

How To format Data for Transcript Explorer

Transcript Explorer supports two primary data formats: **plain text files** and **CSV files**. The choice of format depends on the level of detail you want to include in a transcript and the types of interactions that will be made available in the tool

1. Simple Text File Format

The simplest accepted format is a **plain text file (.txt)**, where:

- Each line represents a **single turn of talk**.
- The **first word** in each line is interpreted as the **speaker's name**.

Example (.txt format):

Alice: Hello, how are you?

Bob: I'm doing well, thanks!

Alice: Great to hear.

 This format is quick and easy but lacks timing data, limiting interactions with video.

2. CSV File Format (Recommended for Full Transcripts)


For richer transcripts and enhanced visualization options, use CSV files (.csv). A valid CSV file **must include at least these two labeled columns**:

- ✓ speaker - The name of the speaker for each turn.
- ✓ content - The transcript text for that speaker's turn.

Additionally, you can also include either or both:

- ✓ start - The start time of the utterance (a number in seconds or HH:MM:SS format).
- ✓ end - The end time of the utterance (a number in seconds or HH:MM:SS format).

speaker	content	start	end
Alice	Hello, how are you?	0	2
Bob	I'm doing well, thanks!	3	5
Alice	Great to hear.	7	9

 Including at least the **“start”** column is necessary to interact with video, which can be loaded as MP4 files in Transcript Explorer. <https://www.transcriptexplorer.org>