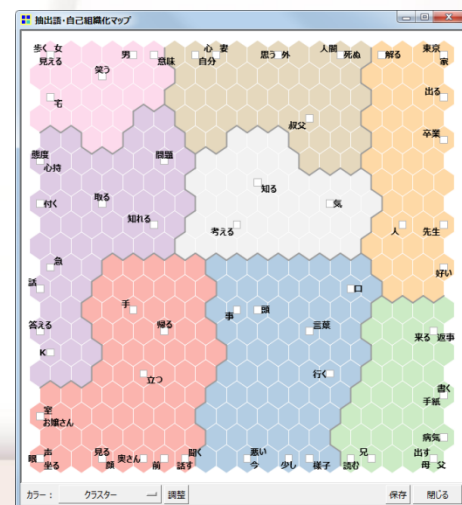
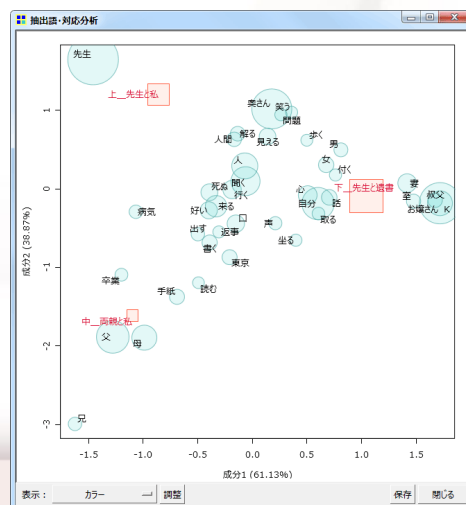
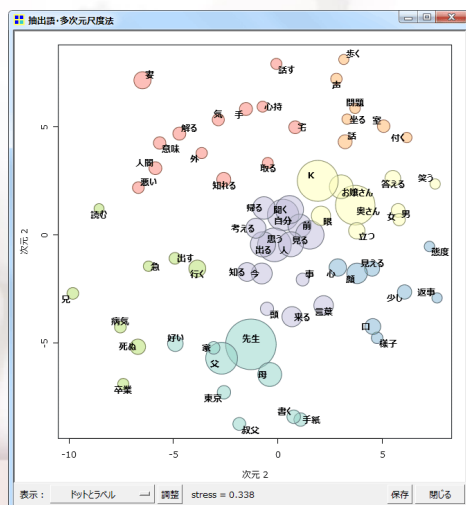


KH Coder 3 チュートリアル

漱石『こころ』を題材に
【スライド版】

本チュートリアルの内容

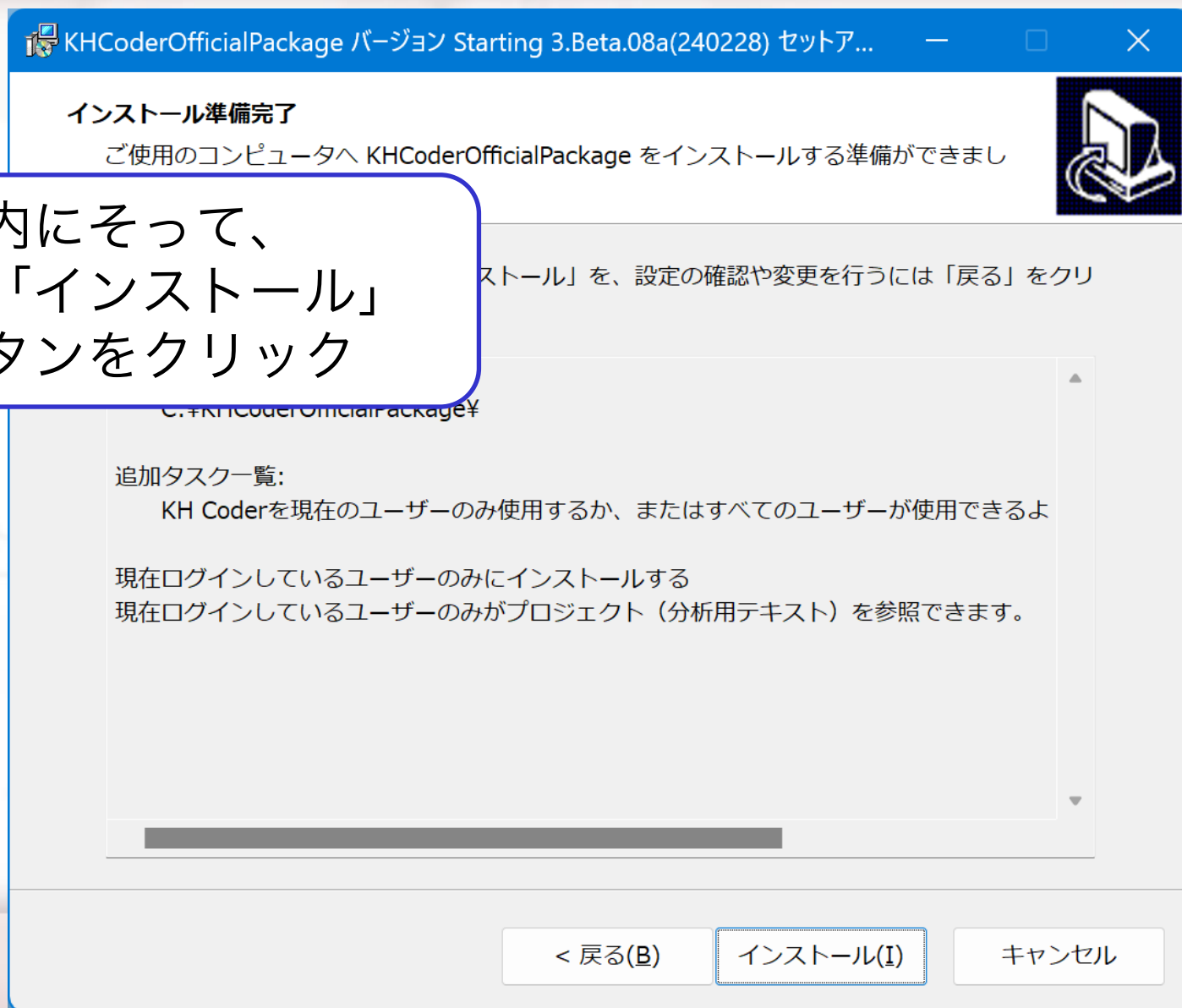
1. KH Coderの準備
2. プロジェクト作成と前処理
3. 頻出語と共起
4. それぞれの部（上・中・下）に特徴的な語
5. コーディングによるコンセプトの抽出



1. KH Coderの準備

1.1 インストール

画面の案内にそって、
「次へ」「インストール」
などのボタンをクリック



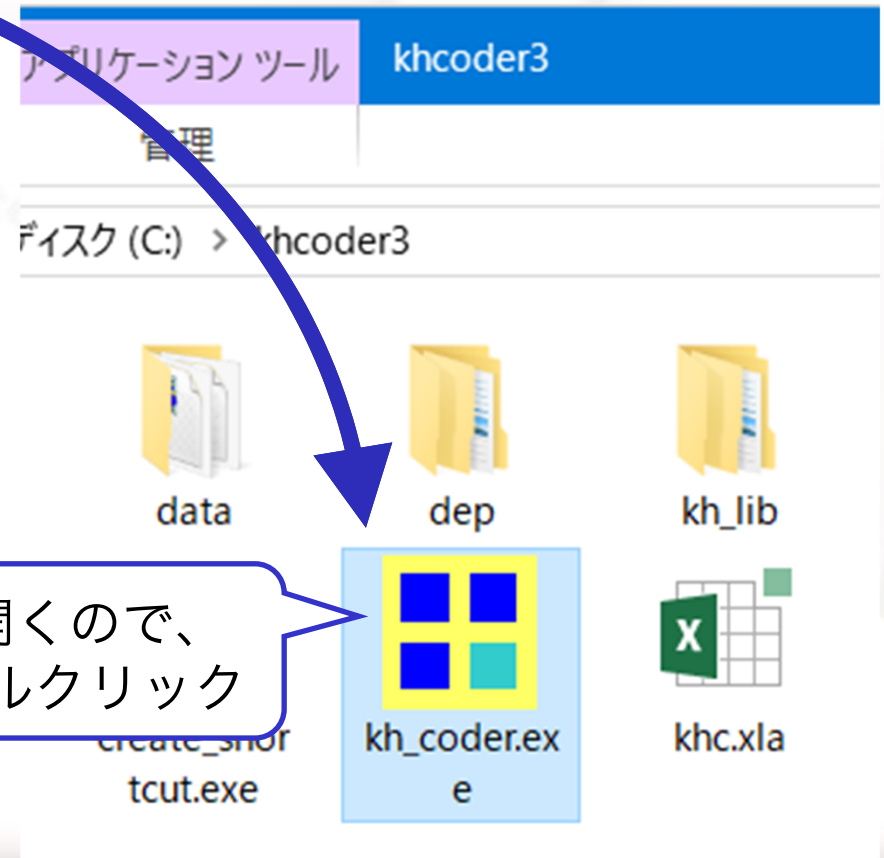
1.2 KH Coderの起動方法



KH Coder 3
Folder

① デスクトップに作成されたアイコンをダブルクリック

② 解凍先フォルダが開くので、「kh_coder.exe」をダブルクリック



※PCの設定によっては、「kh_coder」のように、ファイル名の「.」ピリオド以降を省略した状態で表示されます。

2. プロジェクト作成と 前処理

【解説】 分析のねらいと進め方

本ページのようにタイトルが【解説】で始まるページでは操作不要です。お読みいただくだけで結構です。

■ 漱石『こころ』の探索的分析

- 物語最後の「先生」の自殺が突然で不自然という批判はもっともか？
- 『こころ』の人間観、人間の罪とは？

■ 分析の手順は大きく分けて2段階

- 語を自動的に取り出して集計し、予断を交えずにデータの全体像を探る
- コーディングを行うことで、分析者の観点を活かしてコンセプトを取り出し、分析を深める

【解説】 データの準備

1行目に列の名前を入力

	A	B	C
1	テキスト	部	章
2	私はその人を常に先生と呼んでいた。だ	[1]上 先生と私	1_01
3	私が先生と初めて会ったのは鎌倉で	[1]上 先生と私	1_01
4	学校の校庭にはまだ木が少なかった。	[1]上 先生と私	1_01
5		先生と私	1_01
6	私は毎日海へ行って先生と話をした。	先生と私	1_01
7	私は実際に先生をこの雑沓の間に見付け	[1]上 先生と私	1_01
8	私がその掛茶屋で先生を見た時は、先生	[1]上 先生と私	1_02
9	その西洋人の優れて白い皮膚の色が、	[1]上 先生と私	1_02
10	彼はやがて自分の傍を顧みて、そこにこ	[1]上 先生と私	1_02
11	私は単に好奇心のために、並んで浜辺を	[1]上 先生と私	1_02
12	彼らの出て行った後、私はやはり元の床	[1]上 先生と私	1_02
13	その時の私は屈託がないという上は	[1]上 先生と私	1_02
14		先生と私	1_03
15		先生と私	1_03
16	次の日私は先生の後につづいて海へ飛	[1]上 先生と私	1_03
17	しばらくして海の中で起き上がるように姿	[1]上 先生と私	1_03
18			1_03
19			1_03
20			1_03

2行目からデータを入力

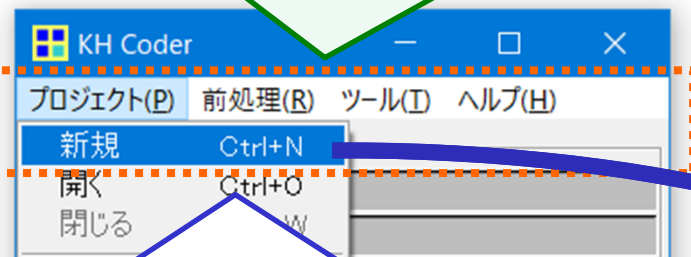
1つ目のシートに入力

※データ量が多くてKH Coderでエラーになる場合、CSV形式で保存してからKH Coderへ

- 『こころ』は上・中・下の3部構成。部の中がさらに一・二・三などの章に分かれる
- Excelで1つの列にテキストを入力し、ほかの列にテキスト以外の情報（今回は部や章の番号）を入力。テキスト以外の情報のことを「外部変数」と呼ぶ。

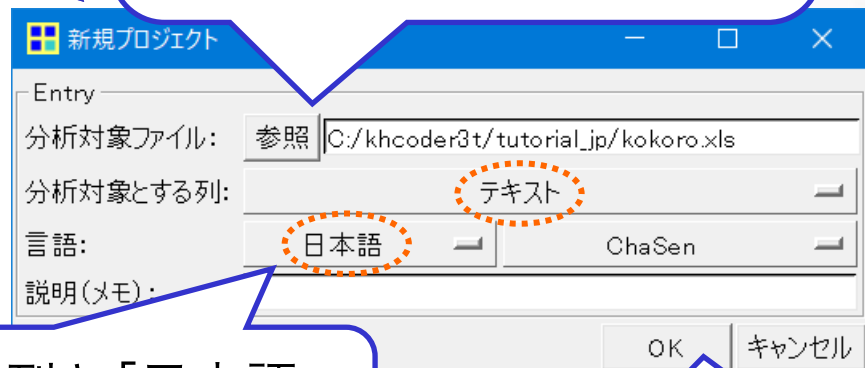
2.1 プロジェクト作成（分析するファイルを指定）

この部分を「メニュー」と呼ぶ



① メニューから「プロジェクト」
「新規」をクリック

② 「参照」をクリックして
「tutorial_jp」フォルダ内の
「kokoro.xls」（Excelアイコン）を開く



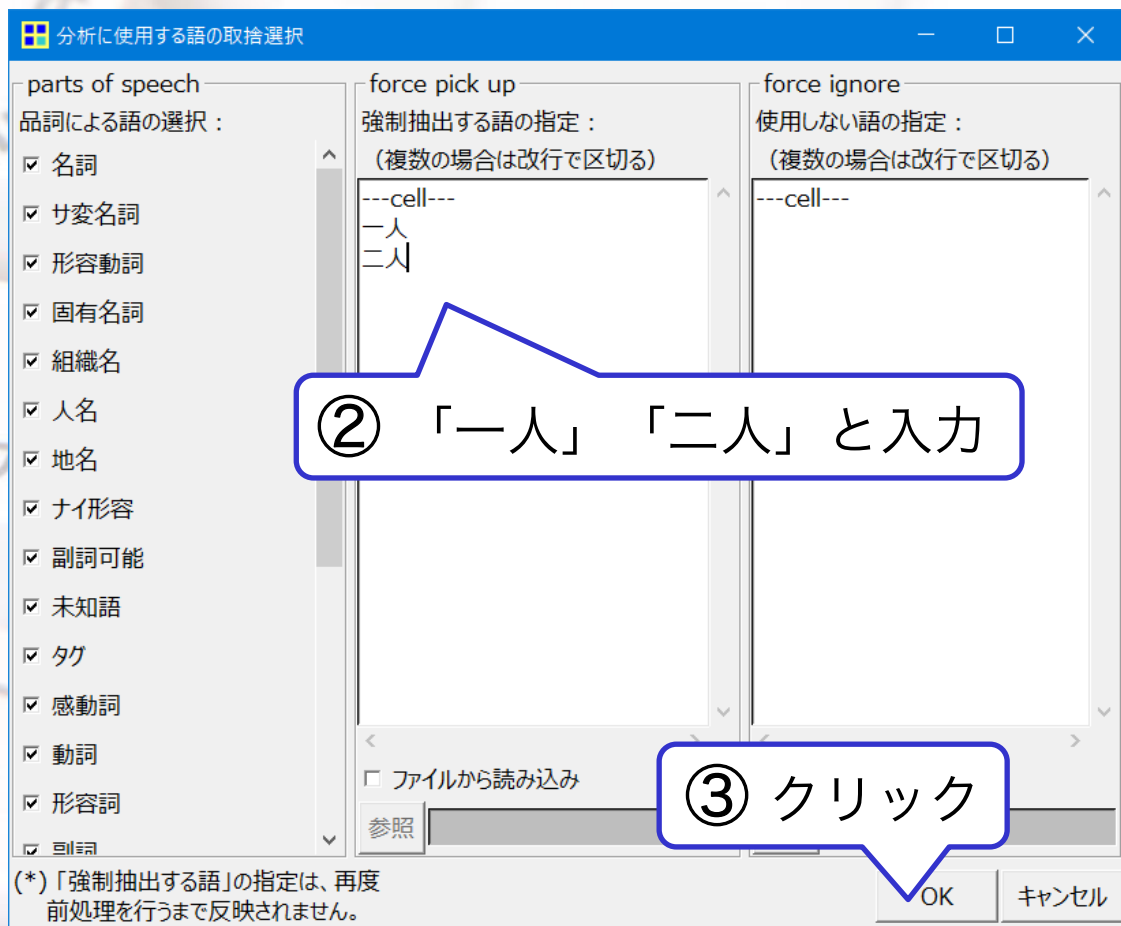
③ 「テキスト」列と「日本語」
が選択されていることを確認

④ クリック

■ 次回KH Coderを起動した時は「新規」ではなく、「プロジェクト」「開く」を選択して、既存のプロジェクトを開く

2.2 強制抽出する語の指定

① メニューから「前処理」「語の取捨選択」を選択



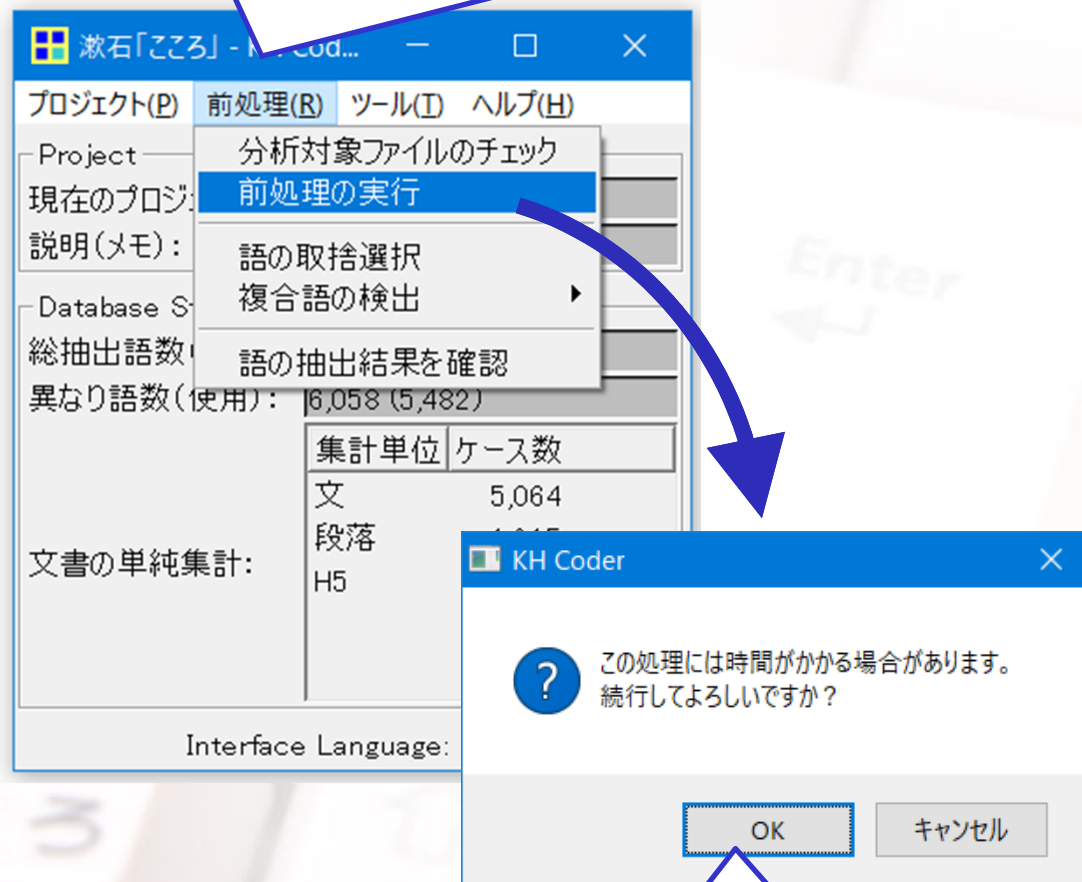
■ 重要な言葉なのに、1語として抽出されない時は「強制抽出」

■ 「一」と「人」に分かれてしまうような、分割が細かすぎる場合にも有効

■ 細かすぎる分割を洗い出すには「前処理」「複合語の検出」コマンドが便利

2.3 前処理の実行

① メニューから「前処理」「前処理の実行」を選択



■ 「前処理」とはテキスト中から自動的に語を取り出して分析の準備をする処理

■ この処理を実行しておかないと分析は行なえない

② クリック

3. 頻出語と共起ネットワーク

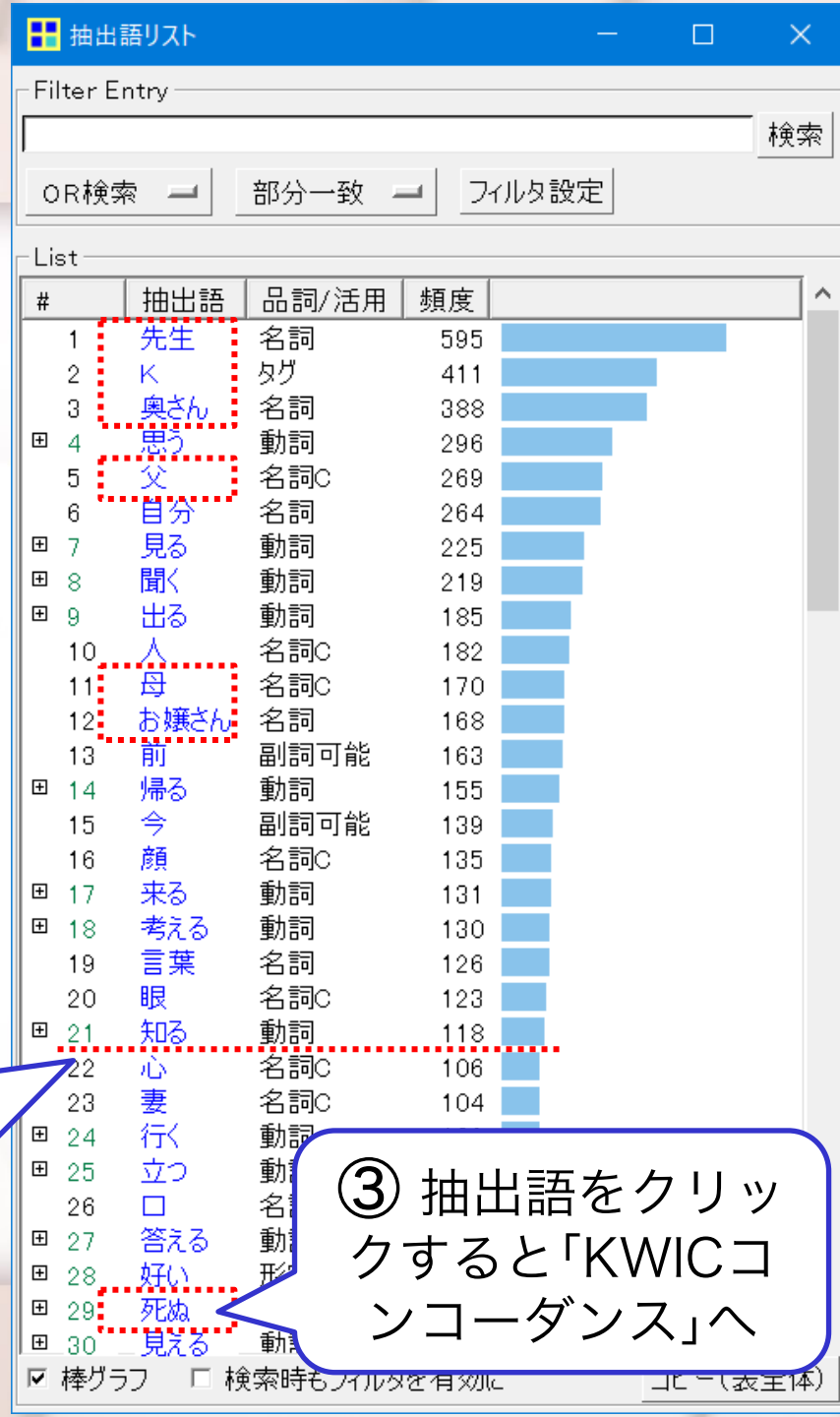
3.1 抽出語リスト

① メニューから「ツール」
「抽出語」 「抽出語リスト」

- 小説なので主要人物が上位に
- 「死ぬ」が89回と多く出現？
- 「一人」より「二人」の方が多い！
(スクロールして確認して下さい)
- 活用のある語は基本形に直して
抽出&カウント
ex. 「知らない」 → 「知る」

② 番号をクリックで活用形を表示

田 21	知る	動詞	118	
├	知ら	未然形!	55	
├	知っ	連用タ接続	39	
├	知り	連用形	14	



③ 抽出語をクリック
すると「KWICコ
ンコーダンス」へ

3.2 「KWICコンコーダンス」 1/2

Search Entry

抽出語: 死ぬ 品詞: 動詞 活用形: 追加条件 検索

ソート1: 出現順 ソート2: 出現順 ソート3: 出現順 (前後24 語を表示) Ready.

Result

① ダブルクリックでさらに広い範囲の文脈を表示可能

② 「集計」をクリックすると「コロケーション統計」へ

- 分析結果中の語をクリックすればこの画面が開く
- 数値だけでなく語の前後の文脈 (context) を見るのが大切
- KWIC = Key Word In Context

3.2 「KWICコンコーダンス」 2/2

① 前ページの手順で「コロケーション統計」が開く

コロケーション統計

Node Word

抽出語: 死ぬ

品詞: 動詞

活用形

Result

N	抽出語	品詞	合計	左合計	右合計	左5	左4	左3	左2	左1	右1	右2	右3	右4	右5	MI3
1	死ぬ	動詞	12	6	6	2	1	0	1	2	2	1	0	1	2	14.501
2	乃木	人名	4	4	0	0	0	4	0	0	0	0	0	0	0	13.221
3	父	名詞C	12	8	4	1	0	0	7	0	0	1	0	0	3	12.000
4	先	名詞C	5	5	0	0	0	0	0	0	0	0	0	0	0	1.000
5	前															
6	母															
7	生きる	動詞	4	2	2	1	1	0	0	0	0	0	0	2	0	1.000
8	大将	名詞	2	2	0	0	0	0	2	0	0	0	0	0	0	1.000
9	暴力	名詞	2	1	1	0	0	0	1	0	0	0	0	0	1	1.000
10	分る	動詞	4	0	4	0	0	0	0	0	2	1	1	0	0	1.000
11	人	名詞C	6	2	4	1	0	0	1	0	2	1	1	0	0	1.000
12	思う	動詞	7	2	5	0	0	0	0	0	0	0	0	0	0	1.000

「左1」とは、左側の1つ目、すなわち直前に出現していた回数のこと

「父」は「死ぬ」の2語前に7回出現した

品詞による語の選択:

- ☒ 名詞
- ☒ サ変名詞
- ☒ 形容動詞
- ☒ 固有名詞
- ☒ 組織名
- ☒ 人名

すべて

既定値

クリア

前後に多く出現していた語を通じて、検索対象語がもとのテキスト中でどう使われていたかという文脈をうかがえる

② クリックすれば特定の品詞の語だけを表示可

3.3 共起ネットワーク

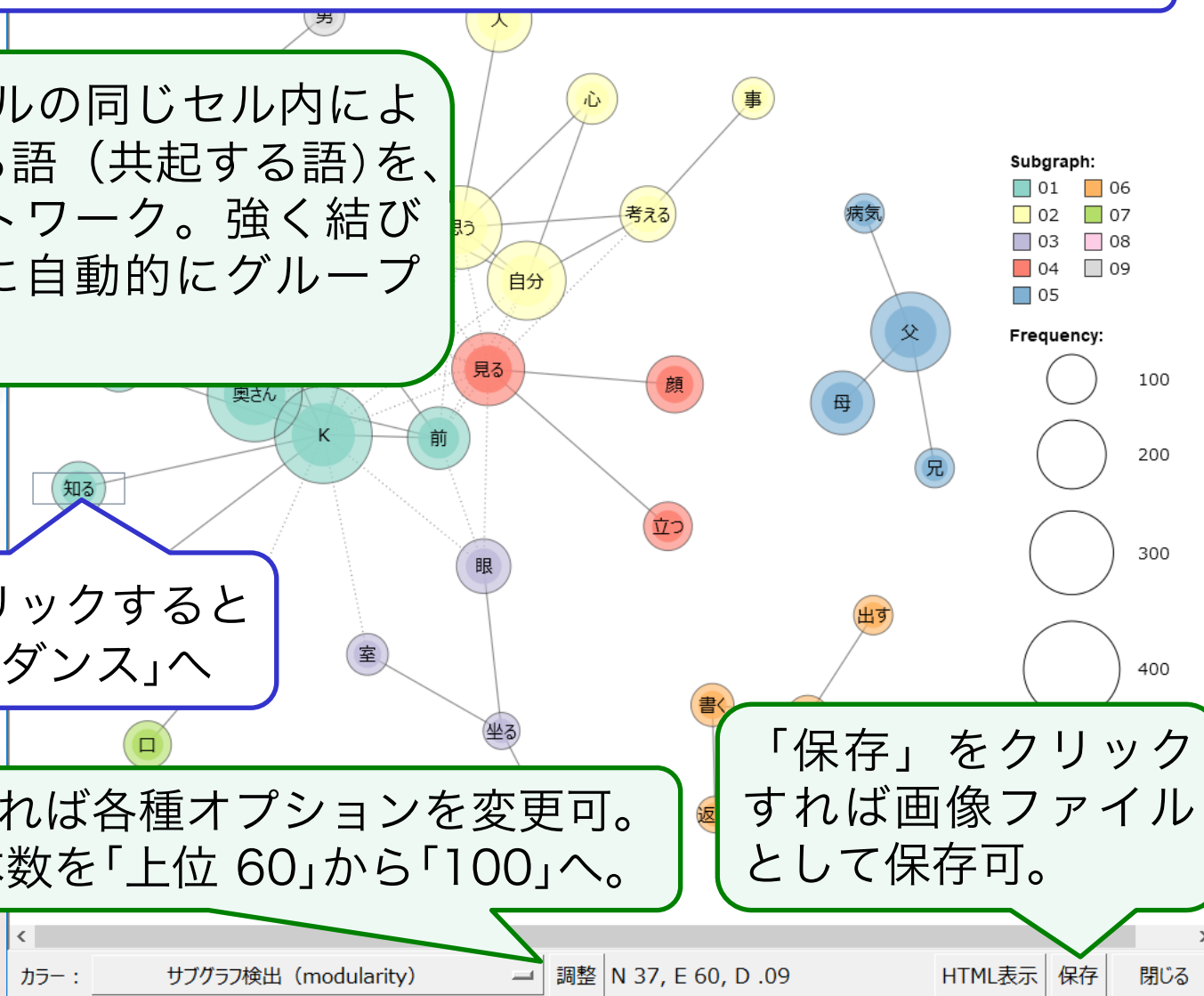
① メニューから「ツール」「抽出語」「共起ネットワーク」→「OK」

入力Excelファイルの同じセル内によく一緒に出現する語（共起する語）を、線で結んだネットワーク。強く結びついた部分ごとに自動的にグループ分け&色分け。

② 抽出語をクリックすると「KWICコンコードダンス」へ

「調整」をクリックすれば各種オプションを変更可。
Ex. 線（edge）の本数を「上位 60」から「100」へ。

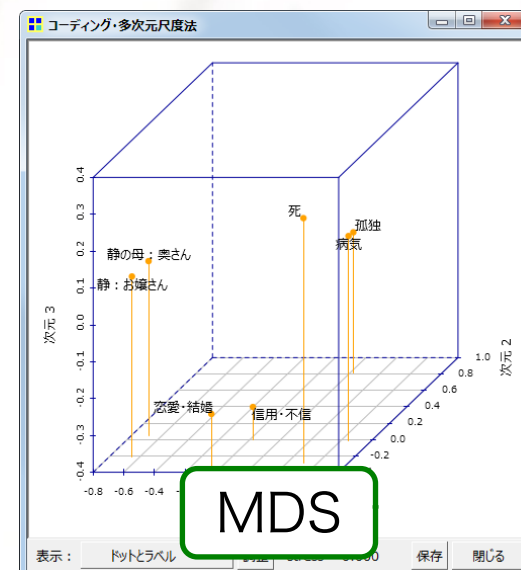
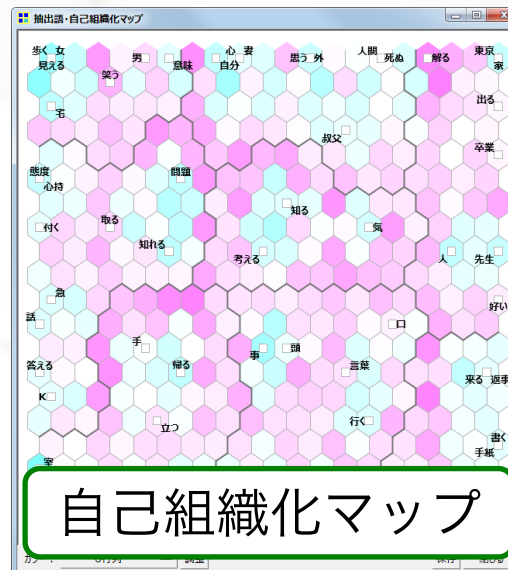
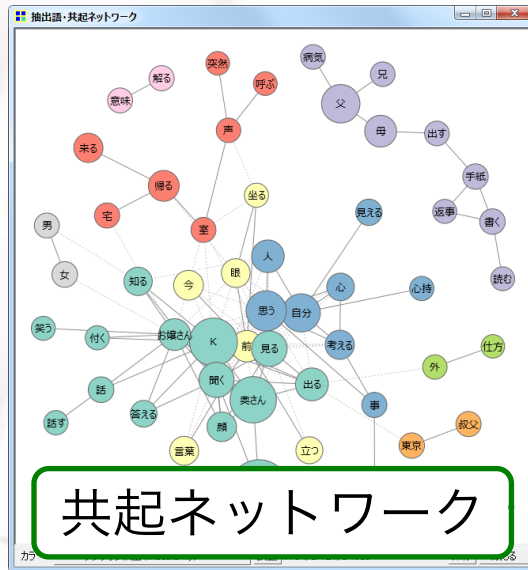
「保存」をクリックすれば画像ファイルとして保存可。



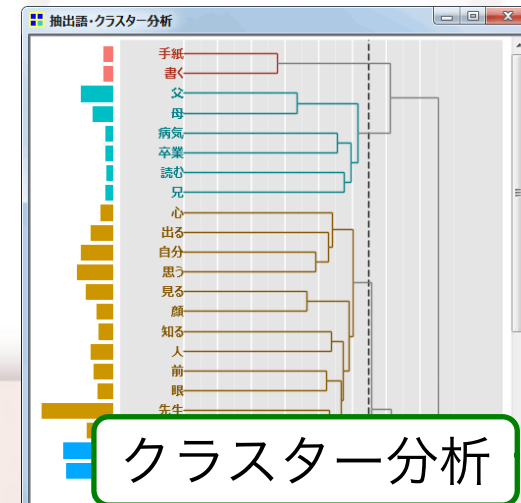
【解説】 語の共起を探索する方法

■ 複数の方法の中から選択可

※自己組織化マップは長時間を要するので注意



- 共起する語のグループから、データ中のテーマないしトピックを探索できる
- これらの分析には内部で「R」という統計ソフトウェアを使用

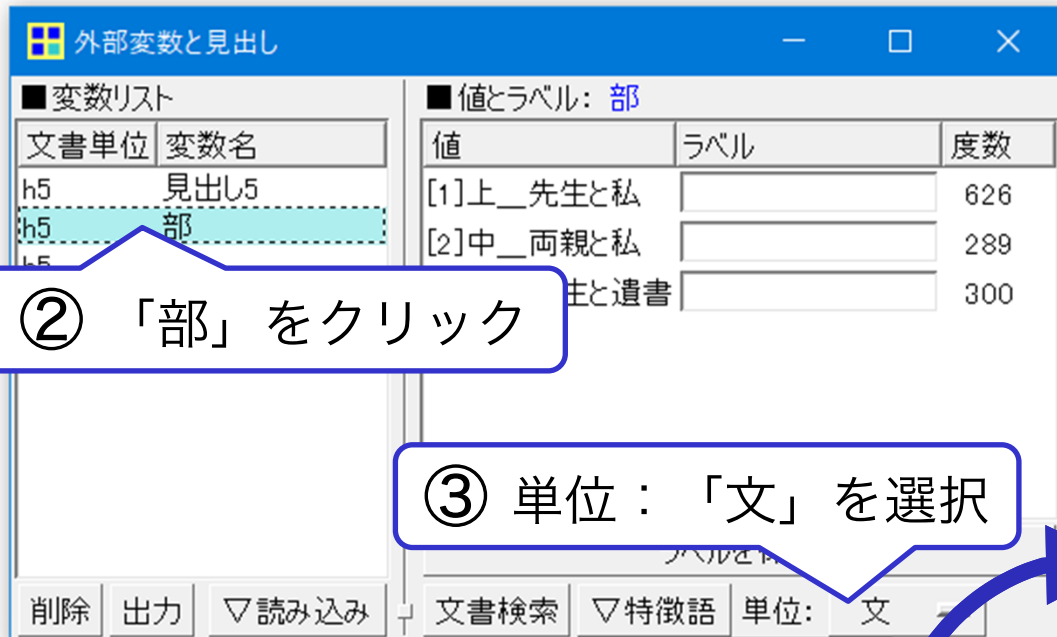


4. それぞれの部（上・中・下）に 特徴的な語

——特徴語 & 対応分析——

4.1 それぞれの部の特徴語リスト

① メニューから「ツール」「外部変数と見出し」



② 「部」をクリック

③ 単位: 「文」を選択

④ 「特徴語」「一覧 (Excel形式)」を選択

	A	B	C	D	E	F	G	H
1								
2		上_先生と私		中_両親と私			下_先生と遺書	
3	先生	.256	父		.191	K		.158
4	奥さん	.090	母		.117	奥さん		.072
5	人	.032	兄		.062	お嬢さん		.067
6	帰る	.030	手紙		.043	思う		.065
7	前	.030	聞く		.041	自分		.061
8	見える	.019	来る		.039	見る		.049
9	解る	.018	卒業		.034	妻		.035
10	少し	.017	出る		.034	出る		.034
11	人間	.016	書く		.031	前		.030
12	死ぬ	.016	東京		.030			.029

3つの部（上・中・下）の特徴語が上位10語ずつリストアップされる。

【解説】 特徴語からみる物語の流れ

	A	B	C	D	E	F	G
1							
2	上 先生と私		中 両親と私			下 先生	
3	先生	.256	父	.191	K		
4	奥さん	.090	母	.117	奥さん		
5	人	.032	兄	.062	お嬢さん		
6	帰る	.030	手紙	.043	思う		
7	前	.030	聞く	.041	自分		
8	見える	.019	来る	.039	見る		
9	解る	.018	卒業	.034	妻		
10	少し	.017	出る	.034	出る		
11	人間	.016	書く		前		

下は「先生」が主人公に送った長い遺書。親友「K」と「お嬢さん」との三角関係などを通じて、上では主人公にも読者にも「解ら」なかった事柄が明らかに。

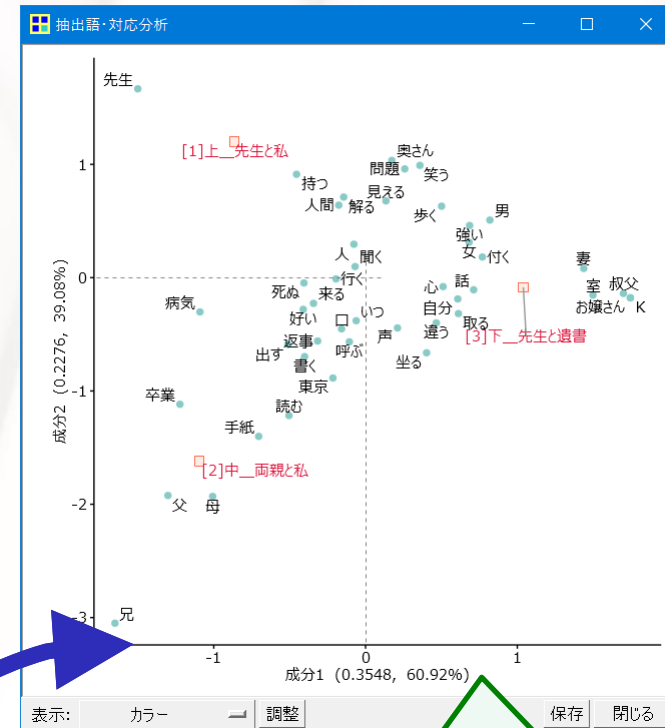
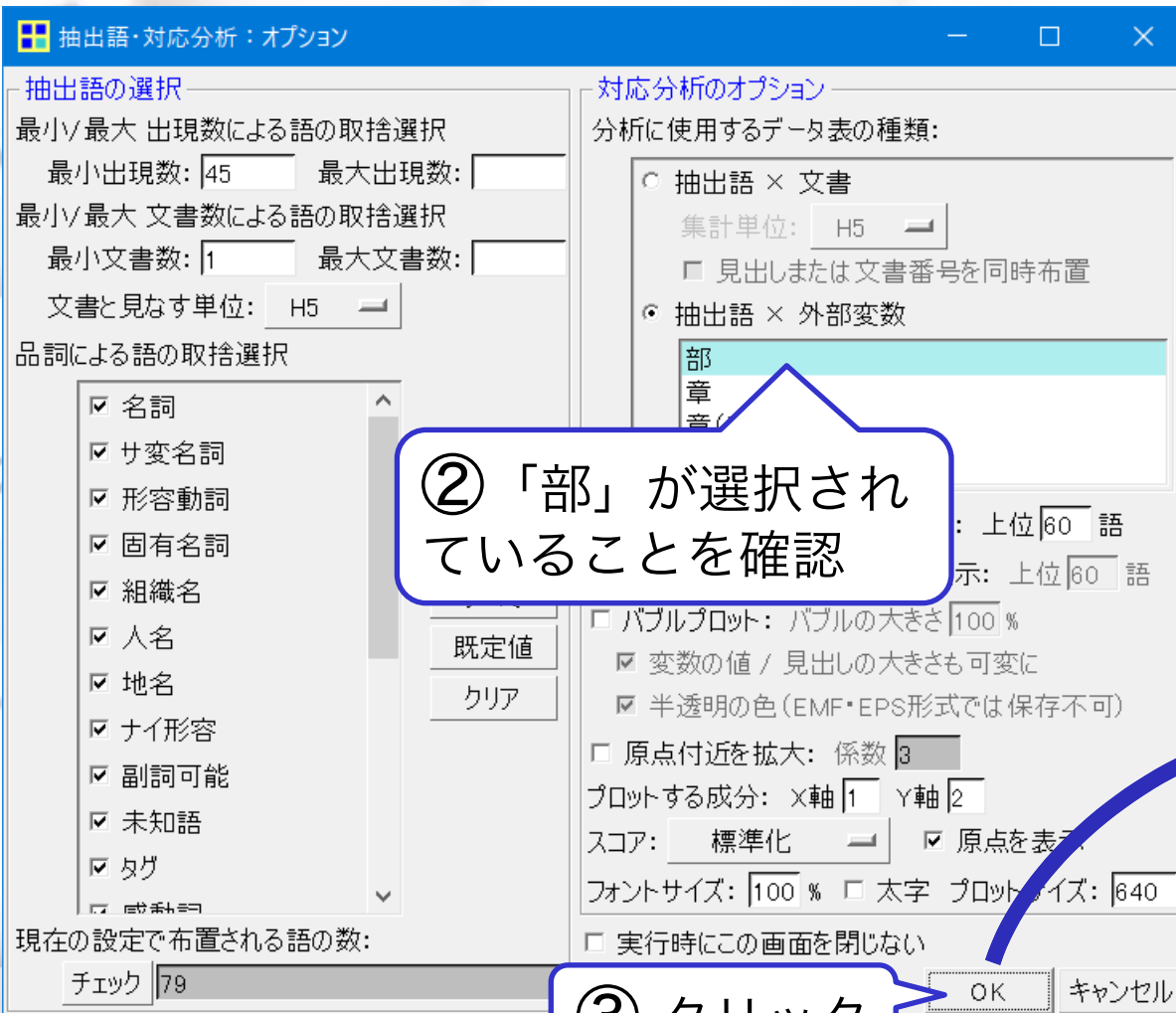
上では主人公と「先生」が交流を深めていく。しかし「先生」の話はよく「解ら」ないことも多い。

中では主人公が実家に戻り「父」「母」との会話が多くなる。「先生」とも「手紙」をやりとり。

Ex. 「先生の話のうちでただ一つ底まで聞いたかったのは、人間がいざという間に、誰でも悪人になるという言葉の意味であった。単なる言葉としては、これだけでも私に解らない事はなかった。しかし…」(上二九)

4.2 「対応分析」による視覚的な探索

① メニューから「ツール」「抽出語」「対応分析」



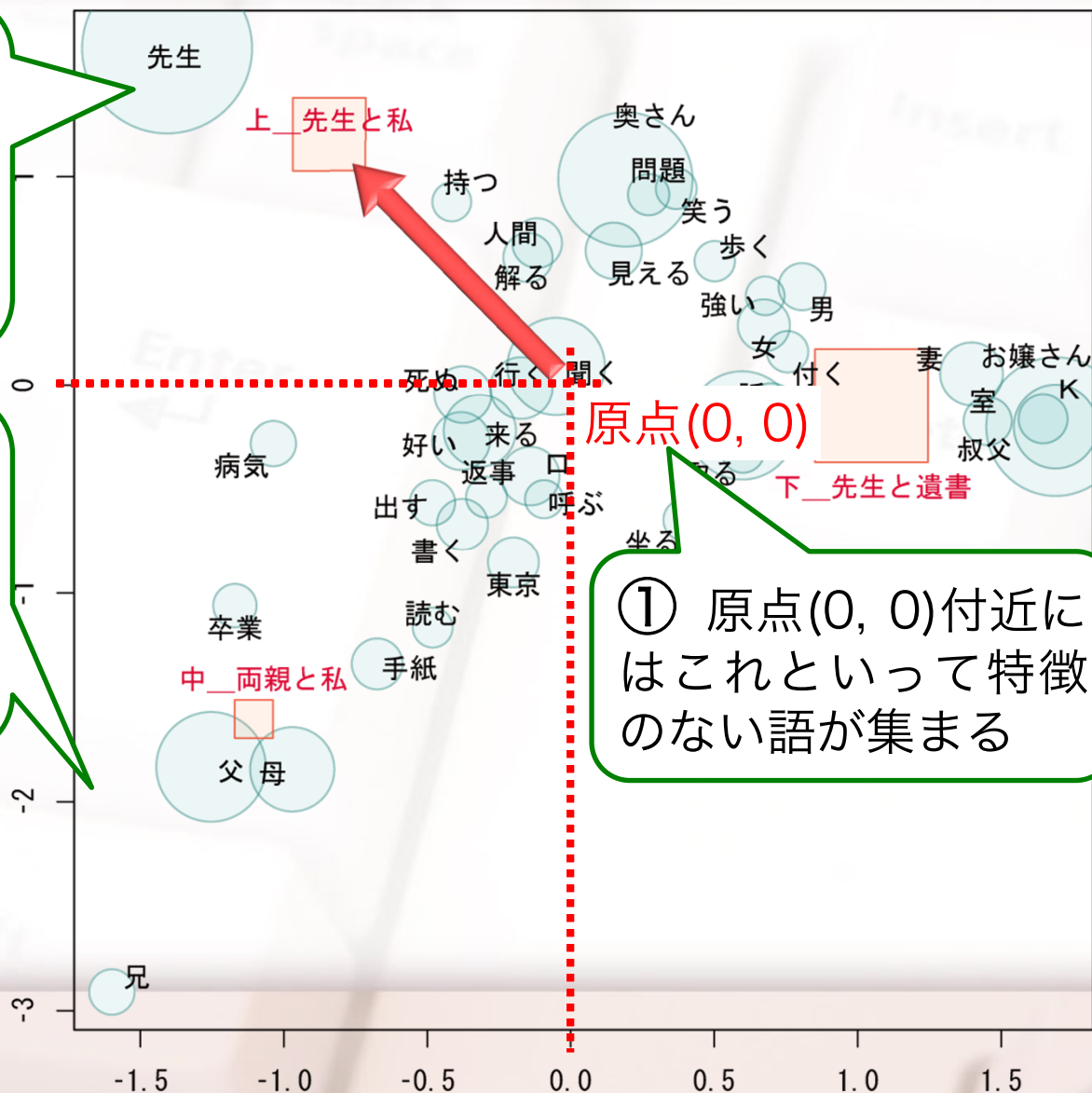
部（上・中・下）の
特徴をグラフィカル
に探索できる

【解説】「対応分析」の見方

② 原点(0, 0)から見て、「上_先生と私」の方向にある語、そして原点から離れている語ほど、上に特徴的！

③ 同様に、原点(0, 0)から見て「中_両親と私」の方に離れている語ほど、中に特徴的！下も同じ。

※特徴語の一覧から読み取れた上・中・下の特徴と、ほぼ同じ特徴を対応分析からも読み取れる。



5. コーディングによる コンセプトの抽出

【解説】 コーディングとは

- 語ではなくコンセプトを数えたい場合もある
- 例えば「人の死」というコンセプトは、「死ぬ」だけでなく「殺す」という語でも表現される
- コンセプトを数えるためのコーディングルール

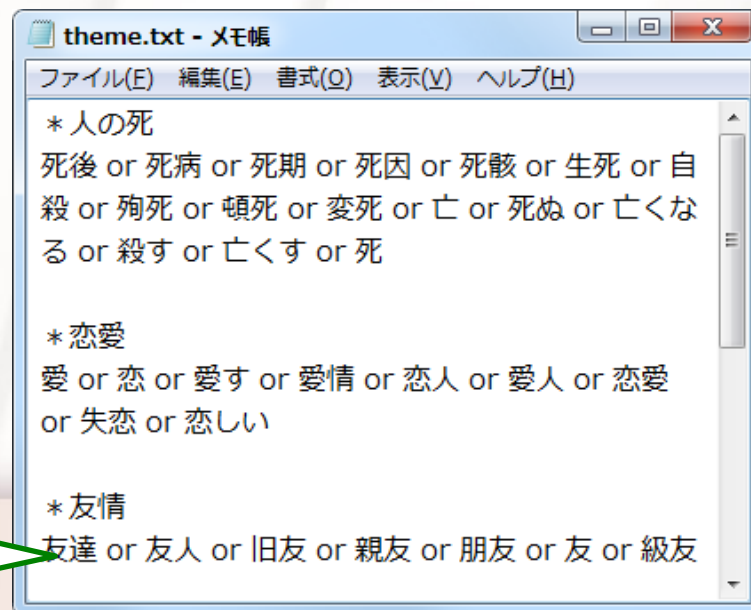
*人の死

コード（コンセプト）の名前

死ぬ or 殺す or 亡くなる

コードを付与する条件。「死ぬ」「殺す」「亡くなる」のどれかが出現している文書は、「*人の死」に言及していたと見なされる。

チュートリアルで使用する「theme.txt」



5.1 コーディングルールによる検索

① メニューから「ツール」「文書」「文書検索」

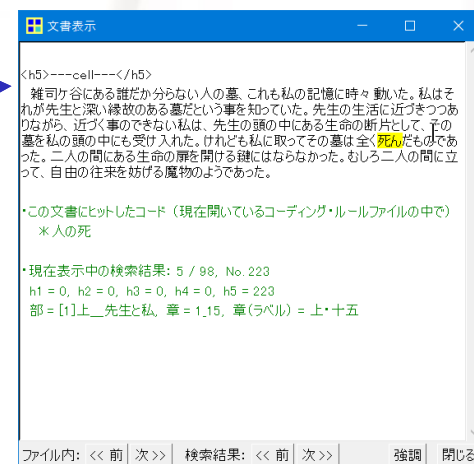
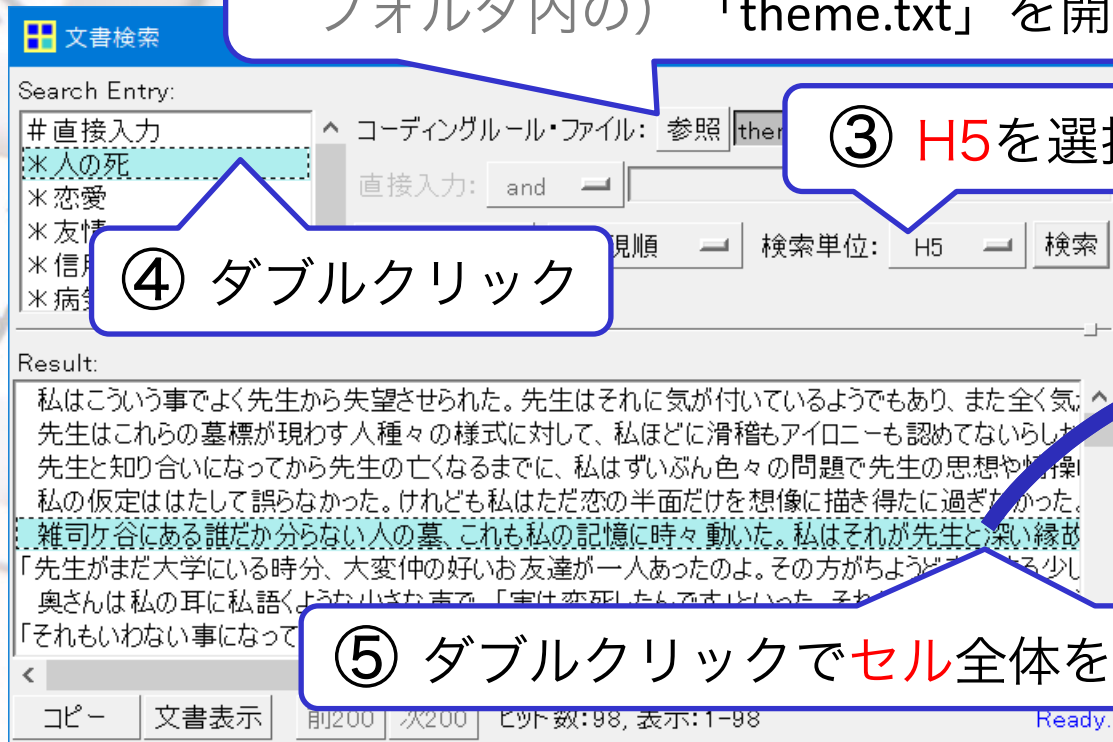
② 「参照」をクリックして (tutorial_jp フォルダ内の) 「theme.txt」を開く

③ H5を選択

Excelの1つ1つのセルをKH Coderは「H5」と認識。H5を選択するとセル単位の検索に。

④ ダブルクリック

⑤ ダブルクリックでセル全体を表示



■ コーディングルール作成時には、どのような文書にコードが付与されているかを検索・確認することが大切

5.2 コードのクロス集計 (部ごと)

① メニューから「ツール」「コーディング」「クロス集計」

② 「参照」をクリックして「theme.txt」を開く
※「theme.txt」とすでに表示されている場合は操作不要

Entry
コーディングルール・ファイル: 参照 theme.txt セル内容: 度数とパーセント
コーディング単位: H5 クロス集計: 部 集計

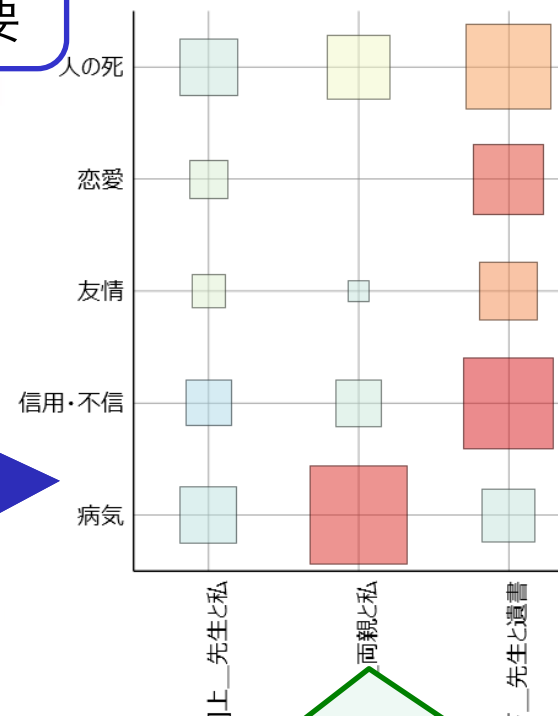
③ H5を選択

④ 「集計」をクリック

「下」では300セルのうち18セル (6%)が
＊友情に言及

⑤ 「バブル」をクリック
してプロットを作成

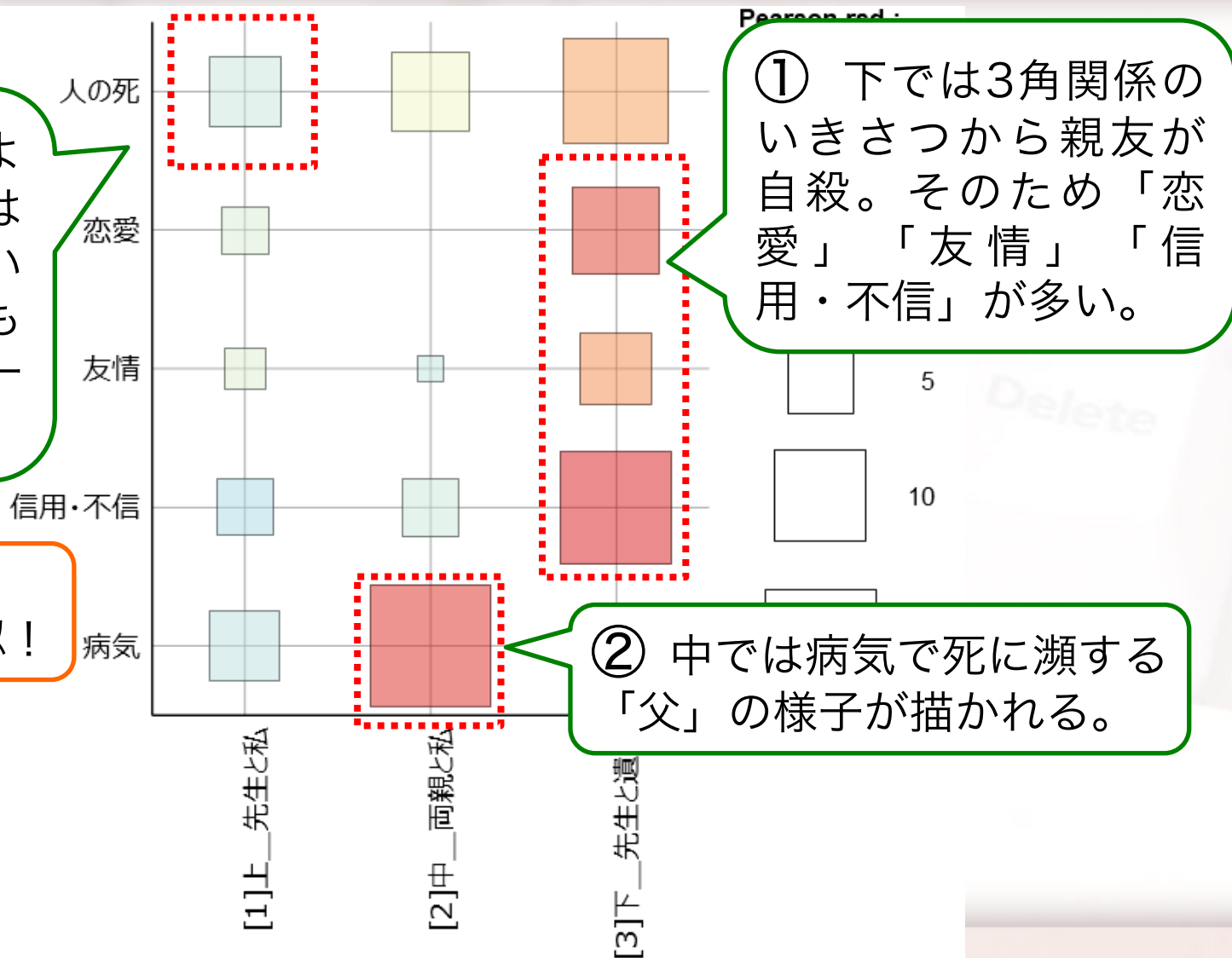
各コードが、どの部で
多く出現していたかを
示すプロット



【解説】 部ごとの集計から見る物語の流れ

③ 中・下のよ
うな人物の死は
描かれていない
のに、上でも
「人の死」が一
定数出現。

なぜ？
次の分析で確認！



5.3 コードのクロス集計（章ごと）

① 5.2に続けて以下の操作を行う

② 「文」を選択

③ 「章」を選択

④ 「集計」をクリック

⑤ 「選択」「人の死」をクリックしてグラフを作成

「人の死」コードがどの章で多く出現していたかを見る折れ線グラフ

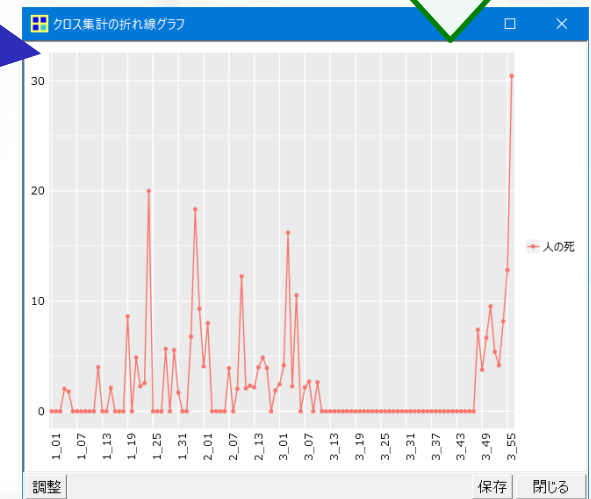
クロス集計

コーディング単位: 文 クロス集計: 章 集計

Result

	* 人の死	* 恋愛	* 友情	* 他
1_01	0 (0.00%)	0 (0.00%)	5 (12.82%)	1
1_02	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)
1_03	0 (0.00%)	0 (0.00%)	0 (0.00%)	1 (2.04%)
1_04	1 (2.04%)	0 (0.00%)	0 (0.00%)	0 (0.00%)
1_05	1 (1.79%)	0 (0.00%)	2 (3.57%)	0 (0.00%)
1_06	0 (0.00%)	1 (2.00%)	0 (0.00%)	0 (0.00%)
1_07	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)

マップ: ヒート バブル 折れ線: すべて 選択 コピー(表全体)

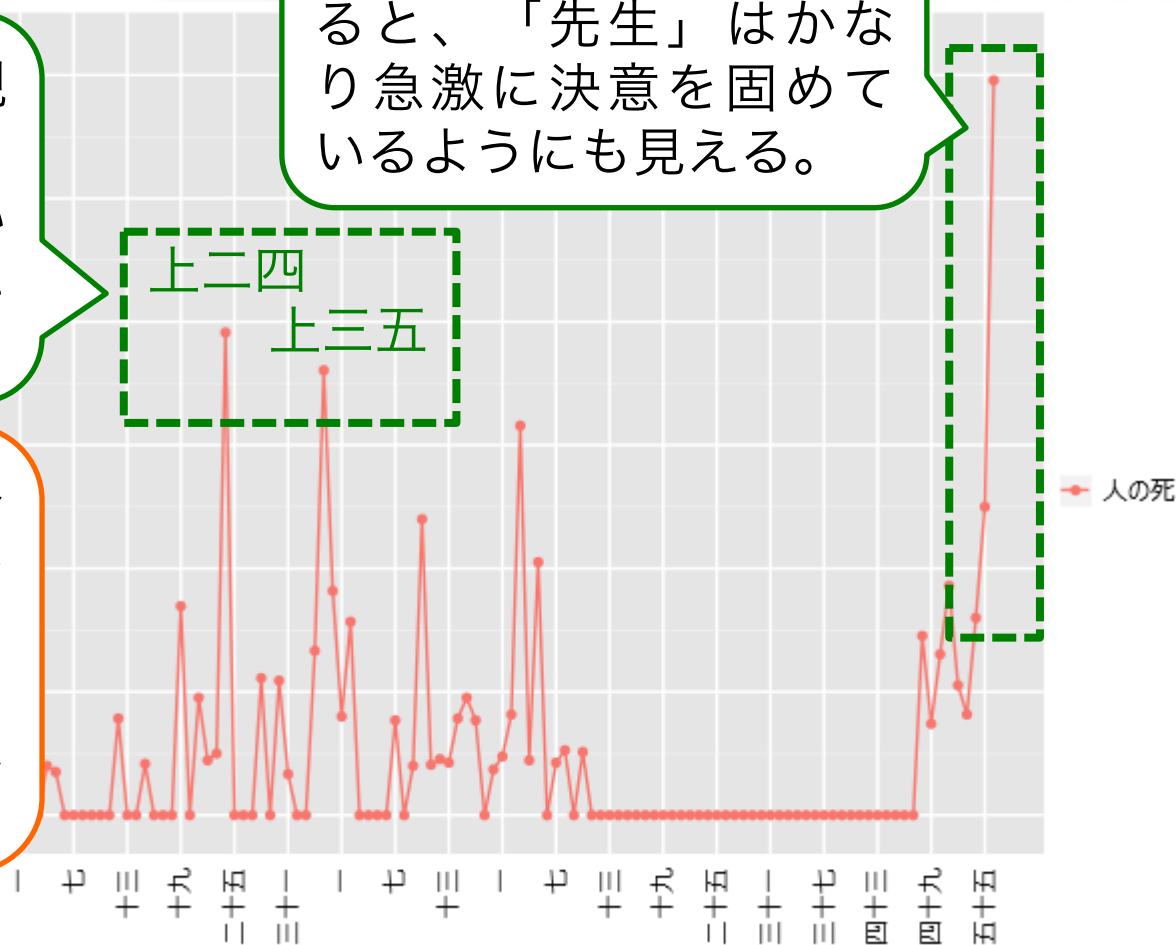


【解説】 「人の死」 の推移から 1/3

② しかし物語全体に視野を広げると、「先生」が死ぬことに執着している様子が、上でも詳細に描かれている。

Ex. 「…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…」 (上三五)

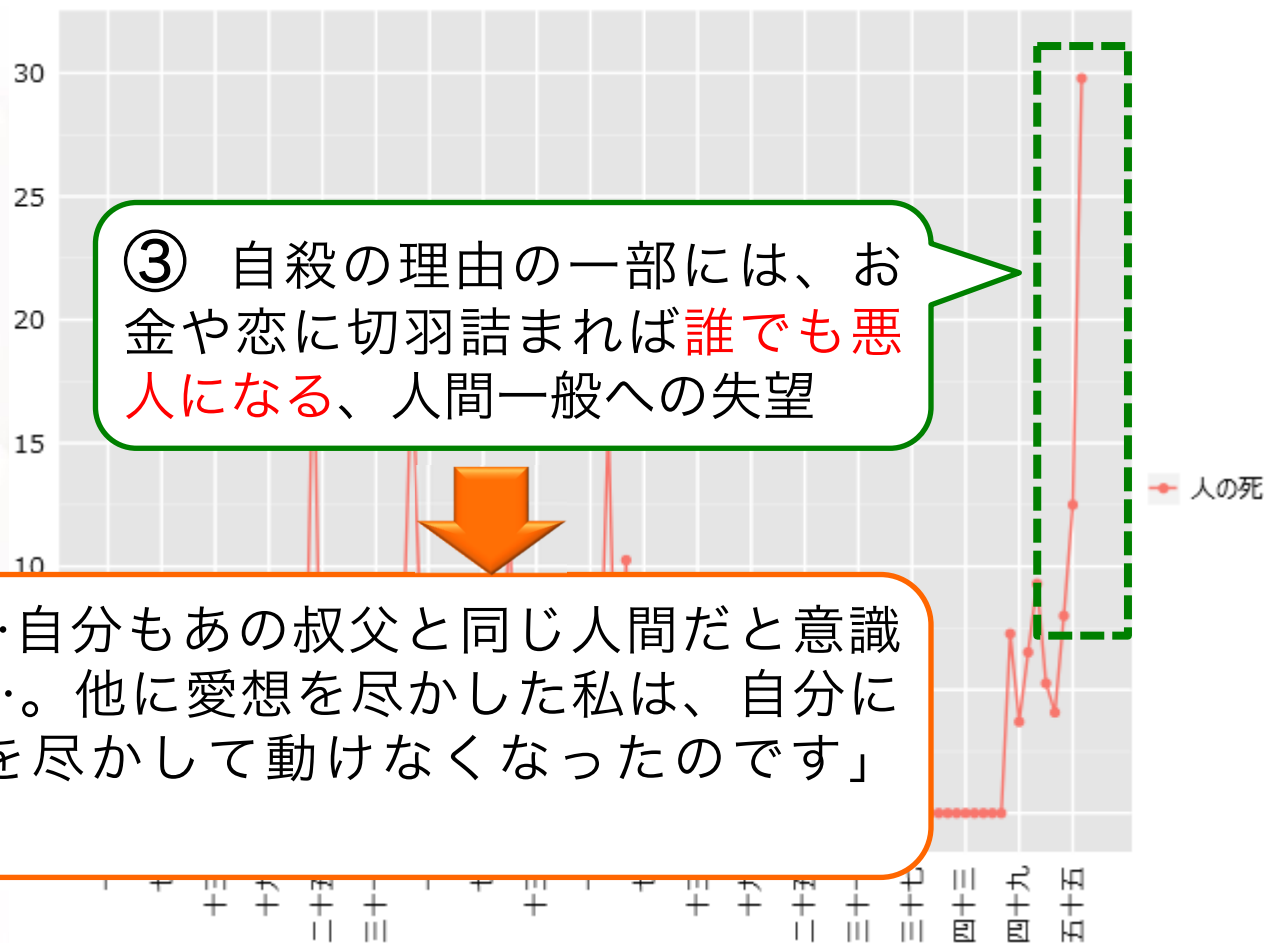
① 自殺の直前だけを見ると、「先生」はかなり急激に決意を固めているようにも見える。



【解説】 「人の死」の推移から 2/3

③ 自殺の理由の一部には、お金や恋に切羽詰まれば誰でも悪人になる、人間一般への失望

Ex. 「…自分もあの叔父と同じ人間だと意識した時…。他に愛想を尽かした私は、自分にも愛想を尽かして動けなくなりました」(下五二)

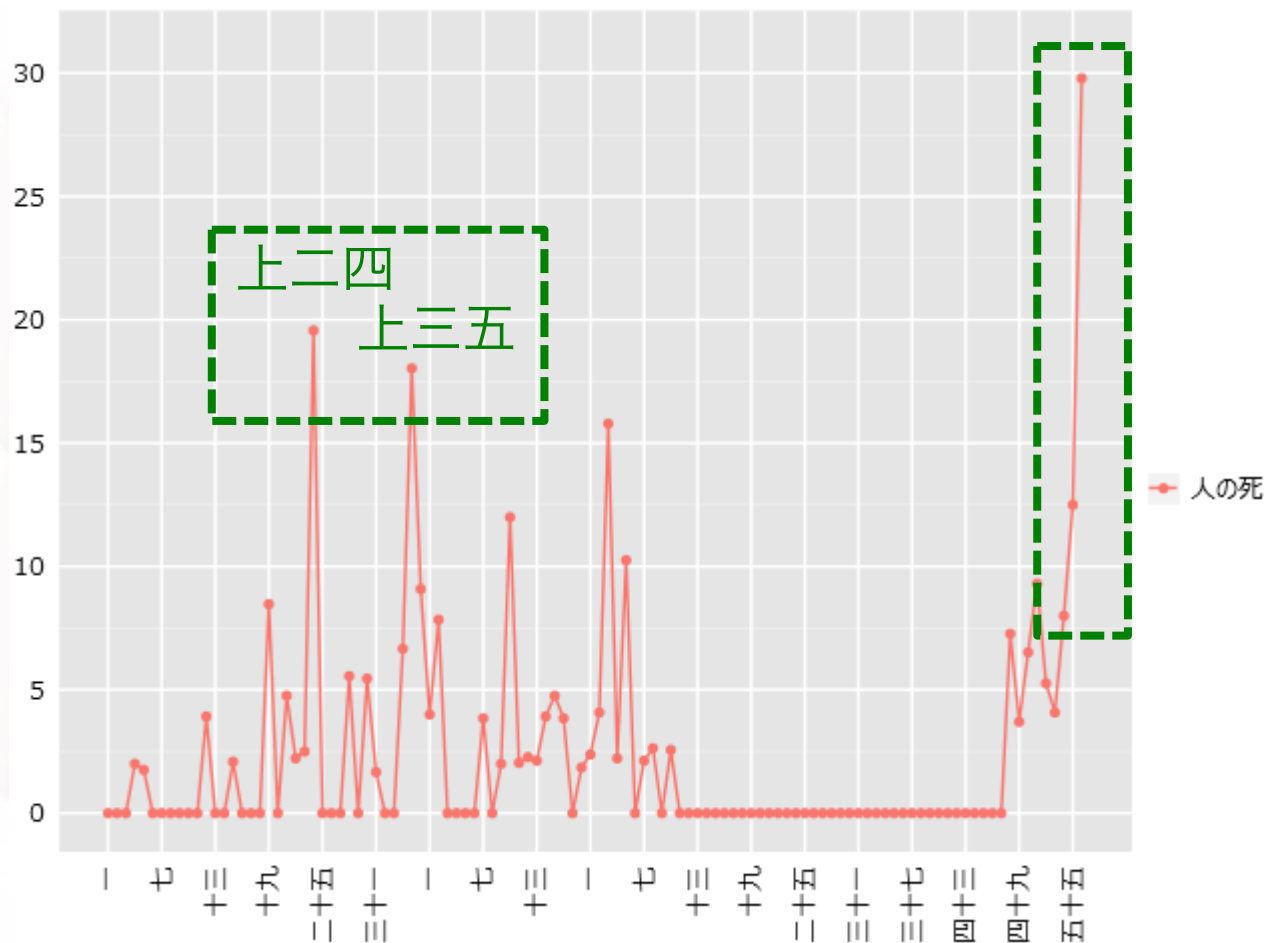


【解説】 「人の死」の推移から 3/3

④ 自殺が突然で
不自然という批判
はちょっと漱石に
対して酷では？

計量テキスト分析
の利点として：

- データ全体を見渡す視点が得られる
- 目で読むべき特徴的な部分はどこか探索できる



【解説】 行なってきた分析のまとめ

■ 抽出語の分析（段階1）

- 多く出現していた語： [抽出語リスト] p. 13
- 共起する語からトピックを探る： [共起ネットワーク] p. 16
- 部分ごとの特徴を探る： [特徴語リスト] p. 19
[対応分析] p. 21
- 語の実際の使われ方を見る： [KWIC] p. 14

■ コーディング（段階2）

- ルール作成と検索： [文書検索] p. 24
- 部分ごとの集計： [クロス集計] p. 26

おわりに

- 各操作の意味や結果の読み取り方についてより詳しくは、「KH Coderの本」第3章をご参照ください。

『社会調査のための計量テキスト分析 第2版』



第2版では、よりよい分析・活用のために研究事例のレビューを増補し、KH Coder 3にも対応しました。

- 謝辞

本チュートリアル作成と改善にあたっては、立命館大学産業社会学部ならびに大学院社会学研究科の皆さまにご協力いただきました。

公式入門書が出ました！



「これだけ押さえておけば大丈夫」という機能を、厳選してご紹介。

初めての方は、こちらからお読みいただくとスムーズでしょう。