

The ALST project

Technologies for audiovisual translation

Anna Matamala (Universitat Autònoma de Barcelona)

anna.matamala@uab.cat

Translating and the Computer 37 (TC37-AsLing 2015)

London, 26-27 November 2015

Funding: FFI-2012-31024, 2014SGR0027

Anna Matamala (2015). "Technologies for audiovisual translation: the ALST project"

As presented at the *37th Conference Translating and the Computer*, London, UK, November 26-27, 2015. © AsLing, The International Association for Advancement in Language Technology, 2015.

Distribution without the authorisation from ASLING is not allowed.

AsLing asking to be informed of such postings, including URLs or URIs where available, to the email address: [<presentations@asling.org>](mailto:presentations@asling.org)."

The ALST project: aims

- Speech recognition
- Machine translation
- Speech synthesis

in audio description (AD)

in voice-over (VO)

Project rationale

- Implementation of existing technologies
 - Two modes chosen as instances of linguistic/sensorial accessibility
 - Oral modes
 - Focus on the professionals and on the audiences
-
- Limitations: 3 years, limited funding
 - Project team and 2 PhD students

Speech Technologies

Speech recognition in voice-over (I)

- Participants: 10 professional transcribers
- 3 conditions:
 - Manual transcription
 - Respeaking
 - Revision of an automatically-generated script (ASR)
- Quantitative data: time ratio, output quality (NER)
- Qualitative data: opinions on usefulness, speed, accuracy, overall quality, effort, boredom.

Source: Matamala, Romero-Fresco & Daniluk (forthcoming)

Speech recognition in voice-over (II)

- Speed: manual, respeaking, ASR revision.
- Accuracy: manual, ASR revision, respeaking.
- Respeaking allowed the highest number of participants to finish.
- Respeaking, better scores in self-reported effort and boredom.
- Manual transcript, better scores in accuracy and overall quality.
- Respeaking: impressed with many possibilities and more job satisfaction but need for specific training and further research.

Speech recognition in audio description

- Soundtrack extraction, speech activity detection, speaker diarization, and speech-to-text transcription
- Film “Closer” (English, Catalan)
- DER and WER measures: low performance because of training conditions of the system and employed AD materials.

Source: Delgado, Matamala & Serrano (forthcoming)

Speech synthesis in voice-over

- On-going

Speech synthesis in audio description (I)

- Pre-test to select the “best” human/natural male/female voices
- Main experiment: 67 blind and visually-impaired volunteers
- Questionnaire inspired by ITU (1994)

Source: Fernández-Torné & Matamala (2015)

Speech synthesis in audio description (II)

- Natural voices have higher values than artificial voices
- No statistical differences between male/female natural voices
- Overall impression/acceptance: mean higher than 3.2 (on a 5-point scale)
- 94% participants: TTS is an alternative acceptable solution although not the preferred one

Source: Fernández-Torné & Matamala (2015)

Machine Translation

Machine translation in VO: effort

- Participants: 12 AVT MA students
- Conditions:
 - human translation
 - post-editing
- Materials: two short wildlife documentary excerpts (En>Es)
- Temporal, technical, cognitive effort using Inputlog (www.inputlog.net/)

Source: Ortiz-Boix & Matamala (forthcoming)

Machine translation in VO: results

- Results obtained for: both excerpts, excerpt 1, excerpt 2
- Temporal effort: post-editing faster although results are only statistically significant in excerpt 1
- Technical and cognitive effort: post-editing requires less effort, although differences are only statistically significant globally and in excerpt 1

Source: Ortiz-Boix & Matamala (forthcoming)

Machine translation in VO: QA

- QA by experts:
 - Pre-correction grading/ Correction based on MQM/ Questionnaire/ Final mark/ Is it a translation or the result of a post-editing?
- QA in Dubbing studio: observational notes, corrections made by dubbing director
- QA by end-users: pre-task questionnaires, viewing, post-task questionnaires (comprehension, self-reported enjoyment, preferences and interest)

Source: Ortiz-Boix & Matamala (2015)

Machine translation in VO: results (I)

- QA by experts: passes round 1 (45/72 translation, 37/72 post-editing), round 2 (41 and 38/72, respectively)
- QA by experts: corrections (12.862 translation vs. 17.957 PE)
- QA by experts: identify translations as such more easily
- Translations are slightly better

Source: Ortiz-Boix & Matamala (2015)

Machine translation in VO: results (II)

- QA by dubbing studio:
 - excerpt 1 (5 changes in translation vs 6 in PE),
 - excerpt 2 (4 in translation, none in PE due to bad quality)
- Translation is better in 2nd excerpt, minimal differences in the 1st

Source: Ortiz-Boix & Matamala (2015)

Machine translation in VO: results (III)

- QA by users:
 - Translated versions slightly better understood than post-edited versions in global analysis and excerpt 1 (post-edited is better in excerpt 2).
 - Better results for translation (enjoyment, preferences and interest)
- Translation quality is slightly higher in all conditions but results are not statistically significant and there are differences according to excerpts/groups.

Source: Ortiz-Boix & Matamala (2015)

Machine translation in audio description (I)

- Pre-test to select the MT engine: Google Translate.
- Participants: 5 volunteers
- 5 MT engines
- Materials: clip from “Closer” (240 words in 14 AD units, 3’09’’)
- Tool: PET

Source: Fernández-Torné & Matamala (2014)

	Automatic	Human
Objective	HBLEU HTER	PE time
Subjective		PE necessity PE difficulty MT output adequacy MT output fluency MT output ranking

Machine translation in audio description (II)

- Main experiment:
 - Participants: 12 AVT MA students
 - Conditions: human creation/translation/post-editing
 - Language pair: English into Catalan
 - Materials: comparable excerpts from “Closer”
 - Tool: Subtitle Workshop
- Temporal, technical and cognitive effort (Inputlog, www.inputlog.net)
- Subjective opinions (pre-task and post-task questionnaires)

Source: Fernández-Torné & Matamala (in prep.)

MT in AD: results

- **Temporal effort:** no statistical differences.
- **Technical effort:** AD creation and AD translation, more keyboard action than post-editing. AD translation and post-editing, higher number of mouse scrolls than AD creation.
- **Cognitive effort:** statistically higher in AD creation.

Source: Fernández-Torné & Matamala (in prep.)

MT in AD: results (II)

- **Subjective data** (on a 10-point Likert scale):

	AD creation		AD translation		AD post-editing	
	Pre	Post	Pre	Post	Pre	Post
Effort	8.25	7.17	6.17	5.58	6.50	7.50
Creativity impairment	3.09	3.82	7.45	7.27	8.45	9.36
Boredom	2.09	1.82	4.18	4.18	6.73	7.27
Calques	1.25	2.00	5.25	5.42	6.93	8.33

Source: Fernández-Torné & Matamala (in prep.)

Conclusions

- Small-scale exploratory project, existing technology
- More research needed with wider samples, more excerpts, other language pairs
- Not only time (productivity) but also opinions of professionals and end users

The ALST project

Technologies for audiovisual translation

Anna Matamala (Universitat Autònoma de Barcelona)

anna.matamala@uab.cat

Translating and the Computer 37 (TC37-AsLing 2015)

London, 26-27 November 2015

Funding: FFI-2012-31024, 2014SGR0027