



Japan Depression Center (JDC)
一般社団法人日本うつ病センター

メール相談テキスト分析

Agenda

- 解析デザイン・データベース化
- 利用者基本情報
 - 年齢・性別
 - 相談者の立場
 - 居住地
 - 流入経路
 - 往復回数
- テキスト分析
 - 前処理方法
 - メール1件あたりの文字数に関する代表値
 - 主題抽出
 - 趣旨抽出

解析デザイン・ データベース化

解析デザインのための問いの設定とDB化作業

問いの設定

- メール相談の存在意義・価値はどこにあるか？
 - 誰に対しての存在意義か？
 - どんな悩みを抱えている人たちの受け皿となっているか？
 - どう対応してもらうことを期待しているか？

データベース化作業

- カラム追加
 - データ取捨選択用
 - Production: テスト／相談
 - Non-duplicated: 重複
 - Acceptance: 適切なメール相談と認定されたもの
 - 分類用
 - Age_seg
 - Prefecture
 - 回数カウント用
 - Reply_count
 - 返信履歴の有無
 - Reply_text
- 表記揺れ修正
 - Age: 全角→半角／XX代→XX

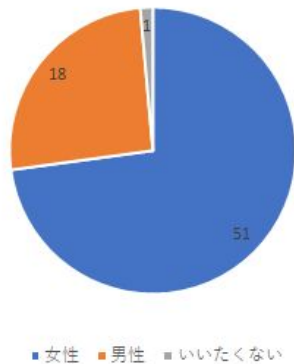
データベース概要

- 対象期間: 2020年11月1日～2021年9月31日
- メール件数
 - 全相談件数: 80件
 - 重複除外後件数: 74件
 - 趣旨に沿った相談件数: 71件
- 利用者数
 - 70名
- データのカラム
 - Timestamp, Agreement, Name, **Sex**, Age, **Age_seg**, Residence, Prefecture, **Position**, Route, Research_use, **Text**, Production, Non-duplicated, Acceptance, **Reply_count**, Reply_text

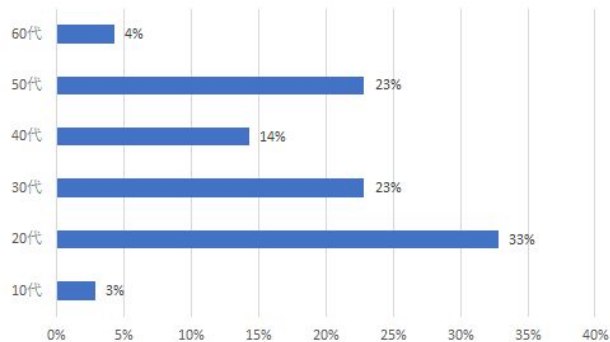
利用者基本情報

利用者基本情報：年齢・性別

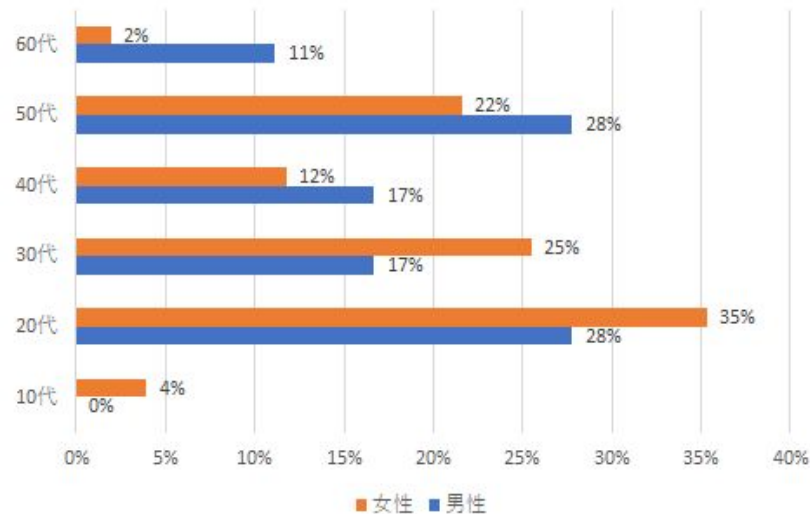
<性別構成比>



<年代構成比>

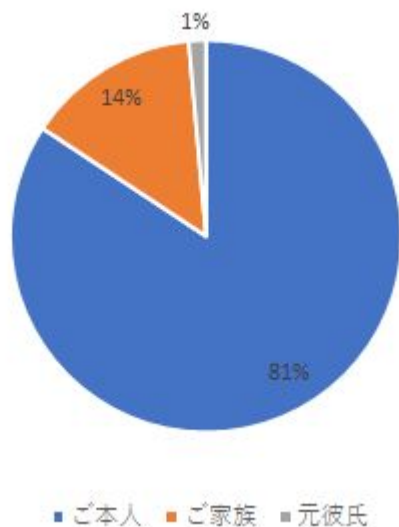


<性別 × 年代構成比>

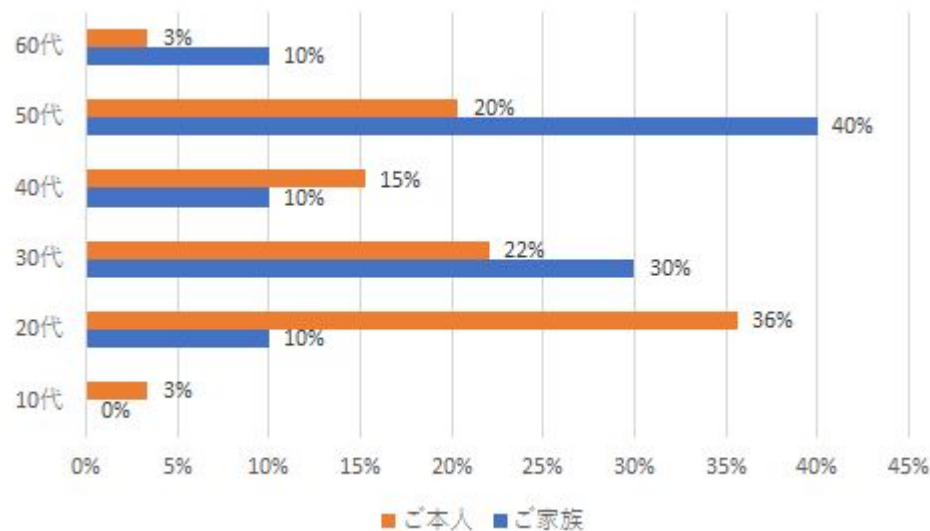


利用者基本情報: 相談者の立場

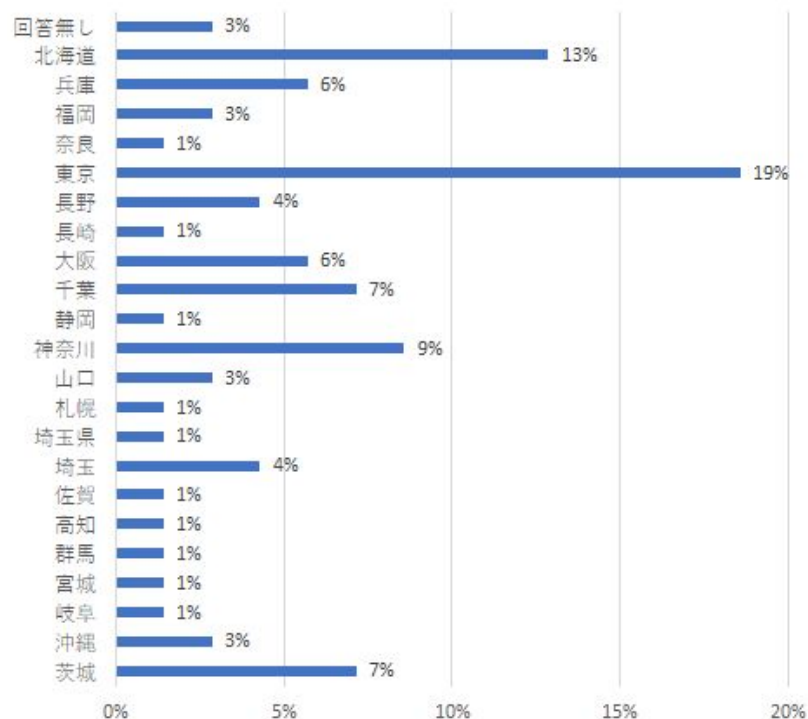
<立場の構成比>



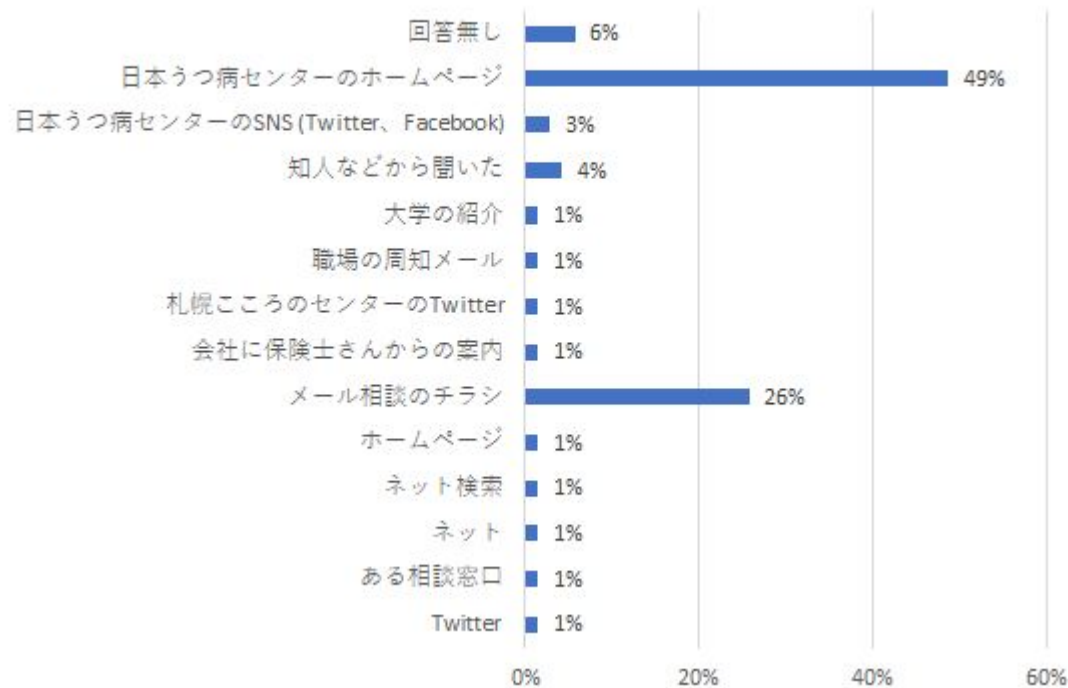
<立場 × 年代構成比>



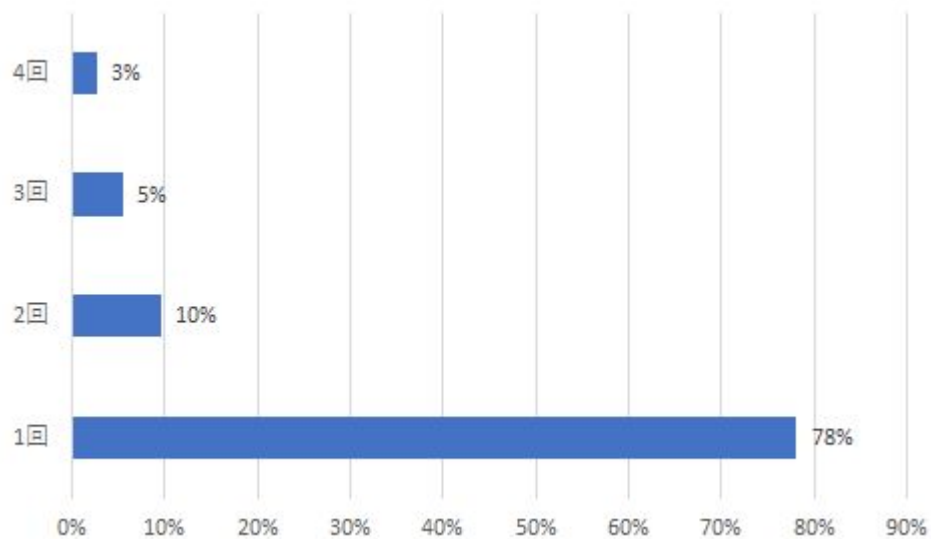
利用者基本情報:居住地



利用者基本情報: 流入経路



利用者基本情報: 往復回数



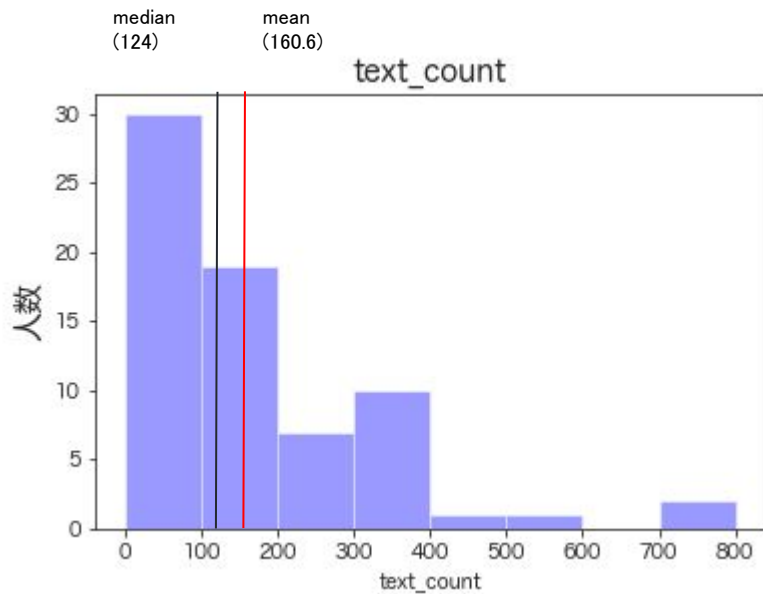
テキスト分析

テキストデータの前処理方法

- クリーニング
 - 改行文字、半角スペース除外
 - 句読点などの記号
 - 名詞: 1632⇒/, 965⇒(, 969⇒), ,1317⇒), 1339⇒?, 1791⇒., 1966⇒[, 2386+2390+2393+2397⇒+, 2762⇒][, 2768⇒], 2837⇒…※
- 形態素解析
- 単語正規化
 - 文字種(半角全角／大文字小文字統一)
 - 数字の削除: 全角・半角数字は削除／漢数字は残す(熟語の一部であることも散見されるため)
 - 単語統一(SONY・ソニー、など): なし
- ストップワード基準の絞り込み
 - 動詞・形容詞・名詞のうち、出現構成比(対象単語の利用あり人数÷利用者数)の降順30単語
 - 仕事／コロナ／気／不安は、相談内容と関係する可能性が高いため、ストップワードから除外

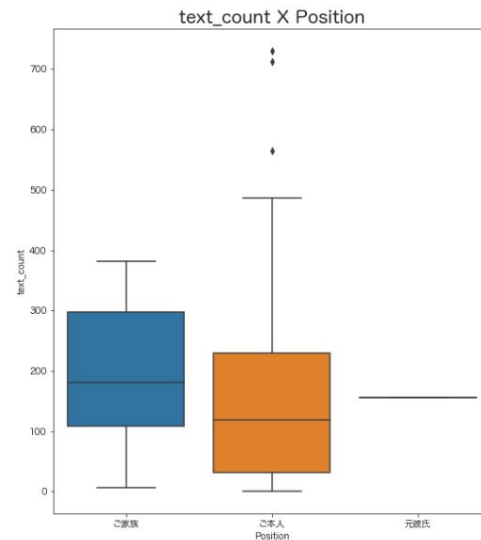
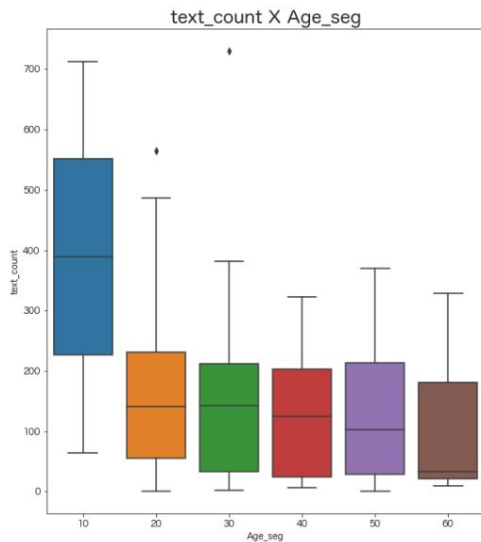
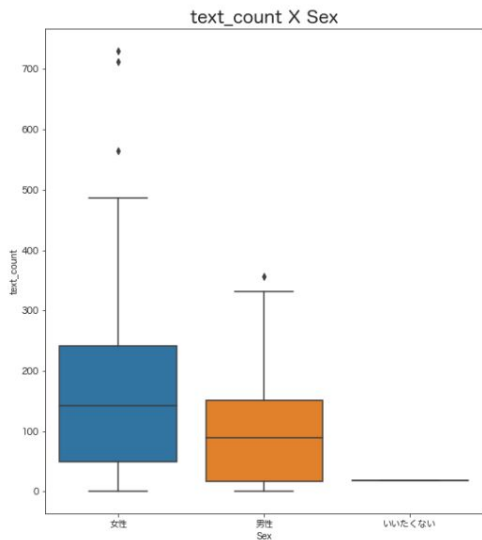
メール1件あたりの文字数

- 中央値:124文字／平均値:160.6文字



メール1件あたりの文字数_属性別

- 1件あたりの文字数は、下記属性で相対的に多い
 - 性別:女性
 - 年齢層:若年層
 - 関係性:ご家族



主題抽出方法

- 悩みの主題を外観的にとらえるべく、各メールを特徴づける重要語(名詞・動詞・形容詞に限定)の抽出を試みた
- 単語の出現頻度ではなく、利用者が意識的に使用した単語を抽出できるよう、TF-IDF法を採用

<TF-IDF値の概要説明>

メール1件

文章1	単語1	2	3	4	...	10
文章2	単語4	5	1	1	...	18
文章3	単語1	20	25	20	...	20
文章4	単語4	12	10	27	...	45
...						
文章24	単語4	12	32	72	...	83
文章25	単語1	4	4	23	...	92

単語4

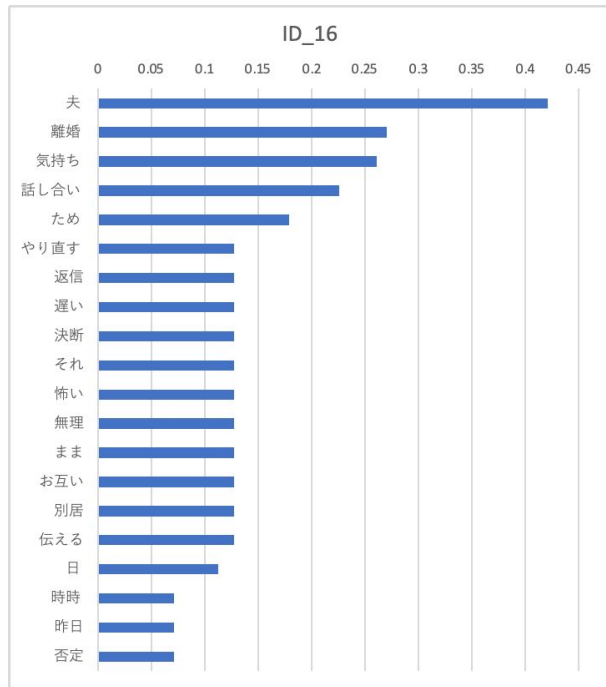
メール1件内で出現する文章数が多く、文章1件あたりで出現する回数も多い
⇒一般的な多用単語とし、重要値を下げる

単語20

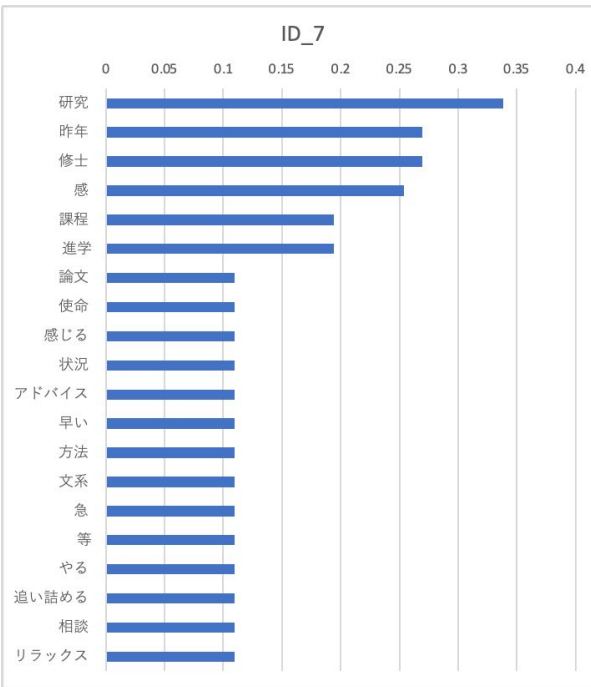
メール1件内で出現する文章数は少ないが、文章1件あたりで出現する回数は多い
⇒一般的ではなく、当メール文章内で主張したい内容にかかわる単語

主題抽出結果:属性「ご本人」から一部抜粋①

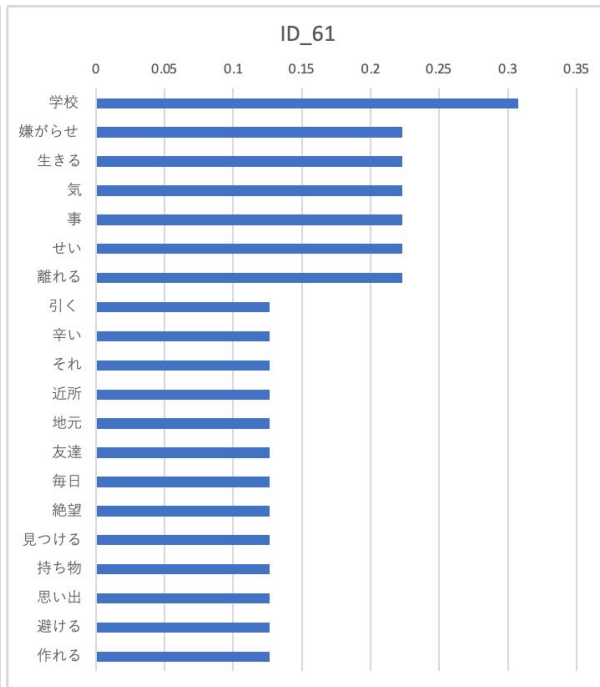
家族関係



学業

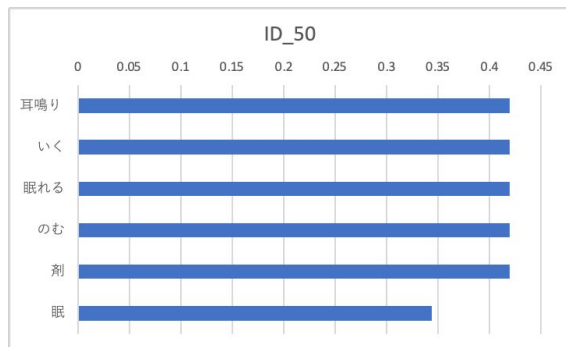


学校での嫌がらせ

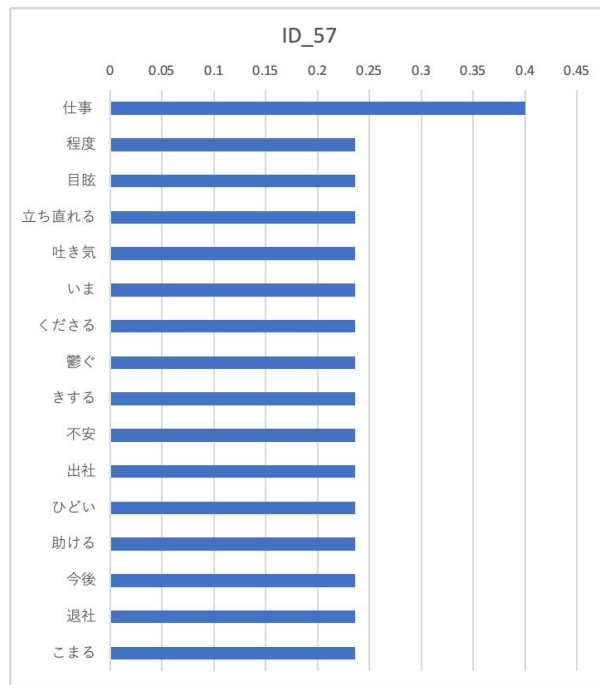


主題抽出結果:属性「ご本人」から一部抜粋②

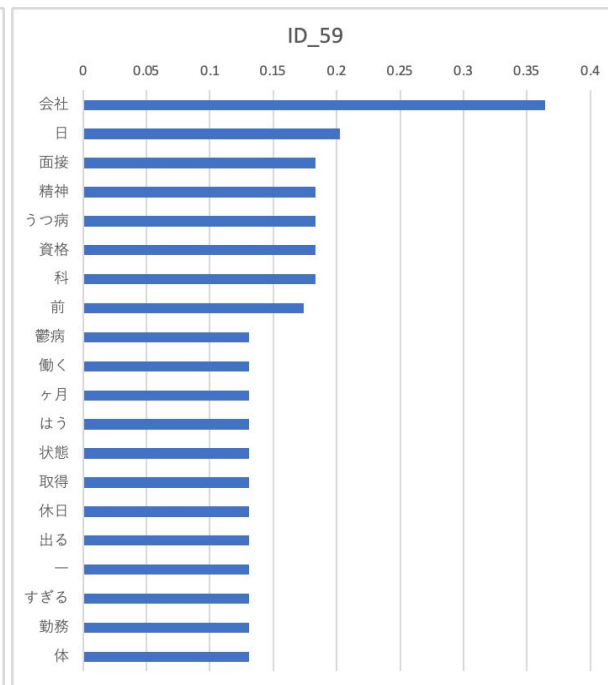
身体症状



仕事

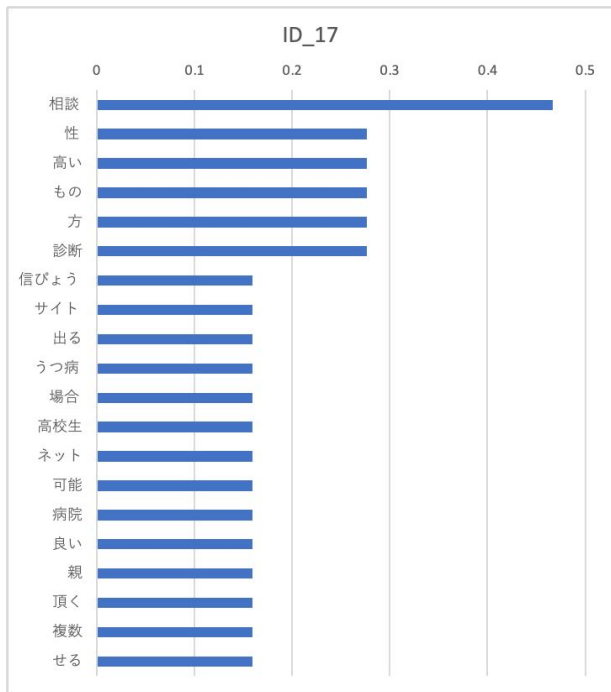


転職

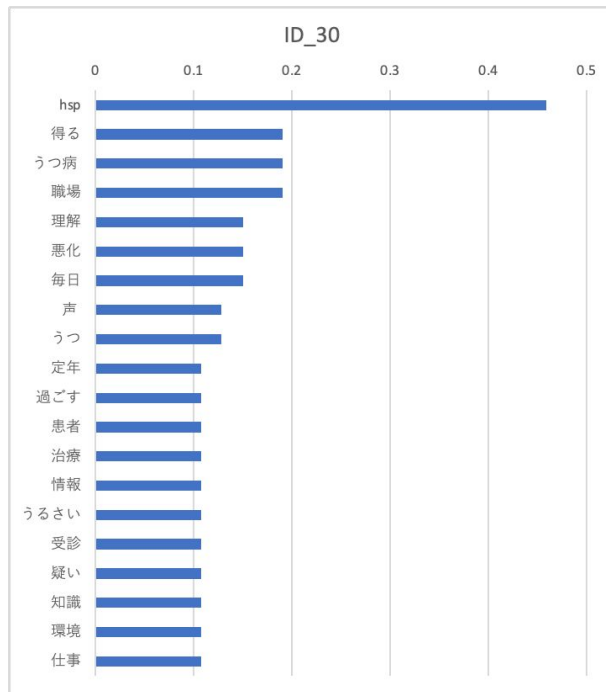


主題抽出結果：属性「ご本人」から一部抜粋③

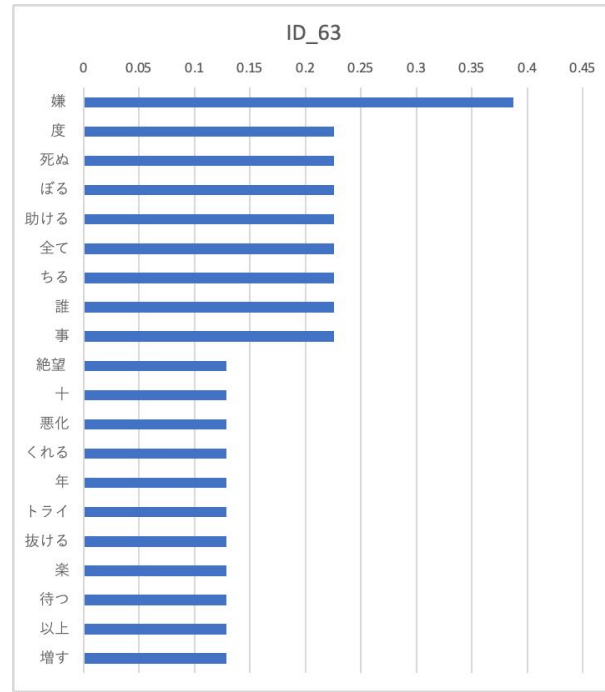
ネット上での相談先



HSPとしての苦悩



希死念慮



主題抽出の活用性

- メール受信後、自動処理による重要語の抽出結果を踏まえ、主題アセスメントとして使用することは可能と評価している
- 主題アセスメントの対応者の割り振りや対応スピード改善などに向けて活用余地があると仮説する
 - メール相談に対応する医師・カウンセラー・PSWの割り振り方を検討する材料とする
 - 対応者が事前に主題を把握することで、本文読解⇒対応スピードを速めるための工夫材料とする
- 今後、割り振り担当者・対応者の意見も踏まえ、より運用プロセスに落とし込んで活用できる方法を協議していく

趣旨抽出の試みと今後の課題

- メール相談をされた目的・どんな情報やアドバイスを期待しているかなどの趣旨について、人間の読解ではなくデータ分析的に抽出することを試みた。
- 結果、相談されている立場によって可否が分かれることが明らかとなった。
 - ご本人からのメール内容は、趣旨の抽出が困難
 - 人間が行間を読むことによって、「気持ちに寄り添った言葉かけがほしい」「何か専門家として思いつくことがあればアドバイスをしてほしい」という期待を読み取ることは可能
 - ただし、テキストで明示的に表現されているわけではないため、顕在情報だけで趣旨を抽出する特徴量を構築しにくい
 - ご家族からのメール内容は、比較的趣旨が文章内で明示されている
 - 文章が起承転結に沿って書かれていることが多い
 - 結に相当する下から1-3行目付近で具体的にどんな情報を必要としているか確認しやすい
- ご本人からのメール内容のように、趣旨の明示がない事例においてもそれを拾い上げ、対応の質を向上させる材料として活用する方法を引き続き模索していく必要がある。