



tokenDetails を用いて、品詞情報を取得したところ、歯科的な専門用語としての名詞が多く、有用な情報が含まれていると考えられました。そこで、解析対象を名詞のみとし、ワードクラウドを作成しました。

## 結果と考察

本プログラムを実行すると、色濃く示された単語は、「咬合」「義歯」「歯」でした。「咬合」= 噛み合わせ、「義歯」= 入れ歯 を指します。その他に大きく示された単語は、「印象」= 型取り、「調整」= 入れ歯の調整、「粘膜」= 入れ歯を支える歯茎や口の中の柔らかい部分、「顎」= 上あごや下あご、などがありました。

これらの単語は、「良い噛み合わせ」や「良い合い具合」といった、使いやすい入れ歯を作るのに必要な治療方法へのフィードバックを含んでいると考えられます。

「入れ歯」に関する治療には、さまざまな診療内容があります。そこで、ワードクラウドを話題（トピック）ごとに分けることのできる「トピックモデリング」を行うこととしました。（TopicModeling.mlx に続く）

## 参考

1) MATLAB Answers - テキスト分析 特定の品詞のみを対象

[https://jp.mathworks.com/matlabcentral/answers/755939-#answer\\_633479](https://jp.mathworks.com/matlabcentral/answers/755939-#answer_633479)

2) ビデオおよび Web セミナーシリーズ - MATLAB によるテキストマイニングビデオシリーズ

<https://jp.mathworks.com/videos/series/text-mining-with-matlab-video-series.html>

## ポートフォリオのインポート | readtable、フィードバックの抽出

```
Portfolio = readtable( 'Portfolio.xlsx' );  
Portfolio = Portfolio.Feedback ;
```

## 各 cell(ポートフォリオ)の最初の"診察お疲れ様でした。"などの削除 | cellfun, eraseBetween

```
Portfolio_erasedFirst = ...  
    cellfun( @(x) eraseBetween( x, 1, "。", 'Boundaries','inclusive'), Portfolio, 'UniformOutput', false );
```

## 引用>>と空白文字(whitespacePattern)の間を削除 | cellfun, eraseBetween, whitespacePattern

```
Portfolio_erasedQuotation = ...  
    cellfun( @(x) eraseBetween( x, '>>', whitespacePattern, 'Boundaries','inclusive' ), Portfolio, 'UniformOutput', false );
```

## ドキュメントトークン (tokenizedDocument 配列) の作成

```
%/ 参考: https://jp.mathworks.com/matlabcentral/answers/755939-#answer\_633479
```

```
rawDocumentToken = tokenizedDocument( Portfolio_erasedQuotation ) ;
```

## ストップワードの除去、句読点の除去 | `removeStopWord, erasePunctuation`

```
DocumentToken = removeStopWords( rawDocumentToken ); %ストップワードの除去  
DocumentToken = erasePunctuation( DocumentToken ); %句読点の除去
```

## レンマ化 | "走れ, 走り -> 走る", 辞書の見出し語のようにする | `normalizeWords`

```
DocumentNormalized = normalizeWords( DocumentToken, 'Style', 'lemma' );
```

## 品詞情報の取得 | `tokenDetails`

```
DocumentDetails = tokenDetails( DocumentNormalized );
```

## `PartOfSpeech` を参照し、名詞のみ抽出（固有名詞は含まない）

```
NounIndex = DocumentDetails.PartOfSpeech == "noun" ;  
OnlyNoun = DocumentDetails( NounIndex, : );
```

## カテゴリー化 | `categorical`

```
CategoricalNoun = categorical( OnlyNoun.Token );
```

## Hello WordCloud !!

```
wordcloud( CategoricalNoun );
```





はじめに、2つのワードクラウドとも **PortfolioCloud.mlx** で強調して表示された単語が中心となってトピックを形成することがわかります。

## 左のワードクラウド

Topic1は「義歯（入れ歯）」を中心に、専門用語ではなく一般的な名詞「適切、状態、必要、大切」などが挙げられており、入れ歯の説明のような文章「適切な入れ歯の状態」「入れ歯は大切」などかもしれません。

Topic 2, 3, 4 はそれぞれ、「入れ歯の合い具合」, 「型取り」, 「噛み合わせ」に関連した単語が表示されました。

なお、これらは周辺の単語を用いて要約することができ、例えば Topic 2 「義歯床の粘膜・顎への適合は支持・負担に必要」は「入れ歯が、上あごや下あごの歯茎に合っていることは、（噛む）力を支えるのに必要」となると考えられます。（歯茎に合わない入れ歯はお食事の際の痛みの原因になります）

## 右のワードクラウド

左のワードクラウドと同様に、Topic 1, 2, 4 はそれぞれ、「噛み合わせ」, 「入れ歯の合い具合」, 「型取り」に関連した単語が表示されました。

Topic 3 は左のワードクラウド Topic 1 に似ていますが、「顎」が中心にあり、近くに「位」があります。合わせて「顎位」とは「上下のあごの位置・噛み合わせの位置」を示すため Topic 2 と同様に「噛み合わせ」に関連した単語が表示されたと考えられます。

こちらも周辺の単語を用いて要約すると、Topic 1 は「咬合は、人工歯と残存歯の接触を診査・確認し、調整を行う」となり「入れ歯の人工の歯と残っている歯とのどちらの当たり具合も確認し調整を行う」であると考えられます。（入れ歯に集中するあまり、入れ歯にある人工の歯ばかりを見てしまうことがあります）

## 左右2つのワードクラウドから

これらの大川が行ったフィードバックは、臨床実習生のポートフォリオに記載された疑問や質問に対して回答した内容になります。

したがって、これらのトピックの中心は「噛み合わせ」, 「入れ歯の合い具合」, 「型取り」であり、**PortfolioCloud.mlx** での結果と同様に、「良い噛み合わせ」や「良い合い具合」といった、使いやすい入れ歯を作るための疑問や質問が多くあったことが推察されます。

さらに、その周辺の単語を用いると「噛み合わせ」, 「入れ歯の合い具合」, 「型取り」を説明する文章に要約されます。もしかすると、この要約される内容が「臨床実習生が治療の前に学習したい・教員が伝えるべき内容」なのかもしれません（各治療の前には、治療に関する予習レポートの提出とディスカッションを行います）。

このように、より多くのポートフォリオをテキストマイニングを用いて解析を行うことで、臨床実習生の予習とディスカッションを充実させることができ、患者様への質の高い治療につながるかもしれません。

また、臨床実習におけるポートフォリオだけでなく、様々な講義や実習で生じた疑問や質問、その教員からのフィードバックに活かせる可能性があると考えられました。

## 課題と解析したいこと

- 臨床実習生の疑問や質問そのものをトピックモデリングしたものではありません。臨床実習生が記載したポートフォリオ全体と大川のフィードバックの内容とが合致しているかを検討する必要があります。
- Portfolio.xlsx を開くと、1 列目は Treatment と書かれた列となっています。この列には、治療の内容（種別）が書かれており、カテゴリとして分けて解析することで、新たな発見があるかもしれません。

## 結論

歯学部臨床実習におけるポートフォリオへのフィードバックに対して、テキストマイニングとしてトピックモデリングを行い、ワードクラウドとして表示を行いました。

その結果、臨床実習生の疑問や質問に対して、どのような回答が多かったかが明らかとなりました。主に、「噛み合わせ」、「入れ歯の合い具合」、「型取り」に関連した単語がワードクラウドに表示されましたが、中心と周辺の単語によってフィードバックが要約されたことから、これらに関連する内容を治療前にディスカッションする必要があると考えられました。

## 謝辞

本解析にあたり、臨床実習に関わる先生方および臨床実習生のみなさまに心から厚く御礼申し上げます。

## 参考

1) MATLAB Answers - テキスト分析 特定の品詞のみを対象

[https://jp.mathworks.com/matlabcentral/answers/755939-#answer\\_633479](https://jp.mathworks.com/matlabcentral/answers/755939-#answer_633479)

2) ビデオおよび Web セミナーシリーズ - MATLAB によるテキストマイニングビデオシリーズ

<https://jp.mathworks.com/videos/series/text-mining-with-matlab-video-series.html>

```
% PortfolioCloud.mlx の続きです。  
% PortfolioCloud.mlx を先に実行しておき、OnlyNoun がワークスペースにあることを確認します。
```

## フィードバックの数を確認

```
NumOfDocuments = max(OnlyNoun.DocumentNumber) ;
```

フィードバックの数と同じサイズの cell 配列を作成し、各 cell にフィードバックごとの名詞を格納 | cell, for, join

```
NewDocuments = cell( NumOfDocuments, 1 ) ;
```

```
for Now = 1 : NumOfDocuments
```

```

NewDocuments{Now} = join( ...
    OnlyNoun.Token(OnlyNoun.DocumentNumber == Now ) ...
);

end

```

## 再トークン化 | tokenizedDocument

```
NounDocumentToken = tokenizedDocument( NewDocuments ) ;
```

## bag-of-words の作成 | bagOfWords

```
bag = bagOfWords(NounDocumentToken);
```

## bag-of-words に空のドキュメントがあれば削除 | removeEmptyDocuments

```
bag = removeEmptyDocuments(bag);
```

## bag-of-words をもとに LDA model の作成 | fitlda

```

numTopics = 4 ; %作成するトピック数
mdl = fitlda(bag, numTopics, 'Verbose', 1);

```

トピックへの初期の割り当ては 0.0138189 秒でサンプリングされました。

Iteration	Time per iteration (seconds)	Relative change in log(L)	Training perplexity	Topic concentration	Topic concentration iterations
0	0.00		4.296e+02	1.000	0
1	0.01	2.7309e-02	3.657e+02	1.000	0
2	0.01	2.0247e-03	3.614e+02	1.000	0
3	0.01	1.6601e-03	3.578e+02	1.000	0
4	0.01	1.9733e-04	3.574e+02	1.000	0
5	0.01	1.1599e-03	3.550e+02	1.000	0
6	0.01	7.7639e-04	3.534e+02	1.000	0
7	0.01	9.5162e-05	3.536e+02	1.000	0

```
disp( mdl ) ;
```

ldaModel のプロパティ:

```

    NumTopics: 4
    WordConcentration: 1
    TopicConcentration: 1
    CorpusTopicProbabilities: [0.2500 0.2500 0.2500 0.2500]
    DocumentTopicProbabilities: [203x4 double]
    TopicWordProbabilities: [1112x4 double]
    Vocabulary: [1x1112 string]
    TopicOrder: 'initial-fit-probability'
    FitInfo: [1x1 struct]

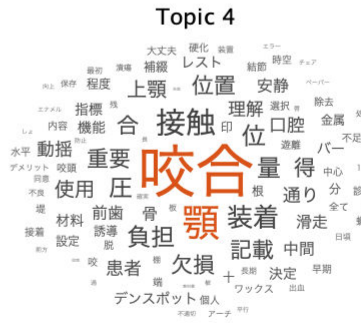
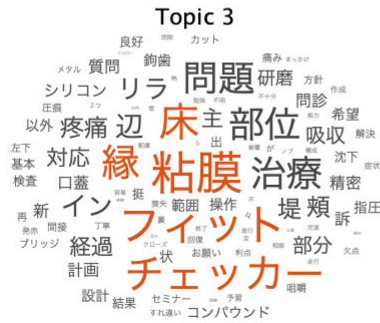
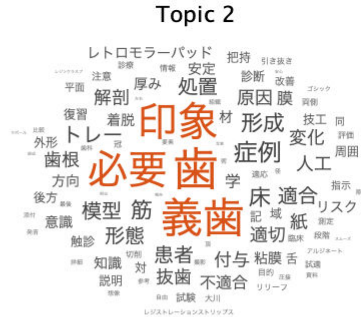
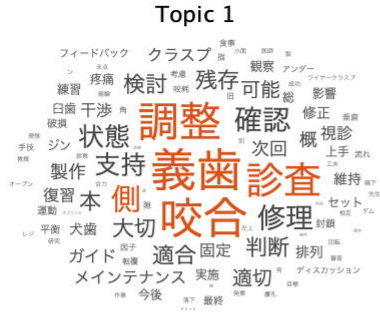
```

## 各トピックの Hello WordCloud !! | subplot

```
figure
```



```
for i = 1:numTopics
    subplot(2,2,i)
```



```
wordcloud mdl, i);
title("Topic " + i)
end
```