



School of Applied Linguistics
IMK Institute of Multilingual
Communication



**University of
Zurich^{UZH}**

An exploratory analysis of LLM-based machine-translated audio description scripts in the French-German language pair

Nicole Lüthi (ZHAW), Alexa Lintner (ZHAW), Lukas Fischer (UZH), Martin Kappus (ZHAW), Sarah Ebling (UZH)

ARSAD 2025 – Panel 2. Technology: research perspectives | 19 March 2025 | UAB, Barcelona



Outline



University of
Zurich^{UZH}



An exploratory analysis of LLM-based machine-translated audio description scripts in the French-German language pair

- Framework
- Related Work
- Methods
- Results
- Conclusions

Framework



University of
Zurich^{UZH}



- Master's Thesis supervised at ZHAW (Prof Sarah Ebling/Dr Martin Kappus and Alexa Lintner/Lukas Fischer)
- Part of the Innosuisse Flagship *Inclusive Information and Communication Technologies* (IICT)
(Grant Agreement Number: PFFS-21-47).
 - **PI:** Prof Sarah Ebling
 - **Project duration:** March 2021–February 2026
 - **Funded by:** [Swiss Innovation Agency](#) (Innosuisse)
 - **Aim:** to develop ICT that provide access services for persons with disabilities by radically changing the production workflows.
- 2024 Supertext Prize for the best Master's Thesis in *Professional Translation*



Overview



University of
Zurich^{UZH}



- **Topic:** Human evaluation of machine translated AD scripts in three different translation versions:
 - Version 1: translated with ChatGPT-4 [input: text only]
 - Version 2: translated with SwissADT [input: text and image]
 - Version 3: SwissADT translation post-edited by ChatGPT-4
- **Language pair:** French – German (high-resource languages) (ChatGPT)
- **Research question:** Which level of quality do LLM-based AD script translations achieve in the French-German language pair?
- **Source material:** *Passe-moi les jumelles*; (Episode: Pierres en équilibre – La Nuit; broadcasted on 19/3/2021 (RTS, 2021)
 - **Video duration:** 01:01:00
 - **AD duration:** 00:22:14 (36.45 %)
 - **AD segments:** 292 (24'743 characters)



Related Work



University of
Zurich^{UZH}



Jankowska, A. (2015). **Translating audio description scripts: Translation as a new strategy of creating audio description.** Peter Lang Edition.

- A three-stage study examined whether translating British AD scripts into Polish is a more time- and cost-effective solution while maintaining quality standards that satisfy target audiences.
- Findings confirm that AD script translation is not only feasible but also beneficial in terms of time efficiency, cost reduction, and quality preservation.

Fernández-Torné, A., & Matamala, A. (2016). **Machine translation in audio description? Comparing creation, translation and post-editing efforts.** *SKASE Journal of Translation and Interpretation*, 9(1), 64–87.

- English-Catalan language pair
- Lower effort in post-editing MT-translated AD scripts than in creating ADs script *ex novo*

Matamala, A., & Ortiz-Boix, C. (2016). **Accessibility and multilingualism: An exploratory study on the machine translation of audio descriptions.** *TRANS*, 20, 11–24.

- Catalan-Spanish language pair
- Evaluation of two MT engines: *Google Translate* (SMT), *Apertium* (RBMT) | Error rates: Google Translate – 42.22%, Apertium – 57.57%

Vercauteren, G., Reviere, N., & Steyaert, K. (2021). **Evaluating the effectiveness of machine translation of audio description: The results of two pilot studies in the English-Dutch language pair.** *Tradumàtica*, 19, 226–252.

- English-Dutch language pair
- Qualitative analysis – DQF-MQM error typology (cf. [Lommel, 2018](#))
- Relatively high error rates, particularly in the categories of *Accuracy* (mistranslation) and *Fluency* (grammar)

Methods



University of
Zurich^{UZH}



Mixed methods approach

Quantitative Analysis

- Source-based direct assessment
- Three participants (AD professionals)
- Pre- and post-study questionnaires
- 200 AD segments evaluated by each participant assigning *adequacy*, *fluency* and *usefulness* scores (1–5) to every AD segment without knowing how each translation version had been produced

Qualitative Analysis

- 20 AD segments (from translation version 3) have been analysed using the harmonized DQF-MQM error typology (cf. Lommel, 2018)
- Scores (critical–neutral) in six out of eight main DQF-MQM error categories (*Accuracy*, *Style*, *Verity*, *Fluency*, *Terminology*, *Other*) [→ 23 subcategories]

Translation Version 1 Prompt



University of
Zurich^{UZH}



Translation version 1 [input: text only]

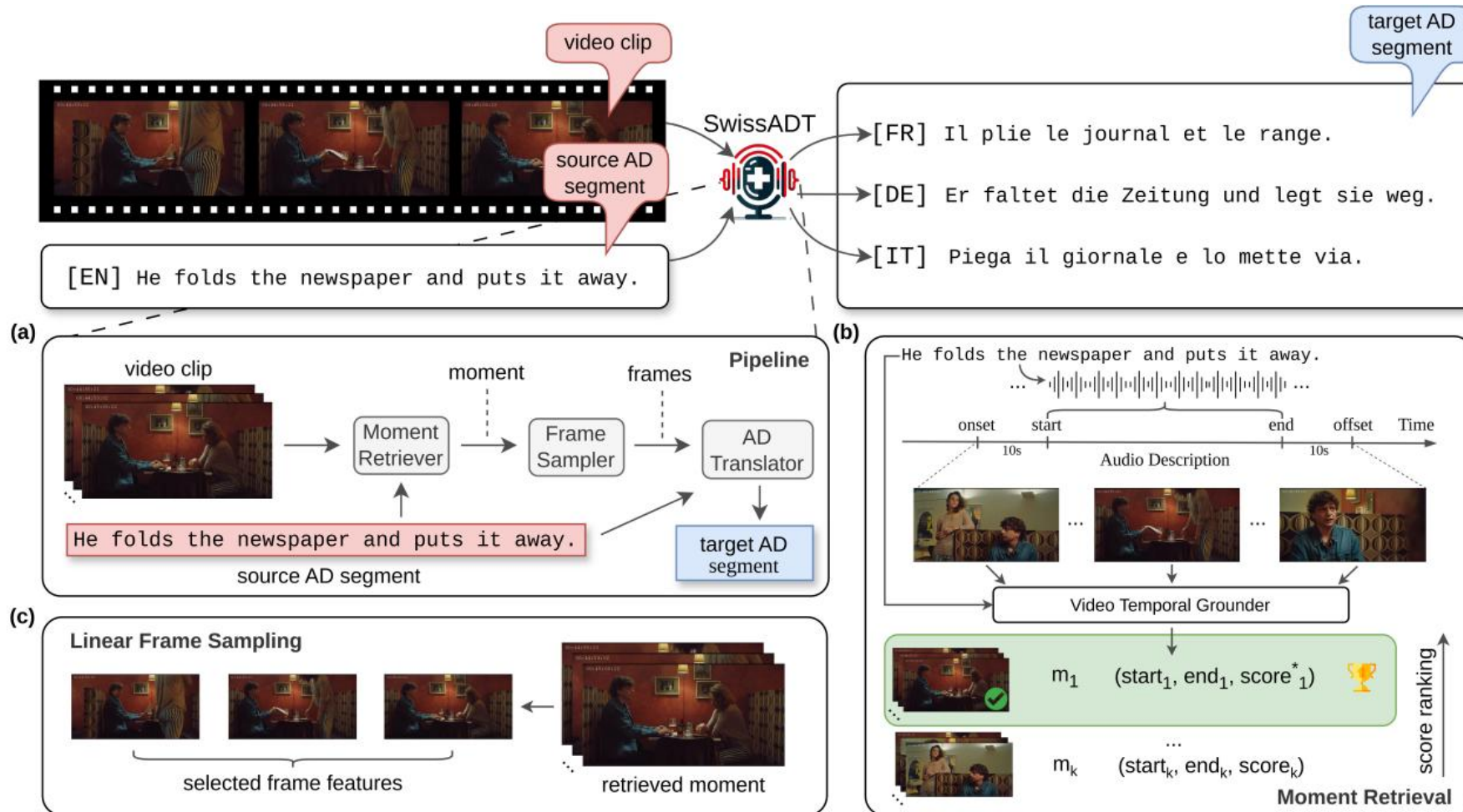
Translate the following audio description from French to German,
only respond with a translation: \n {text}



Translation Version 2

SwissADT: An Audio Description Translation System for Swiss Languages

(Fischer et al., 2024)



Translation Version 3 Prompt



University of
Zurich^{UZH}



Translation version 3 [SwissADT translation post-edited by ChatGPT-4]

As an AI assistant, your task is to evaluate and post-edit an input text. You will identify the following issues in the input text and return a revised version of the text:

- **film terminology:** identify film terminology like "vor der Kamera" or "Nahaufnahme" and either drop them or rephrase the text without them. For example, "Nahaufnahme eines Mannes" could be re-phrased as "Ein Mann".
- **expressions relating to vision:** identify the expressions "wir sehen", "werden gezeigt" and "ist sichtbar" and either drop them or replace them with expression that are appropriate for blind and visually impaired persons. For example, "wir sehen einen Mann vor dem Baum" could be rephrased as "ein Mann steht vor einem Baum". "Blick auf" is an exception and can be kept in the text.
- **grammar:** identify and fix any grammatical errors.
- **fluency:** make sure the text is fluent and easy to understand. make sure words are in correct order and sentences are complete.
- **idiomatic language:** identify nonsensical and non-idiomatic verb-noun combinations like "baum-loser Baum" or "Eiskristalle flackern" and find an alternative. you may drop adjectives, unless they are essential to understand the scene that is described.
- **esszett:** replace the german "ß" with "ss". For example, "straße" should be written as "Strasse".

The input text is this: {text}

Respond only with the post-edited version of the input text. If you did not identify any issues in the input, return the text without changes. Only respond in German.

Examples – translations

(cf. Lüthi, 2024)



University of
Zurich^{UZH}



Segment	Source text	Version 1	Version 2	Version 3
28	Les pas de Jean-Paul Gobet foulent les feuilles mortes dans une forêt.	Die Schritte von Jean-Paul Gobet treten auf die toten Blätter in einem Wald.	Die Schritte von Jean-Paul Gobet treten auf die herabgefallenen Blätter in einem Wald.	Die Schritte von Jean-Paul Gobet treffen auf die herabgefallenen Blätter in einem Wald.
204	Elle est chaudement vêtue et tient un trépied d'appareil photo dans ses mains. Elle s'arrête, le déplie puis s'accroupit à hauteur du trépied pour observer l'horizon, puis reprend sa route.	Sie ist warm angezogen und hält ein Kamerastativ in ihren Händen. Sie hält an , klappt es auf und hockt sich auf die Höhe des Stativs , um den Horizont zu beobachten , dann setzt sie ihren Weg fort.	Sie ist warm gekleidet und hält ein Kamerastativ in ihren Händen. Sie hält an , klappt es auf, hockt sich auf die Höhe des Stativs , um den Horizont zu beobachten , und setzt dann ihren Weg fort.	Sie ist warm gekleidet und hält ein Gerät in ihren Händen. Sie stoppt , öffnet es, duckt sich auf die Höhe des Geräts , um den Horizont zu erkunden , und setzt dann ihren Weg fort.
209	Un cygne plonge son cou élancé dans l'eau, en quête de nourriture.	Ein Schwan taucht seinen schlanken Hals ins Wasser, auf der Suche nach Nahrung.	Ein Schwan taucht seinen schlanken Hals ins Wasser, auf der Suche nach Nahrung.	Ein Schwan taucht seinen schlanken Hals ins Wasser, auf der Suche nach Nahrung.

Results: Quantitative Analysis (I)



University of
Zurich^{UZH}



Overview (cf. Lüthi, 2024, p. 39)

Translation version	<i>Adequacy</i>	<i>Fluency</i>	<i>Usefulness</i>
1) ChatGPT-4 (input: text only)	4.57	3.14	2.93
2) SwissADT (input: text + images)	4.55	3.18	2.94
3) SwissADT version post-edited by ChatGPT-4	4.41	3.10	2.87

Results: Quantitative Analysis (II)



University of
Zurich^{UZH}



Average scores by participant (cf. Lüthi, 2024, p. 43ff)

Translation version 1 [text only]	P1	P2	P3
avg. adequacy	4.40	4.73	4.59
avg. fluency	2.91	2.86	3.65
avg. usefulness	2.34	2.88	3.58

Translation version 2 [SwissADT; text and images]	P1	P2	P3
avg. adequacy	4.24	4.69	4.70
avg. fluency	3.02	2.76	3.76
avg. usefulness	2.33	2.76	3.73

Translation version 3 [version 2 post-edited]	P1	P2	P3
avg. adequacy	4.20	4.52	4.50
avg. fluency	3.00	2.65	3.68
avg. usefulness	2.36	2.36	3.62

Results: Qualitative Analysis



University of
Zurich^{UZH}



Overview (cf. Lüthi, 2024, p. 39)

Main error category	<i>Number of errors identified</i>
Accuracy	10
Fluency	17
Terminology	4
Style	37
Verity	0
Other	0
Tot.	68

Conclusions & Limitations



University of
Zurich^{UZH}



- LLM-based machine translation as a viable option for producing AD scripts
- Translation version with both text and image as input scored higher in the two categories *Fluency* and *Usefulness*
- Harmonized style guide / pre-editing could lead to higher quality output
- Quality of MT output: users'/experts' knowledge of MT technology and post-editing might have an influence on error tolerance level
- Further research: higher number of participants, additional language pairs/genres, prompting for automatic post-editing, evaluation with end users, impact of Pivot-Templates (Jankowska, 2023) on translation quality

References



University of
Zurich^{UZH}



- Fernández-Torné, A., & Matamala, A. (2016). Machine translation in audio description? Comparing creation, translation and post-editing efforts. *SKASE Journal of Translation and Interpretation*, 9(1), 64–87.
- Fischer, L., Gao, Y., Lintner, A., & Ebling, S. (2024). SwissADT: An audio description translation system for Swiss languages. arXiv preprint. arXiv:2411.14967
- Jankowska, A. (2015). *Translating audio description scripts: Translation as a new strategy of creating audio description*. Peter Lang Edition.
- Lommel, A. (2018). Metrics for Translation Quality Assessment: A Case for Standardising Error Typologies. In J. Moorkens, S. Castilho, F. Gaspari, & S. Doherty (eds.), *Translation Quality Assessment: From Principles to Practice* (pp. 109–127). Springer International Publishing. https://doi.org/10.1007/978-3-319-91241-7_6
- Lüthi, N. (2024). KI-gestützte Übersetzung von Audiodeskription. Eine quantitative und qualitative Untersuchung der Sendung Passe-moi les jumelles. (Graduate Papers in Applied Linguistics 26), ZHAW Zurich University of Applied Sciences. <https://doi.org/10.21256/zhaw-2823>
- Matamala, A., & Ortiz-Boix, C. (2016). Accessibility and multilingualism: An exploratory study on the machine translation of audio descriptions. *TRANS*, 20, 11–24.
- RTS. (2021). Passe-moi les jumelles. Pierres en équilibre - La Nuit. Radio Télévision Suisse. <https://www.rts.ch/emissions/passe-moi-les-jumelles/2021/video/pierres-en-equilibre-la-nuit-27168823.html> (10.03.2025).
- Vercauteren, G., Reviere, N., & Steyaert, K. (2021). Evaluating the effectiveness of machine translation of audio description: The results of two pilot studies in the English-Dutch language pair. *Tradumàtica*, 19, 226–252.