

テキストマイニングツール「ユーザーローカル AI テキストマイニング」を使う

KH Coder が有料ツールになったため、今後は「ユーザーローカル AI テキストマイニング」を使って分析する。

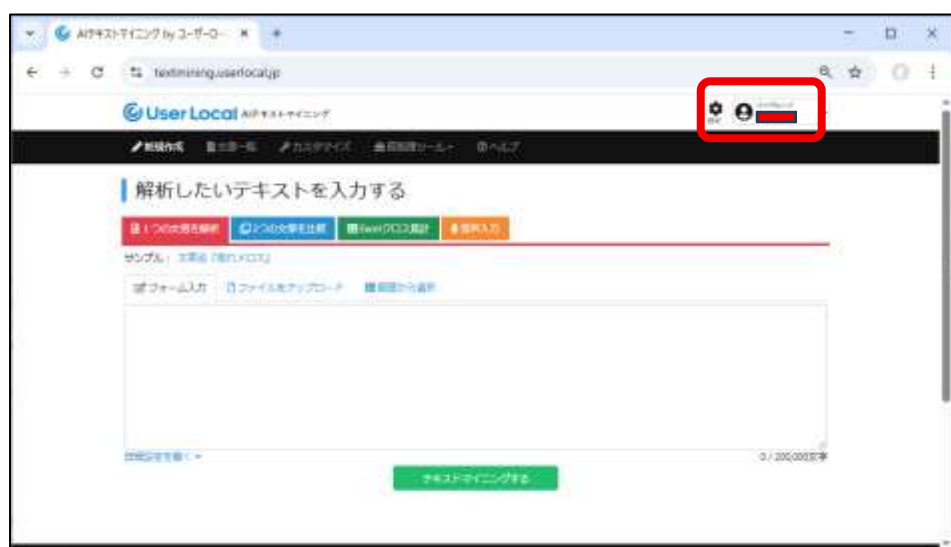
今回は、「さらば黄昏」(長塚圭史著)のテキストデータを題材にして、分析する。

1. User Local AI テキストマイニングに、用意したファイルを読み込ませる。

戯曲の全文章を打ち込んだテキストファイルを User Local AI テキストマイニングに読み込ませる。

User Local(<https://textmining.userlocal.jp/>)

自分のアカウントを作って、ログインしておくで大量の文字が扱えるようになる。



ファイルをドラッグする



詳細設定を開くを押す



辞書の作成・編集はこちら を押す



地名・人名などが複数の単語として分かれずに登録する
※そのまま分析すると踊田は踊と田に分かれる

カスタム辞書を作成する

固有名詞や除外語を設定して、テキストマイニングをカスタマイズできます。

基本設定

辞書名

さらば黄昏

単語設定

固有名詞設定

設定した単語を1つの言葉として認識させます。これにより「株式会社ユーザーローカル」などの固有名詞が分割されずに解析できます。なお、各単語は改行区切りで入力してください。

踊田
竹井大
本町一栄
阪本慎平
阪本雪乃
阪本朝陽
剣持愛一郎
大河原
島村遼子
島村楓

14 / 30語

設定するボタンをクリックする

除外語設定

設定した単語を解析結果から除外します（=ストップワード）。これは、出現頻度が高い割には役に立たない単語（「ない」「思う」など）を除くために行います。なお、各単語は改行区切りで入力してください。

単語1
単語2
単語3

0 / 30語

同義語設定

解析時に同じ意味を持つ単語を1つにまとめます（例：同義語「パソコン」「PC」→代表語「パーソナルコンピューター」）。

同義語

同義語

+ 同義語を追加

代表語

+ 同義語設定を追加

同義語: 0 / 30語

設定する

解析したいテキストを入力する

1つの文書を解析 2つの文章を比較 Excelクロス集計 音声入力

サンプル: 太宰治『走れメロス』

フォーム入力 ファイルをアップロード 履歴から選択

以下の形式がアップロードできます。最大200,000文字、10MBまで。

テキスト形式(.txt)、CSV形式(.csv)、Word形式(.docx)、Excel形式(.xlsx .xlsm)、PowerPoint Text形式(.pptx)、PDF形式(.pdf)

解析対象に名前を設定（任意）

「さらば黄昏」全6場（分析用）.txt

解析するファイル



辞書設定（辞書の作成・編集はこちら）

さらば黄昏

公開設定

☐ 外部に公開

URLを知っている全員が閲覧可能

☒ 公開しない

自分のみ閲覧可能

[詳細設定を開じる](#)

テキストマイニングする

テキストマイニングするボタンを押すと結果が表示される

ワードクラウド

スコアが高い単語を複数選び出し、その値に応じた大きさで図示しています。単語の色は品詞の種類で異なっており、青色が名詞、赤色が動詞、緑色が形容詞、灰色が感動詞を表しています。[【このテキストマイニング結果を利用するには】](#)



※以降、分析結果は

※ユーザーローカル AI テキストマイニングによる分析(<https://textmining.userlocal.jp/>)

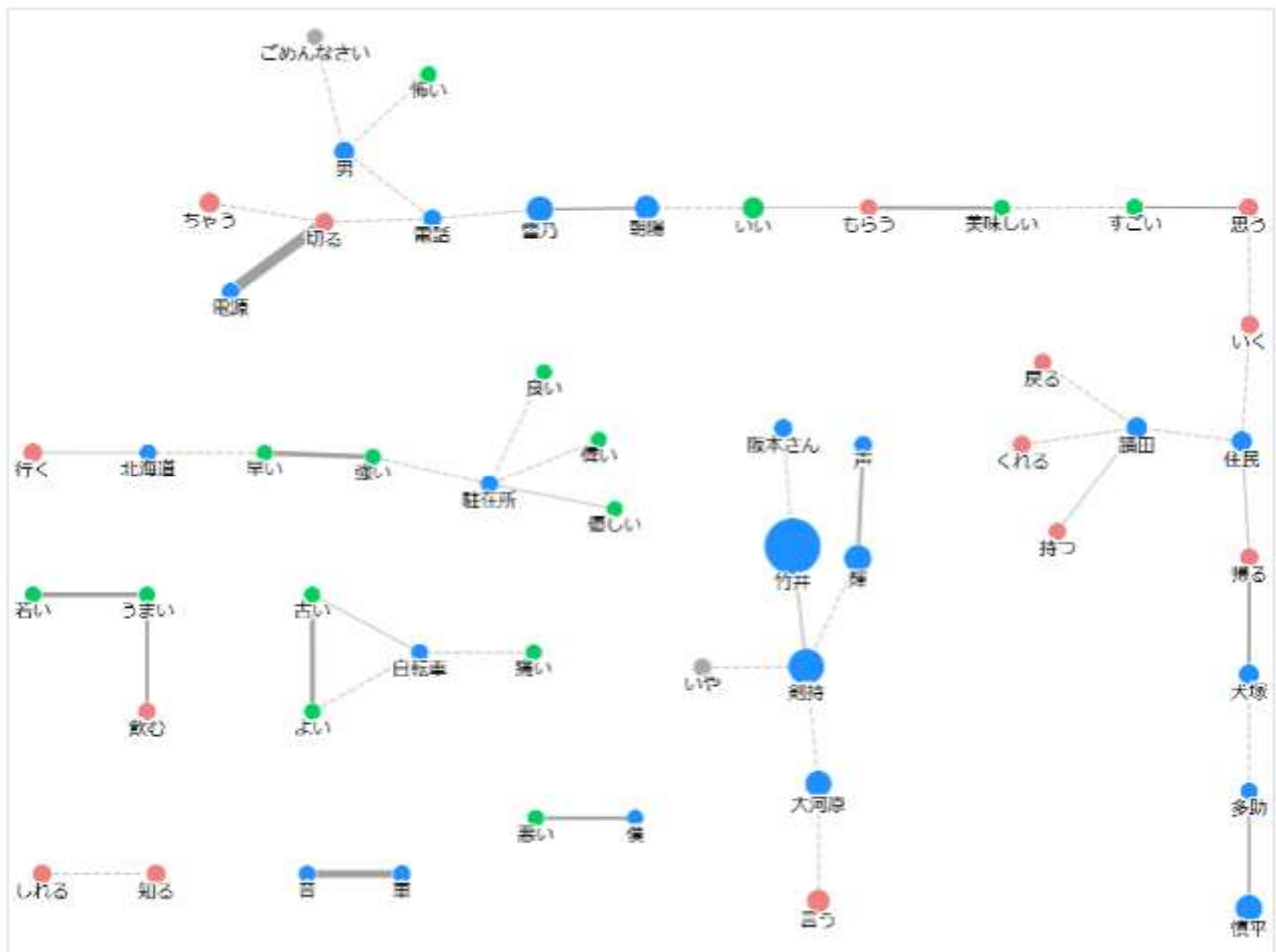
単語出現頻度

文章中に出現する単語の頻出度を表にしています。単語ごとに表示されている「スコア」の大きさは、与えられた文書の中でその単語がどれだけ特徴的であるかを表しています。通常はその単語の出現回数が多いほどスコアが高くなりますが、「言う」や「思う」など、どの文書にもよく現れる単語についてはスコアが低めになります。【[スコアの目的と算出方法について](#)】

■ 名詞	スコア	出現頻度	■ 動詞	スコア	出現頻度
竹井	5979.44	488	言う	3.94	69
剣持	2141.86	241	わかる	6.54	59
輝	707.12	136	ちゃう	3.28	43
慎平	1954.99	133	知る	2.87	34
雪乃	1085.93	131	行く	0.83	33
大河原	1264.73	126	思う	0.58	32
潤子	1340.14	118	しれる	5.18	31
朝陽	898.66	114	切る	10.13	30
一栄	1618.54	112	考える	2.20	28
踊田	828.06	61	いく	1.40	27
住民	218.22	56	帰る	1.60	25
男	19.51	52	聞く	1.40	24
犬塚	441.57	50	くださる	1.05	23
阪本さん	299.02	30	くれる	0.63	23
電話	11.13	30	戻る	3.77	21

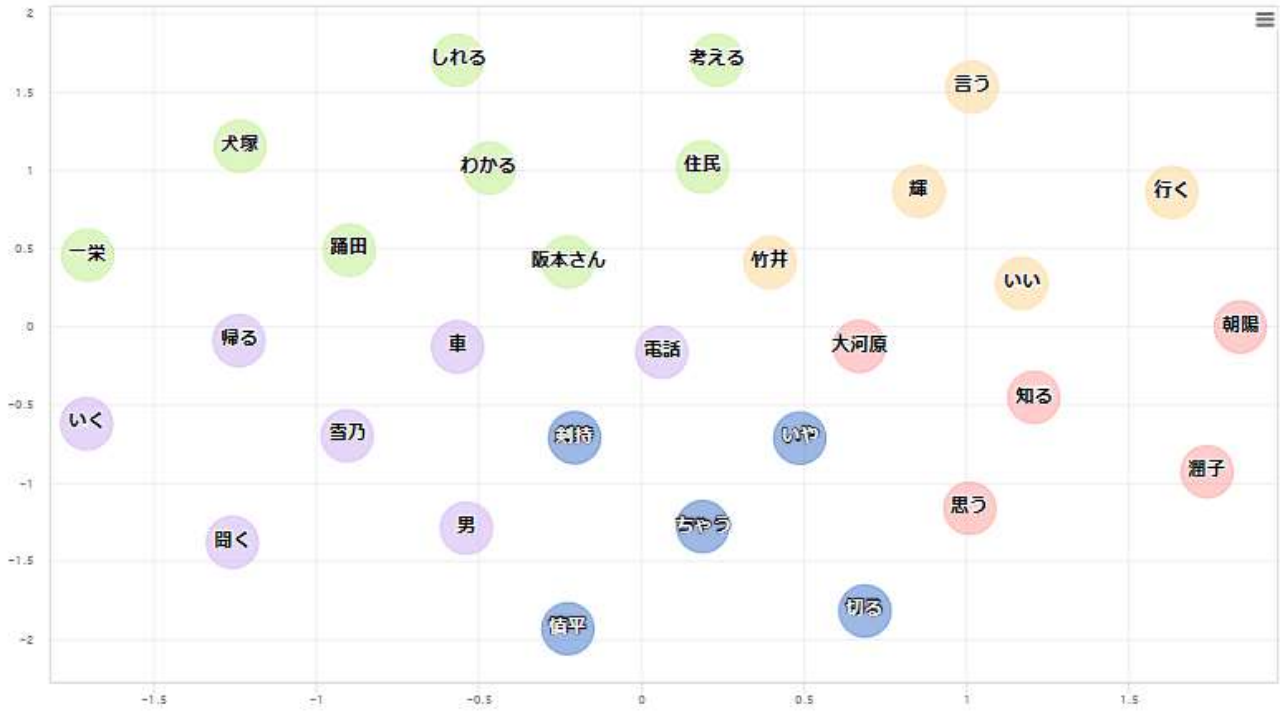
■ 形容詞	スコア	出現頻度	■ 感動詞	スコア	出現頻度
いい	2.27	55	いや	6.70	26
すごい	0.38	13	ごめんなさい	1.20	9
美味しい	0.36	9	ありがとう	0.09	8
悪い	0.30	9	すいません	1.80	7
怖い	0.25	7	おい	0.74	5
よい	0.10	7	じゃあ	0.90	2
若い	0.75	6	いいえ	0.72	2
痛い	0.12	5	うお	0.36	2
強い	0.11	5	そうですね	0.31	2
早い	0.07	5	なるほど	0.05	2
うまい	0.16	4	お疲れ様	0.02	2
良い	0.02	4	ごめんください	4.14	1
古い	0.43	3	はいはい	0.40	1
赤い	0.38	3	さらば	0.29	1
偉い	0.33	3	いえいえ	0.12	1

文章中出现する単語の出現パターンが似たものを線で結んだ図です。出現数が多い語ほど大きく、また共起の程度は強い方から順に 太い実線 > 細い実線 > 破線 で描画されます。【共起とは？】



2次元マップ

文章中での出現傾向が似た単語ほど近く、似ていない単語ほど遠く配置されています。距離が近い単語はグループにまとめ、色分けしています。[【2次元マップの詳しい説明】](#)



係り受け解析

係り受け解析では、「名詞」に係る「形容詞」「動詞」「名詞」についての解析結果を表示します。「スコア」は、出現回数やその係り受け関係が全組み合わせのうちに占める割合などを複合的に判断し、独自に算出した数値です。「スコア」が高いほど、よりその係り受け関係が重要であることを示します。また、単語の後に「(否: 50%)」などとなっている場合、集計された係り受け関係のうち 50%が否定表現(例:「高い」→「高くない」)として使われていることを意味しています。ネガポジは名詞にかかる形容詞がポジティブ(ネガティブ)な単語かどうかを表しています。[【係り受け解析とは?】](#)
(結果の表示に時間がかかる場合があります)

■ 名詞 - ■ 形容詞

名詞 - 形容詞		ネガポジ	スコア	出現頻度
竹井 - いい	Q	ネガティブ	0.36	4
輝 - いい	Q	ネガティブ	0.21	3
雪乃 - 怖い	Q	ネガティブ	0.75	2
僕 - 悪い	Q	ネガティブ	0.60	2
一栄 - すごい	Q	ポジティブ	0.43	2
大河原 - いい	Q	ネガティブ	0.11	2
パトロール - いい	Q	ネガティブ	0.11	2
男 - おぞましい	Q	ネガティブ	1.00	1
頭 - ずるい	Q	ネガティブ	1.00	1
剣持 - 遠い	Q	ネガティブ	1.00	1
ママ - 悲しい	Q	ネガティブ	1.00	1
輝 - 近い	Q	ネガティブ	1.00	1
道内 - あったかい	Q	中立	1.00	1
雪 - 少ない	Q	中立	1.00	1
一栄 - 可愛い	Q	ポジティブ	1.00	1

■ 名詞 - ■ 動詞

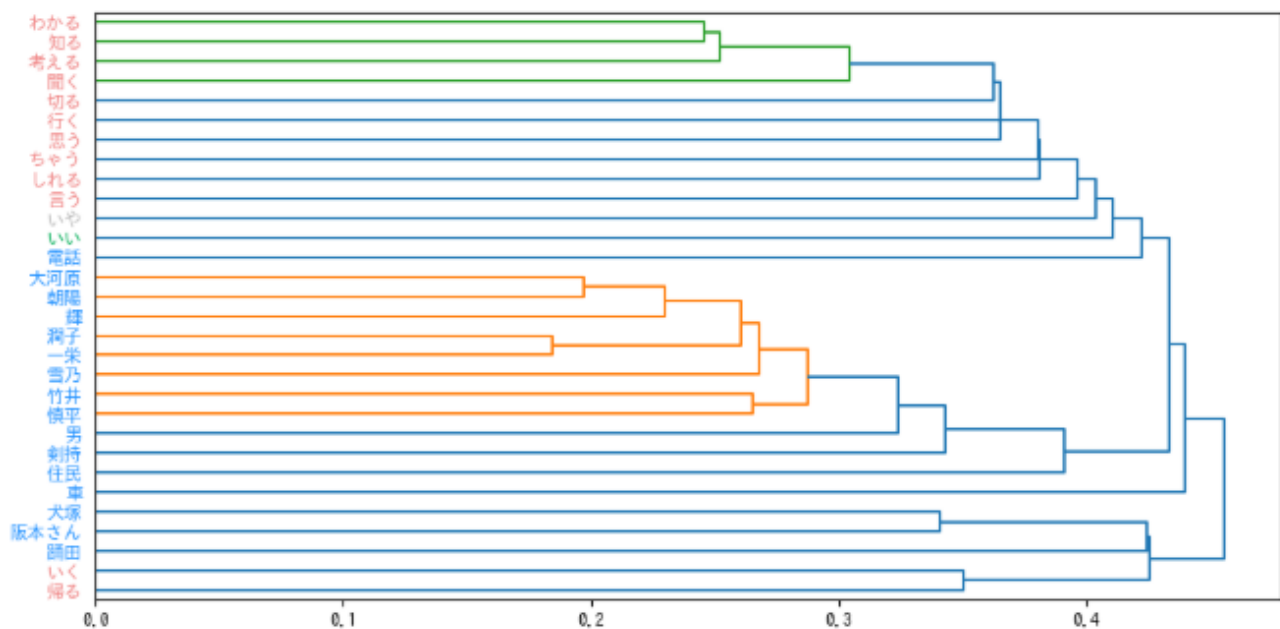
名詞 - 動詞		スコア	出現頻度
電源 - 切る (否: 22.22%)	Q	2.90	9 (否: 2)
竹井 - わかる (否: 42.86%)	Q	0.93	7 (否: 3)
剣持 - 答える	Q	4.20	6
質問 - 答える	Q	4.20	6
竹井 - 行く (否: 16.67%)	Q	1.24	6 (否: 1)
大河原 - 言う	Q	0.60	6
竹井 - 言う (否: 16.67%)	Q	0.60	6 (否: 1)
剣持 - くださる	Q	1.25	5
竹井 - 見る	Q	1.03	5
竹井 - しれる (否: 100.00%)	Q	0.94	5 (否: 5)
剣持 - 行く	Q	0.88	5
輝 - 来る (否: 60.00%)	Q	0.70	5 (否: 3)
竹井 - 出る (否: 20.00%)	Q	0.56	5 (否: 1)
瀬子 - 出る	Q	0.56	5
声 - かける	Q	1.67	4

■ 名詞 - ■ 名詞

名詞 - 名詞	スコア	出現頻度
剣持 - 大丈夫	2.55	7
雪乃 - 朝陽	0.26	5
剣持 - 輝	0.22	5
竹井 - 北海道	1.33	4
輝 - 大丈夫	0.91	4
竹井 - 剣持	0.08	4
剣持 - 仕事	1.20	3
踊田 - 全員	1.00	3
雪乃 - 連絡	1.00	3
竹井 - 連絡	1.00	3
車 - 音	0.75	3
一栄 - 大丈夫	0.55	3
輝 - 声	0.52	3
雪乃 - 電話	0.39	3
竹井 - 電話	0.39	3

階層的クラスタリング

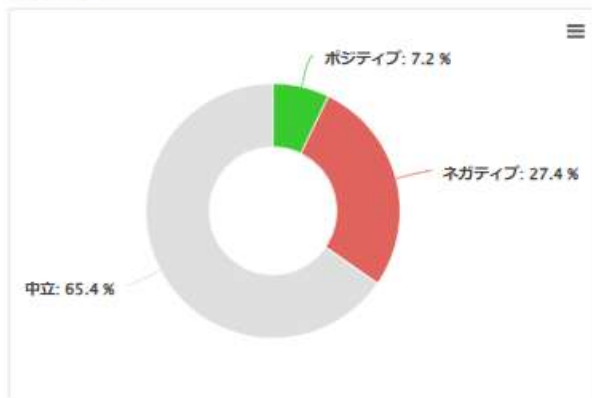
文章中での出現傾向が似た単語をまとまりとしてとらえられるよう樹形図で表したものです。グループは色分けして表示しています。【[階層的クラスタリングの詳しい説明](#)】



サマリー

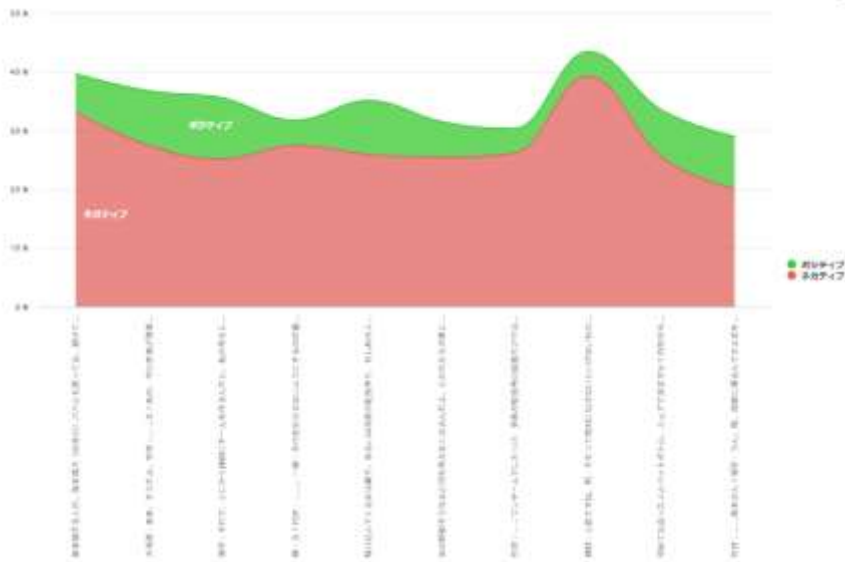
文書全体を分析し、感情の傾向を可視化しています。「ポジネガ」は、文章に含まれるポジティブな感情の文とネガティブな感情の文の存在比を示しています。「感情」は、文章に含まれる各感情の度合いを数値に換算しています。なお、各感情の数値は、全ての感情の平均値を 50%とした偏差値です。

ポジネガ



ポジネガ推移

文書を分割して分析し、ポジティブな感情とネガティブな感情の比率の推移を時系列的に可視化しています。グラフの左端が文書の先頭を表しており、右に進むほど終わりに近づきます。



AI 文章要約

ChatGPT のアルゴリズムと連携し、高精度な要約を生成します。15～60 秒かかることがあります。
※入力されたデータが、AI 学習に利用されることはありません

この作品「さらば黄昏」は、登場人物たちの人間関係や地域社会の問題を描いたドラマです。物語は、ひき逃げ事件をきっかけに、地域住民が集まり、事件の真相やその影響について議論を交わす様子から始まります。1. ****事件の発端****: 竹井大という駐在員が、ひき逃げに遭った阪本朝陽とその母親雪乃から話を聞く場面が描かれ、朝陽が自転車で避けた際に車に巻き込まれたことが語られます。「朝陽はライトつけてますよ。つけてたでしょ？」というやり取りから、事件の詳細が徐々に明らかになります。2. ****地域の反応****: 事件を受けて、地域住民が集まり、竹井がパトロール隊の結成を提案します。「この踊田地区を、このまま過ごしやすい穏やかな集落にしたいじゃないですか」と、地域の安全を守るための行動が求められます。3. ****住民の対立****: 住民たちの間には意見の対立が生じ、「暴力反対！」と叫ぶ声上がる中で、竹井は冷静に地域の分断を防ぐための提案を続けます。「このまま何もせず事態を放置しておけば、やがてその時が来た時、ビクビク怯えて…」と、地域の未来を憂慮する姿が描かれます。4. ****個人の葛藤****: 朝陽は自らの責任を感じ、「僕が悪いんです」と発言し、事件の影響を受けたことを示します。母親の雪乃は感情的になりながらも、息子を守ろうとする姿が印象的です。5. ****結末と未来への希望****: 最終的に、朝陽や他の住民たちがパトロール隊に参加する意志を示し、地域の安全を守るための行動を起こすことが決まります。「僕／私やります」との言葉が、未来への希望を象徴しています。物語は、個々の葛藤や地域の問題を通じて、コミュニティの絆や協力の重要性を描いています。

(以上、User Local によるテキストマイニング結果)

ChatGPT による要約が的外れだったため、この文章と戯曲全文を Gemini に読ませて、概要を批判させる。

■ Gemini による要約

ご提示いただいた要約文を読むと、確かに「的を外している」と感じられるのも無理はありません。その要約は、物語の表面的な言葉(パトロール、安全、希望)だけを拾ってしまい、この戯曲の核心である**「過去の因縁」「暴力の影」「人々の無責任さと欺瞞」といったドロドロとした緊張感**が完全に抜け落ちてしまっています。

特に「未来への希望を持たせる形で締めくくられる」という部分は、あの静かで不穏なラストシーンを読み解けていない印象です。

私(Gemini)が、全六場を通読した上での「物語の本質」を突いた要約を作成します。

戯曲『さらば黄昏』要約

【概要】 本作は、のどかな田舎町「踊田地区」を舞台に、25 年前の凄惨な放火殺人事件の記憶と、その加害者の帰還という影に怯える人々の姿を描いた心理サスペンスである。正義を信じようとする駐在員と、過去に縛られた住民たちの対比を通じ、平穏なコミュニティの裏側に潜む排他性と無責任さを浮き彫りにする。

1. 繰り返される暴力の予兆 物語は、少年・朝陽が「正体不明の車」に煽られ負傷する事件から幕を開ける。駐在員の竹井は、これがかつて自分が逮捕した凶悪犯・犬塚の親族による「警告」ではないかと直感する。竹井は定年退職と同時に被害者の娘・一栄と北海道へ移住する計画を立てていたが、村に忍び寄る不穏な空気が彼を足止めする。

2. 英雄の孤独と住民の無関心 竹井は村を守るため「自警団(パトロール隊)」の結成を住民に促すが、住民たちは竹井を「かつて発砲した危ない男」と揶揄し、自分たちがリスクを負うことを拒絶する。守ろうとする竹井の「正義」は空回りし、住民たちの「人任せ」な態度との溝は深まっていく。

3. 暴かれる「血」の秘密 竹井の後任予定である剣持と、犬塚の元愛人・潤子の娘である輝の結婚話が進む中、衝撃的な事実が明かされる。輝は「犬塚の娘」である可能性が高く、その血縁関係が村のパワーバランスを揺さぶり始める。深夜、駐在所に現れた謎の男は、圧倒的な「悪」の気配で村の平和が砂上の楼閣であることを突きつける。

4. 崩壊する仮面 パトロール隊の結成集会で、住民たちの不満が爆発する。竹井を「セイロン(異分子)」と呼び排除しようとする醜い集団心理が露呈する中、良心的な市民であったはずの慎平が、自身の過去の罪や輝との繋がりを暗示して逃亡するように姿を消す。竹井は、自分が守ろうとした「平和な村」など最初から存在せず、全員が共犯者として秘密を抱えていたことに気づかされる。

5. 結末:沈黙と停滞 数年後。竹井と一栄は北海道へ去り、大河原など古い世代は世を去った。村には静かな日常が戻ったように見えるが、それは「解決」ではなく、単なる「忘却と沈黙」による安定であった。残された慎平と一栄が交わす短い言葉には、癒えることのない傷跡と、一生消えない「罪」の感覚が重く横たわっている。

Gemini の分析:なぜ前の要約はダメだったのか？

前の要約が「的外れ」だった理由は以下の 3 点に集約されます。

1. 「犬塚」という不在の主演を無視した: この劇は「犬塚が帰ってくる」という恐怖がすべての動機になっています。それを見落とすと、単なる「防犯意識向上の物語」に見えてしまいます。
2. ジャンルの誤解: これを「社会派の感動ドラマ」として処理したため、「希望」という言葉を使ってしまった。実際には「ウェスタン風の心理スリラー」であり、結末は「ビター(苦い)」なものです。
3. 慎平の行動の意味を捉えていない: 慎平がトイレに行く嘘をついて逃げたのは、感動的な場面ではなく、この村の「闇」が限界に達した瞬間です。

あなたがタイピング中に感じた**「住民の対比が面白い」「竹井も問題がある」「慎平が不可解」**という感覚こそが、この戯曲の正解だと思います。私の要約では、その「不自然さ」を強調するように構成しました。

(以上、Gemini による要約)

Gemini はただテキストを読ませただけでなく、事前に戯曲について会話をしているため、その情報を汲んで訳している Gemini の方が、よりよい要約になっている。ただし、大河原家の葬式は、おそらく村の最年長の大河原家の姫子さんのお葬式だと思うので、誰が死んだが把握できていない。あと、朝陽のセリフは、男の場合のセリフ／女の場合のセリフとして書いて置き、共通するセリフをそのままつなげていて、機械的に判別しにくい。もしかすると、AI はそれが読み取れず、朝陽を男と認識して回答している可能性がある。

2. テキストデータを Excel で整形する。

ある人物のセリフのみにしたいというときなどは、Excel などの表のデータにしておけば、フィルタ機能を使って、それぞれのセリフを容易に取り出せる。

Excel で役名とセリフに分けやすくするために、役名とセリフの間の全角スペースを「:」(全角コロン)に置換する。メモ帳の置換機能を使い、「役名」を「役名:」とするように置換すればいい。置換後は漏れがないか確認する。

スペース区切り

→

コロン区切り

竹井 無灯火だった？
雪乃 朝陽はライトつけて
朝陽 つけてた。
雪乃 え？ つけてたよね
朝陽 つけてたつけてた。

竹井: 無灯火だった？
雪乃: 朝陽はライトつけて
朝陽: つけてた。
雪乃: え？ つけてたよね
朝陽: つけてたつけてた。

新しく Excel ファイルを作って開き、「データ」メニューの「外部データの取り込み」から、戯曲のテキストファイルを選び、区切り文字を「タブ」とその他「:」にする。ト書きの前にはタブを入れていたのでト書きは B 列に表示される。役名は A 列に、セリフは B 列に来る。文章中に「:」があった場合、区切られるので、整形後、Excel ファイルを一通りチェックする。

テキストファイル ウィザード - 2 / 3

フィールドの区切り文字を指定してください。[データのプレビュー] ボックスには区切り位置が表示されます。

区切り文字

☒ タブ(T)

☐ セミコロン(M)

☐ カンマ(C)

☐ スペース(S)

☒ その他(Q): :

☐ 連続した区切り文字は 1 文字として扱う(B)

文字列の引用符(Q): "

データのプレビュー(P)

「さらば黄昏」長塚圭史

登場人物

竹井太 (ひろし)

キャンセル < 戻る(B) 次へ(N) > 完了(E)

データ形式は念のため文字列にした。今回の戯曲なら標準にしても取り込んだ後、書式がおかしくなる可能性は少ない。「完了」ボタンを押す。

テキストファイル ウィザード - 3 / 3

区切ったあとの列のデータ形式を選択してください。

列のデータ形式

☐ G/標準(G)

☒ 文字列(T)

☐ 日付(D): YMD

☐ 削除する(I)

[G/標準] を選択すると、数字は数値に、日付は日付形式の値に、その他の値は文字列に変換されます。

詳細(A)...

データのプレビュー(P)

文字列

「さらば黄昏」 長塚圭史

登場人物

竹井大 (ひろし)

キャンセル < 戻る(B) 次へ(N) > 完了(F)

そのままそのブックの A1 に結果を入れればいいので、OK ボタンをクリックする。

データの取り込み

このデータをブックでどのように表示するかを選択してください。

☒ テーブル(T)

☐ ピボットテーブル レポート(P)

☐ ピボットグラフ(C)

☐ 接続の作成のみ(O)

データを返す先を選択してください。

☒ 既存のワークシート(E):

= \$A\$1

☐ 新規ワークシート(N)

☐ このデータをデータ モデルに追加する(M)

プロパティ(R)... OK キャンセル

Excel ファイルに戯曲のテキストが読み込まれる

Microsoft Excel 分析用 - Excel

ファイル ホーム 挿入 ページレイアウト 数式 データ 校閲 表示 開発 操作アシスト サインイン 共有

貼付け フォント 配置 数値 スタイル 条件付き書式 テーブルとして書式設定 セルのスタイル セル 編集

A22

	A	B
22		
23	第一場	
24		
25		鍋田地区の駐在所。
26		夜遅く。
27		駐在員で巡査長の竹井大の前に、泥だらけの阪本朝陽が
28		その横に朝陽の母親の雪乃。
29		鍋田まちづくり委員会である大河原もそこに。
30		駐在所の中は片付いている（引越し用の段ボールが積み
31	雪乃	私がすごいなって思うのが、やっぱり怒ってないってこと

元データ

100%

これで、B 列からセリフの前の役名がないデータを取り出すことができ、頻出語も役名が回数が多くなったが、それを回避することができる。

セリフのみのデータの分析結果

ワードクラウド



単語出現頻度

名前	スコア	出題頻度
新井	584.20	80
竹井	610.32	86
朝田	483.03	62
藤田	620.06	67
	24.43	10

行動	スコア	出現頻度
言う	3.94	44
わかる	3.32	38
しゃう	3.28	43
行く	2.93	35
知らない	2.67	34

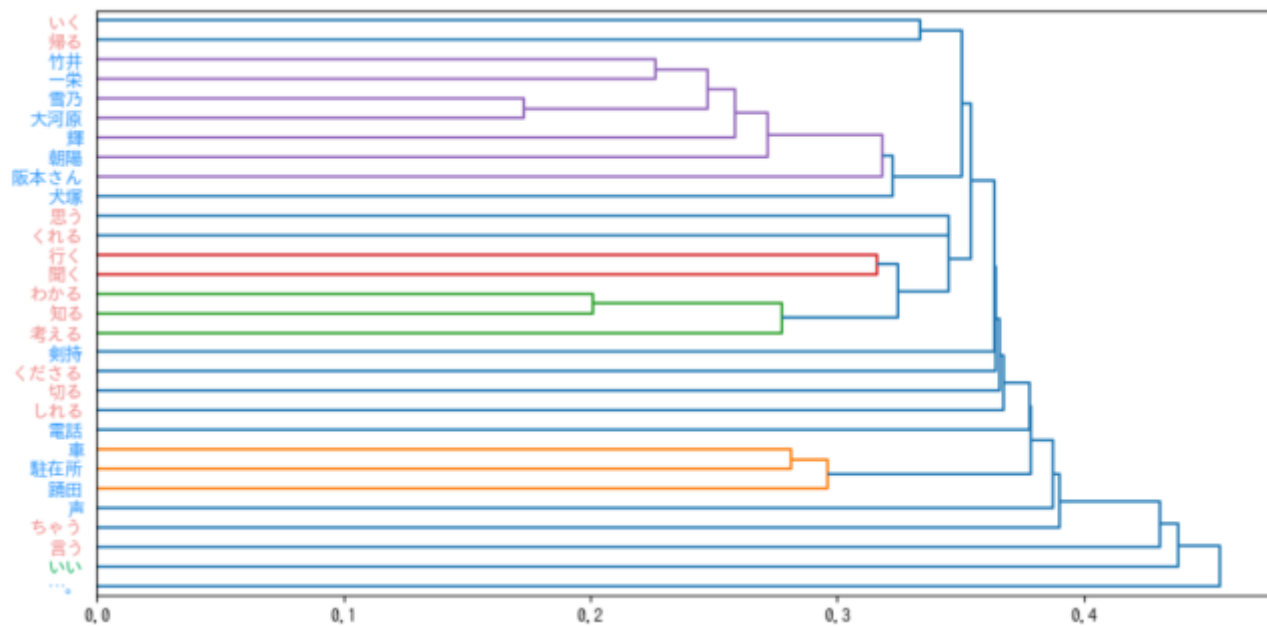
形容詞	スコア	出現頻度
いい	0.25	16
すごい	0.38	13
美味しい	0.28	8
悪い	0.36	7
熱い	0.25	7

評価詞	スコア	出現頻度
ごめんなさい	1.00	9
ありがとう	0.09	8
すいません	1.00	7
いや	0.52	6
おい	0.75	5

共起キーワード

	A	B	C	D
1	単語ペア	単語1	単語2	共起回数
2	切る 電源	切る	電源	17
3	地区 踊田	地区	踊田	9
4	剣持 輝	剣持	輝	9
5	剣持 竹井	剣持	竹井	8
6	北海道 竹井	北海道	竹井	7

階層的クラスタリング



3. 特定の人物のセリフを分析する

2列の表にしてフィルタ機能を使って、A 列のフィルタボタンで登場人物を絞り込めるようにする。

	A	B
1	場・登場人物 ▼	ト書き・台詞
2	第一場	
3		
4		踊田地区の駐在所。
5		夜遅く。
6		駐在員で巡査長の竹井大の前に、泥だらけの阪本朝陽が座っている。
7		その横に朝陽の母親の雪乃。
8		踊田まちづくり委員会である大河原もそこに。
9		駐在所の中は片付いている（引越し用の段ボールが積まれていてもよ
10		
11	雪乃	私がすごいなって思うのが、やっぱり怒ってないってことだと思うん 怒ってないんですよ。ただびっくりしたよって。でも自転車のタイヤ す。ぐにゃんぐにゃんってなっちゃって。だからわあって避（よ）け でも普通あれですよ、自転車が避けるわけですからこっち傾くわけ んですから、向こうは一切ハンドル切ってない、もしくは若干朝日側 それ避ける朝陽の動体視力すごくないかって思っちゃってるんですけ
12	竹井	無灯火だった？
13	雪乃	朝陽はライトつけてますよ。つけてたでしょ？

男のセリフだけ取り出したければ、1 行目のボタンを押して、男だけチェックする。

	A	B
1	場・登場人物 ▼	ト書き・台詞
<div>昇順(S)</div> <div>降順(O)</div> <div>色で並べ替え(I) ▶</div> <div>“場・登場人物” からフィルターをクリア(C)</div> <div>色フィルター(I) ▶</div> <div>テキストフィルター(E) ▶</div> <div>検索 <input type="text"/></div> <div> <input type="checkbox"/> 第一場 <input type="checkbox"/> 第五場 <input type="checkbox"/> 第三場 <input type="checkbox"/> 第四場 <input type="checkbox"/> 第二場 <input checked="" type="checkbox"/> 男 <input type="checkbox"/> 竹井 <input type="checkbox"/> 朝陽 <input type="checkbox"/> (空白セル) </div> <div>OK キャンセル</div>		

	A	B
1	場・登場人物 ▼	ト書き・台詞
1166	男	電話いいんですか切っちゃって。
1168	男	輝さん、まだ話している最中でし
1170	男	はい？
1172	男	名前って？
1174	男	ああ、輝さん。
1176	男	今、あなたが電話でそう呼んだか
1178	男	えー呼びましたよ。おかしいな。
1182	男	あれ、いいんですか切っちゃって
1184	男	輝さんじゃないんですか電話？
	男	わ、怖いですね。いいんですか、
1186		
1188	男	それって絶対に出さないといけな
1190	男	いけませんでしたか？ どなたか
	元データ	フィルタ (+)

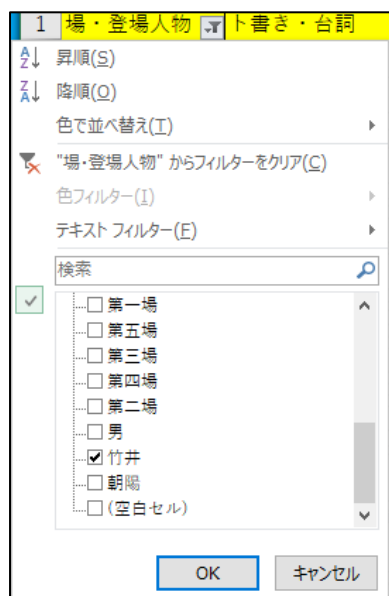
絞り込んだ後、B 列のセリフをコピーして、テキストファイルに貼り付けて読み込ませれば、男のセリフの分析ができる。

男のセリフ.txt

電話いいんですかね
輝さん、まだ話して
はい？
名前って？
ああ、輝さん。

同様に、竹井のセリフだけのテキストファイルを作る。

竹井のセリフに絞り込む



	A	
1	場・登場人物	ト書き・台詞
12	竹井	無灯火だった？
18	竹井	ありがとうございます。朝陽さんちよ
20	竹井	車種とかナンバーはわからなかったよ
25	竹井	うん。
27	竹井	雪乃さんも怖かったですね。
33	竹井	あれでも雪乃さんはそこにいらっしや
35	竹井	車種もわかんないよね？
38	竹井	見てないんですよ雪乃さんは？
55	竹井	あのあたり農道で防犯カメラないんで んに聞きに行ってもらってますから。
57	竹井	じゃ私もいいですか。

この B 列のセリフをコピーして、新しいテキストファイルに読み込ませて、竹井のセリフだけのテキストファイルを作る。

竹井のセリフ.txt

無灯火だった？
ありがとうございます。
車種とかナンバーは
うん。
雪乃さんも怖かった

セリフだけのテキストファイルを User Local に読み込ませる。

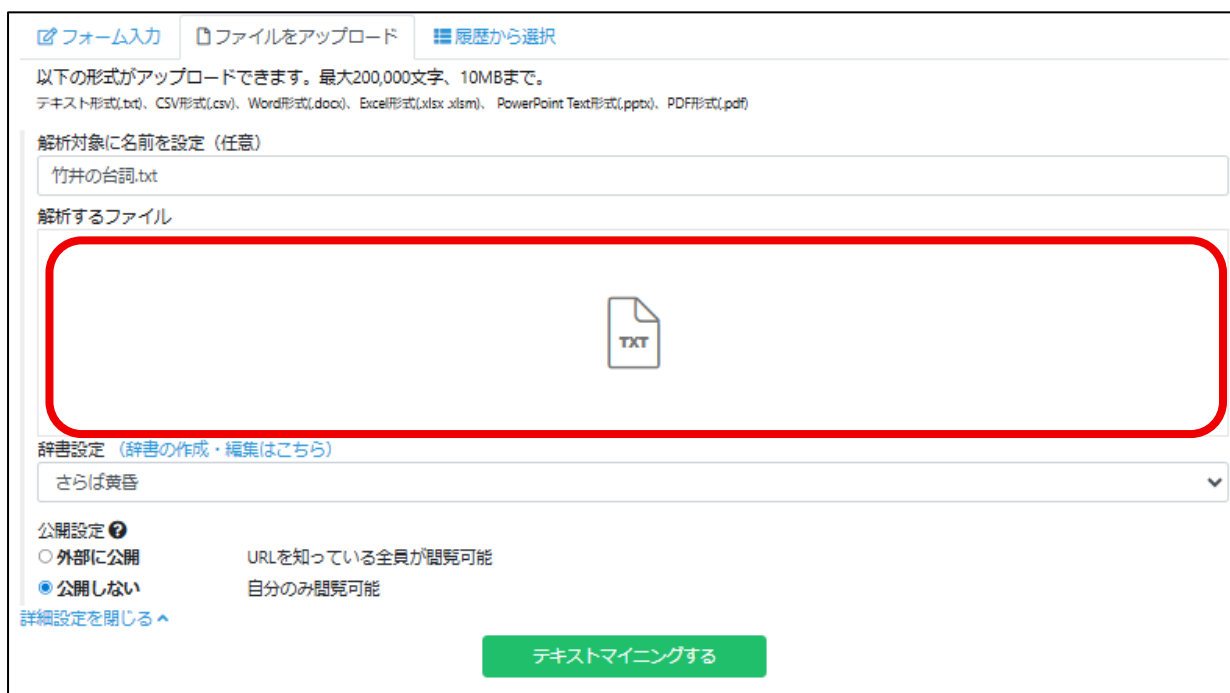
User Local のサイト <https://textmining.userlocal.jp/>

にアクセスし、ログインする。

ファイルをアップロードをクリックする



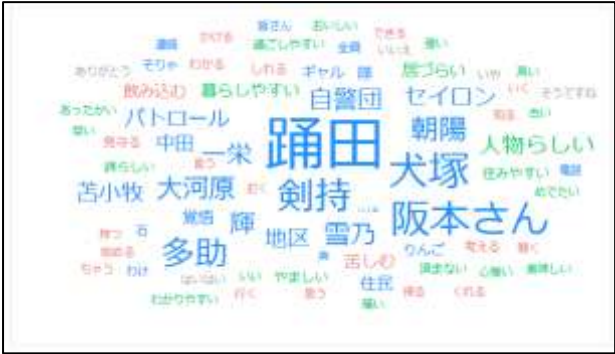
解析するファイルにファイルをドラッグする



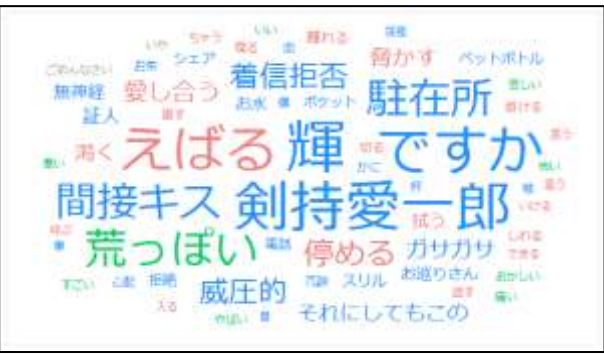
竹井と男の1つの文書を解析の結果。比較をしたいなら、「2つの文書を比較」を使う。

ワードクラウド

竹井



男



単語出現頻度

竹井

■ 名詞	スコア ▲	出現頻度 ▲
...	9.32	33
踊田	334.77	27
剣持	141.11	25
犬塚	168.32	22
阪本さん	166.16	18
輝	38.08	16
朝陽	68.61	14
覚悟	4.99	11
多助	109.67	10
雪乃	46.33	10
地区	13.04	9
大河原	48.64	8
全員	0.89	8
隊	4.18	7
そりゃ	1.08	7

■ 動詞	スコア ▲	出現頻度 ▲
わかる	0.56	17
考える	0.56	14
行く	0.11	12
思う	0.08	12
しれる	0.68	11
言う	0.10	11
知る	0.25	10
いく	0.16	9
くれる	0.08	8
持つ	0.14	7
聞く	0.12	7
かける	0.20	6
帰る	0.09	6
苦しむ	2.30	5
おく	0.15	5

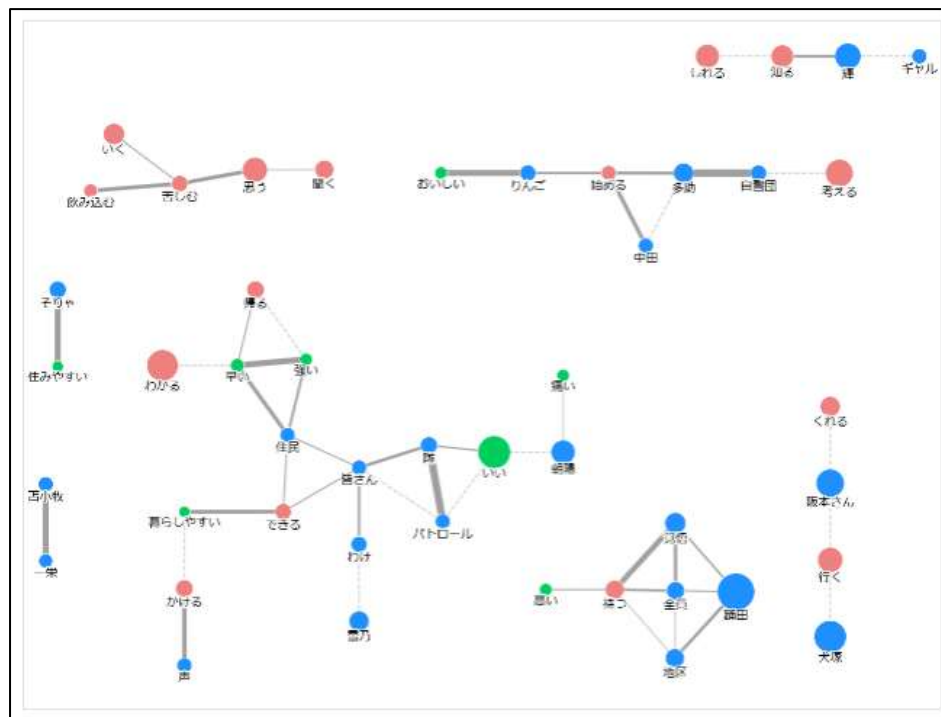
男

■ 名詞	スコア ▲	出現頻度 ▲
僕	0.23	7
輝	8.19	6
電話	0.49	6
音	0.09	3
間接キス	5.23	2
ガサガサ	1.98	2
お水	0.85	2
シェア	0.66	2
ペットボトル	0.60	2
ポケット	0.49	2
冗談	0.26	2
杯	0.26	2
お茶	0.17	2
血	0.14	2
深夜	0.11	2

■ 動詞	スコア ▲	出現頻度 ▲
ちゃう	0.03	4
思う	0.01	4
離れる	0.19	3
切る	0.11	3
しれる	0.05	3
できる	0.01	3
停める	1.24	2
掛ける	0.14	2
直す	0.09	2
喋る	0.09	2
呼ぶ	0.03	2
話す	0.03	2
いける	0.02	2
入る	0.01	2
言う	0.00	2

頻出回数が同じ単語でも、スコアが高いものは、その言葉は重要であると AI が判断しているということ。

竹井



離れる
止める

荒っぽい
疎る
威圧的

ちゃう
切る
いい

話す
駐在所
深夜
お巡りさん

直す
拒絶
掛ける

シェア
ベットボトル
できる

ごめんなさい
ポケット

思う
入る
喉
間接キス
苦しい
すごい
涙く

愛し合う
着信拒否

杯
お水

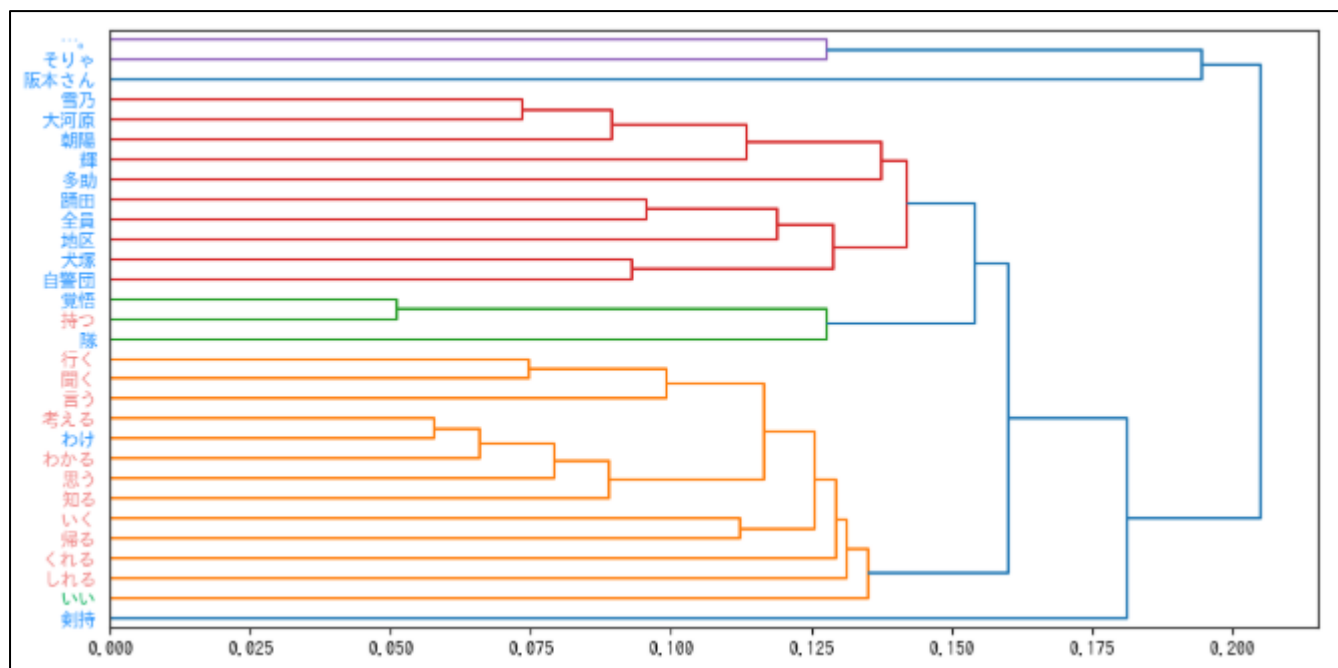
言う
無神経

串
音

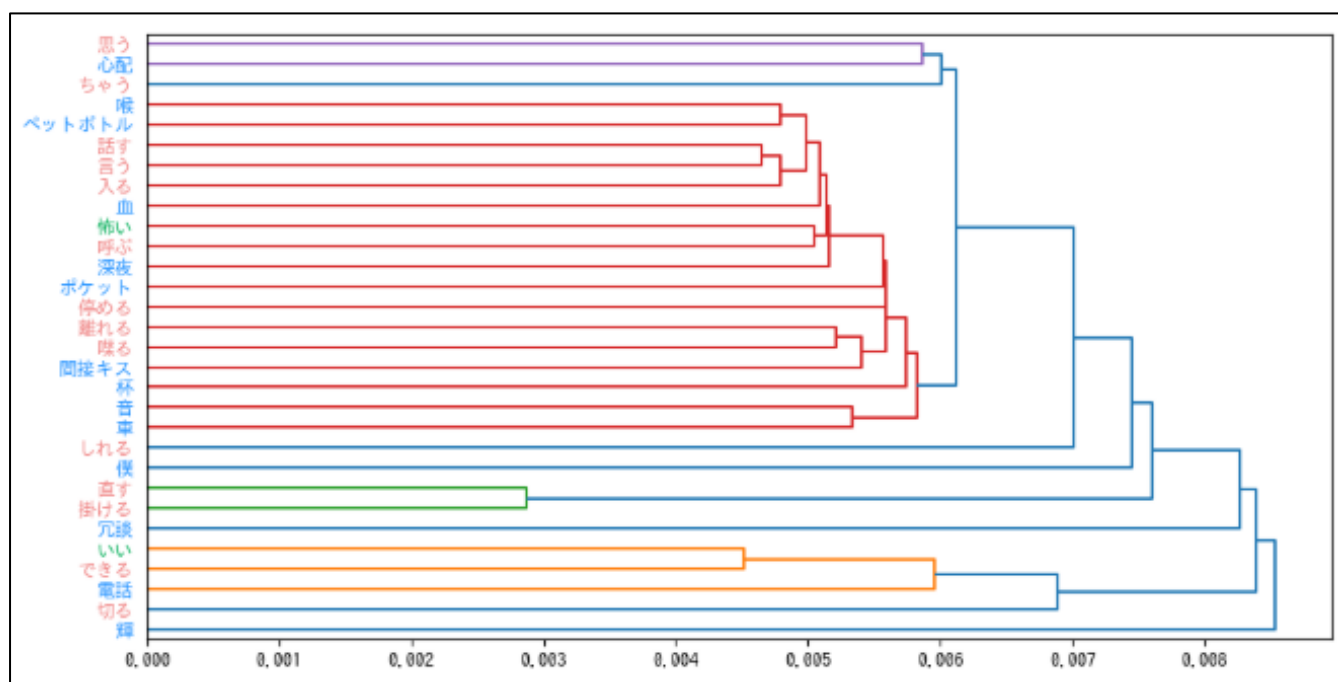
頻出している単語であっても、「繋がり(共起)のパターンがバラバラである」場合は共起ネットワークには出てこない。「輝」は頻出するが同じ言葉を、文の中で言っていたりしない。あとは、輝の名前だけで繰り返しているなどの原因も考えられる。

階層的クラスタリング

竹井



男



竹井のグループは、公的な義務という軸から来る「固さ」が感じられるが、男のグループの方は「個人的な嫌がらせ」から来るものだろうか、論理的ではない一見関係のなさそうなバラバラな言葉が隣り合っている。次にやってみることとしては、二人のデータを合体させて一つのクラスタリングにかけてみたりすると、思わぬ言葉が結びついたりして、戯曲のテーマが浮かび上がったり、二人が危険な領域を共有していることが分かったりする。

「2つの文書を比較」による分析

④ いくつかの文書を解析 **⑤ 2つの文書と比較** **Excelクロス集計** **登録済人件**

サンプル：標準機能の活用例・拡張機能レビューを比較分析

「**ウェブフォーム入力**」 「**ファイルを上アップロード**」 「**画像から読み取**」

以下の形式がアップロードできます。最大200,000文字、10MBまで。
テキスト形式(txt)、CSV形式(csv)、Word形式(docx)、Excel形式(xlsx、xls)、PowerPoint形式(ppsx)、PDF形式(pdf)

解析対象に名前を設定（任意）
資料の台詞.txt

解析するファイル



詳細設定を開く ≡

比較対象に名前を設定（任意）
資料の台詞.txt

比較するファイル



詳細設定を開く ≡

実行結果を表示

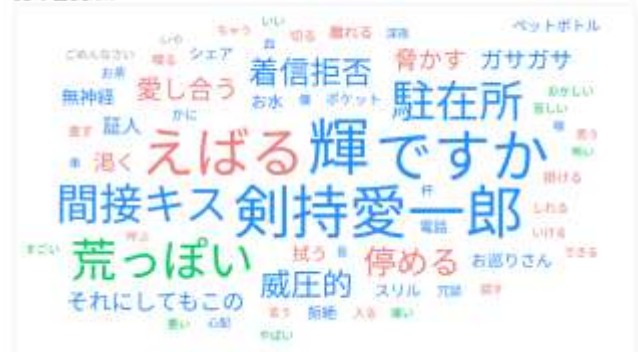
ワードクラウドで比較

スコアが高い単語を複数選び出し、その値に応じた大きさで図示しています。単語の色は品詞の種類で異っており、青色が名詞、赤色が動詞、緑色が形容詞、灰色が感動詞を表しています。

竹井の台詞.txt



男の台詞.txt



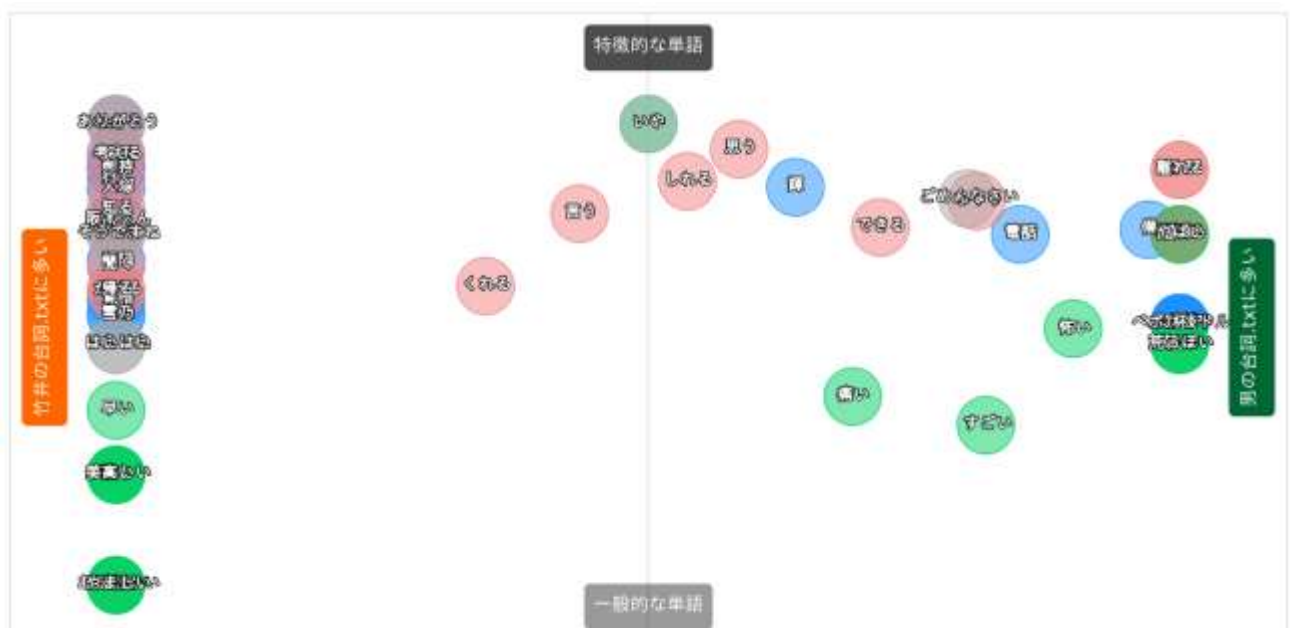
単語分類

2つの文書に出現する単語を、それぞれどちらの文書に偏って出現しているかでグループ分けし、表にしています。グループ中の単語は出現頻度が多い順に並ぶ傾向があります。

竹井の台詞.txtにだけ出現	竹井の台詞.txtによく出る	両方によく出る	男の台詞.txtによく出る	男の台詞.txtにだけ出現
ありがとう ー わかる 考える 踊田 剣持 行く 犬塚 知る 阪本さん いく いいえ そうですね 朝陽 聞く かける 帰る 覚悟 多助 雪乃 苦しむ 地区 いえいえ はいはい 全員 大河原 そりゃ 隊 りんご わけ	くれる	いや いい 思う しれる 輝 言う 持つ おく	ごめんなさい ちゃう できる 僕 電話 入る いける 呼ぶ 話す 怖い 音 車 心配 痛い お水 すごい	切る 離れる やばい 停める 喋る 掛ける 直す お茶 ガサガサ シェア ペットボトル ポケット 冗談 喉 杯 深夜 血 間接キス おかしい 悪い 苦しい 荒っぽい えばる よる 乗り越える 乾く 出会う 切れる 愛し合う 感じる

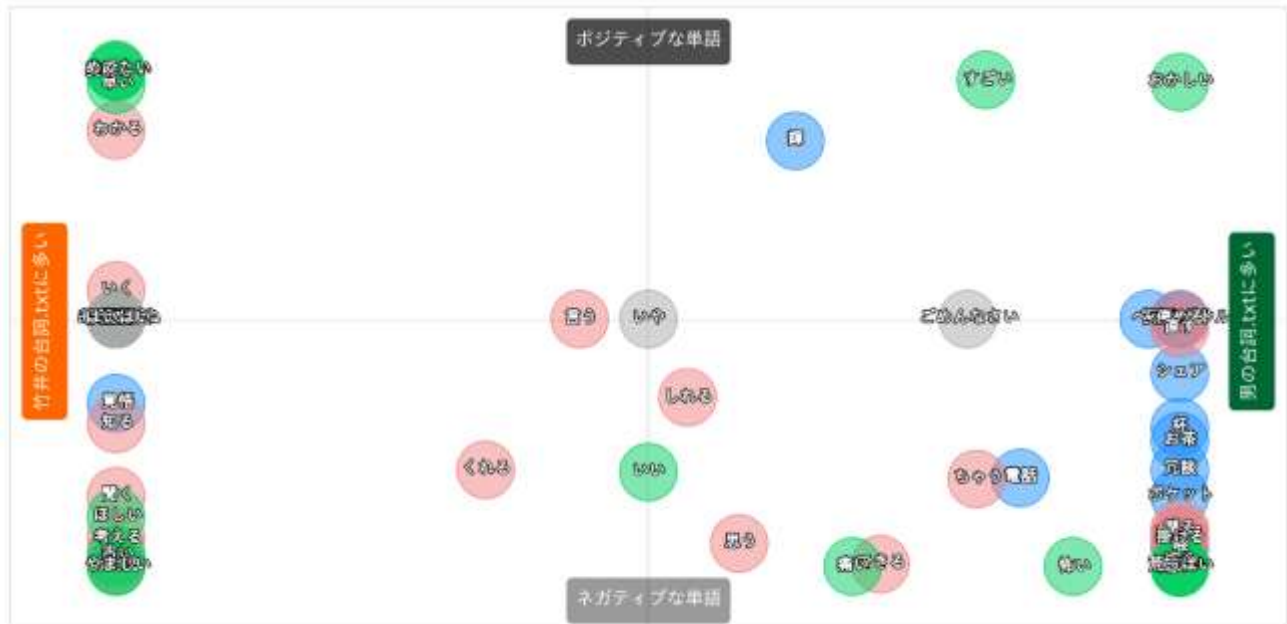
特徴語マップ

それぞれの単語がどちらの文書により多く出現するか、またその単語が文書においてどれだけ特徴的であるかを2次元にマッピングしています。上寄りの単語はこの文書の特徴的な単語であることを、下寄りな単語はどんな文書にも出現する一般的な単語であることを示しています。



ネガポジマップ

それぞれの単語がどちらの文書により多く出現するか、またその単語の意味がどれだけポジティブ・ネガティブであるかを 2 次元にマッピングしています。上寄りな単語はその単語がポジティブな意味を持つことを、下寄りな単語はその単語がネガティブな意味を持つことを示しています。



単語の出現比率

2 つの文書から出現回数の多い単語を選び出し、それらが 2 つの文書においてどれぐらいの比率で出現するかをグラフにしています。

名詞

竹井の台詞.txt	単語	男の台詞.txt
100	一	0
100	随分	0
100	剣持	0
36	輝	64
100	犬塚	0
100	阪本さん	0
2	僕	98
15	電話	85
100	朝陽	0
100	覚悟	0
100	多助	0
100	雪乃	0
0	お茶	100
0	ガサガサ	100
0	シェア	100

動詞

竹井の台詞.txt	単語	男の台詞.txt
100	わかる	0
41	思う	59
100	考える	0
0	切る	100
0	離れる	100
100	行く	0
46	しれる	54
19	ちやう	81
100	知る	0
56	言う	44
100	いく	0
28	できる	72
0	停める	100
0	帰る	100
0	掛ける	100

形容詞

竹井の台詞.txt	単語	男の台詞.txt
50	いい	50
0	やばい	100
10	怖い	90
0	おかしい	100
0	悪い	100
0	苦しい	100
0	荒っぽい	100
30	痛い	70
100	早い	0
18	すごい	82
100	おいしい	0
100	古い	0
100	強い	0
100	美味しい	0
100	高い	0