

The NICT Japanese Learner English (JLE) Corpus

https://alaginrc.nict.go.jp/nict_jle/index.html

概要

- ・ 1,281 名分
- ・ 英語インタビューテスト 15 分
 - ・ 自己紹介・イラスト描写・ロールプレイ・イラストの説明
- ・ 9 段階評価
- ・ 167 ファイルにエラータグ
- ・ 母語話者コーパスあり

データ

ダウンロード

https://alaginrc.nict.go.jp/nict_jle/index.html#download

フォーマット XML 形式

- ・ 書き起こし・基本タグ付与ガイドライン

https://alaginrc.nict.go.jp/nict_jle/src/readme_transcription.pdf

全体構成

発話内タグ

Plain Text 化（学習者の発話部分）

必要な情報

<SST_level>4</SST_level>

 で始まる行

タグの削除