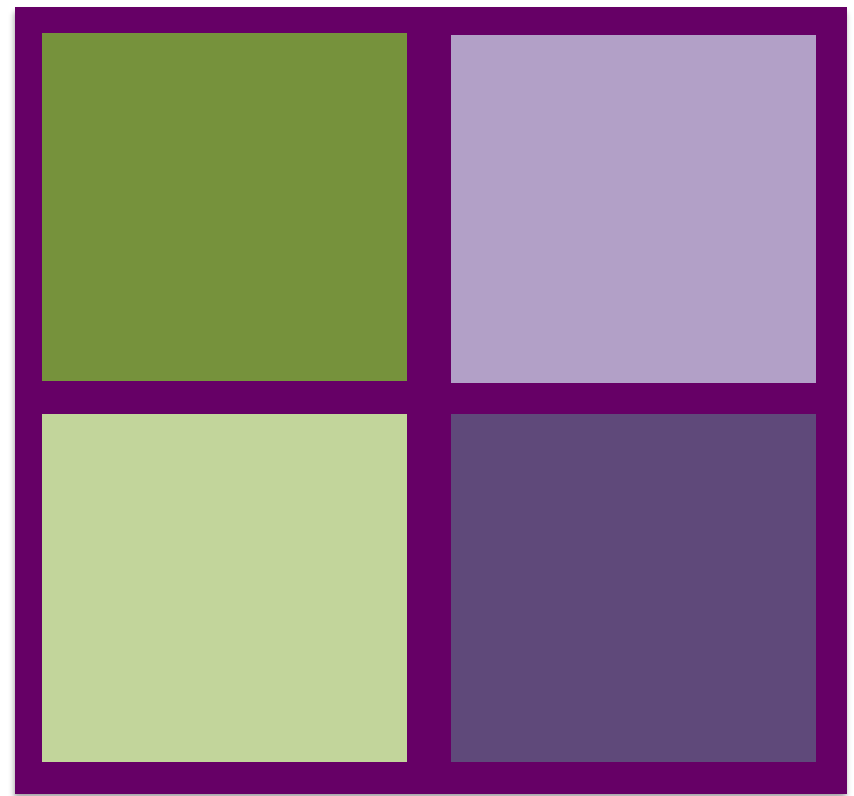




MULTI-TASK
MULTI-LINGUAL
MULTI-MODAL



Funded by the Horizon 2020 Framework Programme
of the European Union



WG2 Efficient Machine Learning Algorithms, Methods, and Applications to Language Generation

Elena Lloret, University of Alicante, Spain (WG coleader)
Aykut Erdem, Koç University, Turkey (WG leader)



MULTI-TASK
MULTI-LINGUAL
MULTI-MODAL



Funded by the Horizon 2020 Framework Programme
of the European Union

Outline

- Theme of WG2
- Results/outcomes
- Planned activities



MULTI-TASK
MULTI-LINGUAL
MULTI-MODAL



Funded by the Horizon 2020 Framework Programme
of the European Union

Outline

- **Theme of WG2**
- Results/outcomes
- Planned activities



MULTI-TASK
MULTI-LINGUAL
MULTI-MODAL



Funded by the Horizon 2020 Framework Programme
of the European Union

Theme of WG2

- WG2 deals with ML machinery behind state-of-the-art LG models
- WG2 focuses tasks
 - multi-task and transfer learning
 - representation learning
 - structured prediction
 - generative models
 - multi-modal models
 - others



MULTI-TASK
MULTI-LINGUAL
MULTI-MODAL



Funded by the Horizon 2020 Framework Programme
of the European Union

Theme of WG2

Communication channels within WG members:

- WG2 mailing list (wg2@sympa.multi3generation.eu)
- Slack channel (#wg2-ml)



MULTI-TASK
MULTI-LINGUAL
MULTI-MODAL



Funded by the Horizon 2020 Framework Programme
of the European Union

Outline

- Theme of WG2
- **Results/outcomes**
- Planned activities



MULTI-TASK
MULTI-LINGUAL
MULTI-MODAL



Funded by the Horizon 2020 Framework Programme
of the European Union

Results/outcomes

Tangible outcomes for WG2

Survey and position papers

Joint journal and conference publications

Cross-topic research roadmap identifying where research effort is most needed and the research challenges that need to be addressed

Repository of open source software for processing language and visual content, inc. a directory of sources of materials or components

Organisation of training and summer schools with courseware, including a short course on Topics on ML for LG for Multi3Generation members, and courses for the wider community (ET, SR)

Reports from WG meetings and STSMs

MoU: https://e-services.cost.eu/files/domain_files/CA/Action_CA18231/mou/CA18231-e.pdf



MULTI-TASK
MULTI-LINGUAL
MULTI-MODAL



Funded by the Horizon 2020 Framework Programme
of the European Union

Results/outcomes

Tangible outcomes for WG2	ACHIEVED?
Survey and position papers	YES
Joint journal and conference publications	YES
Cross-topic research roadmap identifying where research effort is most needed and the research challenges that need to be addressed	PARTLY DONE
Repository of open source software for processing language and visual content, inc. a directory of sources of materials or components	YES
Organisation of training and summer schools with courseware, including a short course on Topics on ML for LG for Multi3Generation members, and courses for the wider community (ET, SR)	PARTLY DONE
Reports from WG meetings and STSMs	YES

MoU: https://e-services.cost.eu/files/domain_files/CA/Action_CA18231/mou/CA18231-e.pdf



MULTI-TASK
MULTI-LINGUAL
MULTI-MODAL



Funded by the Horizon 2020 Framework Programme
of the European Union

Results/outcomes

On the state-of-the-art surveys:

- **Survey 1: The state-of-the-art in Neural Natural Language Generation** (led by Aykut Erdem & Elena Lloret)
 - Investigates several different NLG tasks from 4 unique dimensions: multilinguality, multimodality, controllability, and learning strategies
 - Accepted in Journal of Artificial Intelligence Research (JAIR) (submission done in April 2021; acceptance notification 12 September 2021)
- **Survey 2: Computing-resource efficient ML for Natural Language Generation** (led by Marcin Paprzyck)
 - No progress so far



MULTI-TASK
MULTI-LINGUAL
MULTI-MODAL



Funded by the Horizon 2020 Framework Programme
of the European Union

Results/outcomes

Recent joint publications by WG2 members:

- M. Sercan Amac, Semih Yagcioglu, Aykut Erdem, and Erkut Erdem. **Procedural Reasoning Networks for Understanding Multimodal Procedures**, CoNLL 2019.
- Victor Milewski, Marie-Francine Moens and Iacer Calixto. **Are scene graphs good enough to improve image captioning?**, ACL 2020.
- Ales Zamuda and Elena Lloret. **Optimizing Data-Driven Models for Summarization as Parallel Tasks**, Journal of Computational Science, Vol. 42, April 2020.



MULTI-TASK
MULTI-LINGUAL
MULTI-MODAL



Funded by the Horizon 2020 Framework Programme
of the European Union

Results/outcomes

Cross-topic research roadmap identifying where research effort is most needed and the research challenges that need to be addressed

Table 1: The maturity levels of the NLG applications with respect to the four dimensions explored in the survey.

	Multilinguality	Multimodality	Learning Strategies	Controllability
Machine translation	★★★★★	★★★★★	★★★★★	★★★★★
Description generation	★★★★★	★★★★★	★★★★★	★★★★★
Automatic Speech Recognition	★★★★★	★★★★★	★★★★★	★★★★★
Abstractive Summarization	★★★★★	★★★★★	★★★★★	★★★★★
Text Simplification	★★★★★	★★★★★	★★★★★	★★★★★
Question Answering and Generation	★★★★★	★★★★★	★★★★★	★★★★★
Dialogue Generation	★★★★★	★★★★★	★★★★★	★★★★★



MULTI-TASK
MULTI-LINGUAL
MULTI-MODAL



Funded by the Horizon 2020 Framework Programme
of the European Union

Results/outcomes

On the repository of open-source software:

- Neural Natural Language Generation repository in Github:

<https://github.com/Multi3Generation/neural-natural-language-generation>

- Link available in the Mutli3Generation Webpage:

<https://multi3generation.eu/working-groups/wg2/>

<https://multi3generation.eu/outcomes/software-demos/>



MULTI-TASK
MULTI-LINGUAL
MULTI-MODAL



Funded by the Horizon 2020 Framework Programme
of the European Union

Results/outcomes

Organisation of training and summer schools with courseware, including a short course on Topics on ML for LG for Multi3Generation members, and courses for the wider community (ET, SR)

- Spotlight online talks by WG2 members (February 10, 2021)
- WG2 members participated in Advanced Language Processing Winter School Winter School (ALPS) in January 2021

<http://lig-alps.imag.fr/index.php/sample-page/>



MULTI-TASK
MULTI-LINGUAL
MULTI-MODAL



Funded by the Horizon 2020 Framework Programme
of the European Union

Results/outcomes

Reports from WG meetings and STSMs

- WG2 online meeting (February 10, 2021)

<https://multi3generation.eu/wp-content/uploads/2021/01/WG2-progress-video-presentation.pdf>



MULTI-TASK
MULTI-LINGUAL
MULTI-MODAL



Funded by the Horizon 2020 Framework Programme
of the European Union

Outline

- Theme of WG2
- Results/outcomes
- **Planned activities**



MULTI-TASK
MULTI-LINGUAL
MULTI-MODAL



Funded by the Horizon 2020 Framework Programme
of the European Union

Planned activities

Concrete outputs for the forthcoming years:

- Cross-topic research roadmap identifying where research effort is most needed and the research challenges that need to be addressed
- Organization of training and summer schools (with WG4)
- Reports from WG meetings and STSMs
 - STMS done at the moment by Alkis Katsalis at the University of Alicante (supervisor: Elena Lloret) → topic: Knowledge-grounded NLG



MULTI-TASK
MULTI-LINGUAL
MULTI-MODAL



Funded by the Horizon 2020 Framework Programme
of the European Union

Planned activities

Recent news from WG2 members:

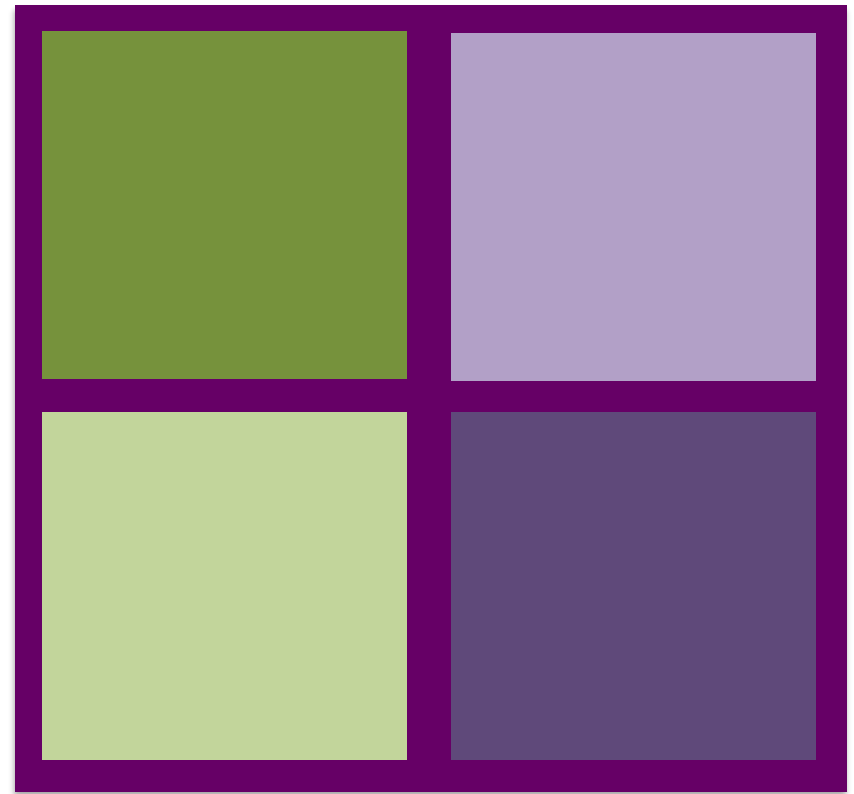
- Boris Assanovich: “We are engaged in the preparation and annotation of a Russian-language dataset based on the National Corpus of the Russian language, including collections of texts on scientific, financial and political topics, as well as authorization of existing corporate reporting documents.”



MULTI-TASK
MULTI-LINGUAL
MULTI-MODAL



Funded by the Horizon 2020 Framework Programme
of the European Union



Thanks for your attention!

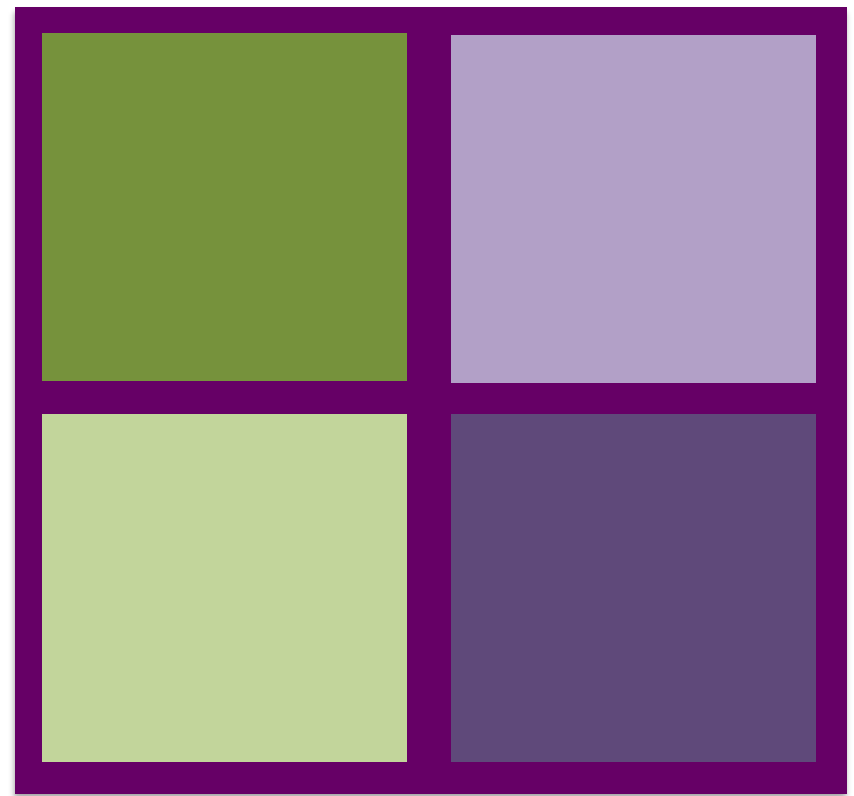
Elena Lloret, University of Alicante, Spain (WG coleader)
Aykut Erdem, Koç University, Turkey (WG leader)



MULTI-TASK
MULTI-LINGUAL
MULTI-MODAL



Funded by the Horizon 2020 Framework Programme
of the European Union



WG2 Efficient Machine Learning Algorithms, Methods, and Applications to Language Generation

Elena Lloret, University of Alicante, Spain (WG coleader)
Aykut Erdem, Koç University, Turkey (WG leader)