This document is part of the Research and Innovation Action "European Live Translator (ELITR)".
This project has received funding from the European Union's Horizon 2020 Research and
Innovation Programme under Grant Agreement No 825460.

Deliverable D6.4

Report on Follow-up SAO Events

Rishu Kumar, Ondřej Bojar (CUNI)

Dissemination Level: Public

Final (Version 1.0), 28th February, 2022

Grant agreement no.	825460
Project acronym	ELITR
Project full title	European Live Translator
Type of action	Research and Innovation Action
Coordinator	Doc. RNDr. Ondřej Bojar, PhD. (CUNI)
Start date, duration	1st January, 2019, 36 months
Dissemination level	Public
Contractual date of delivery	Former: 31st January, 2022; Updated: 28th February, 2022
Actual date of delivery	28th February, 2022
Deliverable number	D6.4
Deliverable title	Report on Follow-up SAO Events
Type	Demonstrator
Status and version	Final (Version 1.0)
Number of pages	9
Contributing partners	AV, PV, CUNI
WP leader	PV
Author(s)	Rishu Kumar, Ondřej Bojar (CUNI)
EC project officer	Luis Eduardo Martinez Lafuente
The partners in ELITR are:	▪ Univerzita Karlova (CUNI), Czech Republic ▪ University of Edinburgh (UEDIN), United Kingdom ▪ Karlsruher Institut für Technologie (KIT), Germany ▪ PerVoice SPA (PV), Italy ▪ alfatraining Bildungszentrum GmbH (AV), Germany
Partially-participating party	▪ Nejvyšší kontrolní úřad (SAO), Czech Republic

For copies of reports, updates on project activities and other ELITR-related information, contact:

Doc. RNDr. Ondřej Bojar, PhD., ÚFAL MFF UK bojar@ufal.mff.cuni.cz
Malostranské náměstí 25 Phone: +420 951 554 276
118 00 Praha, Czech Republic Fax: +420 257 223 293

Copies of reports and other material can also be accessed via the project's homepage:

http://www.elitr.eu/

Contents

1 Executive Summary

This deliverable reports on the technical implementation details of the transcription and translation of events in 2021 which were carried out either fully remote, online or hybrid, namely: EMLCT Summer School & Demo Session, EUROSAI Board Meeting (held at Hotel Lindner, Prague), META-FORUM 2021, ELG Seminar 2021 and a standalone installation at Goethe Institute.

In Section 2, the main presentation view used in the events is briefly described. Then in Section 3 we briefly provide information about the events where we provided transcription and translation, with their technical details followed by the any possible new lessons learned at that event.

2 Presentation Techniques Overview

As described in D6.1 “Publishing Platform”, ELITR has been developing two web platforms for presenting the live output of simulatenous speech translation into text. The “subtitle” view is good for events with transcription only or events which accept a larger delay, otherwise the frequent updates of the subtitles make the output incomprehensible. The “paragraph” view implemented in our tool called Online Text Flow can be easily followed even with a relatively high flicker. All the events used English as the main language and we were testing translation into other languages. In the setting where the audience can follow the source, lower latency is preferred, so we opted to use Online Text Flow for these events.[1]

For the respective events, an endpoint was created on *quest*[2] machine at CUNI premises which provides paragraph view for the users. Optionally, the web server was secured with shared username and password to ensure only authorized users have the access. An overview of the platform has been presented in Bojar et al. (2021), Section 7.2.

quest runs Online Text Flow to present the output which in turn uses Hypercorn to create a server which is available to access via a web-interface. The continuously updated hypothesis of ASR output is first locally processed with Online Text Flow Events which ensures that update messages are aligned with (estimated) sentence boundaries. This step generates outputs with artificial timestamps, where the difference between timestamps for a single hypothesis is one of the three values 1, 10, or 100. The difference of 1 indicates that the input is still coming for that sentence, 10 denotes that the estimated sentence is complete but it still may be updated. This is denoted as dark gray colour on the paragraph view on *quest*. Lastly the difference of 100 denotes that the sentence has been finalized and no further updates are possible. This is shown in black font on the *quest* platform. In our jargon, we call this special format the “evented” output, to indicate that events correspond to sentences.

In the current setup, the presentation of the various languages is generally mutually independent: evented ASR output is sent to Online Text Flow as soon as available, so the source language transcription is visible even while MT workers are translating it. After the translations arrive the output is again immediately passed to the Online Text Flow server on *quest* to present it. This behaviour is expected to introduce a small delay in displaying of the hypothesis in the target language, compared to the source language transcription but we prefer this independent presentation to achieve the highest possible simultaneity.

3 Events

This sections describes the individual test events where we were automatically transcribing and translating the speech after the EUROSAI Congress which is described independently in D6.3. According to ELITR Description of Action, we were expected to run only one such event (described in Section 3.4) but we took advantage of several further occasions and tested our systems in diverse conditions.

Most of the events operated in English and exceptionally Czech. While we tried to solicit also our participation in German-spoken events, we were never successful; the organizers did not want to handle any possible additional burden or security concerns.

In all cases, we presented our job as a **test** rather than a service.

3.1 LCT Demo Session

As a demonstration of active research in the EMLCT consortium[3] universities, an online demo session was held at the EMLCT summer school (30.08.2021). Thanks to the multilinguality and

[1] Note that we did not have the plan or capacity to provide any tighter integration of our tools with the online conferencing tools used by the organizers, namely Microsoft Teams and Zoom. We only collected the sound, processed it with out systems and presented our outputs in Online Text Flow.

[2] https://quest.ms.mff.cuni.cz/elitr/demo

[3] `https://lct-master.org/contents_2014/consortium.php`

multinationality of the audience, the event also served as an opportunity to gain feedback from the students as well as local coordinators from the consortium. Apart from the demonstration, a few other talks were also transcribed and translated during the event.

In total, 40 participants were present for the demonstration and 65 participants interacted with the transcription and translation throughout the event. This event served as a great opportunity for us to check our domain capabilities for computational linguistics as the participants were students of Computational Linguistics. Overall, participants provided us with good feedback and the major drawback noticed was the delay between transcription and translation.

3.2 Installation at Goethe Institute

We provided a round the clock setup of our system as standalone service at Goethe Institute in Prague for Czech → German speech translation from September 2021 to January 2022, as part of an exhibition on artificial intelligence.

The setup consisted of capturing the sound at a booth at the venue and forwarding it to a dedicated virtual machine at CUNI premises via *netcat* to be transcribed and translated. We then processed the audio with CUNI's Kaldi-based Czech ASR worker, and then translated to English with CUNI's MT worker for Czech→English translation. This translation was then used as the input to translate to German using UEDIN's Rainbow worker (simultaneous translation into multiple languages).

This event motivated us to dockerize our individual systems, which is a useful step for deployment at different sites. In this case, we ran all the docker containers on the same machine.

The client side setup consisted of running a systemd service to process input. This helped us monitor when the system has been idle for a predefined time and restart it if any of the component has crashed. Since it was supposed to be an event running for months, monitoring the logs manually was not a feasible option, so we relied on getting informed by the organizers if any intervention from our side would be needed.

The setup was designed for the least possible maintenance and operator involvement. Thanks to remote connection, our assistance, if required, was easy to provide without being present at the venue. Unfortunately, at the end of the exhibition we learnt that in multiple cases, our setup did not work but nobody informed us. For any such future installation automatic full end-to-end tests will be necessary to ensure uninterrupted operation.

3.3 UAB meeting

On 04/10/2021, during our periodic User Advisory Board (AUB) meeting, we demonstrated the full transcription and translation service and also the new dictionary extension feature of KIT ASR technology (Huber et al., 2020).

In preparation for an event, it's very useful to be able to quickly update the ASR model with domain-specific terms that are expected to appear in the speeches, such as the names of the speakers or the name of the event itself. However adapting an ASR model can take several days, while with this functionality, as demonstrated during the demo, it only takes a few seconds.

During the demo it was real-time demonstrated that before the name of the ELITR project was not recognized, as well as the name of participants, then by inserting them in the dedicated memory table, they were recognized perfectly. An improvement of this setup along with an empirical evaluation will be described in the final deliverable on ASR, D2.2.

3.4 SAO-Lindner

In October 2021, the Supreme Audit Office of the Czech Republic (SAO) hosted a hybrid event for EUROSAI Governing Board. SAO decided that this event is suitable for the purposes of ELITR testing and to fulfil the required number of test events, as mutually agreed between CUNI and SAO, so we were invited to take part and provide automatic transcription and translation.

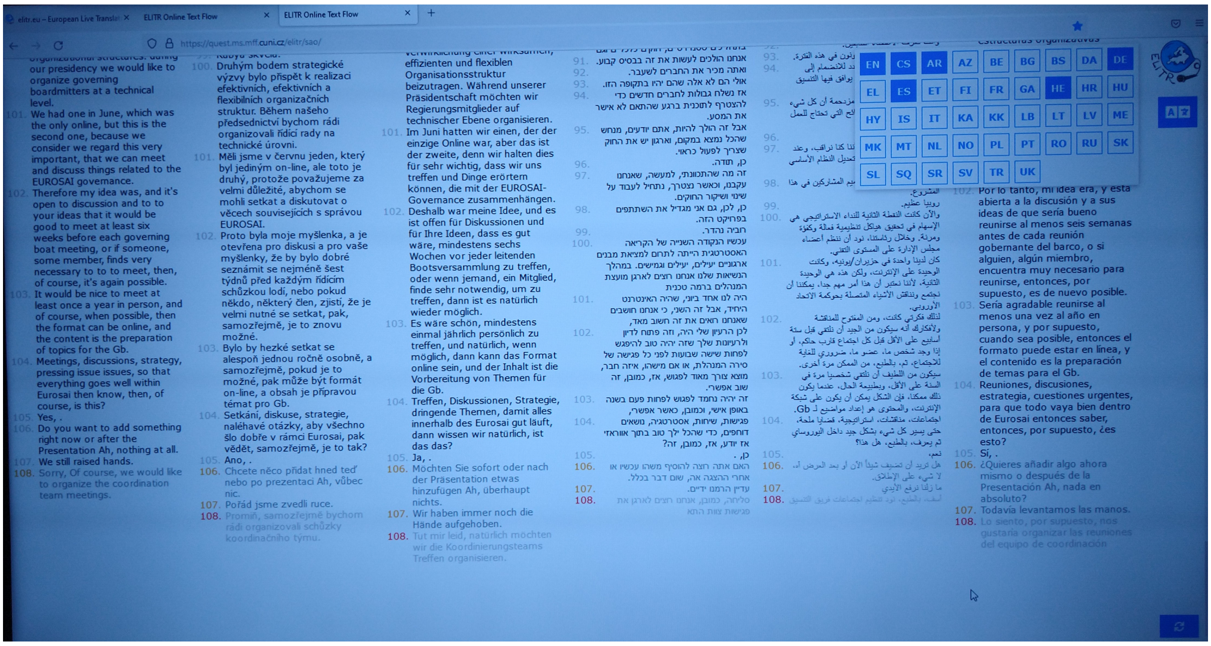

elitr.eu – European Live Translator
ELITR Online Text Flow
ELITR Online Text Flow
https://quest.ms.mff.cuni.cz/elitr/sao/
EN CS AR AZ BE BG BS DA DE
EL ES ET FI FR GA HE HR HU
HY IS IT KA KK LB LT LV ME
MK MT NL NO PL PT RO RU SK
SL SQ SR SV TR UK

Figure 1: Screen setup in EUROSAI event at Hotel Lindner in Prague

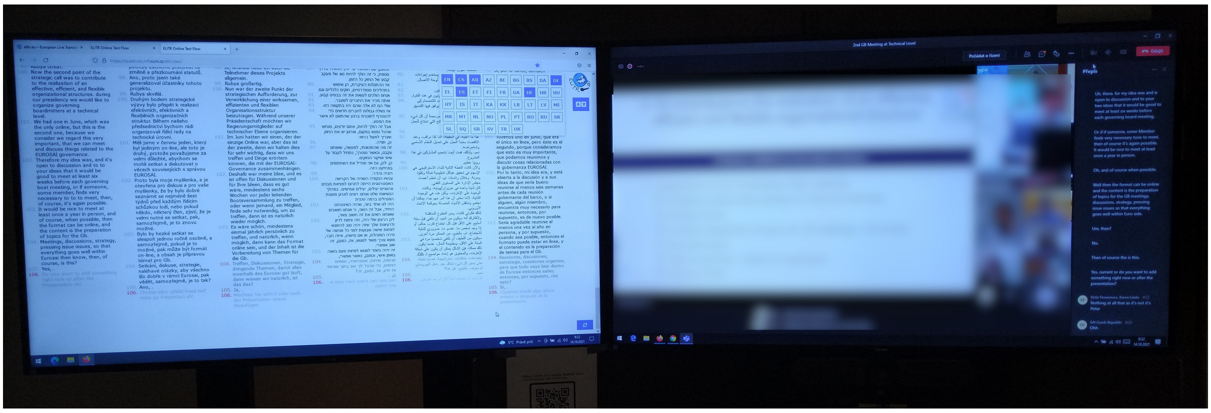
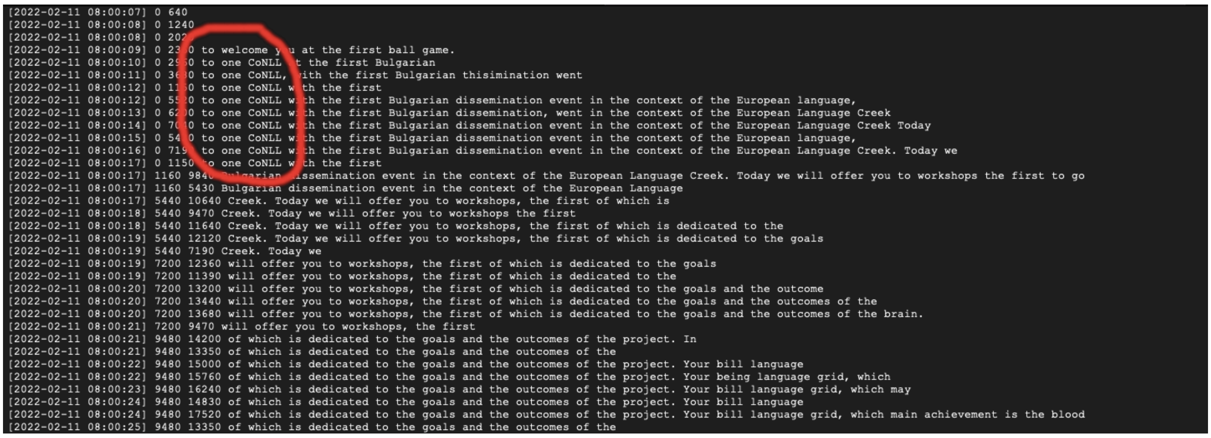


Figure 2: A side by side screen setup for the participants to assess the coherence of transcription and translation in comparison to the automatic transcription by Microsoft Teams. (The main content of the MS Teams session is blurred).

While delegates from some countries attended the event in-person at Hotel Lindner in Prague, other delegates joined via Microsoft Teams. Due to the hybrid nature of event, we decided to only receive one channel of sound, as recorded by MS Teams and kept another channel for a handheld microphone to demonstrate the real-time working of the entire pipeline.

Because the delegates actually did not need to rely on our systems, the were disturbed rather minimally. They were notified about the online service and about the password needed to access it by the moderator and they had a chance to see the system in operation on a big screen next to the coffee area. During breaks, we were standing next to the screen with a handheld microphone, so that anyone could test the service themselves.

Figure 1 shows the screen setup near the coffee booth at the venue. The inbuilt captioning from Microsoft Teams was also displayed on an second screen (Figure 2) to offer a quick quality comparison with our ASR. With real time domain-adaption discussed in Section 3.3, the participants were happy with the results of our systems for ASR as well as they found the translation to be coherent for most of the part.



```
[2022-02-11 08:00:07] 0 640
[2022-02-11 08:00:08] 0 1240
[2022-02-11 08:00:08] 0 2020
[2022-02-11 08:00:09] 0 2330 to welcome you at the first ball game.
[2022-02-11 08:00:10] 0 2950 to one CoNLL at the first Bulgarian
[2022-02-11 08:00:11] 0 3680 to one CoNLL, with the first Bulgarian thisimination went
[2022-02-11 08:00:12] 0 1150 to one CoNLL with the first
[2022-02-11 08:00:12] 0 5520 to one CoNLL with the first Bulgarian dissemination event in the context of the European language,
[2022-02-11 08:00:13] 0 6200 to one CoNLL with the first Bulgarian dissemination, went in the context of the European Language Creek
[2022-02-11 08:00:14] 0 7040 to one CoNLL with the first Bulgarian dissemination event in the context of the European Language Creek Today
[2022-02-11 08:00:15] 0 5440 to one CoNLL with the first Bulgarian dissemination event in the context of the European language,
[2022-02-11 08:00:16] 0 7190 to one CoNLL with the first Bulgarian dissemination event in the context of the European Language Creek. Today we
[2022-02-11 08:00:17] 0 1150 to one CoNLL with the first
[2022-02-11 08:00:17] 1160 9840 Bulgarian dissemination event in the context of the European Language Creek. Today we will offer you to workshops the first to go
[2022-02-11 08:00:17] 1160 5430 Bulgarian dissemination event in the context of the European Language
[2022-02-11 08:00:17] 5440 10640 Creek. Today we will offer you to workshops, the first of which is
[2022-02-11 08:00:18] 5440 9470 Creek. Today we will offer you to workshops the first
[2022-02-11 08:00:18] 5440 11640 Creek. Today we will offer you to workshops, the first of which is dedicated to the
[2022-02-11 08:00:19] 5440 12120 Creek. Today we will offer you to workshops, the first of which is dedicated to the goals
[2022-02-11 08:00:19] 5440 7190 Creek. Today we
[2022-02-11 08:00:19] 7200 12360 will offer you to workshops, the first of which is dedicated to the goals
[2022-02-11 08:00:19] 7200 11390 will offer you to workshops, the first of which is dedicated to the
[2022-02-11 08:00:20] 7200 13200 will offer you to workshops, the first of which is dedicated to the goals and the outcome
[2022-02-11 08:00:20] 7200 13440 will offer you to workshops, the first of which is dedicated to the goals and the outcomes of the
[2022-02-11 08:00:20] 7200 13680 will offer you to workshops, the first of which is dedicated to the goals and the outcomes of the brain.
[2022-02-11 08:00:21] 7200 9470 will offer you to workshops, the first
[2022-02-11 08:00:21] 9480 14200 of which is dedicated to the goals and the outcomes of the project. In
[2022-02-11 08:00:21] 9480 13350 of which is dedicated to the goals and the outcomes of the
[2022-02-11 08:00:22] 9480 15000 of which is dedicated to the goals and the outcomes of the project. Your bill language
[2022-02-11 08:00:22] 9480 15760 of which is dedicated to the goals and the outcomes of the project. Your being language grid, which
[2022-02-11 08:00:23] 9480 16240 of which is dedicated to the goals and the outcomes of the project. Your bill language grid, which may
[2022-02-11 08:00:24] 9480 14830 of which is dedicated to the goals and the outcomes of the project. Your bill language
[2022-02-11 08:00:24] 9480 17520 of which is dedicated to the goals and the outcomes of the project. Your bill language grid, which main achievement is the blood
[2022-02-11 08:00:25] 9480 13350 of which is dedicated to the goals and the outcomes of the
```

Figure 3: False positives in the ASR output because of the manually added words in the memory component

3.5 ELG Workshop Czechia-Slovakia

The European Language Grid (ELG) workshop for Czechia and Slovakia was held on 18th October 2021, moderated by colleagues at CUNI. The talks in the program were held in Czech and Slovak, with an exception of the talk by Georg Rehm (DFKI), whose talk was in English. We used this workshop as an opportunity to increase the language-technology domain related document in ELITR-testset and further improve our live adaptation mechanism as introduced in Section 3.3 and later refined in Section 3.6. We provided transcription from Czech audio and translation into English and Slovak using CUNI's ASR & MT workers.

3.6 ELG Workshop Bulgaria

The European Language Grid workshop in Bulgaria was an event held online on 11th Feb 2021. In this event, we tested two methods of live-adaptation of the English ASR by KIT.

In total, 18 users watched our subtitling and seven of them provided us with a detailed feedback on the working of our system. The first part of the workshop was held completely in English.

We used the manually corrected transcript for Georg Rehm's talk from the previous workshop as a basis to add new words in our dictionary.

One undesired effect of this new method was that we saw some false positives in the transcript, i.e. domain-specific terms instead of common words, such as "*coNLL*" instead of "*welcome*", see Figure 3. Whenever a false positive occurred, the operator then added the misrecognized common word to the dictionary to promote its selection again. The ASR workers then performed reasonably after such a manual correction.

We also noticed few names not getting finalized correctly even after manually getting inserted in memory component, "*Svetla Koeva*" (organizer of the event) being one of them while other names such as "*Katrin Marheinecke*" (DFKI) and "*Penny Labropoulou*" (organizer of the event). We learnt that even though the manual addition of terms such as name goes toward improving user experience, there are cases where the approach does not help. We assume that this primarily happens when the pronunciation of the name differs from what the model "assumes" to the be pronunciation based on the text similarity.

We received 7 responses from 18 participants who used the system throughout the event. Overall, the feedback provided a good impression of the research demonstration, and highlighted known issues such as misrecognition of terms which in turn produced funny translation. It was also mentioned that at times, some common English name was selected instead of a participant's name due to similarity.

4 Conclusion

This deliverable described our presence at test events that followed the EUROSAI Congress.

We used the presentation technique of Online Text Flow which displays the text as paragraphs of growing text. Throughout the events, we were doing smaller fixes and improvements, the primary one was the improvement of ASR recognition of special terms.

The events also tested as a form of dissemination of ELITR results to meeting participants, often colleagues in the area of computational linguistics and natural language processing.

References

Ondřej Bojar, Dominik Macháček, Sangeet Sagar, Otakar Smrž, Jonáš Kratochvíl, Peter Polák, Ebrahim Ansari, Mohammad Mahmoudi, Rishu Kumar, Dario Franceschini, Chiara Canton, Ivan Simonini, Thai-Son Nguyen, Felix Schneider, Sebastian Stüker, Alex Waibel, Barry Haddow, Rico Sennrich, and Philip Williams. ELITR multilingual live subtitling: Demo and strategy. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations*, pages 271–277, Online, April 2021. Association for Computational Linguistics. doi: 10.18653/v1/2021.eacl-demos.32. URL `https://aclanthology.org/2021.eacl-demos.32`.

Christian Huber, Juan Hussain, Tuan Nam Nguyen, Kaihang Song, Sebastian Stüker, and Alex Waibel. Supervised adaptation of sequence-to-sequence speech recognition systems using batch-weighting. In *Proceedings of the 2nd Workshop on Life-long Learning for Spoken Language Systems*, pages 9–17. Association for Computational Linguistics, 2020.