

はじめに

- 音声分析ソフトウェアPraat (Boersma & Weenink 2022)をコーパス構築用ツールとして拡張(Praatスクリプトを使用)
- tsvファイルを直接Praatで読み書きできるようにする
→ TextGrid (Praatアノテーション形式)ファイルが不要となる

「子ども日常会話コーパス」の作業環境

- 転記アノテーションには、PraatとELANを使用(ELANはここでは取り上げない)
- tsv形式ファイルでも転記を保持(それをテキストエディタ上で修正することもあり)
- 転記ファイルはGit(バージョン管理システム)で管理
- Windows/Mac OS

拡張機能

TextGrid形式

Praat独自形式

文字列処理しにくい

差分わかりにくい

標準機能で読み書き

tsv(タブ区切り)

一般的形式

文字列処理しやすい

差分わかりやすい

拡張機能で直接読み書きできるようにする
(TextGridをファイルとして保持しておく必要がなくなる)

Praat TextGridEditor

(tsv形式での保存は、そのためのスクリプトをEditorのLog Script 4に割り当てて実行)

Praat Object Window

独自メニュー

対象の会話を入力(パス等は自動補完)

対象音声ファイルを選択(ローカルかサーバーか)

(内部でtsvからTextGridを生成)

独自ウインドウ

Run script: kodomotsv2Praat

*** AUTO MODE ***

tsvFile: Y002_006

Note: extension (.tsv) not required

target sound: ☐ in the same folder as the tsv ☐ in the server

*** MANUAL MODE ***

tsvFullPath:

soundFullPath:

Note: the absolute paths and the extensions required

Standards Cancel Apply OK

参考文献

小磯他(2022)「子ども日常会話コーパスの構築」『言語資源ワークショップ2022』、西川(2018)『日本語日常会話コーパス』構築におけるPraatの利用』『言語資源活用ワークショップ2018』142-147.

関連URL

Praat <https://www.fon.hum.uva.nl/praat/> ELAN <https://archive.mpi.nl/tla/elan>

謝辞

本研究は、国語研究所共同研究プロジェクト「多世代会話コーパスに基づく話し言葉の総合的研究」の成果である。