



**MeMAD**

Methods for Managing  
Audiovisual Data

memad.eu  
info@memad.eu

Twitter - @memadproject  
Linkedin - MeMAD Project

# MeMAD Project Publications

*Version 1.1, June 16, 2021*

**2021**

## **WP2 AUTOMATIC MULTIMODAL CONTENT ANALYSIS**

Peter Smit, Sami Virpioja, Mikko Kurimo Advances in subword-based HMM-DNN speech recognition across languages Computer Speech & Language, Volume 66, 2021. ([link](#))

Tuomas Kaseva, Hemant Kathania, Aku Rouhe and Mikko Kurimo Speaker Verification Experiments for Adults and Children using a shared embedding spaces Proceedings of the Nordic Conference on Computational Linguistics, NoDaLiDa, 2021 ([link](#))

Aku Rouhe, Astrid Van Camp, Mittul Singh, Hugo Van hamme and Mikko Kurimo End-to-end and HMM/DNN ASR in an equal data setting: A Finnish case study (submitted)

Jean Carrive, Abdelkrim Beloued, Pascale Goetschel, Serge Heiden, Antoine Laurent, Pasquale Lisena, Franck Mazuet, Sylvain Meignier, Bénédicte Pincemin, Géraldine Poels, Raphaël Troncy. Transdisciplinary Analysis of a Corpus of French Newsreels: The ANTRACT Project. Digital Humanities Quarterly. January 2021 ([link](#))

Lisena, Pasquale; Troncy, Raphaël. Face Recognition in Videos Using Images From the Web. The Web Conference 2021, Demo Track (to appear)

Ismail Harrando and Raphael Troncy. Named Entity Recognition as Graph Classification. In 18th European Semantic Web Conference (ESWC 2021), June 7-10, 2021. (to appear)



*MeMAD project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 780069. This document has been produced by the MeMAD project. The content in this document represents the views of the authors, and the European Commission has no liability in respect of the content.*



## WP3 MEDIA ENRICHMENT AND HYPERLINKING

Pasquale Lisena, Jorma Laaksonen and Raphael Troncy. FaceRec: An Interactive Framework for Face Recognition in Video Archives. In 2nd International Workshop on Data-driven Personalisation of Television (DataTV) colocated with the ACM International Conference on Interactive Media Experiences (IMX 2021), June 21-23, 2021. (to appear)

Ismail Harrando and Raphael Troncy. And Cut! Unsupervised Content Segmentation and Alignment. In 2nd International Workshop on Data-driven Personalisation of Television (DataTV) colocated with the ACM International Conference on Interactive Media Experiences (IMX 2021), June 21-23, 2021 (to appear)

Ismail Harrando and Raphael Troncy. Explainable Zero-Shot Topic Extraction Using a Common-Sense Knowledge Graph. In 3rd Conference on Language, Data and Knowledge (LDK 2021), September 1-3, 2021. (to appear)

Dejan Porjazovski, Juho Leinonen and Mikko Kurimo. Attention-Based End-To-End Named Entity Recognition From Speech. Proceedings of 24th International Conference of Text, Speech and Dialogue (TSD2021). 2021 (to appear).

## WP4 MULTIMODAL MACHINE TRANSLATION

Grönroos, Stig-Arne; Virpioja, Sami; Kurimo, Mikko Transfer learning and subword sampling for asymmetric-resource one-to-many neural translation MACHINE TRANSLATION, 2021. ([Publisher site](#), [arxiv preprint](#))

## WP5 HUMAN PROCESSING IN MULTIMODAL CONTENT DESCRIPTION AND TRANSLATION

Kim Starr, Sabine Braun and Jaleh Delfani (2021) 'A Sentient Being's Guide to Automatic Video Description: a six-point Roadmap for Building the Computer Model of the Future', in *Proceedings of the Media for All 9 Conference*, Barcelona [Virtual], 27-29th January. Video presentation publication forthcoming.

Sabine Braun, Kim Starr and Jaleh Delfani (2021) 'When worlds collide: AI-created, human-mediated video description services and the user experience'. *UAHCI Conference, Washington DC/online, July 24-29*. (Accepted).

## 2020





**MeMAD**

Methods for Managing  
Audiovisual Data

memad.eu  
info@memad.eu

Twitter - @memadproject  
Linkedin - MeMAD Project

## WP2 AUTOMATIC MULTIMODAL CONTENT ANALYSIS

Ismail Harrando, Alison Reboud, Pasquale Lisena, Raphaël Troncy, Jorma Laaksonen, Anja Virkkunen, Mikko Kurimo. Using Fan-Made Content, Subtitles and Face Recognition for Character-Centric Video Summarization, Proceedings of the TRECVID 2020 Workshop. (PDF)

Matias Lindgren, Tommi Jauhiainen and Mikko Kurimo Releasing a toolkit and comparing the performance of language embeddings across various spoken language identification datasets Proceedings of Interspeech 2020 ([link](#))

Matias Lindgren. Deep learning for spoken language identification. Master's thesis, Aalto University, 2020. ([link](#))

Abhilash Jain, Aku Rouhe, Stig-Arne Grönroos, Mikko Kurimo. Finnish ASR with Deep Transformer Models Proceedings of Interspeech 2020. (PDF)

Abhilash Jain. Finnish language modeling and ASR with Deep Transformer Models. Master's thesis, Aalto University, 2020. ([link](#))

Aku Rouhe, Tuomas Kaseva, and Mikko Kurimo. Speaker-aware training of attention-based end-to-end speech recognition using neural speaker embeddings. In Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2020. (PDF)

Phuong Anh Nguyen, Jiaxin Wu, Chong-Wah Ngo, Danny Francis and Benoit Huet. VIREO @ Video browser showdown 2020. In 26th International Conference on MultiMedia Modeling (MMM 2020), 5-8 January 2020, Daejeon, Korea. ([Publisher site](#), [preprint](#))

Tzu-Jui Julius Wang, Selen Pehlivan and Jorma Laaksonen. Tackling the Unannotated: Scene Graph Generation with Bias-Reduced Models. In Proceeding of the British Machine Vision Conference (BMVC), Online Conference 7–10 September, 2020. (PDF)

Jorma Laaksonen and Zixin Guo. PicSOM Experiments in TRECVID 2020. In Proceedings of the TRECVID 2020 Workshop. Online Conference 17–19 November, 2020. (PDF)



*MeMAD project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 780069. This document has been produced by the MeMAD project. The content in this document represents the views of the authors, and the European Commission has no liability in respect of the content.*



## WP3 MEDIA ENRICHMENT AND HYPERLINKING

Dejan Porjazovski, Juho Leinonen, and Mikko Kurimo. Named entity recognition for spoken Finnish. Proceedings of 2nd International Workshop on AI for Smart TV Content Production Access and Delivery. ([link](#))

Dejan Porjazovski. End-to-end named entity recognition for spoken Finnish. Master's thesis, Aalto University, 2020. ([link](#))

Lisena, Pasquale; Harrando, Ismail; Kandakji, Oussama; Troncy, Raphaël. ToModAPI: A topic modeling API to train, use and compare topic models. In NLP-OSS @ EMNLP 2020. ([link](#))

Reboud, Alison; Harrando, Ismail; Laaksonen, Jorma; Troncy, Raphaël. Predicting media memorability with audio, video, and text representations. In MediaEval 2020. ([PDF](#))

## WP4 MULTIMODAL MACHINE TRANSLATION

Stig-Arne Grönroos. Machine translation into morphologically rich low-resource languages. PhD thesis. Aalto University, 2020. ([PDF](#))

Mary Nurminen and Maarit Koponen. 2020. Machine Translation and Fair Access to Information. *Translation Spaces* 9(1), pages 150–169. ([link](#))

Jörg Tiedemann. The Tatoeba Translation Challenge – Realistic Data Sets for Low Resource and Multilingual MT. In *Proceedings of the 5th Conference on Machine Translation (WMT 2020)*, pages 1174–1182. Online Conference 19–20 November, 2020. ([PDF](#))

Jörg Tiedemann, Santhosh Thottingal. OPUS-MT – Building open translation services for the World. In *Proceedings of the 22nd Annual Conference of the European Association for Machine Translation (EAMT 2020)*, pages 479–480. Online Conference 3–5 November, 2020. ([PDF](#))

Maarit Koponen, Umut Sulubacak, Kaisa Vitikainen, Jörg Tiedemann. MT for subtitling: User evaluation of post-editing productivity. In *Proceedings of the 22nd Annual Conference of the European Association for Machine Translation (EAMT 2020)*, pages 115–124. Online Conference 3–5 November, 2020. ([PDF](#))

Maarit Koponen, Umut Sulubacak, Kaisa Vitikainen, Jörg Tiedemann. MT for Subtitling: Investigating professional translators' user experience and feedback. In *Proceedings of the 14th Conference of the Association for Machine Translation in the Americas October 6 – 9, 2020, 1st Workshop on Post-Editing in Modern-Day Translation*, pages 79–92. Online, 6 October, 2020. ([PDF](#))

Umut Sulubacak, Ozan Caglayan, Stig-Arne Grönroos, Aku Rouhe, Desmond Elliott, Lucia Specia, Jörg Tiedemann. *Multimodal Machine Translation through Visuals and Speech*. In *Machine Translation*, vol. 34, issue 2–3, pages 97–147. Springer, 13 August, 2020. ([PDF](#))





Raúl Vázquez, Mikko Aulamo, Umut Sulubacak, Jörg Tiedemann. The University of Helsinki Submission to the IWSLT2020 Offline Speech Translation Task. In *Proceedings of the 17th International Conference on Spoken Language Translation (IWSLT)*, pages 95–102. Stroudsburg, PA, 9 July, 2020. ([PDF](#))

Mikko Aulamo, Umut Sulubacak, Sami Virpioja, Jörg Tiedemann. *OpusTools and Parallel Corpus Diagnostics*. In Proceedings of the 12th Language Resources and Evaluation Conference, pages 3782–3789. ELRA, Marseilles, France, 17 May, 2020. ([PDF](#))

Stig-Arne Grönroos, Sami Virpioja, Mikko Kurimo. *Morfessor EM+Prune: Improved subword segmentation with expectation maximization and pruning*. In Proceedings of the 12th Language Resources and Evaluation Conference, pages 3944–3953. ELRA, Marseilles, France, 17 May, 2020. ([PDF](#))

## WP5 HUMAN PROCESSING IN MULTIMODAL CONTENT DESCRIPTION AND TRANSLATION

Kim Starr, Sabine Braun and Jaleh Delfani (2020) ‘Taking a Cue from the Human: Linguistic and Visual Prompts for the Automatic Sequencing of Multimodal Narrative’. *Journal of Audiovisual Translation*, 3(2), pp: 140–169. ([link](#))

Sabine Braun and Kim Starr (2020/21) *Innovation in Audio Description Research*. London: Routledge. ([Publisher site](#))

Sabine Braun and Kim Starr (2020/21) ‘Comparing human and automated approaches to visual storytelling’, in Braun, S. and Starr, K. (eds) *Innovation in Audio Description Research*. London: Routledge. ([Publisher site](#))

Sabine Braun and Kim Starr (2020) ‘Byte-Sized Storytelling: Training the Machine to See the Bigger Picture’. *13th Languages and the Media*, Berlin, December 14-16th. [Accepted]. Conference postponed until September 2021.

## WP6 EVALUATION

Lauri Saarikoski, Dieter Van Rijsselbergen, Maija Hirvonen, Maarit Koponen, Umut Sulubacak, Kaisa Vitikainen. MeMAD project: End user feedback on AI in the media production workflows. In Proceedings of the International Broadcasting Convention (IBC) 2020. ([link](#))

Maarit Koponen, Tiina Tuominen, Maija Hirvonen, Kaisa Vitikainen, and Liisa Tiittula. 2020. User perspectives on developing technology-assisted access services in public broadcasting. *Bridge: Trends and Traditions in Translation and Interpreting Studies* 1(2), 47-67. ([PDF](#))





## 2019

### WP2 AUTOMATIC MULTIMODAL CONTENT ANALYSIS

Peter Smit. Modern subword-based models for automatic speech recognition. PhD Thesis. Aalto University. 2019. ([link](#))

Rao Muhammad Anwer, Fahad Shahbaz Khan, Jorma Laaksonen and Nazar Zaki. *Multi-stream convolutional networks for indoor scene recognition*. In Proceedings of the 18th International Conference on Computer Analysis of Images and Patterns (CAIP2019), pages 196–208, Salerno, Italy, September 2019. ([PDF](#))

Tzu-Jui Julius Wang, Hamed Rezazadegan Tavakoli, Mats Sjöberg and Jorma Laaksonen. *Geometry-aware relational exemplar attention for dense captioning*. In Proceedings of the 1st International Workshop on Multimodal Understanding and Learning for Embodied Applications (MULEA '19) in ACM Multimedia Conference, pages 3–11, Nice, France, October 2019. ([PDF](#))

Tiancai Wang, Rao Muhammad Anwer, Muhammad Haris Khan, Fahad Shahbaz Khan, Yanwei Pang, Ling Shao and Jorma Laaksonen. *Deep contextual attention for human-object interaction detection*. In Proceedings of the International Conference on Computer Vision (ICCV2019), pages 5694–5702, Seoul, Korea, October 2019. ([PDF](#))

Héctor Laria Mantecón, Jorma Laaksonen, Danny Francis and Benoit Huet. *PicSOM and EURECOM Experiments in TRECVID 2019*. In *Proceedings of the TRECVID 2019 Workshop*. NIST, Gaithersburg, MA, USA, November 2019. ([PDF](#))

Jean Carrive. Using Artificial Intelligence to Preserve Audiovisual Archives: New Horizons, More Questions. In Proceedings of the 27th ACM International Conference on Multimedia, Conference proceedings, Pages 1–2, Nice, France, October 2019. Invited keynote. ([PDF](#))

David Doukhan, Géraldine Poels, Zohra Rezgui and Jean Carrive. Describing Gender Equality in French Audiovisual Streams with a Deep Learning Approach. *VIEW Journal of European Television History and Culture*, 7(14), pp.103. 2019 ([PDF](#))

Stefanos Vrochidis, Benoit Huet, Edward Y. Chang and Ioannis Kompatsiaris. Big data analytics for large-scale multimedia search. 2019. Wiley, ISBN: 978-1119376972 ([Publisher site](#))

Danny Francis, Phuong Anh Nguyenn, Benoit Huet and Chong-Wah Ngo. Fusion of multimodal embeddings for ad-hoc video search. In 1st International Workshop on Video Retrieval Methods and Their Limits (ViRaL 2019), co-located with ICCV 2019, Seoul, Korea, October 2019. ([link](#))







# MeMAD

Methods for Managing  
Audiovisual Data

memad.eu  
info@memad.eu

Twitter - @memadproject  
Linkedin - MeMAD Project

Danny Francis, Benoit Huet. L-STAP: Learned Spatio-Temporal Adaptive Pooling for video captioning. In 1st International Workshop on AI for smart TV content production, access and delivery (AI4TV 2019), co-located with the 27th ACM International Conference on Multimedia, October 2019, Nice, France ([Publisher site](#))

Danny Francis, Phuong Anh Nguyen, Benoit Huet and Chong-Wah Ngo. EURECOM at TRECVID AVS 2019. In Proceedings of the TRECVID 2019 Workshop. NIST, Gaithersburg, MA, USA, November 2019. ([link](#))

Phuong Anh Nguyen, Jiaxin Wu, Chong-Wah Ngo, Francis Danny and Benoit Huet. VIREO-EURECOM @ TRECVID 2019: Ad-hoc Video Search (AVS). In Proceedings of the TRECVID 2019 Workshop. NIST, Gaithersburg, MA, USA, November 2019. ([link](#))

Tuomas Kaseva, Aku Rouhe, and Mikko Kurimo. *Spherediar – an efficient speaker diarization system for meeting data*. In 2019 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU), 2019. ([link](#))

Arturs Polis. *Paragraph-length image captioning using hierarchical recurrent neural networks*. Master's Thesis, University of Helsinki, 2019. ([PDF](#))

Héctor Laria Mantecón. *Deep Reinforcement Sequence Learning for Visual Captioning*. Master's Thesis, Aalto University, 2019. ([PDF](#))

Aditya Surikuchi: *Visual Storytelling: Captioning of Image Sequences*. Master's Thesis, Aalto University, 2019. ([PDF](#))

Tuomas Kaseva. *Spherediar – an efficient speaker diarization system for meeting data*. Master's thesis, Aalto University, 2019. ([PDF](#))

Zohra Rezgui. *Détection et classification de visages pour la description de l'égalité femme-homme dans les archives télévisuelles*, INA, 2019. ([PDF](#))

Danny Francis. *Semantic representations of images and videos*. PhD Thesis, Sorbonne University/EURECOM. December 2019.



MeMAD project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 780069. This document has been produced by the MeMAD project. The content in this document represents the views of the authors, and the European Commission has no liability in respect of the content.



## WP3 MEDIA ENRICHMENT AND HYPERLINKING

Mohammad Reza Kavosifard, Daniele Apiletti, Elena Baralis, Paolo Garza and Benoit Huet. Effective video hyperlinking by means of enriched feature sets and monomodal query combinations. *International Journal of Multimedia Information Retrieval*, 2019 ([Publisher site](#))

Lorenzo Canale, Pasquale Lisena and Raphaël Troncy. Une nouvelle méthode ensembliste pour la reconnaissance et la désambiguïsation d'entités nommées en utilisant des réseaux de neurones. In *Journées francophones d'Ingénierie des Connaissances (IC 2019)*, July 2019, Toulouse, France ([PDF](#))

Pasquale Lisena, Albert Meroño-Peñuela, Tobias Kuhn and Raphaël Troncy. Easy web API development with SPARQL transformer. In *18th International Semantic Web Conference (ISWC 2019)*, In-use Track, October 2019, Auckland, New Zealand ([PDF](#))

## WP4 MULTIMODAL MACHINE TRANSLATION

Aarne Talman, Umut Sulubacak, Raúl Vázquez, Yves Scherrer, Sami Virpioja, Alessandro Raganato, Arvi Hurskainen and Jörg Tiedemann. The University of Helsinki submissions to the WMT19 news translation task. In *Proceedings of the Fourth Conference on Machine Translation (WMT): Shared Task Papers*. 2019. ([PDF](#))

Raúl Vázquez, Umut Sulubacak and Jörg Tiedemann. The University of Helsinki submission to the WMT19 parallel corpus filtering task. In *Proceedings of the Fourth Conference on Machine Translation (WMT): Shared Task Papers*. 2019. ([PDF](#))

## WP5 HUMAN PROCESSING IN MULTIMODAL CONTENT DESCRIPTION AND TRANSLATION

Sabine Braun and Kim Starr (2019) 'Finding the Right Words: Investigating Machine-Generated Video Description Quality using a Human-Derived Corpus-based Approach'. *Journal of Audiovisual Translation*, 2(2), pp. 11–35. ([PDF](#))

Sabine Braun and Kim Starr (2019) 'Comparing human and automated approaches to video description'. *Media for All 8*, Stockholm, 19th June.

Sabine Braun and Kim Starr (2019) 'Mind the Gap: An Investigation of Omissions in Audio Description', *ARSAD*, Barcelona. Conference abstract.







## 2018

### WP2 AUTOMATIC MULTIMODAL CONTENT ANALYSIS

Danny Francis, Benoit Huet and Bernard Merialdo. *EURECOM participation in TrecVid VTT 2018*. In 22nd International Workshop on Video Retrieval Evaluation (TRECVID 2018), Gaithersburg, USA, November 13-15, 2018. ([PDF](#))

Mats Sjöberg, Hamed R. Tavakoli, Zhicun Xu, Hector Laria Mantecon and Jorma Laaksonen. *PicSOM experiments in TRECVID 2018*. In 22nd International Workshop on Video Retrieval Evaluation (TRECVID 2018), Gaithersburg, USA, November 13-15, 2018. ([PDF](#))

David Doukhan, Elliott Lechapt, Marc Evrard and Jean Carrive. *INA's MIREX 2018 music and speech detection system*. In 14th Music Information Retrieval Evaluation eXchange (MIREX), September 2018, Paris, France.

Zhicun Xu, Peter Smit, and Mikko Kurimo. *The Aalto system based on fine-tuned audioset features for DCASE2018 task2 – general purpose audio tagging*. In Detection and Classification of Acoustic Scenes and Events Workshop (DCASE 2018), Surrey, UK, November 2018. ([PDF](#))

Zhicun Xu. *Audio Event Classification Using Deep Learning Methods*. Master's thesis. Aalto University, 2018. ([Permanent link](#))

Olfa Ben-Ahmed and Benoit Huet. *Deep Multimodal Features for Movie Genre and Interestingness Prediction*. In International Conference on Content-Based Multimedia Indexing (CBMI 2018), La Rochelle, France, 4-6 September 2018. ([PDF](#))

### WP3 MEDIA ENRICHMENT AND HYPERLINKING

Julien Plu, Giuseppe Rizzo and Raphaël Troncy. *ADEL: ADaptable Entity Linking. A Hybrid Approach to Link Entities with Linked Data for Information Extraction*. In Semantic Web Journal, IOS Press, (to appear) 2019. ([PDF](#))

Lorenzo Canale, Pasquale Lisena and Raphaël Troncy. *A Novel Ensemble Method for Named Entity Recognition and Disambiguation based on Neural Network*. In International Semantic Web Conference (ISWC 2018), Monterey, CA, USA, 8-12 October 2018. ([PDF](#))





**MeMAD**

Methods for Managing  
Audiovisual Data

memad.eu  
info@memad.eu

Twitter - @memadproject  
Linkedin - MeMAD Project

## WP4 MULTIMODAL MACHINE TRANSLATION

Umut Sulubacak, Aku Rouhe, Jörg Tiedemann, Stig-Arne Grönroos and Mikko Kurimo. *The MeMAD Submission to the IWSLT 2018 Speech Translation Task*. In 15th International Workshop on Spoken Language Translation (IWSLT 2018); Bruges, Belgium. pp. 89–94. ([PDF](#))

Stig-Arne Grönroos, Benoit Huet, Mikko Kurimo, Jorma Laaksonen, Bernard Merialdo, Phu Pham, Mats Sjöberg, Umut Sulubacak, Jörg Tiedemann, Raphaël Troncy and Raul Vázquez. 2018. *The MeMAD Submission to the WMT18 Multimodal Translation Task*. In 3rd Conference on Machine Translation (WMT 2018); Brussels, Belgium. Association for Computational Linguistics, pp. 609–617. ArXiv: 1808.10802. ([PDF](#))

Stig-Arne Grönroos, Sami Virpioja and Mikko Kurimo. *Cognate-aware morphological segmentation for multilingual neural translation*. In 3rd Conference on Machine Translation (WMT 2018); Brussels, Belgium. Association for Computational Linguistics, pp. 390–397. ([PDF](#))

Franck Burlot, Yves Scherrer, Vinit Ravishankar, Ondřej Bojar, Stig-Arne Grönroos, Maarit Koponen, Tommi Nieminen and François Yvon. *The WMT18 Morpheval test suites for English–Czech, English–German, English–Finnish and Turkish–English*. In 3rd Conference on Machine Translation (WMT18); Brussels, Belgium. Association for Computational Linguistics, pp. 550–564. ([PDF](#))

## WP5 HUMAN PROCESSING IN MULTIMODAL CONTENT DESCRIPTION AND TRANSLATION

Sabine Braun and Kim Starr (2018) 'From Slicing Bananas to Pluto the Dog: Human and Automatic Approaches to Visual Storytelling', *Languages & the Media*, Berlin, 10/2018. Conference abstract. ([PDF](#))



MeMAD project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 780069. This document has been produced by the MeMAD project. The content in this document represents the views of the authors, and the European Commission has no liability in respect of the content.