Khamseen
Islamic Art History Online

On Closed Captions, Access, and Teaching: Khamseen Tales and Suggestions

Yasemin Gencer

Spoken and written language are the keys to communicating ideas and information, which makes both invaluable tools for teaching. Miscommunication thus can pose hurdles to the reliable and accurate transmission of information from teacher to student or from person to person.

As educators and mediators, members of *Khamseen: Islamic Art History Online* help produce short video presentations about a range of subjects related to Islamic art, architecture, and visual culture, including a new Glossary of key terms in the field. They take seriously both the matter of providing a wide variety of perspectives and clarity in communicating them in text, image, and speech. For these reasons we provide closed captioning for all our audio-visual presentations, which ensures the highest level of accuracy and accessibility for manifold audiences.

The decision to add closed captions to our video collection was a conscious one. It was precipitated by our desire to make our content accessible to hearing impaired and/or non-Anglophone students and teachers using our digital platform in both virtual and in-person learning environments. Yet, the process of creating captions revealed a greater need for intelligibility for a much broader audience than previously anticipated.

Most captioning efforts facilitated by video editing or video hosting services begin with convenient machine-generated text, which we proceed to edit for typos, mistakes, and punctuation. Overall, these services provide an accuracy rate of about 90-95%. While this is an impressive feat for auto speech recognition software (ASR), it is far from perfect since 5-10% inaccuracy represents a noteworthy percentage of mistaken words, missed expressions, and jumbled sentence structures.

Moreover, our realization that speech recognition technology cannot always understand what we are saying begs this critical question: how much of what we say in our classrooms is unintelligible to our students? While this might be a query for a different essay, it is something to consider when we deliver live lectures during which the spoken word quickly evaporates into thin air. It is here one second and gone the next, with no transcript left as a counterbalance.

Many of us have relied on a variety of supplemental materials such as textbooks, PowerPoints, and handouts, but these text-based resources still cannot fill the gap between written and spoken word. That is, even if a student "did the reading" and reviewed their handout, there is no guarantee that our students are able to fully hear what we say or understand what they read. And if they have not prepared ahead of time, their chances of getting lost in specialist terminology are even greater.

In-class challenges aside, the issue of legibility and comprehensibility is a matter that *Khamseen* has tackled actively by editing dozens of multimedia presentations over the course of our first year of operation. Our efforts have revealed a major blind spot in academic and pedagogical communication on a purely linguistic level. This is no more abundantly visible than in the editing process that takes place after captions are auto-generated for our videos. Let's elaborate a bit more.

Benefits of Captioning Videos

Closed captioning is great. If you have not yet thought a lot about it, we would like to take this opportunity to enumerate the reasons why pairing audio-visual content with text is a game-changer for everyone. Firstly, our current pandemic and the remote academic year of 2020-21 has opened many of our eyes to the need for accessible content for everybody. Indeed, this was one of the prime motivators behind the idea of launching *Khamseen: Islamic Art History Online*. The benefit of making information accessible and catering to the needs of all individuals, not just our most able-bodied students, is similar to ramps alongside stairs: they help all seekers on an upward trajectory. Ultimately, everyone benefits when the broadest set of audience demands and needs are taken into consideration and accommodated.

The anticipated audience for *Khamseen* videos are at present college-level students and interested adults. Our subject matter is Islamic art history, and currently our main language of communication is English. However, discussing the art, architecture, and visual culture of the Islamic world is a multi-lingual endeavor and most experts in our field require fluency in at least one other language, mainly Arabic, Persian, or Turkish. As a result, the terms we use to discuss the visual and material cultures of the Islamic world do not always have English equivalents. Even when they do, we strive to introduce our students to emic terminologies whenever possible to avoid presenting a strictly Western or Eurocentric approach or methodology.

These words, in addition to plain English, are often subject to misinterpretation by the speech recognition software used by auto-captioning services. A careful editing of closed captions is therefore of paramount importance.

The process of adding captions to videos involves ordering an automatically generated script, which is then checked and edited for optimal accuracy. The editing process is not difficult, but it is quite time-consuming. The frequent mistakes in the auto-captions reveal how certain words may also be "heard" by students. For instance, across several videos, the word "Qur'anic" was mistranscribed as "chronic." The term "mihrab," which refers to a concave niche found in the walls of mosques, was erroneously recorded in a number of ways, including: MATLAB, the throb, microbe, mic drop, minitab, mirror up, and meat rub. Some of the mistakes sound close to the original word, while others are wildly inconsistent. For instance, the name of the Umayyad caliph "al-Walid" was transcribed in a number of ways, such as "I believe," "all will eat," "all the lead," and "all while lead." While these examples seem harmless and humorous, less amusing ones also arise. In one instance, the geographic area of the "Hijaz" was mistaken for "Hijackers"—two words that hardly sound the same and bear very different meanings than that which was intended. A similar comparison can be made with the spoken word "wares" somehow auto transcribing as "weapons" and "chahar-taq" as "Jihad talk." Curiously, some of these AI errors mimic Islamophobic rhetorical slights and slippages.[1]

These mistakes are shortcomings of the imperfect speech recognition technology created and tested by a class of developers whose homogeneity and therefore lack of diversity is a known problem in the tech industry. This is the technology available to us for now, however, and human cognition and live editing resultant transcriptions are the only way to remedy errors. It is this laborious process that has opened our eyes to the broader issue at hand: namely, that speech comprehensibility and the proper transference of information inside and outside the classroom are not uniform from one student to another. As educators, we have a basic responsibility to our students: in a digital environment in particular, we must communicate clearly and captions prove a major equalizer for a mixed audience.

[1] Examples of this sort abound in the transcripts generated for *Khamseen* and for the author's own lectures on Islamic art history. Some stunning mishaps include: stepped pulpit becoming "sex pulpit," minbar becoming "minibar," calligraphic becoming "kill a graphic," inscriptional band becoming "inscriptional bomb," dikka becoming "deep gun," mihrab becoming "meet drugs," sacrality becoming "sexuality," Suhrawardi becoming "shooter awardee," Iranian becoming "Uranium," terrace becoming "terrorists," and Qur'ans becoming "crimes."

As an informal assessment exercise, we have been compiling a running list of auto-caption mishaps or "bloopers." And as the transcripts grew, so did the list of errata and so did Team Khamseen's collective realization that a lot more is at stake when it comes to what our students at all levels of comprehension could be (mis)hearing when we speak. If the speech recognition mishaps are any indicator, it is clear that even students who are fluent in English are probably hearing a range of things that we are not saying and perhaps missing out on up to 10% of the information communicated in speech. If this is the case, then what kinds of difficulties are our ESL students facing? And what about those among us who are hearing-impaired?

Written transcripts function, essentially, as buttresses to the spoken word. However, the benefits of captioning audio-video content do not end there. Captions and seeing speech in written form helps with note-taking and some transcripts are searchable, meaning a student can search for specific terms in a given video recording. This last advantage is currently not available with *Khamseen* videos, but is something that can be made available with podcasts and course lectures depending on the captioning service and/or interface being used. Nevertheless, many interfaces will generate a transcript that, once edited, can be exported and posted as a text-document alongside videos (if captioning services are not available).

On a pedagogical level, transcripts and closed captions also bridge spoken and written word in a way that reinforces proper spelling of both specialist terminology and basic English vocabulary through the visual repetition of words alongside proper pronunciation. The availability of a text-based rendition of audio-visual aids also may serve as a substitute for assigned readings and articles. With captions and transcripts, students can see how these terms look and be able to better recreate the sounds in speech that they hear, and the words in writing that they see. This reiterative process can bear cumulative benefits for student learning, writing, and conceptual retention.

Khamseen hopes to include foreign-language subtitles to videos so that our content can continue to be viewed by the widest possible audience. Its audio-visual presentations already attract a global audience and we are committed to making this content widely accessible to viewers worldwide. Multilingual translations—even more than transcripts—require substantial time and expertise. Currently, *Khamseen* has recruited a small team of volunteers who devote time and effort to translating the titles and synopses of the short-form presentations into a range of languages such as Arabic, Persian, Turkish, Spanish, French, and German. Translations of longer videos, however, will require a greater commitment of hours and funds.

How to Caption Audio-Video Content

Most universities will have at least one (if not several) types of video recording and editing software available for faculty, and these platforms will often provide some form of transcription or captioning services for videos and lectures. Most captioning/transcription journeys will begin with the very convenient and time-saving feature of the auto-generated texts that use speech recognition software to provide an initial—and rather raw—script template. For *Khamseen*, this usually ranges between 93-97% in accuracy. But the devil is in the detail, and we aim for a 100% accuracy rate by manually editing these captions.

Meme circulated on Twitter on June 14, 2021 created by Aaron Gabriel at aaron@philosophequeer.

The process of editing auto-generated captions is indeed time-consuming. Expect to spend 2-4 hours editing the captions/transcripts for every 15 minutes of content. It is a significant time investment, but one well worth the effort and commitment.

Programs and Software

University systems offer a range of options. The University of Michigan uses MiVideo for editing and hosting *Khamseen's* content. Some universities have subscriptions with Echo360, which is a recording and hosting interface that offers transcription-requesting and -editing services. Contact your university's Office for Teaching and Learning (or its equivalent) for a consultation and they will be happy to help you navigate the basics of these processes. You may also have an Adobe Creative Cloud subscription through your university where you can access the video editing software such as Premier Pro, which offers captioning functions as well.

Additionally, YouTube provides free subtitling/captioning services and a venue for hosting videos. This is a useful resource for anyone without an academic affiliation. It is also recommended for those with academic affiliations as the interface is user-friendly, free, and it also allows the storage of large video files as well. YouTube offers the option to keep videos private (unlisted) so that they are only accessible to those with whom you share the link. They will not appear in random searches. YouTube will auto-generate transcripts that you can both edit and export to other platforms as a separate file at any time. Conveniently, the University of Washington has compiled a very helpful guide to captioning videos for free; see: https://www.washington.edu/accessibility/videos/free-captioning/

Tips for Correcting and/or Enhancing Captions

In most cases, creating transcripts and captions for videos begins with ordering auto-generated text. The next step is to check this transcript for mistakes and edit it carefully. This is usually done through an interface that allows you to listen to the audio alongside a ticker of each text separated into 2-6 second intervals, much like subtitles in a movie. You thus edit captions line-by-line. Depending on the content of your lecture/video, expect to find mistakes and edit at least one in every 3 lines. Below is a list of anticipated corrections as well as tips to make your captions clearer and ways to sync spoken and written words with greater precision.

- Watch for homophones (words that sound the same but are different in meaning and spelling). In the case of Islamic art, the oft-occurring mistake we see is "Prophet" mistaken for "profit."[2] Tellingly, this error is often seen in students' written work as well.

- You also can place translated words, short side-notes, and tangents in parentheses. In fact, you can add parenthetical afterthoughts as a clarification for viewers even if you didn't necessarily say it. For example, for the spoken term "Shahnama," you can insert "(Book of Kings)" in the transcript, especially if it is not written in the slideshow.

- Alternately, you can clean up repeated words/phrases by omitting the repetition. This will make your expression clearer in text at times when it falls short of ideal in speech.

- Minor transition words can be added to the text to supplement and refine your speech.

- You can add emphasis by using punctuation like quotes, all caps, or asterisks. Unfortunately, many transcription/caption services do not allow for stylistic variants like italics or bold font.

- Most programs will accommodate special characters so long as you cut and paste them into the text box.

- You can decide which Arabic numerals to keep, and which to spell out depending on the intended impact or purpose of the numbers used. Captions are especially helpful in elucidating exact dates and measurements.

- Most transcription services will provide 2-3 lines for the captioned text. Since the timing of each portion of text will accompany the speech, you can delineate sentences or subjects by separating them into a second or third line in the ticker.

- You can "search and rescue" oft-mistaken words. For instance, if you notice that "Qur'anic" is showing up as "chronic" more than once, you can search and replace all the instances of "chronic" with "Qur'anic," thus saving time.

- Since one word can be mistranscribed in many different ways, the above "search and rescue" method does not save you from proofing the rest of the transcript. For instance, notice all the different ways "al-Buraq" was misheard by speech recognition software in a very short 3-minute video: All Brock, well Brock, Albert AAC, Oberon, Iraq, Barack, Albert rock, and I'll block. It is

[2] There are also hidden or near-homophones such as: ***or/our/are/were*** as well as ***in/an/and*** or even ***as/is*** that are commonly misheard by speech recognition software. Other common examples include: versus/verses, root/route, right/write, seen/scene, edition/addition, and wear/ware.

thus impossible to anticipate the breadth of mistakes AI is capable of generating. This symphony of errors vitiates against a swift editorial switcheroo.

- Occasionally proper nouns and multi-word terms are broken up between captions, which is a function of the random time slice created by the caption generator. Some interfaces will permit the adjusting of the timing of the caption intervals, in which case it is possible to modify the minutes to allow the whole name or term to appear uninterrupted in the same caption. This process, however, is time-consuming. And if done more than a few times, it will slow down the editing process significantly.

- Make sure that for special terminology, the spelling of the word in the captions matches how it appears in any accompanying visual aids, such as the PowerPoint slides used in the video. This will reduce confusion and reinforce how these new or foreign words are learned and retained by viewers and students.

- Check your transcripts for proper punctuation and capitalization. Auto-generated transcripts will provide a great deal of punctuation, some correct, others not. Also, some sentences are complete, others fragments. Adding or omitting punctuation will mould statements into correct grammatical form, adding further intelligibility.

- Finally, it is imperative to play back the video and check your transcripts/captions a second time. This is because fixing one mistake often distracts from additional errors that may be present on a single line and which, ergo, are overlooked. At times, mistakes will be caught much later, after days or even weeks of optical disengagement.

Cautionary Tales from Khamseen

Caption your videos and lectures and check them twice… or else:[3]

> Your “bismillah” may become “business law.”
>
> Your “İbrahim Müteferrika” may become “Miss America”
>
> Your “al-rawda al-mubaraka” may become “I’ll rub the Elmo Baraka”
>
> Your “Dalā’il al-Khayrāt” may become “dialogue hideout.”

[3] All of these examples are mistakes from auto-generated captions for *Khamseen* videos and the author’s own lecture transcripts.

Your “caliphate” may become “Kayla fit.”

Your “Mahmoud Mukhtar” may become “mama looked hard.”

Your “Dhu’l Fiqar” may become “dolphin car.”

Your “Shah ‘Abbas” may become “shot a boss.”

Your “Gunbad-i Qabus” may become “goomba the caboose.”

Your “Suleymaniye Mosque” may become “the silly ammonia mosque.”

Your “Nuruosmaniye” may become “the Euro US money.”

Your “Siyer-i Nebi” may become “CNN Maybe.”

Your “exegesis” may become “accept Jesus.”

Your “Souk al-Khamis” may become “so-called homies.”

Your “the Iskandarnama of Nizami” may become “extend our number of moles Army.”

Your “Haram al-Sharif” may become “how to machete.”

Your “Khanqah of Baybars” may become “fun café bars.”

Your “Persian bowls” may become “passion balls.”

Your “the Rassulid princes” may become “there are 3D printers.”

Your “Fath Ali Shah and his sons” may become “Fact that he shot his sons.”

Your “aniconism” may become “an icon is um.”

Your “Nizami’s Khamsa” may become “Miserables homesick.”

Your “al-Muqtataf” may become “and I walk the talk.”

Your “Bagh-e Fin in Kashan” may become “Bug, caffeine and caution.”

Your “al-isrā’ wal-miʿrāj” may become “throttling that Ouch.”

Your “Jabal al-Tariq” may become “Jabaal, alcoholic heartache.”[4]

[4] Ironically, not misheard as “Gibraltar.”

Citation:

Yasemin Gencer, "On Closed Captions, Access, and Teaching: Khamseen Tales and Suggestions," *Khamseen: Islamic Art Online,* published 3 November 2021.