How to keep your sanity when preparing a transcript of an online interview for publication

Barbara Beeton

Abstract

Three interviews conducted during TUG 2021 online were transcribed and edited for publication in the conference proceedings. Accomplishing this to the desired quality proved far more difficult than anticipated. The reasons for this are presented here, along with lessons learned that might enable cooperation of future interview participants, both interviewer and subject, in making this process more straightforward and less painful for the editor.

Background

Over many years I've had direct or indirect experience with these transcription mechanisms:

- transcription by experienced court stenotypists;
- transcription directly from an audio recording (tape or computer, including cell phone);
- editing an auto-transcription from an online video service (Zoom and YouTube).

In most cases, I was present (either in person or online) at the original presentation, so I was familiar with the subject matter, or had at least heard it myself. In all but the first situation, I was the individual responsible for preparing the transcript for publication. It's not for the faint of heart.

On the assumption that most interviews these days are conducted online, or the result is posted and viewable there, most of what follows will be specific to that medium.

Overview

The TUG 2021 interviews were conducted electronically, over Zoom, with an additional transmission via YouTube, and recorded for future viewing. The fact that interviewer and subject, and in some cases attendees asking questions, were not in the same location raised some complications with respect to communication.

Only one of the interviewees was a native speaker of (British) English, as was one of the individuals involved in later discussion. The other participants were from varied linguistic backgrounds; all of them are fluent in English, but most accents were quite distinct from the U.S. English norm. A built-in complication was the subject matter, which was highly technical, and not all familiar to me.

Both online services provided auto-transcriptions as starting text. Neither was ideal, but the difficulties were quite different between the two.

The "automatic" transcriptions

Since participants in the Zoom thread had to be registered, their names were known, and were present in the transcript. A new "paragraph" was started with each change of speaker, with a blank line between two entries. If a speaker continued talking for a relatively long time, or the discourse was interrupted by a brief silence, the contribution might be broken by additional blank lines. These segments were numbered consecutively in the transcript, and for each segment, the starting and ending times (relative to the start of the file at 00:00) were given. Occasionally, at a transition, a word or two would be assigned to the wrong speaker, but in general, as long as more than one person wasn't talking at once, the speaker identification was accurate. The accuracy of the text, on the other hand, left much to be desired. More about that later.

The YouTube auto-transcript was quite different. The identity of the speakers wasn't known, and no attempt was made to mark a change of speaker. The text was presented like a "stream of consciousness" in strings of irregular length separated by blank lines. On my monitor, the edit windows are usually set to a width of 80 characters, and a run-on line is ended with a (meaningless) backslash, for an effective length of 79 characters per line. At least one "line" in one transcript was 88 lines long, or nearly 7,000 characters. No useful punctuation (except for an occasional period in a url). And the accuracy of the text was no better in general than what was provided by Zoom.

Aside from the "flow" and presence or absence of speaker identification, the content was far from identical. Both systems were equally unfamiliar with the specific technical environments represented by the three interviewees. The only way to obtain a script worthy of publication was to start with the video recordings and listen carefully. In that respect, there was not much difference (although the YouTube recording of one interview was lost when the session was not closed before 12 hours had elapsed).

Consistent lapses

The terms TEX and LATEX were spoken frequently throughout the interviews, but appeared almost nowhere in the transcripts. Instead, "tech" was a frequent substitution, as were "later", "late tech", and "late hack". Preliminary edit passes, searching for "tech" and "late", were effective in eliminating these

doi.org/10.47397/tb/42-3/tb132beeton-transcribe

misinterpretations, along with "TEX Live", "MiKTEX", "WriteLATEX", "ShareLATEX", and a few other compounds.

Company names were also consistent failures. "Overleaf" sometimes survived, but also occurred frequently as "overly"; another search, for "over", took care of that. "Fujitsu" didn't fare so well, and couldn't be cleaned until a comparison was made with the recording; I think my favorite miscue was "42" instead of "Fujitsu".

A few other terms occurred frequently enough to be attacked by a global search and explicit replace, but mostly they weren't discovered until the voice/text comparison. When such a case did arise, the word-for-word comparison was interrupted for a more targeted cleanup.

To be able to recognize and correct such lapses, it's highly desirable that the person editing the transcript be familiar with the speaker and the general subject of the interview; without this knowledge, it will likely be necessary to ask the interviewee to provide the needed corrections. Another problem area is people's names; here again a close personal knowledge of the interviewee is useful.

Mechanical considerations — know your equipment

Now it's time to attack the details of the text, so that the final transcript records the interview accurately. Unless you can type as fast as people talk, it will be necessary to stop the audio from time to time in order to catch up. Other reasons to stop are so that you can listen again to something that isn't clear the first time, or to verify a passage against another source.

How to reset the position in the audio file may be a puzzle. It took me several tries to find out that the back arrow on my keyboard could move the recording back in 15 second increments. This was far more efficient than trying to position the slider to align with the timing reported in the Zoom transcript, although using the timer was effective when the goal was to review a larger section. A few minutes of practice before starting can pay off handsomely later.

Details of the text

As noted earlier, the texts of the Zoom and YouTube transcriptions were not alike physically:

- Zoom: segmented, timed, speakers identified, sentence structure marked.
- YouTube: run-on, no speaker ID, no punctuation or case differentiation, interminable strings of words.

The textual content was far from identical as well. When a word (often a technical term) was unknown to the system, its representation in one text might be quite different from what appeared in the other. This turned out to be useful when the version chosen as the starter text made no sense, and the likely meaning couldn't be determined from the audio; it was usually possible to check what was in the other version and come to a sensible conclusion. For technical terms, YouTube was slightly better. However, homing in on the same passage was not easy; finding a matching term near the questioned material that could be used to search in a file that is just a jumble of words involves careful guessing.

Another weakness is the possibility that the interview participants are not skilled at this activity. An unscripted, unrehearsed interview may be littered with repetitions, meaningless interjections ("uh", "I mean"), and even an occasional interruption. While the primary goal of a published transcript is to record the content accurately, the result should also make sense if read without prior exposure to the event. Here is where editorial intervention is required. Consider carefully whether that "I mean" is just filler, or does in fact mean that the speaker is trying to clarify a particular point.

Occasionally, especially in an online interview, there may be unexpected interruptions. If an interruption is relevant to the topic of the interview, it can be worthwhile to include the details in the transcript. However, if it isn't relevant, and the interruption is short enough, it can be omitted; a longer interruption can be noted briefly in a [bracketed comment]. The choice depends on an estimation of whether noting or omitting would be more disruptive to someone reading the transcript and watching the interview at the same time.

Some details, finally, will require explanation or confirmation by the speakers themselves. An instance in the TUG 2021 interviews was when one of the interviewees referred to colleagues by only their first names. Since I don't know these individuals, direct contact was necessary. That said, it's always a good idea to ask a subject to review the transcript before publication to avoid surprises.

Examples of misinterpretation

As mentioned earlier, familiarity with the subject matter is a great advantage. I encountered this long ago, when observing the result of a symposium on mathematical physics, recorded by experienced court stenotypists. The transcribed phrase "brownie in motion" was determined to mean "Brownian motion". (The stenotypists would undoubtedly have produced

letter-perfect transcriptions for medical terminology.) Be warned.

Here is a not entirely random sample of terms that led to head-scratching in the TUG 2021 interviews.

spoken	Zoom	YouTube
a local editor	a low planetary	a local editor
BachoTEX	Bangladesh	bob attack parody
`biblatex`	the block back	people attack
CTAN	Stacy time and tse-tung	say see town sita ctan
Fujitsu	42	fujitsu
John Lees Miller	john these Miller	john lee's miller
LaTeX	late night	latex, later late act
Overleaf	overly	overly
Overleaf usage	obese usage	overleafs usage
quarantine	current time	current time
ShareLaTeX	show a tech	chelatec
TeX Live	deck live tech live	deck live tech live
WriteLaTeX	right lasik right later	right latex

Suggestions for a prospective editor

0a. Make sure your equipment and support software are in good working order. You should be a competent user of your editing software, and ideally, this software should be designed for use with TeX text files.

0b. If you haven't already listened to the session, do so, completely, before starting to work on the transcript. Becoming even slightly familiar with the individuals involved, their manner of speaking, and the subject matter is worth the time and effort.

1. Collect all recordings (audio or video) and text files in a convenient area. If the interview is part of a larger recording, remove any unrelated material from beginning or end, so that only the relevant content will be part of the working set.
2. Create a "working" copy with a new name. If more than one auto-transcript is available, choose the one that provides the text in the form closest to the final product. "Lock" all original files so that they can't be changed.
3. Analyze the text for consistently misinterpreted items. Fix these globally in whatever manner is most efficient and accurate. There may be a function available with your chosen editor, or a "search-and-replace" utility (such as `sed` on Unix).
4. Now you are ready to compare the text file to the recording. Review the mechanism for stopping the recording quickly and backing up just a few seconds.
5. Update the text. Listen to the recording while reading the corresponding text. Make corrections as necessary. If something is unclear, go over it again, and if it still doesn't make sense, refer to an alternate transcript if you have one, or leave a comment in the file for later attention, and keep a separate list of questions.
6. It was suggested by a reviewer that "waypoints" (timing indicators) be inserted in the transcript so that readers can find locations in the video if they want. Since this was not done in the transcripts that led to this article, I'm unable to offer specific suggestions on how to do this in a way that doesn't detract from the natural flow.
7. Process the file to "final" form, and reread it, preferably while listening to (and watching) the recording. Make additional corrections as needed, and clean up stammering, repetitions, etc., that would cause confusion for someone reading (but not watching) the interview for the first time.
8. Ask the participants to review the result, being specific about questions that arose during the editing. After approval, make any final corrections and process for the final release.

There are organizations that offer transcription services for a fee. It might be worth considering use of such a service. Even if the transcriptionist is not familiar with the technical details discussed in the interview, the resulting text is likely to be much closer to what was actually said than what is produced by Zoom or YouTube.

References to the three interviews, both printed transcripts and videos, that led to this article can be found at `tug.org/TUGboat/tb42-2`.

◇ Barbara Beeton
`https://tug.org/TUGboat`