泰國學生Praat多層次華語學習語料庫處理作業程序

[] 作業程序 \*\*\* 人工程序 >>> 自動化程式程序

以下黃底為主要人工程序，請老師幫忙審查；

**粗體為新增自動化程式程序，紅字請博士幫忙評估是否可行**

[1] \*\*\*Praat逐字稿+斷詞+Interval標註 (部分語料有第3、4層，目前先保留)

|  |  |  |
| --- | --- | --- |
| Tier | Title | Content |
| 1 | Word/teacher | 斷詞逐字稿 |
| 2 | Word/student | 斷詞逐字稿，聲調偏誤標註 ^ ，發音偏誤標註 \* |
| 3 | IU/tone | 目標詞彙聲調 |
| 4 | EU/tone | 發音人詞彙聲調 (標註**實際**聲調) |

[2]\*\*\*標註每句語流的斷句Interval (部分語料有第5、6層，目前先保留)

|  |  |  |
| --- | --- | --- |
| Tier | Title | Content |
| 1 | IU/teacher | 內容空白只有Interval |
| 2 | Word/teacher | 一開始保留斷詞逐字稿以便輸入以下分層，最後對照**知網**詞性標註 |
| 3 | IU/student | 內容空白只有Interval |
| 4 | Word/student | 一開始保留斷詞逐字稿以便輸入以下分層，最後對照**知網**詞性標註 |

[3]>>>將Word層斷詞詞彙複製至IU層，移除原Word層Interval，只保留[2]增加的Interval，

再加(Title欄粗體)層次如以下

|  |  |  |
| --- | --- | --- |
| Tier | Title | Content |
| 1 | IU/teacher | **單句語流逐字稿，自Word層複製詞彙至IU層，僅保留[2]增加的語流Interval** |
| 2 | **English/teacher** | **單句語流英文註譯** |
| 3 | **Word/teacher** | **對照知網詞性標註** |
| 4 | **IU/phone** | **詞彙IPA (對照Tier3，自動標出IPA，每個符號間加Interval**，人工校正切齊**)** |
| 5 | **IU/syllable** | 詞彙音節結構CGVN (對照Tier3) |
| 6 | **IU/tone** | 發音人詞彙聲調 (對照Tier3) |
| 7 | IU/student | **單句語流逐字稿，自Word層複製詞彙至IU層，僅保留[2]加的語流Interval** |
| 8 | **English/student** | **單句語流英文註譯** |
| 9 | **Word/student** | **對照知網詞性標註** |
| 10 | **EU/phone** | **詞彙IPA (對照Tier9自動標出IPA，每個符號間加Interval，人工校正切齊)** |
| 11 | **EU/syllable** | 詞彙音節結構CGVN (對照Tier9) |
| 12 | **EU/tone** | 發音人詞彙聲調 (對照Tier9) |
| 13 | **EU/type** | 對照Tier9的Interval，**[1]標註的符號移到對應Interval，原Tier9符號不保留** |
| 14 | **Subject** | 對照Tier7的Interval |

[4]\*\*\*對齊IPA的Interval、標註EU層、語誤類別及語者資訊

完成學習語料庫標記；前6層為老師層次；後8層為學生層次

|  |  |  |
| --- | --- | --- |
| Tier | Title | Content |
| 1 | IU/teacher | 單句語流逐字稿，自Word層複製詞彙至IU層，僅保留[2]增加的語流Interval |
| 2 | English/teacher | 單句語流英文註譯 |
| 3 | Word/teacher | 斷詞逐字稿 |
| 4 | IU/phone | 詞彙IPA (對照Tier3，自動標出IPA，每個符號間加Interval，**人工校正切齊**) |
| 5 | IU/syllable | 詞彙音節結構CGVN (對照Tier3) |
| 6 | IU/tone | 發音人詞彙聲調 (對照Tier3) |
| 7 | IU/student | 單句語流逐字稿，自Word層複製詞彙至IU層，僅保留[2]加的語流Interval |
| 8 | English/student | 單句語流英文註譯 |
| 9 | Word/student | 斷詞逐字稿 |
| 10 | EU/phone | **標註發音人實際IPA**，**人工校正切齊** |
| 11 | EU/syllable | **標註發音人實際CGVN** |
| 12 | EU/tone | **標註發音人實際聲調** |
| 13 | EU/type | **將符號改為對應語誤類別** |
| 14 | Subject | **#[語者編號]C[男M女F]\_[情緒類別] (一定要有底線)** |

範例：

