plain X – AI Supported Multilingual Video Workflow Platform

Carlos Amaral
Priberam
Lisbon, Portugal
carlos@priberam.pt

Peggy van der Kreeft
Deutsche Welle
Bonn, Germany
peggy.van-der-kreeft@dw.com

Abstract

The plain X platform is a toolbox for multilingual adaptation, for video, audio, and text content. The software is a 4-in-1 tool, combining several steps in the adaptation process, i.e., transcription, translation, subtitling, and voice-over, all automatically generated, but with a high level of editorial control. Users can choose which translation engine is used (e.g., MS Azure, Google, DeepL) depending on best performance. As a result, plain X enables a smooth semi-automated production of subtitles or voice-over, much faster than with older, manual workflows. The software was developed out of EU research projects and has recently been rolled out for professional use. It brings Artificial Intelligence (AI) into the multilingual media production process, while keeping the human in the loop.

1 Introduction

plain X has been built by and for the media industry, although its use can be extended to other sectors as well. A key driver is the growing amount of content which needs language adaptation., based on user or market needs, for enhanced accessibility or to comply with regulation. Feature development is based on the needs from Deutsche Welle (DW), a world broadcaster producing in over 30 languages. The plain X platform is the result of a partnership between DW as user partner and Priberam, a Lisbon-based natural language processing developer.

The platform simplifies the multilingual adaptation process to a large degree, enabling easy subtitling in source and any target language requirement. After a full year of preparation, we are currently rolling out the platform for daily use in Deutsche Welle. Some other organizations are trialing the tool. In the future the software will be available to others, based on a software-as-a-service subscription model.

2 Challenges

The concept for plain X originated from the need to produce more with less, i.e., to use automation in the production process, so media producers can increase the volume of certain target languages, distribute content in more languages, or use synthetic voice, allowing to reach more people in their own spoken tongue, including in specific African or Asian regions.

As DW produces content in so many languages, it is essential to cover as many languages as possible, in the best possible quality, through a combination of engines from carefully selected providers, for instance for transcription or translation. In plain X users can freely switch between different translation engines. The software allows for the inclusion of additional engines in the future.

As the tool was – and is – co-developed by user partner Deutsche Welle, direct access to user requirements and feedback is ensured. This revealed that integration with internal systems and customization is a must to reach the highest level of user acceptance.

3 Origin

plain X initially came out of the SUMMA multilingual media platform, funded by the European Commission's H-2020 project as a basic prototype for controlled transcription and translation.

This prototype was then further developed and funded through the Google Digital News Initiative projects speech.media and news.bridge.

Finally, Deutsche Welle, world broadcaster in need of such platform, and Priberam, a natural language processing developer, decided to turn the prototype into a scalable, fully operational multilingual platform for wider use, supporting the needs of broadcasters and other multilingual content producers. That was the birth of plain X, a platform which turns content from virtually any language into almost any target language.

4 Workflow

The task-based workflow is easy to use, but very powerful, offering editorial users the comfort of their familiar workflow, yet encompassing advanced automated technologies to support them in the creative process.

The first step is *ingestion* of content, be it video, audio, or text, with many input formats.

The next step for audiovisual content is *transcription*, through speech-to-text in the source language. That could be an end-goal, for instance for interviews.

This also allows for a primary output of automatically generated *source-language subtitles*, which can be used as open or closed captions.

The next step is automated *translation* to a selected target language, which can be post-edited to any level. Again, the translation can be an end-goal on its own, and used as input text for respeaking, for example. One file can be translated to multiple languages.

However, it can also generate automated *subtitling* in the same *target language*.

As a final step, the translation can be used for *voice-over*, by converting text to speech in the target language after selecting a synthetic voice.

Post-editing and review by colleagues can be added in every step, as required. Subsequently, other target languages can be added and produce equivalent content.

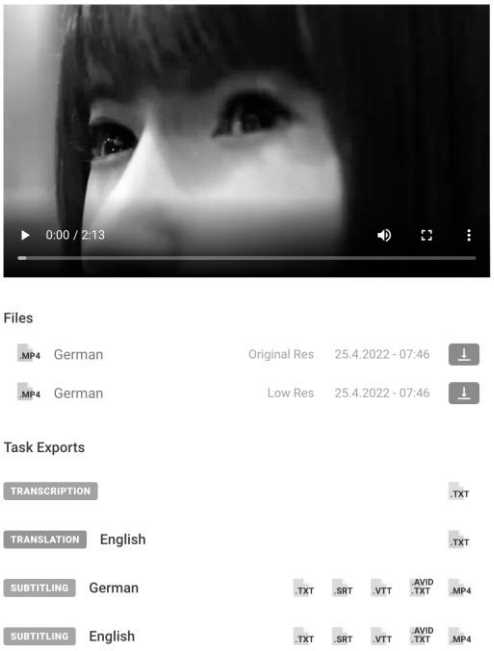

Figure 1: plain X Workflow Tasks

5 Integration

It was vital to integrate this tool into the existing workflow infrastructure at Deutsche Welle and to allow for customization. This meant connecting it to input platforms for a smooth ingestion, as well as output tools for an efficient post-production and publication in the company style and branding.

Subtitle templates help to prepare the output in a particular house format. Other customizations include library management and access, setting subtitling rules, assigning roles to users, keeping track of usage and billing. It is possible to create fully automated processes for subtitling.

Working directly in a user environment from the start, with user input and feedback at every stage, allowed us to build a user-oriented platform to support editors in their adaptation process with the help of AI, while minimizing the feeling of insecurity and threat coming from automated processing.

More enhancements are planned to cater for different use cases, improve the quality of the output and strengthen post-editing options.

Acknowledgments

This work has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 957017, Project SELMA (https://selma-project.eu/).