

## 表計算ソフト 形態素解析

## おしながき

宿題提出

- 表計算ソフトの使用
- タグ付きコーパス
- 形態素解析

## 表計算ソフト

## 表計算ソフト

- 表形式のデータを扱う
- 見つかった例を手作業でチェック・分類・集計するうえで最適
- タグ付きのデータを閲覧・検索するときに便利
- テキストエディタは、表形式のデータを確認・編集するには向かない

## 表計算ソフトの使用

## 表計算ソフト

- 表形式のデータを扱う
- 例えば、見つかった例を分類したり、集計したりするうえで最適
- とくに統計的な処理をしない場合でも、見つかった例を表にまとめて整理するには最適
- テキストエディタは、表形式のデータを確認・編集するには向かない

## 表計算ソフト

- Excel など
- 表形式のデータの閲覧・編集には最適

## 表計算ソフト

- 悪い例
  - 空白に意味を持たせている
  - セルの結合
  - データとコメントなどが分かれていない
  - 色がデータの一部

	A	B	C	D	E	F
1						
2						
3					黄色は××さんの情報による	
4						
5		名詞				
6			kanamaI	頭		
7			kanmai			
8			mipana	顔		
9			mii	目		
10			pana	鼻		
11			mim	耳		
12			fucl	口 (※ほんと?要確認)		
13		形容詞				
14			gaba:gaba	大きい		
15			mma:mma	おいしい		
16						
17						
18						

## 表計算ソフト

	A	B	C	D	E
1	単語	品詞	意味	備考	
2	kammai	名詞	頭		
3	kanamai	名詞	頭		
4	mipana	名詞	顔		
5	mii	名詞	目		
6	pana	名詞	鼻		
7	mim	名詞	耳	× ×さんによる	
8	fucl	名詞	口	要確認	
9	gaba:gaba	形容詞	大きい		
10	mma:mma	形容詞	おいしい		
11					
12					
13					
14					
15					
16					

- 良い例

## 表計算ソフトの実際

- 検索結果の読み込み
- 見出し行
- タグ付け作業の実際
- 条件付き書式
- ソート、フィルタ、ピボットテーブル

## タグ付きコーパス

## タグ付きコーパス

- コーパスに、単純な文字列だけでは分からない情報を付与したもの
- 単語境界、品詞、活用形、発音、意味カテゴリ、発話時間、誤りかどうか etc.
- 検索できる情報が増える。しかし、作成作業や検索は複雑になる。タグが誤っている場合もある

## 形態素解析

- 英語の場合、主に品詞のタグ付け作業（ほかに活用形の情報など）
- 日本語の場合、文章が単語ごとに分かち書きされていないので、品詞などを決める以前に単語境界を決定する必要がある

## 形態素解析

雨に降られた。

雨	に	降ら	れ	た
名詞	助詞	動詞	助動詞	助動詞
一般名詞	格助詞	ラ行五段	連用形	終止形
あめ	に	未然形	れ	た
		ふら	れる	た
		降る	れる	た
		ふる		

## 形態素解析

- 形態素解析をすると可能になることの例
  - 助詞の「と」のみ検索、動詞の「いる」のみ検索
  - 名詞のあとが「すぎる」が後続しているところのみ検索
  - 利用可能なタグによっては「和語のみ検索」「アクセントが平板型の単語のみ検索」「言い間違いのみ検索」なども可能に

## 形態素解析

- 英語の品詞タグ付けツール
  - [CLAWS tagger](#)
  - GoTagger

## 形態素解析

- 日本語の主要な形態素解析ソフトウェア
  - ChaSen (茶釜)
  - JUMAN
  - MeCab
- 単独では使いこなすのが難しい。これらのツールを簡単に使うための便利なツールがある

## 簡単に使える 形態素解析ツール（日本語）

- WinCha
- Chamame

## 統語解析（係り受け解析）

- CaboCha
- KNP
- ただし、形態素解析と比較してもエラーが多いうえ、手作業による修正も難しい

## コメントシート

- 質問・感想を書いてください。