NICE.tips

学習者コーパス NICE, NICER, NICEST を使った分析の際のちょっとしたこと

・サンプルデータとスクリプト
・keyness

・CHILDES の CHAT フォーマット形式のデータ（テキストファイル）から、
　・データの本文部分だけを抜き出し（ヘッダー部分は削除）
　・行頭の話者記号（*JPN...:\t もしくは *NS...:\t）を削除し
　・全部小文字にして、
　・句読点・スペースを削除し（英数文字のみに）
　・単語の並びを返すスクリプト

```
nice.body <- function(){
  lines.tmp <- scan(choose.files(), what="char", sep="\n")
  data.tmp <- grep("\\*(JPN|NS)...:\t", lines.tmp, value=T)
  body.tmp <- gsub("\\*(JPN|NS)...:\t", "", data.tmp)
  body.tmp <- body.tmp[body.tmp != ""] # 空行を削除。
  body.tmp <- tolower(body.tmp)
  word.tmp <- unlist(strsplit(body.tmp, "\\W+"))
  return(word.tmp)
}
```