# Tentative workflow for Shawi corpus

There are two scenarios to be considered: (a) texts exist as word and (b) texts have to be transcribed from wavs.

## Option a

|  |  |
| --- | --- |
|  | Options |
| Word 🡪 (XML) |  |
| Alignment texts with wavs | Exmaralda/Elan |
| Conversion to tei |  |

## Option b

|  |  |
| --- | --- |
|  | Options |
| Transcription of audios |  |
| Conversion to tei | Exmaralda/Elan |
|  |  |

## Main processing steps

|  |  |
| --- | --- |
|  | Options |
| Tokenisation |  |
| Annotations |  |