の記述、すなわち質的にしか表現できない部分もより明確になる。また、いったん言葉を細やかな文脈から切り離して集計することで、分析者はインタービューを実施するときに要求される共感的・共同的な視点ではなく、より客観的にデータを眺める視点を得ることができる。こういった方法でインタビュー・データの分析を行った研究としてはKawabata & Akiba[1]や景山[7]などが挙げられる。(2)

コンピュータ・コーディングの実践　以上、コンピュータ・コーディングの概要とその利点をかいつまんで述べてきたが、実際にこれを行うためには、コーディングのためのプログラム（ソフトウェア）が必要となる。だが、英語をはじめとする欧米語のデータをコーディングするためのソフトウェアほどには、日本語を扱えるソフトウェアは豊富に準備されていなかったので、独自のコーディング・プログラム「KH CODER(3)」が開発された（樋口[17]）。本稿の目的は、KH CODERを用いてコンピュータ・コーディングを実践する方法と手順を詳しく記述し、これを通じて方法論とプログラムを紹介・提案することにある。

なお本稿の記述は、各自のパソコン上で手順を追うことができるチュートリアルとなっており、題材として用いるデータは夏目漱石「こころ」である。言うまでもなく、「こころ」は漱石の代表作であり、名作の誉れ高い作品として知られている。以下のチュートリアルを通して、単にコンピュータ・コーディングの方法・手順を知るだけでなく、「こころ」という作品の鑑賞をさらに深めることもできるだろう。その過程で、コンピュータ・コーディングの利点を実感していただくことができれば、望外の喜びである。

2　プログラムとデータの準備

2・1　プログラムの準備

2・1・1　KH CODERのインストール

KH CODERはフリー・ソフトウェアであり、次のホームページから入手・ダウンロードすることができる。

http://koichi.nihon.to/psnl

http://hey.to/KO-ichi

なおKH CODERを用いるためには、文章から単語を切り出すためのソフトウェア「茶筌」（松本ら[18]）が必要となるので、事前にインストールしておくとよい[4]。

KH CODERをインストールするためには、上記のホームページから配布されているファイルをダウンロードし、KH CODERをインストールするフォルダ、例えば「C:¥khcoder」などにコピーする。そしてこのファイルを実行（ダブルクリック）すると、図1のような画面が開く。ここで「UNZIP」ボタンをクリックすれば、「kh_coder.exe」という実行ファイルと、その他のファイルがいくつか解凍・生成される。解凍が完了すれば、「7 file(s) unzipped successfully」というメッセージ・ボックスが表示されるので、「OK」ボタンをクリックし、図1の画面を閉じる。

解凍・生成された「kh_coder.exe」というファイルを実行（ダブルクリック）して、KH CODERを起動する。ここで、KH CODERのメニューから「プロジェクト」、「設定」と順にクリックすると図2に示すような設定画面が開く。この画面の「Chasen.exeのパス」という部分に、chasen.exeのフルパス[5]（デフォルトではC:¥Program Files ¥chasen¥chasen.exe）を入力してもよいし、「参照」ボタンをクリックしてマウス操作で「chasen.exe」を探してもよい。この操作の後に設定画面下部の「OK」ボタンをクリックすればKH CODERのインストールは完了である。

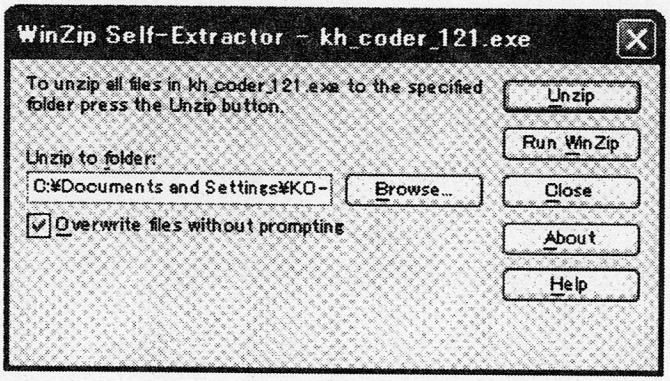

図1：KH Coder のセットアップ（解凍）

2・1・2　EXCEL用アドイン

KH CODERが作成した集計表を、EXCEL上で整形したりグラフを作成する際に、場合によっては、いくぶん面倒な操作が必要になる。最新版（Ver. 1.30）のKH CODERには、そういった操作を自動的に行うためのEXCEL用アドインが添付されている。この節では、アドインをEXCELで利用するための手順を述べる。

まずEXCELを起動して、メニューから「ツール」、「アドイン」と順にクリックすると、図3に示すような画面が表示される。この画面で「参照」ボタンをクリックして、KH CODERをインストールしたフォルダの中にある「khc.xla」というファイルを開く。これで「有効なアドイン」の一覧に「KH Coder - Excel AddIn」が追加されているはずだ。あとは「OK」ボタンをクリックすれば、アドインの準備は完了である。

EXCELのメニューの右端に「KH Coder (K)」というメニューが追加されていることが分かるだろう（図4）。このインストール作業は一度行うだけでよく、次回からは普通にEXCELを起動するだけで、図4に示した「KH Coder (K)」のメニューが追加されている。

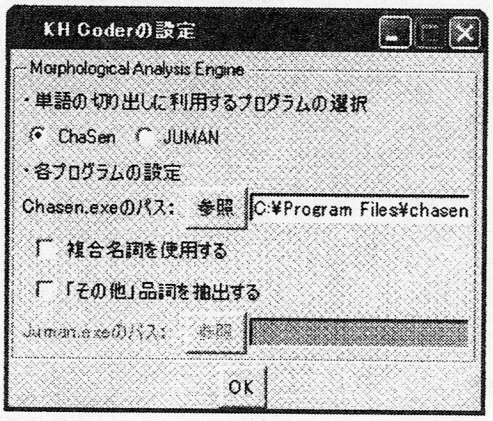

図2：KH Coder のセットアップ（設定）

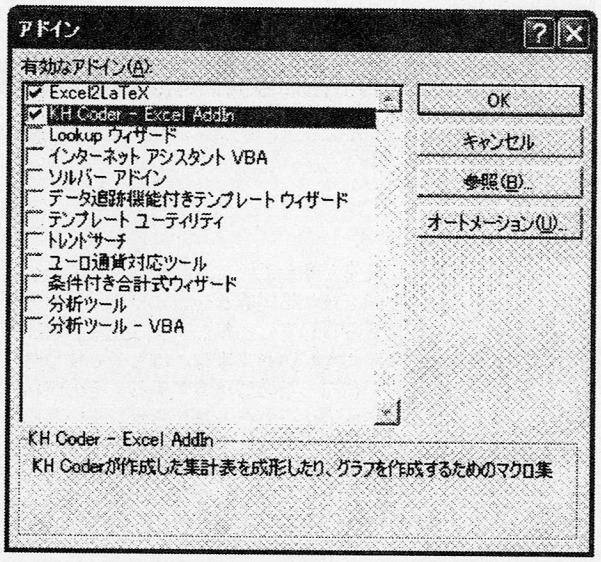

図3：Excel にアドインを登録

2・1・3　チュートリアル用データ

このチュートリアルで用いる夏目漱石「こころ」のテキストデータや、サンプルのコーディング・ルールファイルなどは、2・1・1節に示したKH CODERのホームページから入手できる。KH CODERのホームページから、「チュートリアル用データ」と記述されたファイルをダウンロードする。このファイルをKH CODE

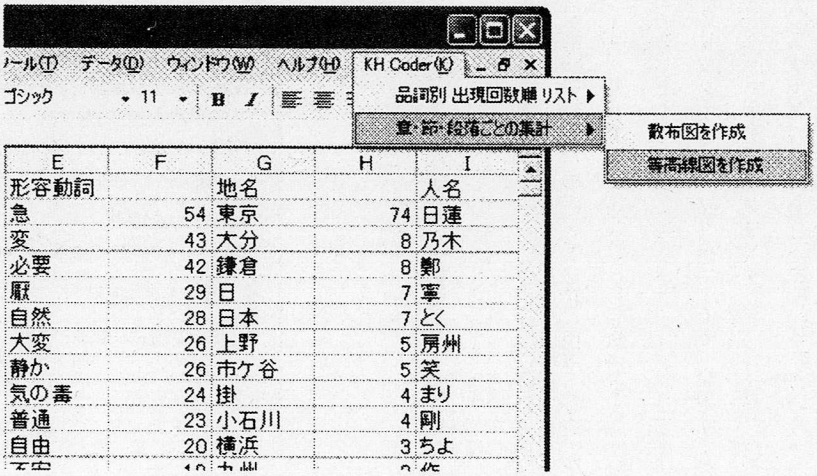

図4：Excelに追加されるメニュー

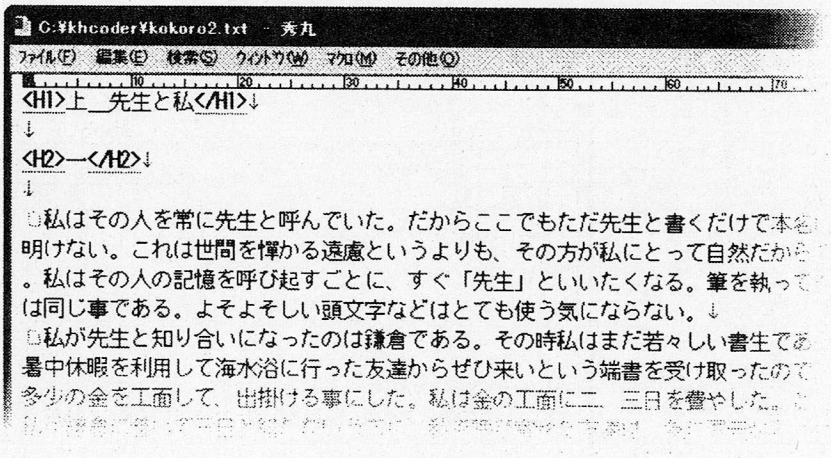

図5：マーキングされたデータ（kokoro2.txt）

Rをインストールしたフォルダにコピーして、実行（ダブルクリック）すると、図1と同様の画面が開く。ここで、「UNZIP」ボタンをクリックすれば、「tutorial」というフォルダが作成され、その中にチュートリアル用のファイル群が解凍される。

チュートリアル用ファイル群の中には、「kokoro1.txt」および「kokoro2.txt」という2つのテキストファイルが含まれている。kokoro1については、小説「こころ」の全文が入力された通常のテキストファイルである[6]。このkokoro1をそのまま分析に用いると、どこからどこまでが1つの章なのかといったことが、KH CODERには区別できない。そこで、図5に示すように、<H1>および<H2>タグを使って見出し部分をマーキングしたものが、kokoro2である。

このマーキングを「手作業」で行うのはいくぶん面倒であるが、多くの場合は、作業のほとんどを自動的に行うことができる。今回のチュートリアル用データに関しては、テキストエディタ「秀丸」が持つ検索と置換の機能を用いて自動的にマーキングを行った[7]。

2・2　KH CODERによるデータの前処理

2・2・1　「プロジェクト」の新規作成

KH CODERを用いてテキストデータの分析を行うためには、まず最初に、分析を行いたいテキストファイルを「プロジェクト」としてKH CODERに登録せねばならない。

これを行うためには、KH CODERのメニューから「プロジェクト」、「新規作成」を順にクリックする。「新規プロジェクト」というタイトルがついた画面が開くので、この画面の「分析対象ファイル」という部分で、分析を行いたいファイルを指定する。今回のチュートリアルでは、「参照」ボタンをクリックして、「tutorial」フォルダを開き、「kokoro2.txt」を選択すればよい。なお、「説明（メモ）」の部分には、わかりやすいメモ書きを入力しておけばよい。このチュートリアルでは、例えば「漱石『こころ』」のように入力しておけばよい。あとは、「OK」ボタンをクリックすれば、プロジェクトの新規作成は完了である。

なお、この作業は一度行うだけでよく、次回からはメニューの「プロジェクト」、「開く」を順にクリックすれば、すでに作成されたプロジェクトの一覧が表示される。その中から、分析を続けたいプロジェクトを選択するだけでよい。

2・2・2　単語切り出しなどの前処理

次に、実際に分析を行うための準備として、いくつかの前処理をKH CODER上で行わなければならない。

語の切り出し　KH CODERのメニューから「基礎処理」、「前処理一括実行」をクリックすると、「前処理一括実行」画面が開くので、図6と同じように選択・入力を行う。すなわち、「強制抽出・強調する語の指定」部分には全角の「K」を入力し、「H1単位」、「H2単位」、「段落単位」にチェックを入れればよい。この状態で「O

K」ボタンをクリックすると、文章からの単語の切り出しをはじめとする、一連の前処理が実行される。この処理が完了するまでには数分を要するだろう。

なお、強制的に抽出する語として指定している「K」は、「こころ」の重要な登場人物である。しかし、何も指定をしないと、KHCODERは「K」のような一語のアルファベットを抽出しない。助詞や助動詞などと同様に、分析には利用しにくい語として無視してしまうのだ。

このように、重要な言葉であるにもかかわらず、無視されてしまったり、あるいは、単語としてうまく抽出されなかった場合には、その言葉を「強制抽出・強調する語」として指定し、再び前処理を行うとよい[8]。

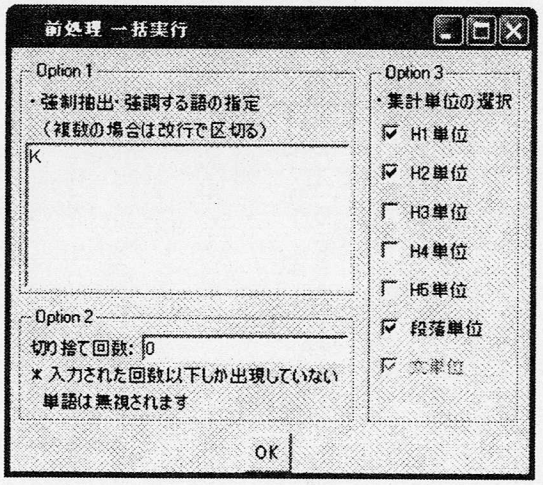

図６：前処理の実行

外部変数　次に、今回のチュートリアルでは「外部変数」を作成しておく必要がある。KHCODERのメニューから「ツール」、「外

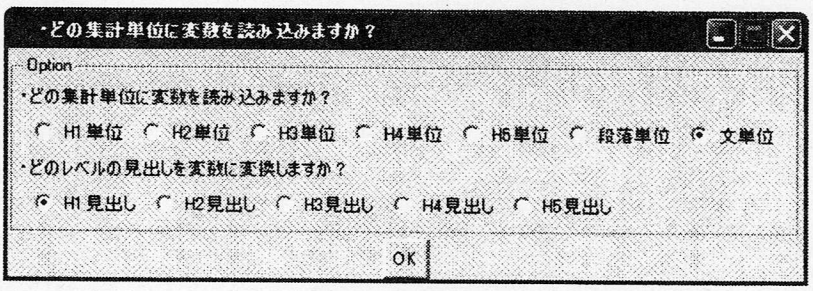

図７：外部変数の作成

表１：変換する組み合わせ

No.	読み込み単位	変換する見出し
1	文	H1
2	文	H2
3	段落	H1
4	段落	H2

部変数」、「上位見出しを変数に変換」をクリックして、開いた画面で図7のように単位と見出しの指定をしてから「OK」ボタンをクリックする。図7では、変数を読み込む単位として「文」を、変換する見出しとして「H1見出し」を選択している。H1というのは、上・中・下という3つの部の見出しを括ったタグであった（図5）。

この指定で「外部変数」を作成することで、単に「ある言葉が含まれている文を探す」といったことだけでなく、「ある言葉が含まれていて、なおかつ、上・中・下3部のうち『下』に含まれている文を探す」といったことが可能になる[9]。

図7に示した指定は、表1に示した1つめの組み合わせ（No. 1）にあたる。同様の操作を繰り返して、残る2～4までの組み合わせでも、「外部変数」を作成しておく必要がある。

3　分析のための操作

3・1　多く出現している言葉

3・1・1　全体での頻出語

まず、「こころ」の中で多く出現している言葉にどんなものがあったのかを確認しておきたい。KH CODERのメニューから「ツール」、「抽出語」、「品詞別出現回数順リスト」をクリックする。するとEXCELが起動し、図4に一部が見えているような、品詞別の頻出語リストが表示される。

ここで、品詞別に分けずに、出現数の多い語から順に並べたければ、EXCELに追加されているメニューの「KH Coder (K)」、「品詞別出現回数順リスト」、「一列に成形」を順にクリックすればよい。また、特に多く出現している語を提示したい場合には、同じくEXCELに追加されているメニューの「KH Coder (K)」、「品詞別出現回数順リスト」、「頻出150語の表を作成」を順にクリックすればよい。この操作によって、表2が得られる。

表2を見ると、まず、「先生」、「K」、「奥さん」、「父」といった主要な登場人物を表す言葉が非常に多く出現していることが分かる。それに次いで多く出現しているのは一般的な動詞で、「思う」（296回）、「考える」（130回）や「見る」（225回）、「聞く」（219回）といったものである。他には、特殊でインパクトの強い語であるにもかかわらず、「死ぬ」という語が89回と多く出現していることが目にとまる。「こころ」においては、主人公の父が病気で死に瀕している様子や、「K」と「先生」の自殺が描かれているが、それにしても89回という出現数は多いといえよう。

ただし、KH CODERは、例えば「死ぬ」、「死ねば」、「死んで」、「死にたい」、「死なれた」などはすべて「死ぬ」という一種類の語として抽出する。このように、活用を持つ語は動詞、形容詞、形容動詞にかかわらず、すべて基本形としてカウント・抽出することに注意が必要である。また、主人公である「私」については、KH CODERによって抽出・カウントされていない。何も設定を変更しない場合、KH CODERは代名詞を抽出しないのである。よって、「私」という語を抽出したり、「死ぬ」としてカウントせずに「死に

表２：頻出150語

抽出語	出現数	抽出語	出現数	抽出語	出現数
先生	597	兄	62	入る	36
K	410	出す	62	医者	35
奥さん	388	病気	62	驚く	35
思う	296	様子	61	若い	35
父	269	声	60	書物	35
自分	264	外	59	信じる	35
見る	225	卒業	58	与える	35
聞く	219	話す	58	頼む	35
出る	185	取る	57	過去	34
人	182	心持	57	決して	34
母	170	急	54	新しい	34
お嬢さん	168	坐る	54	身体	34
前	163	態度	54	世の中	34
帰る	155	付く	54	他	34
今	139	問題	54	変る	34
顔	135	笑う	53	気分	33
来る	131	返事	53	記憶	33
考える	130	歩く	52	起る	33
言葉	126	読む	51	向う	33
眼	123	強い	49	済む	33
知る	118	持つ	49	得る	33
心	106	分る	49	開ける	32
妻	104	違う	48	繰り返す	32
行く	102	呼ぶ	48	生きる	32
立つ	97	動く	48	想像	32
口	93	置く	47	打ち明ける	32
答える	92	突然	47	長い	32
好い	91	仕方	45	無論	32
死ぬ	89	見せる	43	会う	31
見える	87	変	43	例	31
少し	81	落ちる	43	感じる	30
知れる	79	実際	42	結果	30
話	78	眺める	42	寝る	30
女	77	必要	42	進む	30
書く	75	風	42	聞こえる	30
手紙	74	先	41	厭	29
東京	74	全く	41	掛ける	29
頭	73	傍	41	心配	29
解る	71	一つ	40	尋ねる	29
手	71	胸	40	注意	29
事	70	黙る	40	友達	29
叔父	70	気が付く	39	淋しい	29
人間	70	調子	39	恋	29
宅	70	場合	38	下	28
家	68	学校	37	何とも	28
室	68	行う	37	機会	28
意味	67	上げる	37	財産	28
男	66	忘れる	37	自然	28
気	65	向う	36	受ける	28
悪い	62	時々	36	世間	28

たい」という表現・活用のままカウントしたい場合には、２・２・２節にて前述した「強制抽出」の指定を行う必要がある。

３・１・２　それぞれの部に特徴的な言葉

次に、もう少し詳しく「こころ」の構成を調べるために、上・中・下それぞれの部において、特に多く出現している言葉、すなわちそれぞれの部を特徴づけるような言葉をみていきたい。

ＫＨＣＯＤＥＲのメニューから「ツール」、「探索・コーディング支援」、「抽出語連関規則」を順にクリックすると、集計単位を選択するための画面が開く。ここでは「文単位」を選択して「ＯＫ」ボタンをクリックする。すると図８に示すような画面が開く。この画面の「・コード選択」部分にある「開く」ボタンをクリックして、

「tutorial」フォルダの中の「part.cod」を開く。さらに、「・品詞の選択」部分で、「ナイ形容」、「否定」および品詞名の最後に「B」がついた品詞の選択を解除しよう[10]。選択を解除するためには、「Ctrl」キーを押しながら、それぞれの品詞をクリックすればよい。あとは、「・コード選択」部分で「上　先生と私」といった部の見出しをクリックして選択し、「計算」ボタンを押せばよい。この操作によって、上・中・下それぞれの部において、全体と比較して、より高い確率で出現している言葉をリストアップすることができる。

なお、ここで用いた「part.cod」を、「メモ帳」や「秀丸」などのテキストエディタで開くと、内容は図9のようになっていることが分かるだろう。「＊」で始まる部分がコードの名前で、その下の行が、そのコードを与えるための条件となっている。そして、条件部分では「＜＞変数名--＞値」という、外部変数を用いた条件の指定方法が用いられている。このコーディング・ルールでは、2・2・2節で作成しておいた外部変数を用いて、「上　先生と私」といったコードをそれぞれの文にあたえている。そして図8の画面の裏では、それぞれのコードが与えられた文と与えられなかった文とをまず選り分けて、コードが与えられた文の中により高い確率で出現している単語はどれかという計算が行われている。よって、ここでリストアップされている言葉は、ただ単に頻出しているというだけではなく、それぞれの部を特徴づけるような言葉であるといえよう。この計算

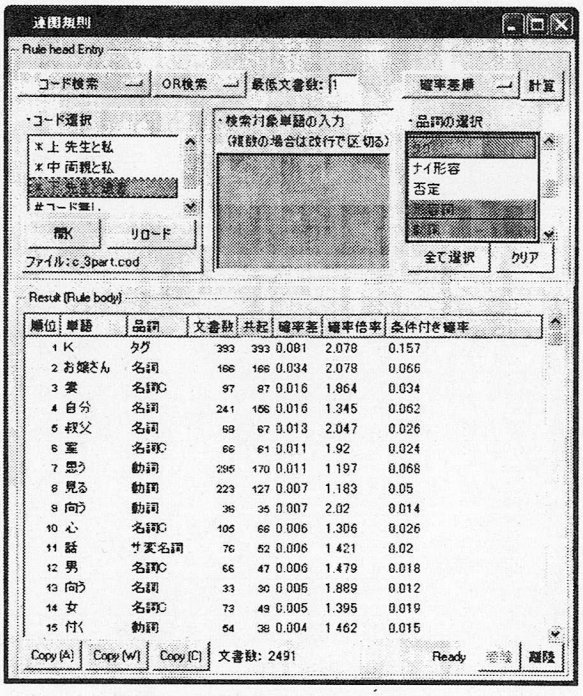

図8：特徴的な言葉の検索

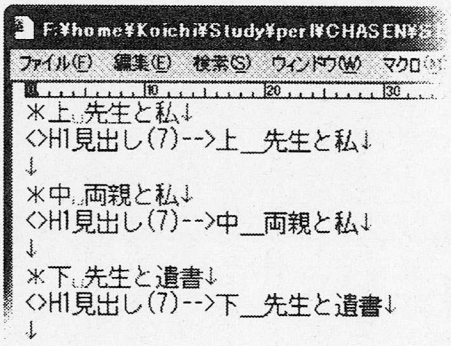

図9：コーディング・ルール・ファイルの内容

の結果をまとめ、全体と比べてより高い確率で出現している語から順に10ずつをリストアップしたものが表3である。

表3より、まず「上・先生と私」においては、主人公が鎌倉の「海」で知り合ってから深い交流を持つことになった「先生」と、その「奥さん」などの語が挙がっている。「解る」についても特徴的な語として挙がっており、文中では次のような用いられ方をしている。

私はなぜ先生に対してだけこんな心持が起るのか**解ら**なかった。それが先生の亡くなった今日になって、始めて**解っ**て来た。

（「こころ」上・四、強調は筆者による）

このように、「上・先生と私」では、主人公が「先生」のことについて「解った」り、「解らなかった」りしたという記述が多かったのである。また、「罪悪」については、「恋の満足を味わっている人はもっと暖かい声を出すものです。しかし……しかし君、恋は罪悪ですよ。解っていますか」（上・十二）という「先生」の発言に始まって、すべて恋が罪悪であることを示している。この、恋が罪悪であるという点も、「上・先生と私」においては主人公にも読者にも、よく「解らなかった」事柄である。

表3：上・中・下それぞれの部を特徴づける言葉

部	特徴的な語
上・先生と私	先生、奥さん、持つ、解る、向く、事実、人間、海、罪悪、下
中・両親と私	父、母、兄、手紙、卒業、書く、来る、東京、病気、読む
下・先生と遺書	K、お嬢さん、妻、自分、叔父、室、思う、見る、向う、心

「中・両親と私」では、「父」の「病気」がいよいよ重くなり、「卒業」した主人公が実家で過ごす様子が描写されている。そのため、「父」だけでなく「母」、「兄」といった家族を表す語が挙がっている。ただし実家にいても、時折「手紙」を「書く」ないし「読む」ことで、「東京」にいる〝先生〟と主人公は交流を絶やさなかった。

次に「下・先生と遺書」は、〝先生〟が自殺にあたって、主人公に書き送った長い遺書となっている。そして、〝先生〟の書生時代の親友であり、また同じ女性「お嬢さん」に恋した男でもある「K」などが挙がっている。この「お嬢さん」は、後に〝先生〟の「妻」となり、「K」は自殺したわけである。なお、遺書ということもあってか、「自分」自身をかえりみて、ものを「思う」といった描写も多いことがわかる。そういった〝先生〟の思いが語られる中で、「上・先生と私」では主人公にとっても読者にとってもよく「解らなかった」事柄の多くが、解きほぐされていくのである。

以上、わずか10ずつではあるが、それぞれの部に特徴的な言葉を、本文中での用いられ方と併せて確認することによって、「こころ」という小説の構成を再確認することができたのではないだろうか。なお、特徴的な語として挙がっていながらも、紙幅の都合から触れられなかったものや、実際にその語が使われている文を引用できなかったものについては、次節に述べる方法を用いて、各自のパソコン上で、本文中での使われ方などを検索・確認していただきたい。

3・1・3　言葉の用いられ方を調べる

前節で行ったように、リストアップされた特徴的な語を確認していくといった場合には、それらの語が、もとの文の中でどのように用いられていたのかを把握することが極めて重要であり、不可欠である。不可欠であるばかりか、頻出語、ないし特徴的な語としてリストアップされた語が、本文中でいかに用いられていたかを調べる作業を通して、データ特性（この場合は「こころ」の特徴）を再確認、再発見することが、コンピュータ・コーディングの目的の1つであるとも言える。

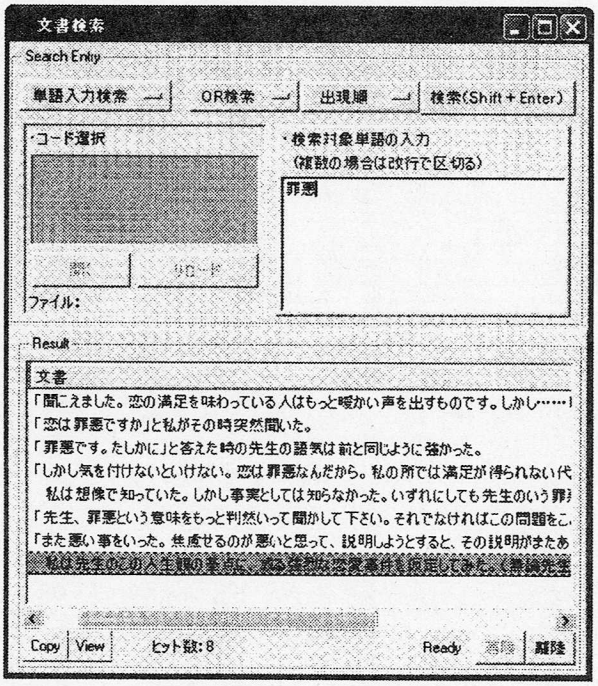

図10：KH Coder の検索画面

これを行うために、KH CODERには検索機能が用意されている。KH CODERのメニューから「ツール」、「探索・コーディング支援」、「文書検索」を順にクリックする。ここで検索を行う単位を選択するための画面が開くので、「段落」を選択して「OK」をクリックする。すると、図10のような画面が開く。この画面で、「コード検索」となっている部分をクリックして、「単語入力検索」を選択する。また、「tf／idf順」となっている部分をクリックして、「出現順」を選択しておく。あとは、検索したい語を入力して「検索」ボタンをクリックすれば、入力した語を含む段落のリストが表示さ

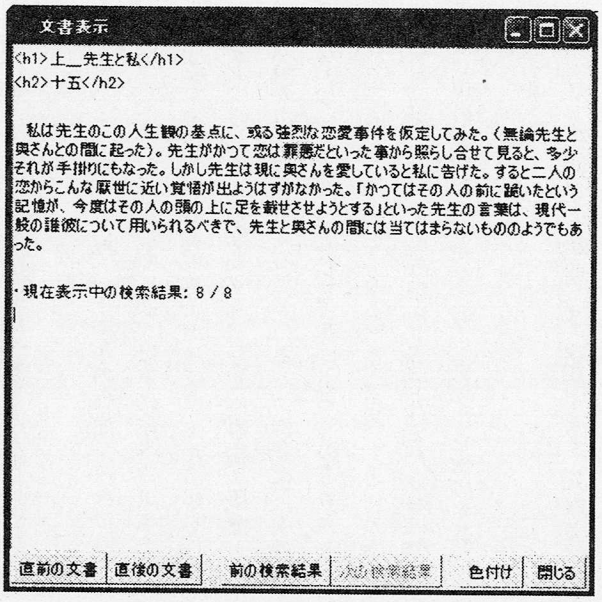

図11：検索した段落の全体を表示

れる。なお、検索の単位として「段落」を選ぶのは、検索結果として1つの文が表示されるよりも、段落が表示された方が、前後の文脈を捉えやすいであろうからだ。

検索結果のリストには、各段落の最初の数十文字しか表示されていないが、リスト部分をダブルクリックすれば図11のような画面が開き、段落全体を見ることができる。この画面では、検索の対象となった語が黄色の背景で強調されているので、その語がいかに用いられているかをすぐに確認できるだろう。また、より広く前後の文脈を調べたい時には、「直前の文書」、「直後の文書」ボタンをクリックすることで、直前ないし直後の段落が表示される。

3・2　テーマの流れ

「こころ」という作品の中では、「死ぬ」という言葉が89回と多用されており、また、「先生」の自殺の原因にかかわる事柄が、「上・先生と私」においては主人公にとっても読者にとっても不可思議なものとなっていた（3・1節）。これらの点から、「先生」の自殺や、人の生死が、作品の主要なテーマとなっていたことを再確認できよう。

しかし、前節で述べた方法で「死ぬ」という言葉が使われている文や段落を検索・確認しても、「死ぬ」という言葉は非常に多く出現しているので、この言葉の用いられ方や作品中での役割を把握することが難しい。そこで本節では、この点を明らかにするために、人の死や、その原因となり得るいくつかの事柄が、作品全体のどの部分で頻繁に言及されているのかを調べていきたい。

3・2・1　コーディング・ルールの作成

そういったことを行うためには、以下に述べるコーディング・ルールを作成する必要がある。コーディング・ルールを作成する目的は、1つ1つの言葉というよりも、むしろ事柄・概念・カテゴリといったものの出現数を数えることにある。例えば、人の死という事柄は、必ずしも「死ぬ」という言葉によってのみ表されるわけではない。「死ぬ」の他に、「亡くなる」や「自殺」といった言葉があったときにも、「人の死」という事柄として数えたいわけである。このためには、次のようなコーディング・ルールを作成すればよい。

＊人の死
死後 or 死病 or 死期 or 死因 or 死骸 or 生死 or 自殺 or 殉死 or 頓死 or 変死 or 亡 or 死ぬ or 亡くなる or 殺す or 亡くす or 死

このコーディング・ルールをKH CODERに読み込ませれば、死後、死病、死期、、といった言葉が含まれる文や段落に「＊人の死」というコードが与えられ、コードの出現数や出現率などを調べることができるようになる。

コーディング・ルールを作成する際には、おもむろに言葉を挙げていってもよいのだが、データ中に出現していない言葉、あるいはKH CODERによって語として抽出されていない言葉を指定して

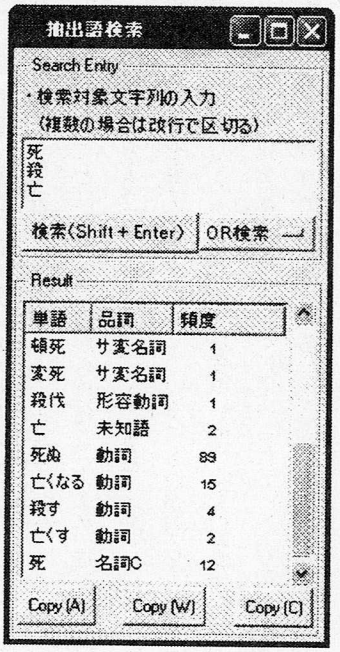

図12：抽出語検索画面

も意味がない。よって、図12に示したKH CODERの抽出語検索という機能が便利であろう。図12では、「死」、「殺」、「亡」といった文字を含む言葉を検索することで、小説「こころ」の中に出現している、人の死を表す言葉をリストアップしている。

なお、以下の集計で用いるコーディング・ルール・ファイルは、チュートリアル用データとともに「theme.cod」として添付しているので、「メモ帳」や「秀丸」などのテキスト・エディタでこのファイルを開き、内容を一度ご確認いただきたい。

3・2・2　部ごとの集計

まずは、上・中・下という3つの部ごとに、それぞれのコードの出現数を集計して、部ごとの推移を確認する。KH CODERのメニューから「ツール」、「コーディング」、「章・節・段落ごとの集計」を順にクリックする。「コーディング単位の選択」画面が開くので、コーディング単位として「段落」を選択し、集計を行う単位の選択で「H1単位」を選択する。そして、「参照」ボタンをクリックして「tutorial」フォルダの中の「theme.cod」を開く。最後に、「表のセル内容の選択」部分では「パーセンテージのみ」を選択してから、「OK」ボタンをクリックすればよい。インターネット・エクスプローラの画面が開き、上・中・下それぞれの部について、各コードが与えられた段落の割合（パーセント）が表示される。

もちろん、この集計表をそのまま読んでもよいのだが、特にコード数が多い場合には、グラフを作成した方が格段に集計結果を読みとりやすくなる。このためには集計表全体を選択してコピーし、EXCEL上にペースト（貼り付け）する。そしてEXCELのメニューから「KH Coder (K)」、「章・節・段落ごとの集計」、「等高線グラフ」を順にクリックすれば図13のようなグラフが得られる。この等高線グラフは、四角形（ひし形）が何重にもかさなっているほど、また、ともに2重であるといった同じかさなり方であれば、四角形が大きいほど、コードの出現率が高いことを示す。折れ線グラフのように線が重なって複雑になってしまうこともなく、また3次元棒グラフのように、奥の棒が隠れてしまうこともないので、集計するコードの数が多い場合も、集計結果を一望することができる。

図13を見ると、コード「＊人の死」は「上・先生と私」と「中・両親と私」に同程度の割合で出現しており、「下・先生と遺書」で大きく増加していることがわかる。また、死の原因として作成したコードについては、「＊病気」が「中・両親と私」で、また「＊信用・不信」と「＊恋愛」が「下・先生と遺書」で、それぞれ多く出現し

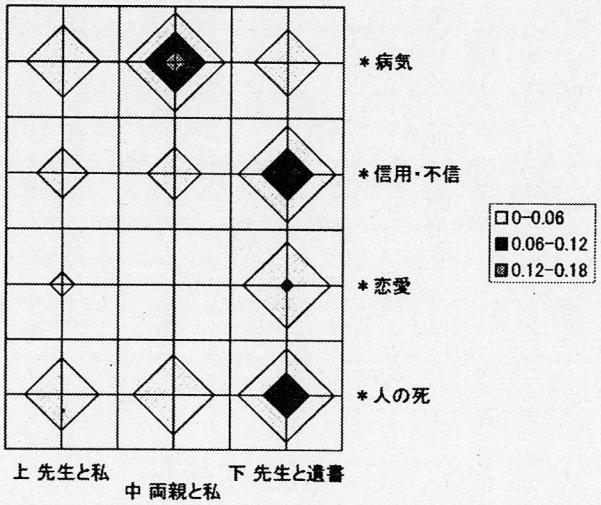

図13：コード出現率・部ごとの集計

ていることがわかる。

「中・両親と私」では病気のために死に瀕した「父」の姿が描かれ、「下・先生と遺書」では「K」と「先生」が恋愛や、信頼ないし友情といったことに関わる原因から自殺したことが描かれていたことを再確認できよう。その一方で、「上・先生と私」には「中」と同程度に人の死を表す言葉が出現しているにもかかわらず、「中」での病気や、「下」での恋愛や信頼・友情にあたるものがない。「上」において、人の死を表す言葉がいかに用いられていたのか、あるいは、死を表す言葉がいかなる役割を果たしていたのかを、図13から推しはかることは難しい。

3・2・3　章ごとの詳細な集計

「上・先生と私」において人の死を表す言葉がいかに用いられているのかという、前節では確認できなかった点に留意しつつ、ここでは人の死を表す言葉が作品中でいかに増減しているかをより詳細に見ていきたい。前節と同じ手順で、KH CODERによる「章・節・段落単位の集計」を行う。ただし、より細かな増減を確認できるように、コーディング単位として「文」を、集計単位として「H2単位」を選択する。この操作によって章ごとに、各コードが与えられた文がどれだけ出現していたのかという集計が得られる。そして得られた集計表をEXCEL上にコピーし、コード「＊人の死」だけの推移を折れ線グラフに表したものが図14である。

なお、EXCELの画面上に作成されたグラフを一目見るだけでは、それぞれの点が、どの章の出現率を示しているのか分かりにくいことがある。この場合、マウスカーソルを点の上に移動すれば、1秒足らずで図15のように情報画面がポップアップ表示される。この情報画面で「要素」とされているものが、章の見出しである。

図14を見ると、作品の最後の2～3章で、人の死への言及が急激に増加していることがわかる。このうち最後の2章、すなわち下・五十五と下・五十六は、「先生」が自らの自殺を決意する部分であり、その直前、下・四十八からの数章では「K」の自殺が扱われている。この部分だけを見ると、「先生」の自殺の決意は、ラスト2章で急激

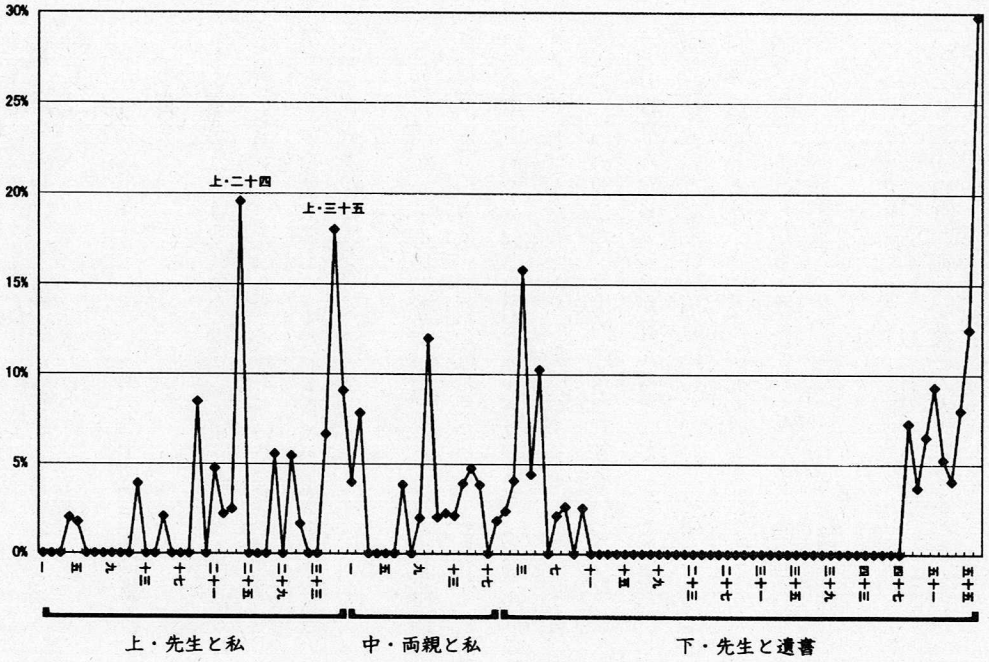

図14：「人の死」への言及数の推移

に固められているように見える。

次に、前節からの懸案であった「上・先生と私」に注目すると、特に上・二十四と上・三十五で、人の死を表す言葉が多く出現していることがわかる。これらの章を順に確認していきたい。

「しかし人間は健康にしろ病気にしろ、どっちにしても脆いものですね。いつどんな事でどんな**死に**ようをしないとも限らないから」

「先生もそんな事を考えてお出ですか」

（中略）

「よくころりと**死ぬ**人があるじゃありませんか。自然に。それからあっと思う間に**死ぬ**人もあるでしょう。不自然な暴力で」

「不自然な暴力って何ですか」

「何だかそれは私にも解らないが、**自殺**する人はみんな不自然な暴力を使うんでしょう」

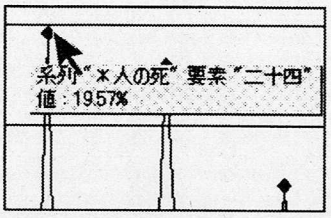

図15：章の確認（EXCEL）

「すると**殺さ**れるのも、やはり不自然な暴力のお蔭ですね」
「**殺さ**れる方はちっとも考えていなかった。なるほどそういえばそうだ」

（上・二十四、強調は筆者による）

この上・二十四における会話からは、「先生」がいつ途切れるかわからない儚いものとして人の命をとらえていたことや、命が途切れる原因として「不自然な暴力」、すなわち自殺のことを念頭に置いていたということが読みとれる。

先生はいくらともいわなかった。けれども先生の話は、容易に自分の**死**という遠い問題を離れなかった。そうしてその**死**は必ず奥さんの前に起るものと仮定されていた。奥さんも最初のうちは、わざとたわいのない受け答えをしているらしく見えた。それがいつの間にか、感傷的な女の心を重苦しくした。
「おれが**死ん**だら、おれが**死ん**だらって、まあ何遍おっしゃるの。後生だからもう好い加減にして、おれが**死ん**だらは止して頂戴。縁喜でもない。あなたが**死ん**だら、何でもあなたの思い通りにして上げるから、それで好いじゃありませんか」

（上・三十五、強調は筆者による）

この上・三十五の記述からは、自分は確実に奥さんよりもはやく死ぬということや、自分の死後に奥さんがどうするのかといったことに、「先生」が執着していたことが読みとれる。このように「先生」の死や自殺への執着が描かれていたために、「上・先生と私」では主要な登場人物の死が描かれていないにもかかわらず、人の死を表す言葉が頻出していたことが分かる。「下・先生と遺書」の中だけでなく、「上・先生と私」においても「先生」が自殺のことを常に念頭に置いていた様子、あまつさえは自殺の決意を徐々に固めつつあったかもしれない様子が、描写されていたのである[11]。

4　結びにかえて

「先生」の自殺　「先生」の自殺が不自然であるという批判は、漱石の生前からなされてきたという（仲[15]）。実際、「下・先生と遺書」に注目するならば、確かに「先生」は最後の2章であまりに急激に自殺の決意を固めていると言えよう。しかし、作品全体にまで視野を広げれば、そもそも「上・先生と私」の中でも「先生」は自らの死や自殺のことを常に念頭においた人物として描かれており、自殺の決意が突然に固められたものとは言い難い（3・2・3節）。そして、自殺が突然なされたものでなかったとしても、自殺の理由そのものが不自然だと言うならば、それは「自殺する訳が明らかに呑み込めないかも知れませんが、もしそうだとすると、それは時勢の推移から来る人間の相違」（下・五十六）といった、漱石と読み手の哲学・個性にかかわる問題ではないだろうか。

コンピュータ・コーディングの利点　「先生」の自殺が、突然で

不自然なものかどうかといった問題を考える際に、文脈の細部に注目することは、もちろん、重要である。一度しか出現していない表現と言えども、「明治の精神に殉死するつもりだ」（下・五十六）と「先生」がいうときの、「明治の精神」とはいったい何かを追求することも重要であろう。しかし、そういった細部にひたすら焦点を合わせれば、当然、同時に作品全体を見渡すことは難しくなる。

この点を補うために、コンピュータ・コーディングによってグラフやマップの形でデータの全体像を掴もうとするのである。作品の全体像をはっきり把握しておくことができれば、文脈の細部を追うときにも、解釈や分析がより容易になることは間違いない。そして、この方法が有効である点を、上述の「先生」の自殺という問題において、部分的にでも例証できたのではないだろうか。

さらに分析を進めるために　以上のチュートリアルでは扱えなかった内容として、まずKH CODERの「コード間関連」コマンドを用いて、コード間の関連を探っていくことが挙げられる。このコマンドでは、2つのコードが頻繁に同じ段落や章の中に出現（共起）しているほど、それらのコードは強く関連しているものと推定される。関連の強いコードの組み合わせや、逆に関連の無いコードの組み合わせを調べることで、扱っているデータの特徴を新たに発見できることもあるだろう。また、関連の強いコードの組み合わせが見つかれば、3・1・3節で触れた文書検索機能を使って、それらのコードが両方とも与えられている段落・章をリストアップし、順に確認していくこともできる。

次に、統計ソフトウェアを別途準備しておけば、多変量解析を探索のために利用できる。例えば川端[8]は、多数のコード間の関連を把握・提示するために、多次元尺度法を用いている。また、コード／コーディング・ルールを作成する前に、多変量解析を用いてデータ概要を把握しておくこともできる（川端・樋口[10]）。

これらの方法の詳細な手順に触れることはできなかったが、いったんこのチュートリアルを終えれば、KH CODERに添付されているマニュアルを必要に応じて参照しつつ、これらの方法を実践することは容易であろう。

注

(1) 分析手順・アプローチの詳細は異なるが、計量的かつ探索的なテキストデータ分析を目指すという点で、本稿で述べる分析の方法と、大隅・Lebart[4]の方法は似通っている。また、大隅らの方法を実現するためのソフトウェアとして、「WORD MINER」が日本電子計算株式会社から発売されている。

(2) 教育学における授業研究という分野では、計量的な分析が必ずしも最終的な目標とならない局面においてそれを利用する点で、似通った方法がはやくから提案されている（大谷・松原[6]、柴田[12]）。

(3) KH CODERは、既存のコーディング・プログラムであるKT CODER／KT2システム（谷口[14]）から、単語を自動的にデータ中から切り出すという発想を受け継いでいる。また同時に、AUTOCODE（佐藤[11]、田中・大郎丸[13]）から、元のテキストデータを検索・表示したり、コーディング基準を明示するた

めの仕組みなど、優れた部分を取り入れている。ＫＨ ＣＯＤＥＲを用いて行われた研究には、大瀧[5]、川端・樋口[10]、樋口[16][17]などがある。

(4) 茶筌は現在http://chasen.aist-nara.ac.jp/から入手できる。

(5) ファイルやフォルダの所在を示す文字列。ファイルやフォルダのコンピュータ内での住所に当たる。

(6) このファイルは青空文庫（http://www.aozora.gr.jp）から入手した。

(7) テキストエディタ「秀丸」はhttp://hidemaru.xaxon.co.jp から入手できるシェアウェアであり、実際の置換・マーキング時には正規表現を用いた。正規表現の詳細については秀丸のヘルプ等を参照されたい。

(8) 例えば「卒業証書」という言葉が、「卒業」と「証書」の2語に分けて抽出されるなど、ＫＨ ＣＯＤＥＲ（茶筌）による単語の切り出し方が細かすぎる場合にも、必要であれば「強制抽出・強調する語」の指定を行えばよい。もし、こういった細かすぎる分割の組み合わせを洗い出したければ、ＫＨ ＣＯＤＥＲの設定画面で「複合名詞を使用する」にチェックを入れてから前処理を行い、「複合名詞」という品詞名で抽出された語を確認すればよい。「複合名詞」とは、名詞が連続して抽出されていたときに、それらの語を連結して1つの語としたものである。もちろん、無理のある連結も多数なされてしまうのだが、出現数の多いものは納得できる連結であることが多い。連結した状態で抽出した方が良い言葉を確認したら、「複合名詞を使用する」のチェックをはずしてから、それらの語を「強制抽出」するよう指定して、再度前処理を行うとよい。

(9) 上・中・下といった、もともとテキストファイルの中にある見出しのことを、「外部変数」と呼ぶのは違和感があるかもしれない。だが、通常は、ある言葉が文の中に含まれているかどうかという条件（の組み合わせ）でしか、文を検索・コーディングすることができないのに対して、含まれている語以外のもの（見出し）を条件にできるという意味で「外部変数」と呼ばれる。

(10) 品詞名の最後にＢがついているものは、ひらがなだけからなる語で、例えば「する」、「なる」、「そう」といったように、一般的で分析には使いにくい語が多い。

(11) この点については、例示している部分にいくぶん違いはあるものの、「私との交流の中で、先生は常に死や自殺の意識にとりつかれていたことを読書は想起できる。雑司ヶ谷の墓地でのエピソード、不自然な暴力をめぐる会話、私の父の病気から死についての対話などである」として仲[15: p. 155]も指摘している。

文献

[1] Kawabata, A. & Akiba, Y. 2001 "Deep into the Shinnyo Spritual World" *International Journal of Japanese Sociology* Vol. 10 pp. 5-15

[2] Krippendorff, K. 1980 *Contnt Analysis: an Introduction to its Methodology*. London: Sage. （三上俊治他訳『メッセージ分析の技法-「内容分析」への招待』勁草書房 1989）

[3] Seale, C. 2000 "Using Computers to Analyse Qualitative Data" Silverman, D. *Doing Qualitative Research: A Practical Handbook* London:Sage pp. 154-174

[4] 大隅昇 Lebart, L. 2000 「調査における自由回答データの解析-InfoMinerによる探索的テキスト型データ解析」『統計数理』48(2) pp. 339-376

[5] 大瀧友織 2001 「夫婦間における『性格の不一致』という問題-身の上相談を資料として」 第10回日本家族社会学会研究報告9月8日
[6] 大谷尚 松原伸一 1984 「出現語の頻度分布にもとづく授業の特徴化について」『電子通信学会技術研究報告』 ET-84-8 pp.1-6
[7] 景山佳代子 2002 「週刊誌『アサヒ芸能』にみる性風俗生成の仕組み」『マス・コミュニケーション研究』 61 pp. 107～120
[8] 川端亮（編著） 1999 『非定型データのコーディング・システムとその利用』 平成8年度～10年度科学研究費補助金（基盤研究(A)(1)）課題番号08551003）研究成果報告書
[9] 川端亮 2001 「コンピュータ・コーディングによる宗教的ライフヒストリーの記述」 『宗教と社会』 7 pp. 133-154
[10] 川端亮 樋口耕一 2003（近刊） 「インターネットに対する人々の意識-自由回答の分析から-」 『大阪大学大学院人間科学研究科紀要』 No. 29
[11] 佐藤裕 1993 「部落問題に関する表現の構造」 『解放社会学研究』 7 pp. 63-86
[12] 柴田好章 1996 「逐語記録にもとづく授業分析の諸手法の検討-質的分析への量的分析の統合をめざして」 『名古屋大学教育学部紀要』 43(2) pp. 217-228
[13] 田中重人 太郎丸博 1996 「文章データのコンピュータコーディング：プログラムとその応用」『第22回数理社会学会大会研究報告要旨集』 pp. 20-23
[14] 谷口敏夫 1999 「全文からの『位置情報付き用語』の抽出」 川端亮（編著）『非定形データのコーディング・システムとその利用』 平成8年度～10年度文部省科学研究費補助金（基盤研究(A)(1)）課題番号08551003）研究成果報告書
[15] 仲秀和 2001 「『こゝろ』の不思議-運命の変化の臨界点・『偶然』と『突然』-」 仲秀和 『漱石-『夢十夜』以後-』 和泉書院
[16] 樋口耕一 2001 「電子コミュニティにおけるメディア特性の影響-同期メディアと非同期メディア」 『年報人間科学』 22 pp. 91-106
[17] 樋口耕一 2002 「コンピュータ・コーディングによる社会情報の分析」 大阪大学大学院修士論文
[18] 松本裕治 北内啓 山下達雄 平野善隆 松田寛 浅原正幸 1999 「日本語形態素解析システム『茶筌』version 2.0 使用説明書第二版」 NAIST Technical Report NAISTIS-TR99012
[19] 見田宗介 1979 『現代日本の社会意識』 弘文堂
[20] 安田三郎 原純輔 1982 『社会調査ハンドブック（第3版）』 有斐閣

Practice of Computer Coding: A Tutorial Using Soseki's "Kokoro"

HIGUCHI Koichi

Computer Coding is a relatively new automated approach to code and analyze qualitative data which will be obtained through sociological study. In this approach, qualitative data such as newspaper articles, open-ended questions in social surveys, transcripts of interviews or conversations will be analyzed quantitatively and, in most case, exploratory.

The purpose of this paper is to introduce the approach and the essential software "KH Coder" which is originally developed free software by describing exactly how the analysis in this approach will be done using it. The description in this paper is a tutorial which anyone will be able to do the same thing on his/her computer. And the material used is the masterpiece written by Natsume Soseki, "Kokoro".

In the tutorial, frequently appearing words in the whole piece and distinctive words which are used especially frequently in each part will be inquired to explore the contexture of the piece. Then, where and how words, which describe men's death and cause of the death, appear in the whole piece will be inquired and described in the graphic form to explore how human's death is depicted or treated in the piece. As a result, the author argue that it can be fairly certain that the criticism about suddenness and/or unreasonableness of the suicide commited by the dramatispersonae called "Sensei" is unreasonable.

Key Words

Computer Coding, Content Analysis, Qualitative Data Analysis, CAQDA, KH Coder