ChatGPT

Plan for a Reading Transcript Processing Model

Overview of the Task

We need to analyze a reading session where a **student (Reader)** reads a given **passage text** aloud with guidance from a **tutor (Agent)**. The goal is to identify **reading miscues** – deviations or disfluencies in the student's reading – and derive **Key Performance Indicators (KPIs)** that reflect the reader's fluency and accuracy. By comparing the **transcript of the reading** (a dialogue between Reader and Tutor) against the **original passage text**, we can detect where the student omitted words, added extra words, misread/substituted words, repeated words, self-corrected errors, or hesitated/paused. These events will be annotated in a "cleaned" version of the passage using tags, and we will calculate quantitative KPIs (like error counts and reading speed). This helps an AI Reading Assistant measure the reader's performance and improvement over time.

Key Markers of Reading Miscues in Transcripts

Research in reading assessment and automated tutors shows several common types of miscues (errors or disfluencies) to watch for [1] [2]. In the transcript text, we will look for markers of the following key events (and any others relevant):

- **Omissions (Skipped words):** The reader fails to read a word that is in the passage (also called deletions). Omitting a word is considered a reading miscue [3]. In the transcript, an omission might be inferred if a passage word has no corresponding spoken word. For example, if the passage is "The **cat** is orange" but the transcript shows the reader skipped "cat," this would be an omission. We might tag the passage word with an `[omission]` marker to indicate it was skipped. Omissions are less frequent than some other errors but do occur [1].

- **Additions/Insertions (Extra words):** The reader inserts a word that is not in the passage text. This could be an extra word spoken, sometimes a repetition or a random addition. Insertions are also a type of miscue [3]. In transcripts, an insertion appears as a word or phrase the student said which does not match any word in that position of the passage. For example, if the student read "The cat is **very** orange" when the passage was "The cat is orange," the word "very" is an insertion. We could tag this in the cleaned passage as `[insertion]` (optionally including the inserted word). Insertions (additions) tend to be less common than substitutions or repetitions [1].

- **Substitutions (Misread words):** The reader reads a word incorrectly, saying a different word than what's in the text. This is one of the most common miscues [1]. For example, if the passage says "The cat is **orange**" and the student read "The cat is **apple**," that is a substitution error. In the cleaned passage, we mark where the wrong word occurred with a `[substitution]` tag, and ideally note that the tutor or student corrected it to the proper word. Substitutions often indicate the reader is guessing or mis-decoding the word. We will treat any wrong word (real word or not) that replaces the correct word as a substitution. (If the student's transcript shows a mispronunciation that the speech recognizer recorded as a different word, it also appears as a substitution in text form.)

- **Repetitions:** The reader repeats a word or phrase, usually as they self-monitor or struggle. Repetition is very common in oral reading [1]. In the transcript, a repetition is evident if a word or part of the sentence is said more than once in a row. For example, the transcript might show: *Reader:* "The... The cat is orange." The word "The" was repeated. We will tag the passage text at that point with `[repeats]` to indicate the repetition. (Repetitions are often not counted as **errors** in formal assessments since the student eventually reads the correct word; they are seen as a **monitoring behavior** [4], but we still want to track them as they affect fluency.)

- **Self-Corrections:** The reader corrects their own mistake, rereading a word or phrase correctly after initially making an error [5]. In transcripts, a self-correction might appear as the student saying a word wrong, then immediately saying it again correctly. For example: *Reader:* "The cat is **apple...** (pause) **orange**." Here the student said "apple" wrong and then self-corrected to "orange." We tag the point of correction with `[self-correction]` in the cleaned passage (usually right before the finally corrected word). Like repetitions, self-corrections are a positive sign of the reader's monitoring and are **often not counted as errors** in accuracy scoring [6] [7], but we will track them to understand the child's reading behavior.

- **Hesitations/Pauses:** The reader hesitates, pauses, or stretches sounds due to uncertainty. Hesitation is considered a disfluency in reading miscues [2] and can indicate processing difficulty [4]. In transcripts, hesitations might be marked by explicit filler words or sounds like "um", "uh", "...", or an obvious long silence before a word. For example, *Reader:* "The cat is ... orange" or "The cat is **uh** orange." We will tag such occurrences with `[hesitation]` in the cleaned passage to mark where a significant pause or filler happened. (If the transcription service provides timestamps, a long gap between words can also signal a pause. If the service explicitly transcribes fillers like "um", we will treat those as hesitation markers.)

- **Student Questions:** If the reader asks a question to the tutor (agent) during reading, it indicates they are struggling and seeking help (for example, asking how to pronounce a word). In the transcript, this appears as a question from the Reader, often containing a question mark or phrased as a request. For instance: *Reader:* "How do I pronounce this word?" or "What's this word?" This will be captured and tagged in the passage as a `[question: "..."]` at the appropriate point. We include the content of the question in the tag for clarity (e.g. `[question: "how do I pronounce this word?"]`). This gives context on what help was requested.

- **Tutor/Agent Interventions:** When the AI tutor (agent) intervenes, for example by providing the correct word or a hint, it signifies the student needed assistance. In the transcript, any utterance by the Tutor (Agent) during the reading of the passage is typically an intervention. Commonly, the agent might supply the correct pronunciation of a word the student got wrong or stuck on. We will insert an `[agent_input]` tag in the cleaned passage at the point where the tutor stepped in. For example, if the tutor had to say the word "orange" for the student, we would mark that point as `[agent_input]`. Tutor-supplied words are considered a category of miscue ("teacher-supplied word") in reading assessments [8], and they count as errors from the student's perspective because the student could not read the word independently.

- **Other Possible Markers:** We should remain aware of other phenomena that might appear. For instance, **reversals** (saying words or letters out of order) and **ignoring punctuation** (not pausing at a period) are sometimes noted in miscue analysis [8]. In our context of a single passage sentence, these might be less common, but if a child, say, reads *"on"* instead of *"no"* or swaps word order, it could be tagged (perhaps as a substitution or separate tag like

`[reversal]`). Punctuation ignorance (reading through a full stop without pause) is more of a prosody issue; our focus is primarily on word-level miscues and hesitation, but such behavior can be noted if needed (e.g., not pausing might be inferred if no hesitation at a period – however, since we are likely dealing with one passage at a time, this may not be prominent).

Each of these markers in the transcript provides insight into the reader's fluency. **Substitutions, omissions, and insertions** are clear **errors** (words not read correctly) [9]. **Repetitions and self-corrections** are usually seen as the reader monitoring their reading and are often **not penalized as errors** (since the correct word is eventually produced) [4] [7], but they do slow reading down and are important to track. **Hesitations** likewise are not "wrong words" but indicate lack of fluency. By tagging these in the passage text, we create a "cleaned" passage annotation that shows exactly where and what type of issues occurred.

Heuristics for Detecting Miscue Markers and KPIs

To extract these KPIs and annotate the passage, we will use a series of heuristic steps, leveraging both text alignment and pattern matching on the transcript:

1. **Align Transcript with Passage Text:** We compare the words spoken by the reader (from the transcript) with the expected words in the passage, in order. A robust way to do this is via a sequence alignment (similar to the Viterbi alignment used in reading tutor research [10] [3]) which pairs each spoken word to a corresponding passage word or marks it as an insertion. This alignment lets us detect:
2. **Omissions (Deletions):** Any passage word that the alignment finds no matching spoken word for (i.e., it was skipped) is an omission [3]. We will record an omission event for that word. In the cleaned passage, we can insert an `[omission]` tag to flag that the student did not read this word.
3. **Insertions (Additions):** Any extra words from the student that do not align to the passage text are insertions [3]. Each such word is an insertion event. We might not include the exact inserted word in the final passage (since it's not part of the original), but we will note an `[insertion]` tag at that point in the sequence. (Alternatively, we could include it like `[insertion: word]` for clarity – design choice.)

4. **Substitutions:** If a student's word aligns to a passage word but is different (and not just a small pronunciation variation), that is a substitution error [3]. Essentially the alignment will match the student's word to a specific target word position but they don't match exactly. We tag the passage at that word with `[substitution]`. Typically, after a substitution, either the student or tutor will correct it; the presence of a tutor correction right after can also confirm a substitution occurred.

5. **Scan for Repetitions:** Using the transcript text (especially the reader's utterances), detect any immediate repetitions. If the same word or phrase is spoken twice in a row by the reader (with no different word in between), mark it as a repetition. For example, if the transcript shows *Reader:* "The the cat…", we identify "The" was repeated. In the passage annotation, we insert `[repeats]` after the first occurrence of "The" to indicate the repetition on that word. (The alignment from step 1 might flag the second "The" as an insertion, but we specifically categorize it as a repetition for clarity, since the content is the same word repeated rather than a random extra word.) We increment a repetition count for KPI.

6. **Identify Self-Corrections:** We look for patterns where a word was initially read incorrectly and then corrected by the student without the tutor directly giving the answer. A self-correction often appears as the student immediately repeating a section of text correctly after an error. For example, transcript: *Reader:* “The cat is **apple**, orange.” Here “apple” was wrong and the student corrected to “orange” by themselves. The presence of an error followed by the correct word from the student indicates a self-correction. Another pattern is the student stopping mid-word or after an error, then trying again correctly. If the transcript has the student saying, “The cat is app… (pause) orange,” this also is a self-correction. We will tag `[self-correction]` at the point of correction in the passage (just before the correct word). We also log a self-correction count. (Note: If the tutor had to intervene with the correct word, we consider that *assisted correction*, not pure self-correction – that scenario involves an agent_input, described next.)

7. **Detect Pauses and Hesitations:** We examine the reader’s transcript for any signs of hesitation:

8. **Filler words:** If words like "um", "uh", "er" appear in the transcript from the reader, or if phrases are drawn out like "The c-c-cat…", these indicate hesitation. We tag `[hesitation]` in the passage at that position. For drawn-out or partial words, the ASR transcript might show repeated letters or fragments, which we can interpret as the student sounding out or hesitating on the word. We count each noticeable hesitation.

9. **Long pauses:** If we have timing information (e.g., from AWS Nova or another source), we can detect a silence longer than a certain threshold (say >2 seconds) during which the student is not speaking. A long pause before a word would also merit a `[hesitation]` tag before that word. If timing info is not directly available, a tutor intervention or student question often implicitly signals a pause occurred (the student stopped reading). We will use such clues as well (e.g., if the student suddenly asks a question or the tutor speaks up, a pause likely preceded it).

10. **Identify Student Questions:** Parse the transcript turns to see if the student (Reader) asked a question. This could be recognized by a question mark in the transcript or key phrases (like "how do I…", "what is this…", "can you tell me…"). For instance, *Reader:* "I don't know that word. What is it?" or "How do I pronounce this word?" are clear questions requesting help. When a question from the reader is detected, we tag it in the passage with `[question: "..."]` containing the question content. The tag position would typically be right where the student paused in the text to ask the question (before the difficult word). We also increment a counter for questions asked.

11. **Mark Tutor/Agent Interventions:** By analyzing the dialogue turns, we note any tutor (Agent) utterances that occur while the passage is being read. Typically, if the tutor speaks, it means the student needed help or feedback at that moment. Commonly the tutor might supply the correct word or prompt the student. For each tutor interjection, we insert an `[agent_input]` tag at the corresponding point in the passage. For example, if the tutor says the word that the student missed or mispronounced, the `[agent_input]` tag goes before the correct word in the passage annotation. Every such occurrence is counted as an agent intervention KPI. (In alignment terms, if the student made an error and did not correct it, the tutor-supplied word effectively “fills in” the gap – our tagged passage will show `[agent_input]` where the tutor provided assistance. This is akin to the **teacher-supplied word** category of miscue [8].)

12. **Compile the Cleaned Passage with Tags:** Using all the info above, we construct the passage text with inline tags. The original passage text will be preserved in order, and wherever an event happened, the appropriate tag(s) are inserted in square brackets. Multiple tags can appear at one location if several events coincided (for example, a hesitation followed by a tutor help on the same word). We ensure the tags clearly describe what happened. For instance:

13. If a word was substituted and the tutor helped, we might annotate: *"The cat is [substitution] [agent_input] orange."* (Meaning the student said the wrong word for "orange" and the agent had to input the correct word "orange").
14. If the student hesitated and asked a question before a word, with tutor help and then self-correction, it might look like: *"The cat is [hesitation] [question: \"how do I pronounce this word?\"] [agent_input] [self-correction] orange."*
15. If a word was repeated, e.g. the student said "The The cat...", we annotate: *"The [repeats] cat is orange."*

These inline tags make the **cleaned passage** easy to read with all miscues marked in context. (They are not part of the original passage, but indicate the reading process.)

1. **Calculate Quantitative KPIs:** Finally, we calculate various metrics from the aligned and tagged data:
2. **Total Words in Passage:** number of words in the passage text (for reference).
3. **Words Read Correctly:** how many words the student read correctly without omissions or substitutions (this can exclude words that were errors unless self-corrected). This can be used to compute accuracy.
4. **Total Errors:** count of all miscue events that are considered errors (typically substitutions + omissions + insertions + any agent-supplied words). Repetitions and self-corrections usually are **not counted as errors** in final accuracy, as long as the word was eventually correct [6]. So, an **accuracy rate** could be calculated as (Words Correct / Total Words). We can derive Words Correct = Total Words – (omissions + substitutions + insertions + words that required agent help).
5. **Count of Each Miscue Type:** We will output the counts for omissions, insertions (additions), substitutions, repetitions, self-corrections, hesitations, student questions, and agent interventions. These counts quantify the reading performance. For example, how many words were skipped, how many times the child had to repeat or self-correct, etc.
6. **Reading Speed (Words Per Minute):** We calculate the reading rate for the passage. If the total time taken to read the passage is known (from transcript timestamps or session data), we compute **words per minute (WPM)**. Typically, if using standard fluency measures, we use **Words Correct Per Minute (WCPM)** [11] [9] – meaning we count only the words read correctly in that minute. However, for simplicity we can report raw WPM as total words read (including errors) per minute, and possibly WCPM as well. For example, if the passage had 20 words and the student took 1 minute, WPM = 20. If 3 errors were made (thus 17 correct words), WCPM = 17. WCPM is a key fluency metric in reading assessment [11]. We should clarify which we output. We can include one or both; likely **"words_per_minute"** (implicitly WCPM if we subtract errors from total words).
7. **Self-Correction Rate:** an additional informative metric could be the ratio of self-corrected errors to total errors. For instance, if the student made 5 errors but self-corrected 2 of them, that's a good sign. We can output something like **self_correction_rate** = 2/5 = 0.4 (40% of errors were fixed by the student without help).
8. **Accuracy Percentage:** Another metric is accuracy % = (words read correctly / total words) * 100%. This is essentially the percentage of the passage read correctly on the first attempt (with perhaps credit for self-corrected words). This can be derived from the counts we have, and is useful to track improvement.
9. **Other**: We might also include **total reading time** (in seconds) if available, number of tutor interventions (already counted), and perhaps an **independent reading percentage** (how many words were read without any help or error). Given the context of improving reading, tracking if certain words were trouble (could list the specific words that were missed or asked about) might also be useful, but the prompt focuses on KPIs so a summary count is likely enough.

All these KPIs will be packaged into a JSON output along with the cleaned passage text, providing a structured result that the application can use to evaluate the reader's performance.

Output Format and JSON Schema

The output will be a **JSON object** containing the annotated passage and the calculated KPIs. The structure might look like this:

- `cleaned_passage` – A string of the passage text with inline tags marking miscues (as discussed). This shows exactly where each event occurred relative to the original text.
- `kpis` – An object with fields for each of the relevant metrics and counts:
- `omissions` – number of words omitted (skipped by the reader).
- `insertions` – number of extra words the reader added.
- `substitutions` – number of words misread/substituted.
- `repetitions` – number of times the reader repeated a word or phrase.
- `self_corrections` – number of errors the student self-corrected (this might overlap with substitutions count in terms of events, but we count a self-correction specifically when they fixed it themselves).
- `hesitations` – number of notable hesitations/pauses.
- `questions` – number of questions the student asked for help.
- `agent_interventions` – number of times the tutor/agent had to step in (e.g. provide a word or correction).
- `words_per_minute` – the reading speed computed (likely words correct per minute).
- `accuracy` – (optional) accuracy rate or percentage.
- `total_time` – (optional) total time taken, if known, which can be used for WPM.
- (Any other metric we find useful, such as `self_correction_rate` or similar, can be included as well.)

Here is an **example** JSON output for a hypothetical reading of the passage "The cat is orange." where the student had some difficulties:

```
{
  "cleaned_passage":
"The [repeats] cat is [hesitation] [question: \"how do I pronounce this word?
\"] [agent_input] [self-correction] orange.",
  "kpis": {
    "omissions": 0,
    "insertions": 0,
    "substitutions": 0,
    "repetitions": 1,
    "self_corrections": 1,
    "hesitations": 1,
    "questions": 1,
    "agent_interventions": 1,
    "words_per_minute": 30.0,
    "accuracy": 75.0
  }
}
```

Explanation: In this example scenario, the student repeated the word "The" (hence `[repeats]` after "The"), read "cat is ..." then hesitated on "orange" and asked a question ("how do I pronounce this word?") – indicated by the `[hesitation]` and `[question: "..."]` tags before "orange". The tutor

then provided help `[agent_input]`, and the student finally self-corrected and said "orange" correctly `[self-correction]`. The counts in KPIs correspond to these events. There were no omissions, insertions, or substitutions in this particular example (so those counts are 0). The repetition, hesitation, question, agent intervention, and self-correction each happened once (count 1). The *words_per_minute* is given as 30.0 (this would be calculated based on the time it took to read; for instance, if it took the student 8 seconds to read 4 words with those hesitations, that's 30 WPM). The *accuracy* is 75% in this example – since out of 4 words, one word ("orange") was not read independently (needed help), we might consider 3/4 = 0.75. (This accuracy definition can be adjusted based on whether we count a self-corrected word as correct; here we treated the needed help as an error for accuracy purposes.)

Note: The exact numeric values are illustrative. In a real scenario, we would compute WPM from timing and determine accuracy more precisely. Also, the schema is flexible – for instance, *words_per_minute* could be an integer, and *accuracy* could be given as a percentage or fraction. We will preserve whatever format is needed.

Prompt Design for Model (if using an AI/LLM)

If we plan to use a language model (like AWS Nova or GPT) to perform this analysis, we will craft a prompt that provides the transcript and passage and asks the model to output the JSON in the specified schema. The prompt should clearly instruct the model to align the transcript with the passage and insert the appropriate tags and counts. For example, a prompt could be:

System/Instruction: *"You are an AI reading assistant. Analyze the following reading session transcript between a Reader and a Tutor, and compare it to the given passage text. Identify any word omissions, insertions (additions), substitutions (misread words), repetitions, self-corrections by the reader, hesitations/pauses, reader questions, and tutor interventions. Mark up the passage text with inline tags [omission], [insertion], [substitution], [repeats], [self-correction], [hesitation], [question: "..."], [agent_input] at the appropriate locations to indicate these events. Then, output a JSON object with a "cleaned_passage" field containing the tagged passage, and a "kpis" field containing counts of each event type and relevant metrics (omissions, insertions, substitutions, repetitions, self_corrections, hesitations, questions, agent_interventions, words_per_minute, accuracy, etc.). Ensure the JSON format is correct."*

We would then **insert the passage text and transcript** in the prompt, for example:

- *Passage:* "The cat is orange."
- *Transcript:*
- **Tutor:** "Please read this sentence."
- **Reader:** "The... The cat is ... (long pause)... how do I pronounce this word?"
- **Tutor:** "\"orange.\""
- **Reader:** "\"orange.\""

The model, following the instructions and using the heuristics we outlined, should produce the JSON with the cleaned passage and KPIs as shown above. We will test the prompt with various transcripts to ensure the model reliably identifies all markers.

Conclusion and Additional Considerations

This plan combines straightforward **text alignment algorithms** with **pattern-based heuristics** to capture reading miscues. By leveraging known categories from reading research (omissions, insertions, substitutions, repetitions, self-corrections, etc. [1] [8]) and by tagging the passage text, we create a

clear visualization of the reading performance. The quantitative KPIs (like error counts and WPM) will help track the reader's fluency and accuracy improvements over time.

In summary, the processing model will provide both an annotated passage (qualitative insight) and numerical metrics (quantitative measures). This aligns with educational best practices, where teachers consider not just the final accuracy but also how the reader arrived there – e.g., **repetitions and self-corrections** are seen as signs of active self-monitoring [7], and **hesitations** signal where the student had difficulty [4]. By capturing all this, the AI Reading Assistant can give targeted feedback and measure progress (for example, a decreasing number of hesitations or an increasing WPM over successive readings would indicate improvement).

Finally, beyond the scope of the immediate question, we could consider extending the model to analyze **prosody** (expression, phrasing) in the future, or to identify specific phonetic difficulties. For now, the focus is on the core reading fluency indicators that we have detailed above. Each of these has been grounded in reading research and can be detected with the transcript + passage input using the described approach, providing a comprehensive view of the reading session's quality and the reader's performance [9] [8].

Sources:

- Lee, K. et al. *"Analysis and Detection of Reading Miscues for Interactive Literacy Tutors."* Coling 2004. (Defines common miscues like substitutions, repetitions, omissions, insertions, self-corrections) [1] [3].
- Mostow, J. *et al. "Predicting Oral Reading Miscues."* (Notes that miscues include repetitions, insertions, substitutions, omissions, and hesitations) [2].
- Hasbrouck, J. *"Understanding and Assessing Fluency."* Reading Rockets, 2006. (Explains WCPM calculation and which miscues count as errors; notes that self-corrected words, repetitions, and insertions are not counted as errors for scoring) [9].
- Hasenpflug, C. *"The Value of Repetition and Self-Correction in Reading Assessments."* (Lists nine miscue categories: omission, insertion, substitution, **reversal**, teacher-supplied word, repetition, ignoring punctuation, hesitation, self-correction, and discusses their significance) [8] [7].
- *Miscue Analysis* – Grokipedia entry. (Notes that repetitions and self-corrections are seen as monitoring behavior, not simply errors, and hesitations indicate processing difficulty) [4].

---

[1] [3] [10] Paper template for Coling 2004, Geneva
https://aclanthology.org/C04-1182.pdf

[2] ri.cmu.edu
https://www.ri.cmu.edu/pub_files/pub3/mostow_jack_2002_1/mostow_jack_2002_1.pdf

[4] Miscue analysis
https://grokipedia.com/page/miscue_analysis

[5] [7] [8] The Value of Repetition and Self-Correction in Reading Assessments | Chelsea Hasenpflug's bPortfolio
https://chelsearoseh.wordpress.com/2010/10/12/the-value-of-repetition-and-self-correction-in-reading-assessments/

[6] [9] [11] Understanding and Assessing Fluency | Reading Rockets
https://www.readingrockets.org/topics/assessment-and-evaluation/articles/understanding-and-assessing-fluency