Guidelines for Annotating Hindi Speech Data Using Vāgyojaka Tool

1. **Transcript Editing**: Correct any errors in the transcript, either through additions, deletions, or substitutions.
2. **Transcribing Numbers**: If the number is being spoken, transcribe it using words (as spoken), but not digits.
3. **Audio Speed Adjustment**: If the speech rate(number of words per minute) in the audio is too fast, reduce the speed to ensure accurate listening.
4. **Timestamp Management**: For audio chunks exceeding 15 seconds, insert a new timestamp at a logical line break.
5. **Validating Timestamps**: To ensure timestamp accuracy, hold the Ctrl key and left-click on the desired line. This allows you to verify the alignment between the text and the corresponding audio for that line. If there is a mismatch, you can adjust by editing the timestamp and recheck for accuracy.
6. **Speaker Boundary Validation**: Ensure speaker boundaries are accurate. If incorrect, adjust the timestamps to separate different speakers precisely.
7. **Handling Extended Silence, Music, or Noise**: Mark any extended silence, music, or significant noise that obscures the audio with an empty line containing three dots . . . and put the associated timestamp.
8. **Code-Mixing Transcription**: There may be sentences mixed with English in between. Transcribe English words in Roman letters when code-mixing occurs between English and the given Indian language.
9. **Handling disfluencies**: If there is an overlap between two speakers (usually with agreeing words or disfluencies such as हा हा, अच्छा अच्छा, ok ok, etc.), then put those words of the secondary speaker's words inside square brackets.
   Example : for this audio [link : https://drive.google.com/file/d/1EMXPiKujTAAQgE8qIbXSFblgZ6FeLG0U/] file, transcript should look like this :-
   जैसे की बहुत अधिक से अधिक अगर धूप पड रही हो [हा हा हा], सूखा पड रहा हो [अच्छा अच्छा] तो मान लीजियेगा आप तीन से चार पानी लगाना होगा [अच्छा अच्छा]
10. **File Upload**: After completing the annotation for each file, upload the saved XML file using the provided Google Form.