

データマイニング

第13回 テキスト分析

2023年春学期

宮津和弘

本日の講義・演習

日付	講義・演習内容
04/14/23	(1) イントロダクション
04/21/23	(2) ビジネスシミュレーション
04/28/23	(3) ID-POSデータ分析
05/12/23	(4) 対応分析
05/19/23	(5) クラスター分析
05/26/23	(6) 自己組織化マップ
06/02/23	(7) 線形判別分析
06/09/23	(8) 非線形判別分析
06/16/23	(9) ツリーモデル
06/23/23	(10) 集団学習
06/30/23	休講 (※黒門祭のため)
07/04/23	(11) サポートベクターマシン
07/14/23	(12) 共分散構造分析
07/21/23	(13) テキスト分析
07/28/23	(14) まとめ (ポートフォリオ)

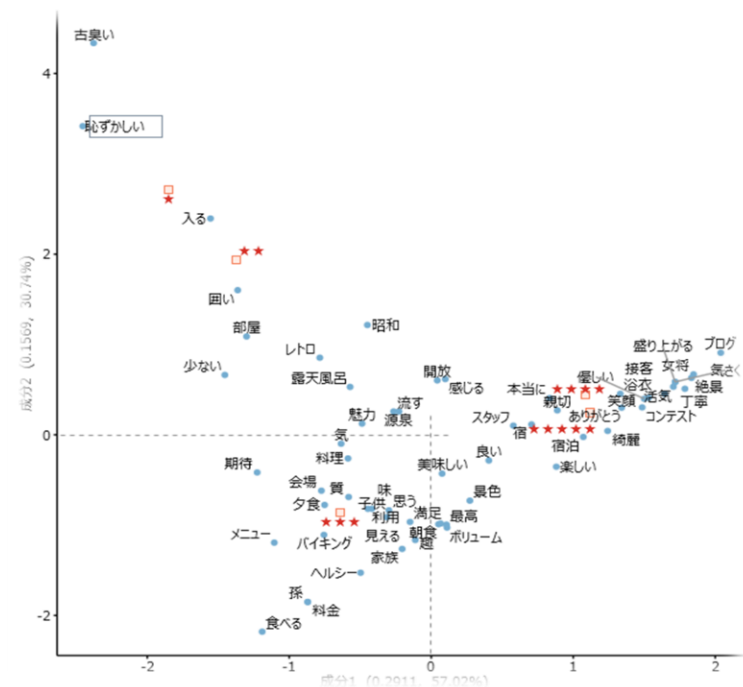
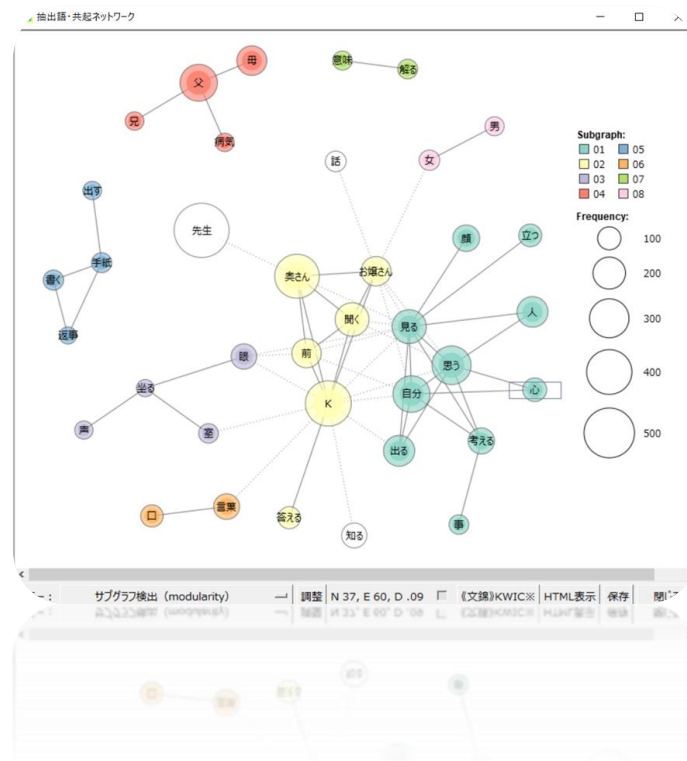


本日の演習概要とポイント

■ テキスト分析

→ 共起ネットワーク

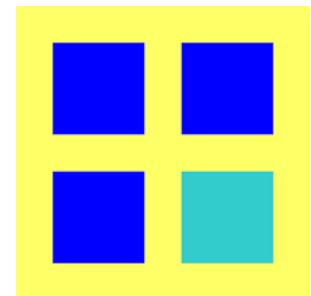
→ 対応分析



テキストマイニングとは

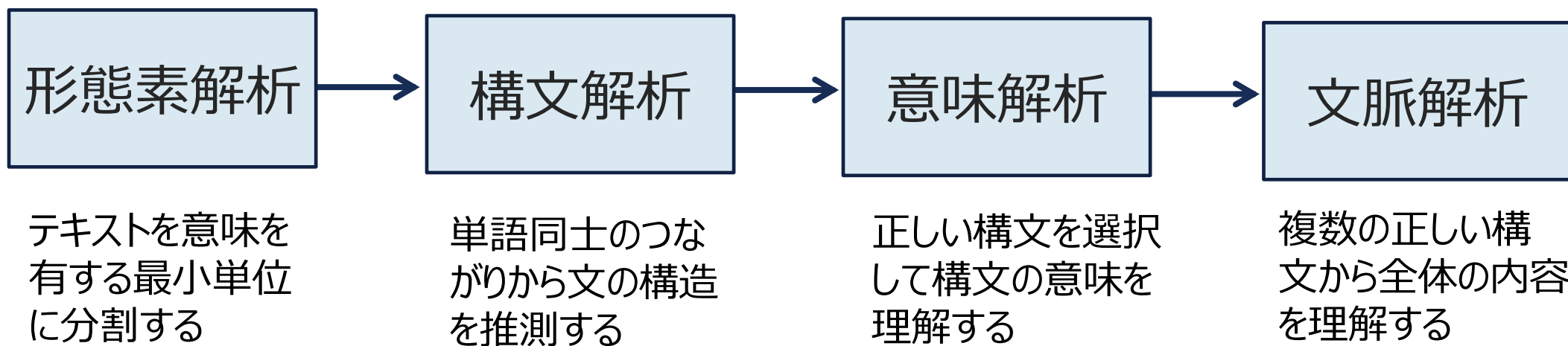
- テキスト・文章を分析対象とする**データマイニングの手法**
- テキストをデータとしてあつかうために、自然言語処理の手法を用いて、データ化してからデータ分析する
- テキストの内容から傾向や特徴を捉えて、新たな知見を見出す

本演習では、**KH Coder** を用いてテキスト分析する



自然言語処理のプロセス

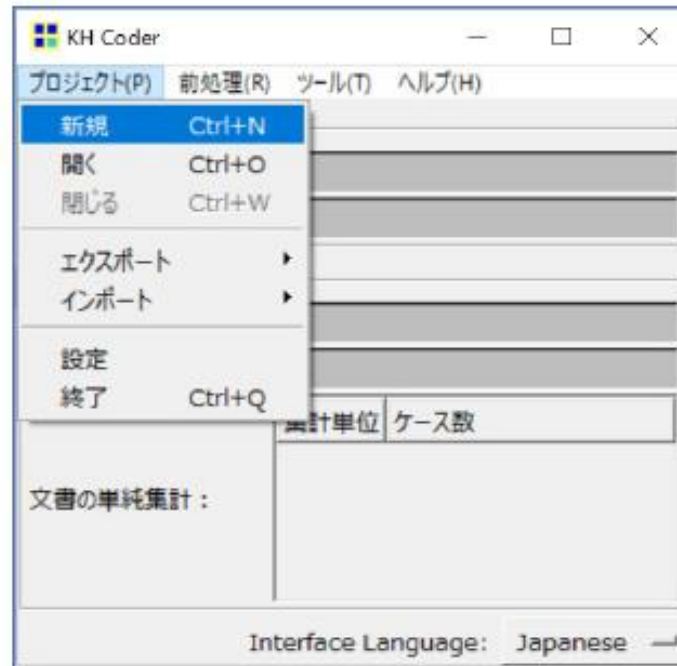
自然言語処理（**NLP** : Natural Language Processing）では、以下の4つのプロセスによりコンピュータに情報処理させる



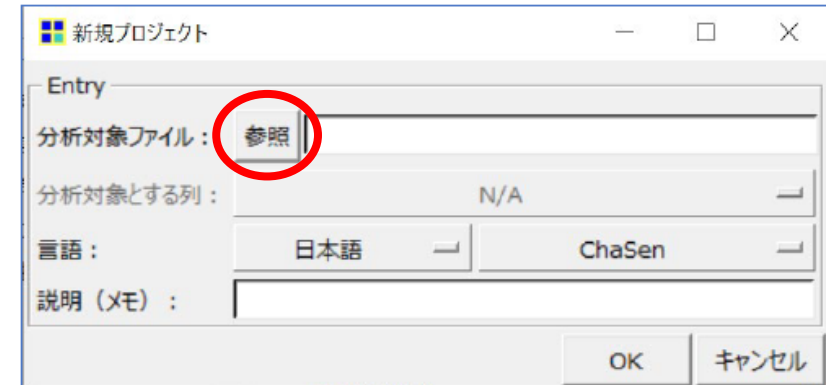
KH CODERの起動とファイル読み込み



①アプリケーション
アイコンをダブルクリック

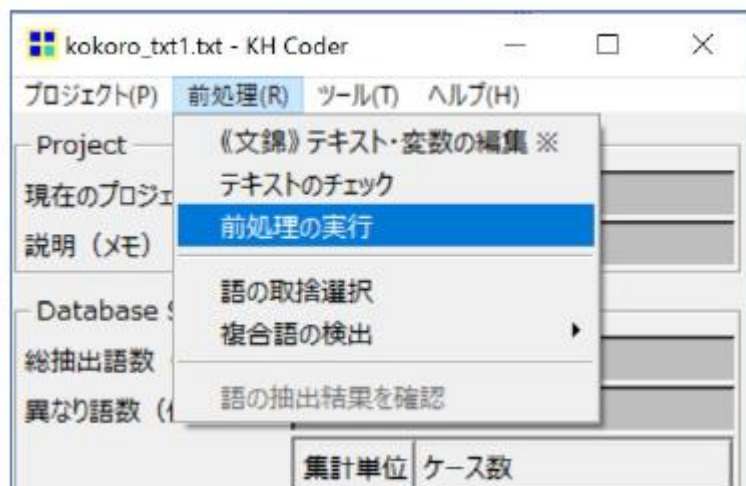


②新規プロジェクト（日本語）
を作成する



③新規プロジェクトに参照をクリックして
“kokoro.xls”を読み込む

テキストデータの前処理



④テキストの前処理を実行する
⇒時間かかると言われるが無視する！

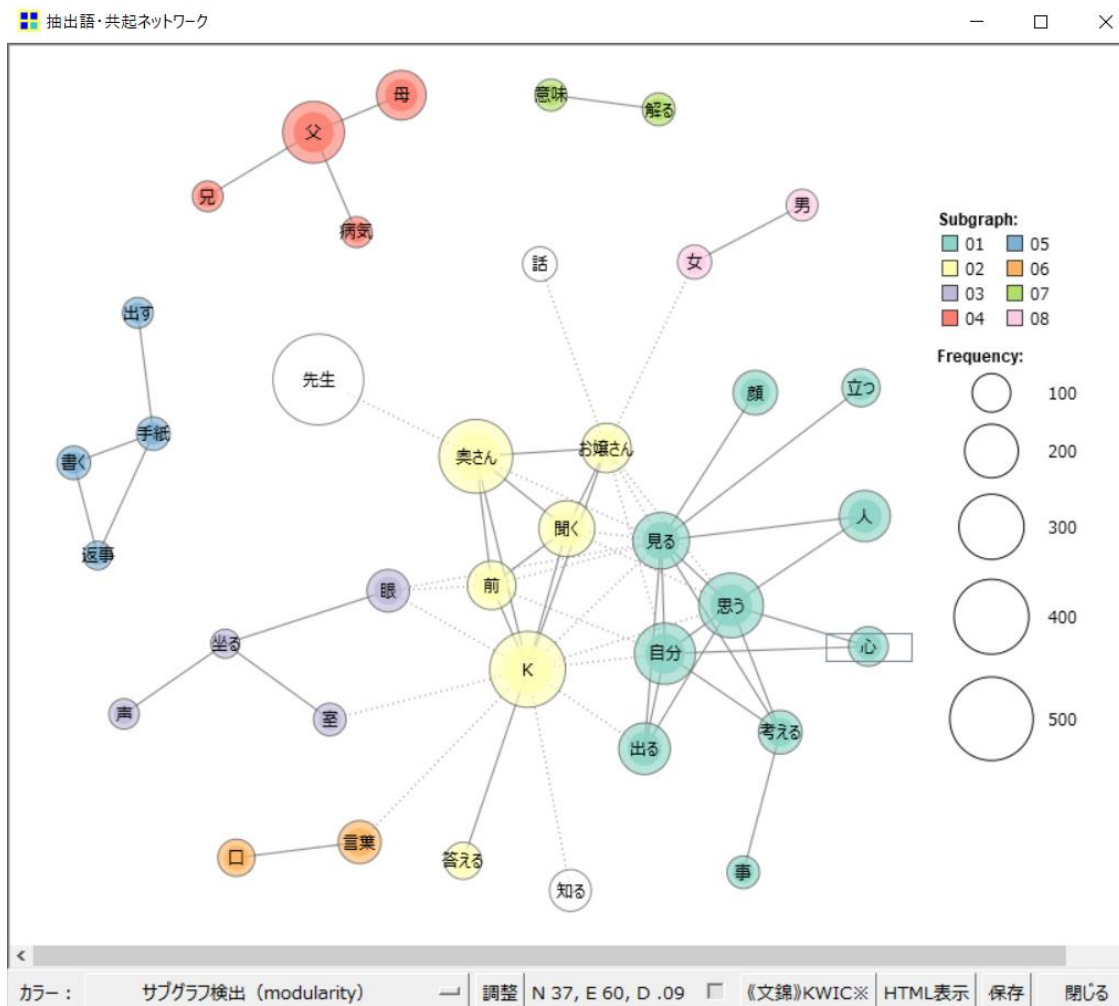
⑤抽出後の前処理が完了したら
“ツール”から処理結果をみる



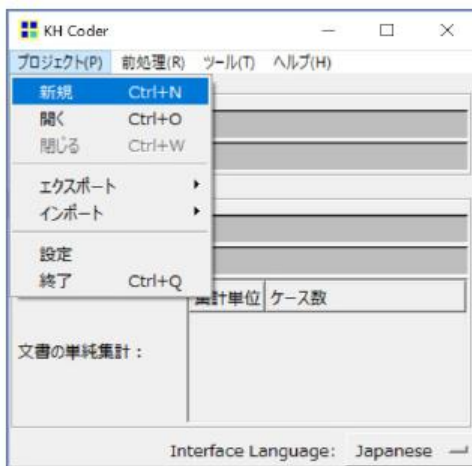
⑤-1 抽出後リスト

⑤-2 共起ネットワーク

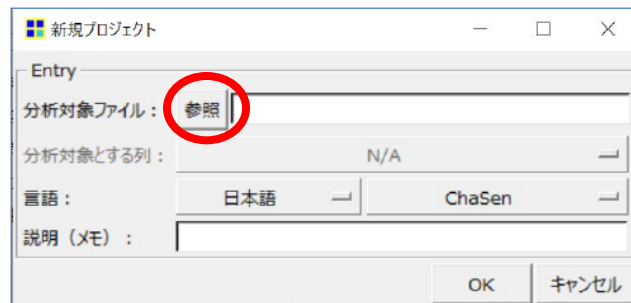
テキストデータの結果表示



“新型コロナ専門家会議提言”を読込んで、以下の結果を確認する！



新規プロジェクトを
作成する



参照をクリックして“新型コロナ
専門家会議提言”を読み込む



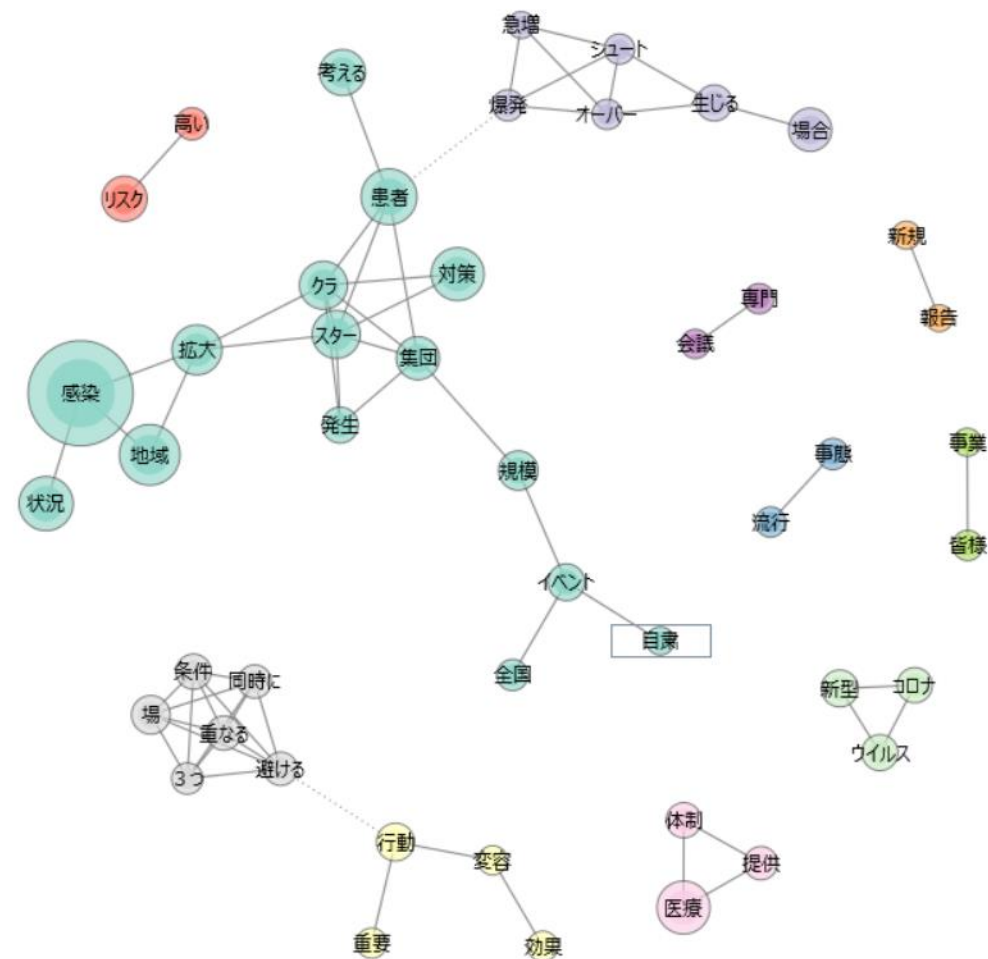
抽出後の**前処理**が完了したら
“ツール”から処理結果をみる

> 抽出後リスト

> 共起ネットワーク

“新型コロナ専門家会議提言”を読込んで、以下の結果を確認する！

List				
#	抽出語	品詞/活用	頻度	
1	感染	サ変名詞	178	
2	地域	名詞	57	
3	患者	名詞	48	
4	状況	名詞	45	
5	必要	形容動詞	44	
6	医療	名詞	43	
7	対策	サ変名詞	43	
8	拡大	サ変名詞	38	
9	クラ	固有名詞	35	
10	スター	名詞	35	
11	人	名詞C	34	
田 12	考える	動詞	32	
13	リスク	名詞	31	
14	集団	名詞	28	
15	場合	副詞可能	28	
田 16	行う	動詞	26	



抽出語がなんかおかしくない？

List				
#	抽出語	品詞/活用	頻度	
1	感染	サ変名詞	178	
2	地域	名詞	57	
3	患者	名詞	48	
4	状況	名詞	45	
5	必要	形容動詞	44	
6	医療	名詞	43	
7	対策	サ変名詞	43	
8	拡大	サ変名詞	38	
9	クラ	固有名詞	35	
10	スター	名詞	35	
11	人	名詞C	34	
田 12	考える	動詞	32	
13	リスク	名詞	31	
14	集団	名詞	28	
15	場合	副詞可能	28	
田 16	行う	動詞	26	

「クラ」と「スター」は、本文脈では“**クラスター**”
として抽出すべき！

その他、「**緊急事態宣言**」は 1 語としたい、
「**新型コロナウイルス**」も 1 語としたい など

抽出語リストで検索に「クラ」と入力すると・・・

「**クラ**」が使われている箇所が示され、「**クラ**」の後には「**スター**」が続いてる！

Search Entry

抽出語: 品詞: 活用形: 追加条件

ソート1: ソート2: ソート3: (前後 語を表示) Ready.

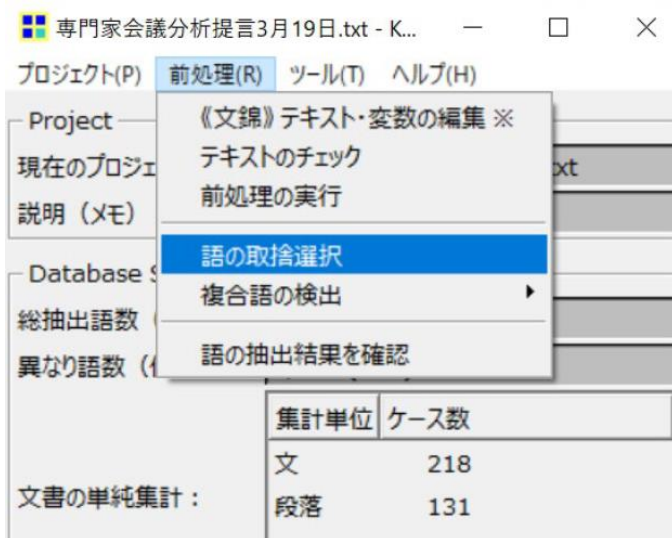
Result

。この見解は、新型コロナウイルス厚生労働省対策本部	クラ スター対策班が分析した内容等に基づき、専門家会議にお	↑
の方針を続けていく必要があると考えています。そのため、「	クラ スター（患者集団）の早期発見・早期対応」、「患者の耳	
新型コロナウイルスへの対応や、国内においても、後述する、	クラ スター（患者集団）の感染源（リンク）が追えない事例が	
に発生していることなどへの対策は依然として必須であり、	クラ スターの早期把握とともに、地域ごとの状況に応じた「市民の	
で収束を図っていけるかが重要になってきています。」2.	クラ スター対策の現状について世界保健機関（WHO）の	
月13日の事務局長のステートメントにおいて、日本が「	クラ スター（患者集団）の早期発見・早期対応」という戦略を	
善を引き起こしたものと考えられますが、日本では少人数の	クラ スター（患者集団）から把握し、この感染症を一定の制御	
ながっていると判断しています。」これまで、厚生労働省の	クラ スター対策班では、感染者、濃厚接触者、保健所、地方	↓

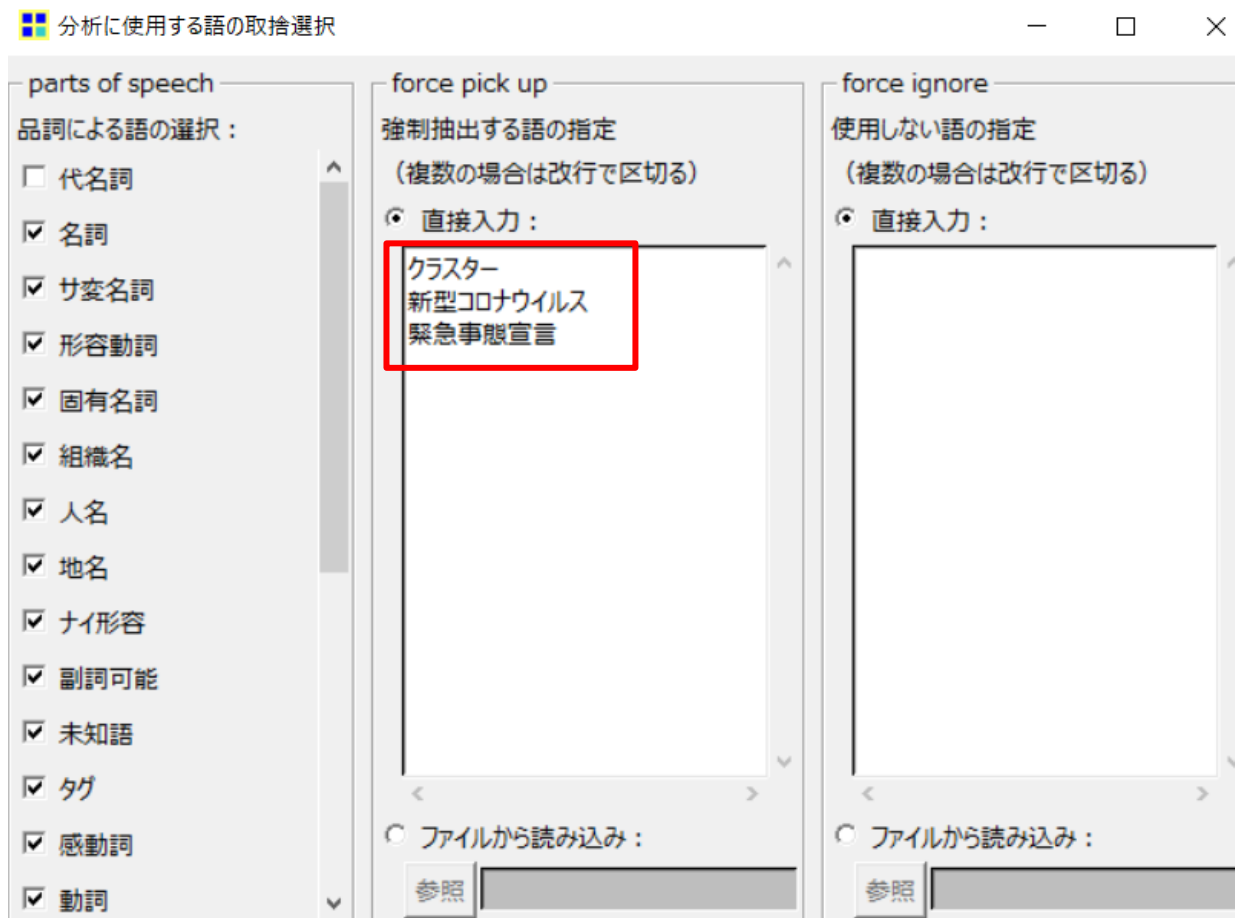
< >

コピー 文書表示 表示単位: 前200 次200 ヒット数: 35, 表示: 1-35

「クラスター」として1語で抽出する



前処理 ⇒「語の取捨選択」の
強制抽出する語の指定に **1語**
で表したい語を入力



強制抽出する語の指定結果

List				
#	抽出語	品詞/活用	頻度	
1	感染	サ変名詞	178	
2	地域	名詞	57	
3	患者	名詞	48	
4	状況	名詞	45	
5	必要	形容動詞	44	
6	医療	名詞	43	
7	対策	サ変名詞	43	
8	拡大	サ変名詞	38	
9	クラスター	タガ	35	
10	人	名詞C	34	
田 11	考える	動詞	32	
12	リスク	名詞	31	
13	集団	名詞	28	
14	場合	副詞可能	28	
田 15	行う	動詞	26	
16	対応	サ変名詞	24	

強制抽出する語を指定したら、
再度前処理を実行する！

⇒ その後、抽出語リストで検証

「クラ」と「スター」ではなく、“**クラスター**”
として抽出された！

「クラスター」として 1 語になっていることの確認

「**クラ**」が使われていた箇所が、「**クラスター**」として認識されている！

Search Entry

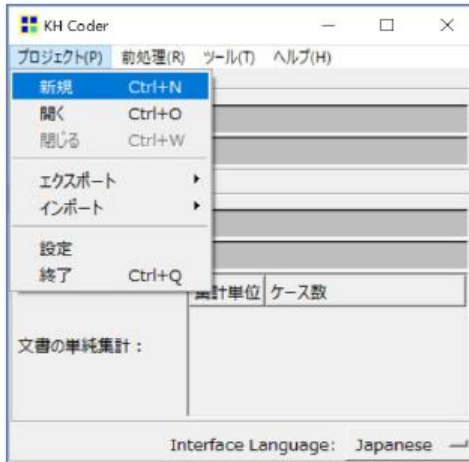
抽出語: 品詞: 活用形:

ソート1: ソート2: ソート3: (前後 24 語を表示) [Read](#)

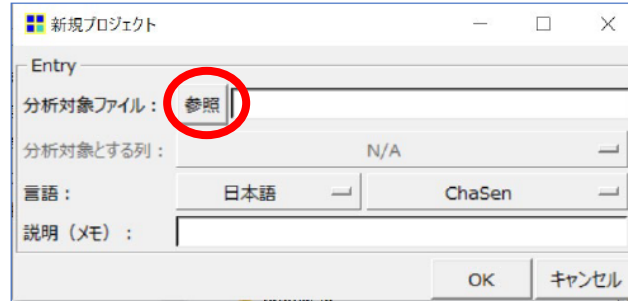
Result

この見解は、新型コロナウイルス厚生労働省対策本部	クラスター	対策班が分析した内容等に基づき、専門家会議におい
方針を続けていく必要があると考えています。そのため、「	クラスター	(患者集団)の早期発見・早期対応」、「患者の早期
コロナウイルスへの対応や、国内においても、後述する、	クラスター	(患者集団)の感染源(リンク)が追えない事例が散
発生していることなどへの対策は依然として必須であり、	クラスター	の早期把握とともに、地域ごとの状況に応じた「市民の行
束を図っていけるかが重要になってきています。⇨ 2.	クラスター	対策の現状について⇨世界保健機関(WHO)のテ
1 3 日の事務局長のステートメントにおいて、日本が「	クラスター	(患者集団)の早期発見・早期対応」という戦略をとっ
引き起こしたものと考えられますが、日本では少人数の	クラスター	(患者集団)から把握し、この感染症を一定の制御下
っていると判断しています。⇨これまで、厚生労働省の	クラスター	対策班では、感染者、濃厚接触者、保健所、地方公

“旅館口コミ”のテキストデータを読み込む



新規プロジェクトを作成する



“参照”をクリックして“ryokan_table”を口コミ指定して読み込む

旅館口コミデータの構成

外部変数

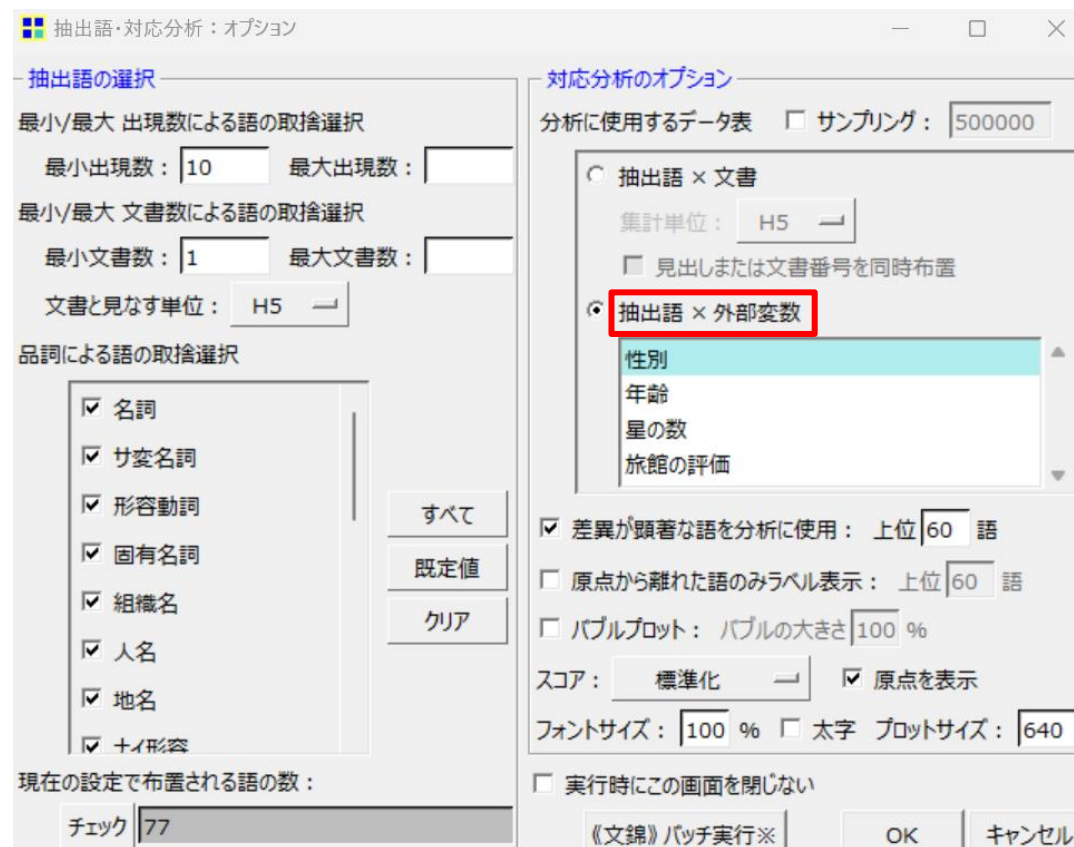
性別	年齢	星の数	旅館の評価	年代-性別	クチコミ
男性	48	★★★★	高評価	中年-男性	女将さんのブログをみて、源泉かけ流しに魅力を感じて予約しました。大きめの湯船に、自然に包まれる開放感、海を一望できる絶景、なんとも贅沢な
女性	24	★★	低評価	若年-女性	泉質もよく開放感たっぷりの露天風呂ですが、囲いはいまいちで裸が見られ
女性	65	★★★	中評価	高齢-女性	2年ぶりに利用しました。露天風呂は開放感があるのがいいですが、角度によっては通行人に見られてるのではない
男性	69	★★★★★	高評価	高齢-男性	朝食は、地元食材中心のヘルシーメニューが用意してあったので、安心していただきました。スタッフの方々はきめ細やかな心遣いとサービス精神に
男性	72	★★★★★	高評価	高齢-男性	当日の予約にも関わらず、スタッフの方に丁寧に接客して頂き、お部屋もきれいに掃除されていて良かったです。
男性	21	★★★	中評価	若年-男性	露天風呂は周囲の自然と相まって開放感抜群ですが、恥ずかしくて行く気に

※ 口コミごとに5つの属性(外部変数)が付記される

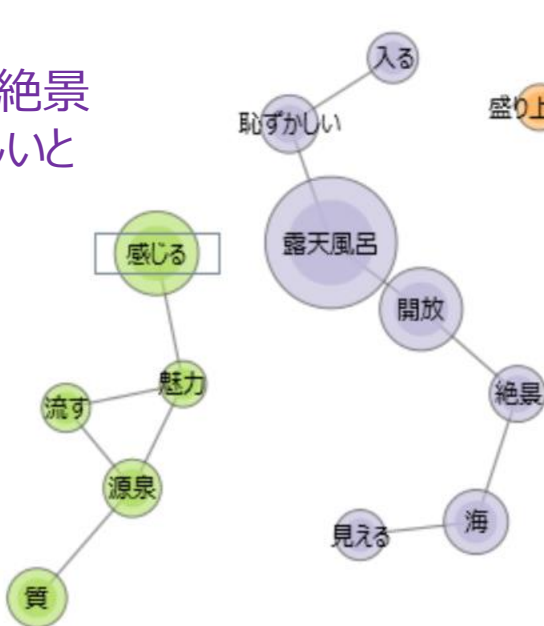
口コミに対する対応分析



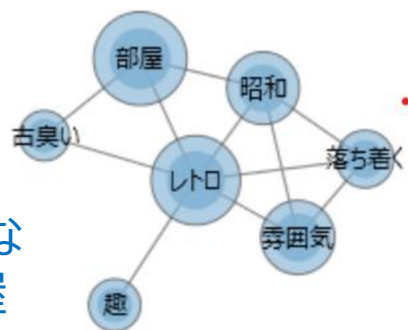
抽出後の前処理が完了したら
“ツール”から**対応分析**を選択



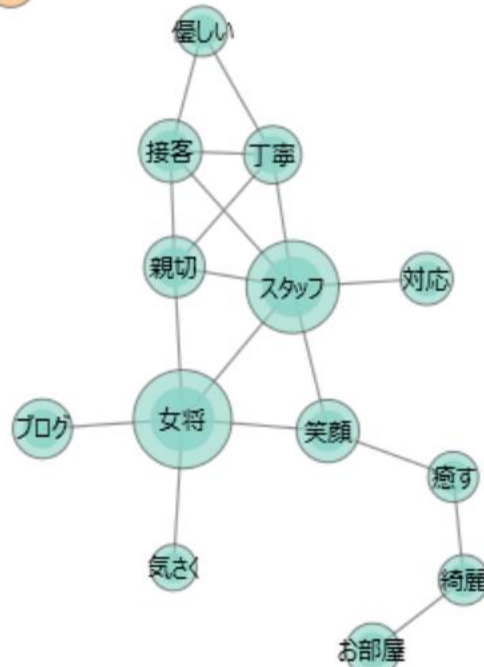
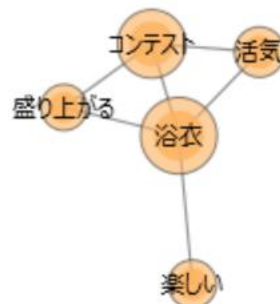
露天風呂は海が見える絶景
で解放的だが、恥ずかしいと
感じる顧客もいる



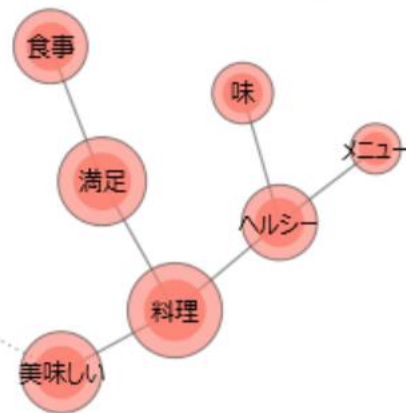
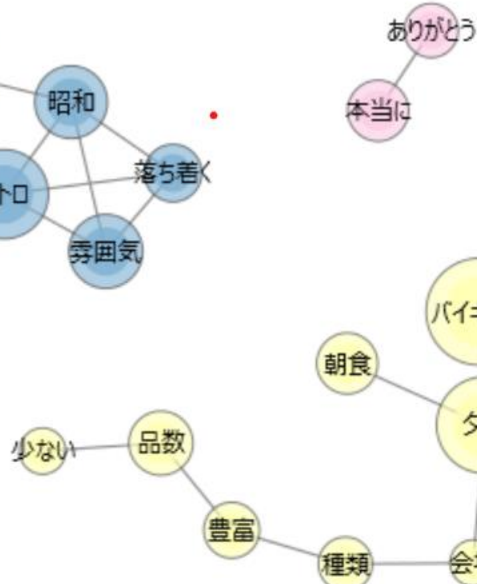
部屋は昭和時代のレトロな
雰囲気（趣ある）の部屋



浴衣コンテストもあったようで活気がある

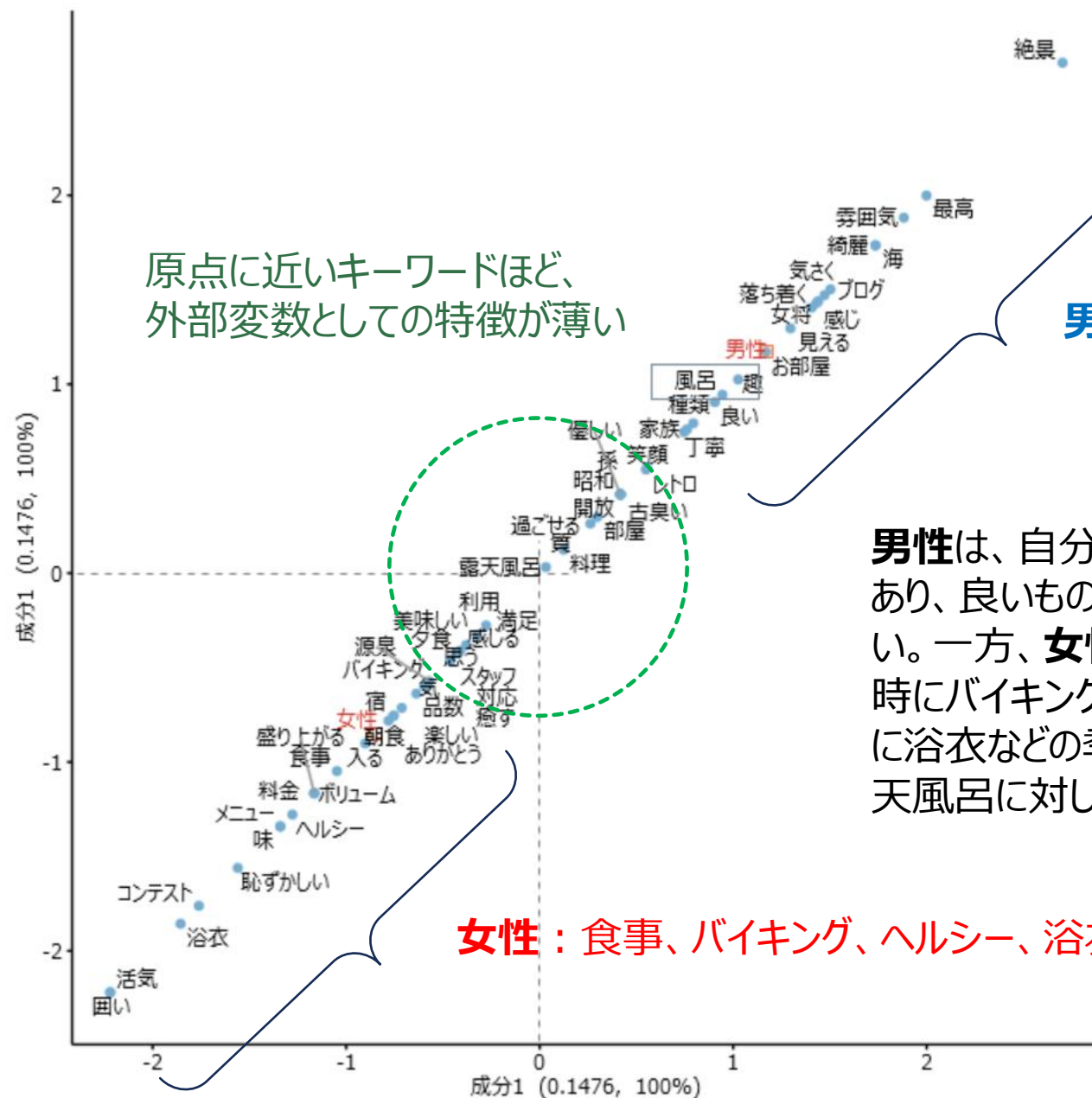


宿の女将やスタッフの接客が
に対するコメント
→ワードから好印象が分かる

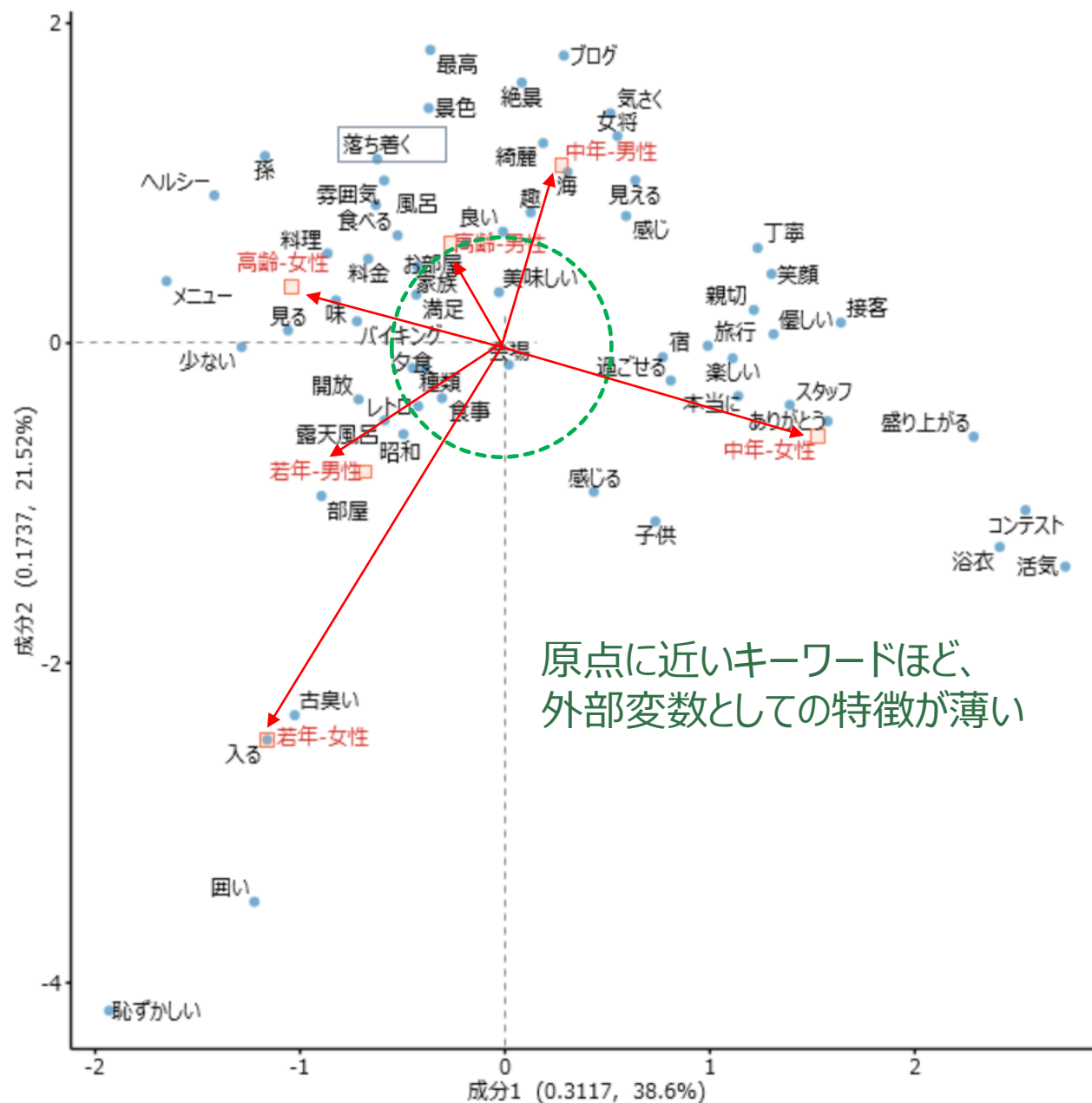


料理が美味しくて満足である
ことが分かる（特に、ヘルシー
であったようだ！）

食事形式はバイキングのようだ！



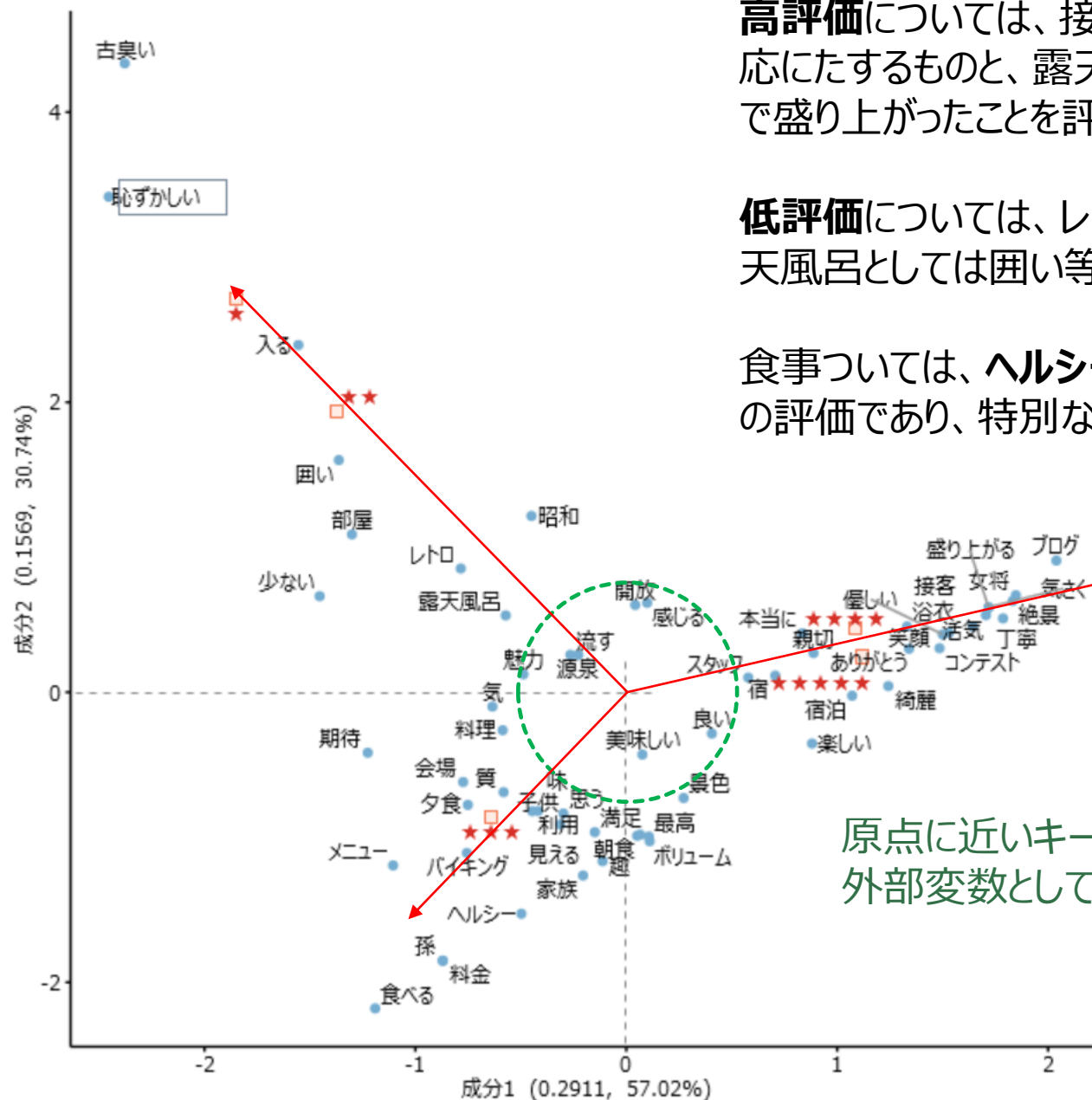
男性は、自分自身が楽しめてくつろげる宿が重要であり、良いものなら料金については大きな要因ではない。一方、**女性**は料金は特徴的な要因であり、同時にバイキングなどの食事は重要な要素で、その他に浴衣などの季節的な体験が注目される。但し、露天風呂に対しては恥ずかしいとの思いがある。



中年男性は、前頁の男性を象徴するようなキーワードが並んでおり、海や絶景といった自分自身が楽しめてくつろげることが宿に結びつく

料金や料理が重要な要因となるのは**高齢女性**であり、**中年女性**は楽しくて盛り上がる活気あることが宿選びには宿が結びつく

若年層は、男女ともにレトロで、古臭く昭和的なイメージが宿には連想される。女性は開放的は絶景露天風呂に対しては恥ずかしいとの思いが強い

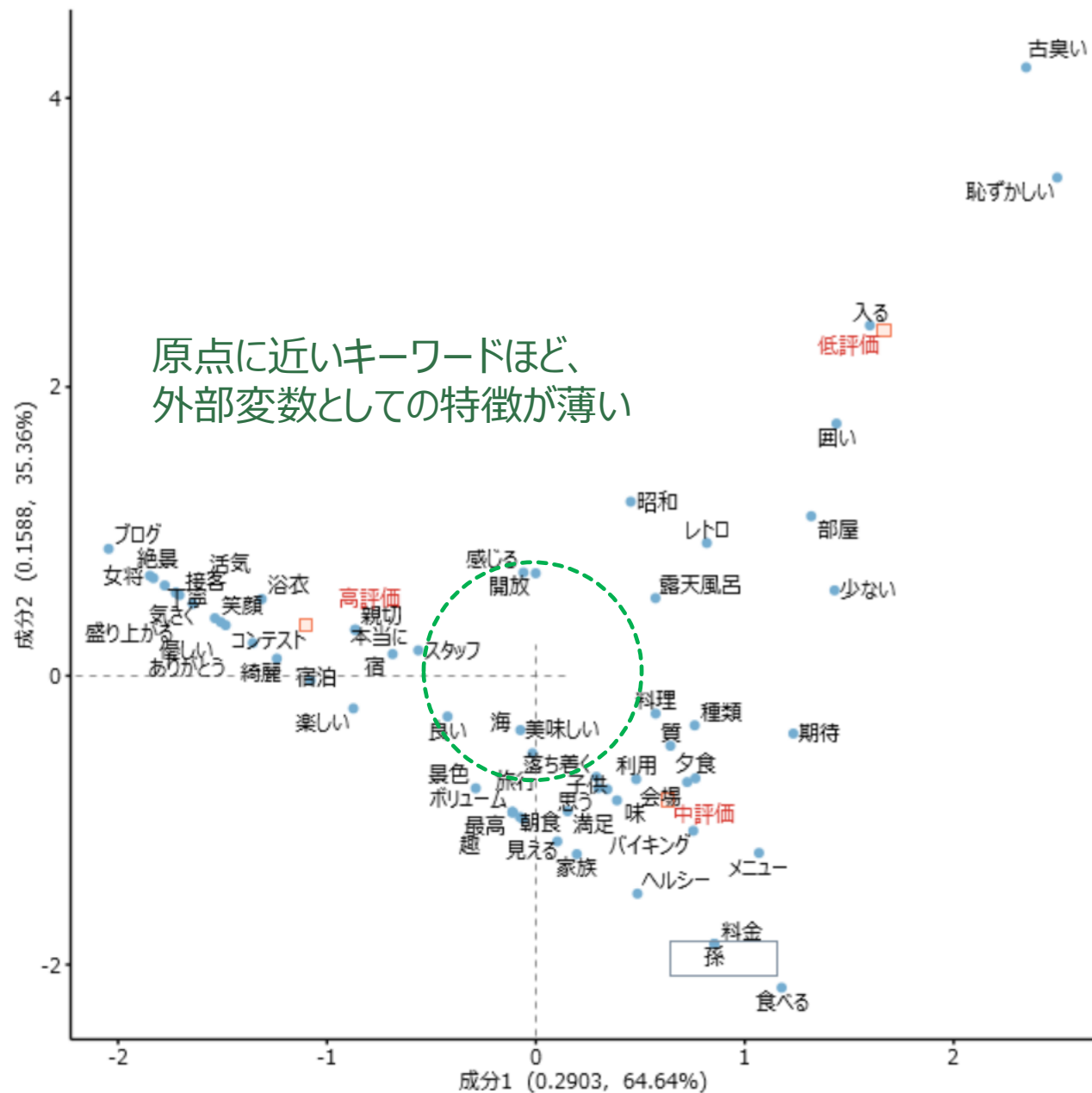


高評価については、接客、優しい、新設、楽しい、気さくといった対応にたするものと、露天風呂の絶景で、浴衣コンテストなどのイベントで盛り上がったことを評価している。

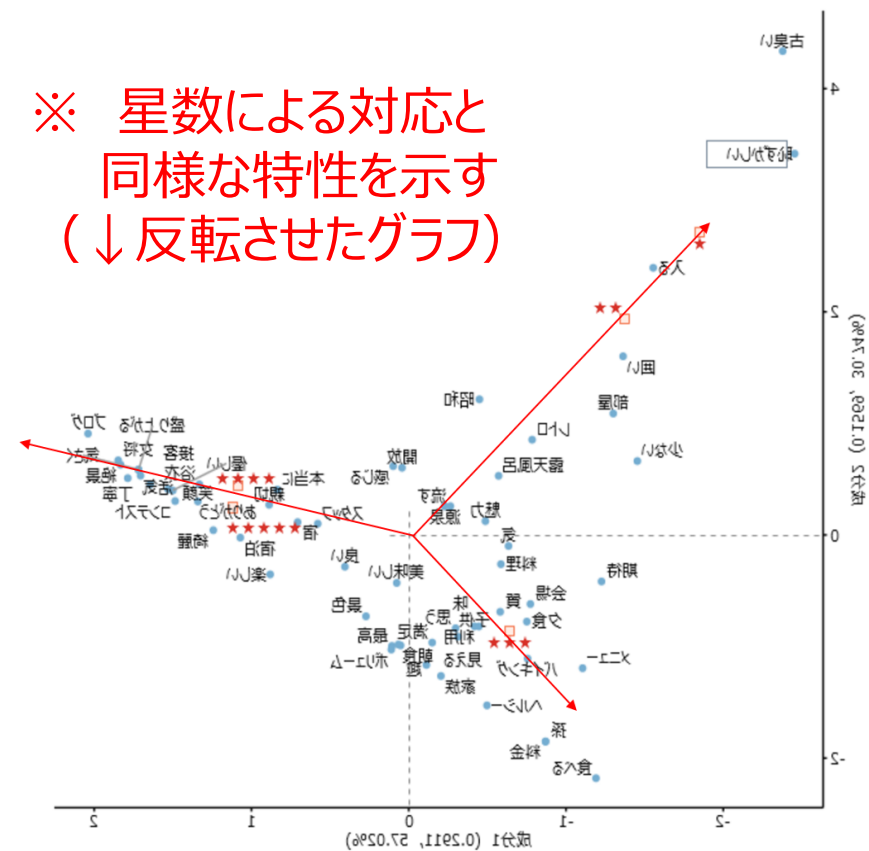
低評価については、レトロな昭和時代の部屋を挙げることも多く、露天風呂としては囲い等の点で評価が低いと思われる。

食事については、**ヘルシーな食事**で**バイキング形式**であるが、中程度の評価であり、特別な評価レベルではない

原点に近いキーワードほど、
外部変数としての特徴が薄い



※ 星数による対応と
同様な特性を示す
(↓反転させたグラフ)



課題：商品レビューとレーティング

“Rating_Review.xlsx”ファイルの商品レビューとそれに対する星の数(レーティング)に関するテキストデータを読み込み、共起ネットワークおよび対応分析により、この商品に対して、どのようなことが言えるかをレポートしなさい。

星数/評価

Rating
5つ星のうち5.0
5つ星のうち4.0
5つ星のうち5.0
5つ星のうち3.0
5つ星のうち5.0
5つ星のうち5.0
5つ星のうち1.0
5つ星のうち5.0
5つ星のうち4.0
5つ星のうち4.0
5つ星のうち5.0
5つ星のうち4.0
5つ星のうち3.0
5つ星のうち4.0
5つ星のうち1.0
5つ星のうち5.0
5つ星のうち3.0
5つ星のうち5.0
5つ星のうち1.0
5つ星のうち4.0
5つ星のうち2.0
5つ星のうち4.0
5つ星のうち1.0

クチコミ/レビュー

Review Content
言われていたよりも早く届きました。フリースは薄いですが、コタツはしっかりしていて足が折れるなどのレビューが多かったので、すぐチェックしてみましたが、特に問題ない1人で使うサイズとしてはちょうどいいと思いますヒーターもかなり暖かいです
こたつユニットのメーカーが日本国内は1社だけなので、国産ユニットのコタツはワット子供部屋に購入。小さいけどまあまあ大きさです。パーツも全部しっかりしてて問題立ち組みは簡単です、想像より大きい、とても暖かいです！買ったよかった！
コタツが届きました。!?こたつ布団が無いセットなのだから、忘れ無いで欲しい。子供が最高です。無地と迷いましたがこちらで良かったです。かわいいのとサイズも一人暮らし冬場は床に直接座ってパソコン作業や食事をしたくなり購入。コンパクトなサイズや軽何もかもチャッチイですね皆さんが言われる通り天板は弱いしフリースは薄くてとてもコメントで天板や布団が薄いとの情報があった為、正直買う時に他のものと迷いましたこの値段で、これなら満足です。掛け布団は薄いので、薄手のニトリ毛布、ブランケット思ってた以上に布団がぺらぺらで使いものにならない。こたつじたいはおしゃれな感じ温度スイッチを入れたて最大にすると直ぐに暖まります買って良かったです
商品の到着は早かったのでありがたかったのですが、スイッチONにしてしばらくしてもコンパクトで掃除もしやすく見た目もかわいいです。お手頃価格で買って良かったです組み立て簡単。女性1人でも出来ると思います。10%OFFで8099円で買いました。まあ一人で使うにはちょうど良い大きさ。軽いので掃除のときの移動が簡単です。
確かに掛け布団の布地はフリースみたいです。あとヒーター強にしてもぬるくて寒いです他の方の情報通りで布団は薄くて頼りありませんが、暖かさは大丈夫でした。もっと貧電気を付けると変な匂いがします。付属の掛け布団は使えない
旦那さんがこたつが欲しいとのことでレビューの良かったこちらを購入しました。丸か四こたつ布団がぺらぺらなのは想像通りでした思ったより暖かくなりませんがもう一



データマイニングを楽しもう！