

**WebNLG 2020**

**Proceedings of the  
3rd International Workshop  
on  
Natural Language Generation  
from the Semantic Web  
(WebNLG+)**



**18 December 2020  
Dublin, Ireland (Virtual)**

## Sponsors

The WebNLG+ challenge was supported by the XNLG AI Chair on Multi-lingual, Multi-Source Text Generation, an initiative funded by the French National Research Agency, Facebook AI Research and the Grand East Region (Gardent; award ANR-20-CHIA-0003, XNLG "Multi-lingual, Multi-Source Text Generation").



©2020 The Association for Computational Linguistics

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)  
209 N. Eighth Street  
Stroudsburg, PA 18360  
USA  
Tel: +1-570-476-8006  
Fax: +1-570-476-0860  
[acl@aclweb.org](mailto:acl@aclweb.org)

ISBN 978-1-952148-59-0

## Preface

Welcome to the third edition of the “Workshop on Natural Language Generation from the Semantic Web” (WebNLG+).

Initiated in 2015 and tightly linked to the eponymous WebNLG Challenge, the WebNLG workshops target two goals: to investigate the micro-planning tasks involved in generating from RDF triples (document structuring, referring expression generation, aggregation, lexicalisation, surface realisation); and to explore the potential of NLG to verbalise the Semantic Web.

Different from the WebNLG 2017 challenge, WebNLG+ includes not one but four tasks: Generating from RDF triples into English and into Russian; and parsing English and Russian text into a set of RDF triples. The challenge attracted participation from 16 teams, two of which subsequently withdrew yielding a final total of 15 submitted systems and 14 participants: 6 from the private sector (Amazon Shanghai AI Lab with two systems, Facebook, Google, Huawei Noah’s Ark Lab, Orange, Sber AI Lab) and 8 from universities or public research labs (Montréal, Prague, Catalunya, São Paulo, Ireland, Ohio State, AIST). The proposed approaches include template-based methods, mono-task neural approaches, bilingual approaches and multi-task bi-directional, bilingual models. Pre-training is widely used and various data transformation (delexicalisation) and data augmentation techniques were explored.

The technical program of the WebNLG+ 2020 workshop consists of 2 long papers, 4 short papers and 15 WebNLG+ system descriptions, all of which will be presented COVID-style, either through live videos or through pre-recorded videos. The program also includes an overview of the WebNLG+ 2020 participants and the results, which will be presented by the challenge organisers.

WebNLG+ (workshop and challenge) was organised by an international group of young and less young researchers: Thiago Castro Ferreira (Federal University of Minas Gerais, Brazil), Claire Gardent (CNRS / LORIA, Nancy, France), Nikolai Ilinykh (University of Gothenburg, Sweden), Chris van der Lee (Tilburg University, The Netherlands), Simon Mille (Universitat Pompeu Fabra, Barcelona, Spain), Diego Mousallem (Paderborn University, Germany) and Anastasia Shimorina (Université de Lorraine/LORIA, Nancy, France).

We would like to thank all the people who contributed to this workshop: the authors for their submissions; the program committee for their prompt and effective reviewing; the INLG 2020 organising committee, especially the workshops chairs, Daniel Braun; the publication chair, Laura Perez-Beltrachini, and the Programme chairs, Brian Davis, Yvette Graham, John Kelleher and Gowri Sripada Pighin; and the workshop participants for sharing time and thoughts on this increasingly important research topic. Special thanks go to Yannick Parmentier (Université de Lorraine, Nancy, France) for the short notice processing of these proceedings into an ACL anthology compliant format.

Wishing you all an enjoyable, informative read!

Thiago, Claire, Nikolai, Chris, Simon, Diego and Anastasia

# Committees

## Program Committee

- Jose Maria Alonso, University of Santiago de Compostela, Spain
- Emmanuel Ayodele, Peoples' Friendship University of Russia, Russia
- Anja Belz, University of Brighton, UK
- Bernd Bohnet, Google Research, The Netherlands
- Thiago Castro Ferreira, Federal University of Minas Gerais, Brazil
- Benoît Crabbé, University of Paris, France
- David Dale, Independent researcher, Russia
- Stamatia Dasiopoulou, Independent researcher, Spain
- Henry Elder, ADAPT, Dublin City University, Ireland
- Claire Gardent, CNRS/LORIA, Nancy, France
- Sebastian Gehrmann, Google Research, USA
- David M. Howcroft, Heriot-Watt University, UK
- Nikolai Ilinykh, University of Gothenburg, Sweden
- Guy Lapalme, RALI-DIRO, Université de Montréal, Canada
- Simon Mille, Universitat Pompeu Fabra, Barcelona, Spain
- Chris Kedzie, Columbia University, USA
- Chris van der Lee, Tilburg University, The Netherlands
- Diego Moussalem, Paderborn University, Germany
- Dang Tuan Nguyen, Saigon University, Vietnam
- Alina Petrova, University of Oxford, UK
- Abhishek V. Potnis, IIT Bombay, India
- Marco Roberti, University of Turin, Italy
- Aleksander Shvets, Pompeu Fabra University, Spain
- Anastasia Shimorina, Université de Lorraine/LORIA, Nancy, France
- Marco Antonio Sobrevilla Cabezudo, University of São Paulo, Brazil
- Ashish Upadhyay, Robert Gordon University, UK
- Leo Wanner, Pompeu Fabra University, Spain
- Xiang Yu, University of Stuttgart, Germany
- Rui Zhang, Yale University, USA

## Organising committee

- Thiago Castro Ferreira, Federal University of Minas Gerais, Brazil
- Claire Gardent, CNRS/LORIA, Nancy, France
- Nikolai Ilinykh, University of Gothenburg, Sweden
- Chris van der Lee, Tilburg University, The Netherlands
- Simon Mille, Universitat Pompeu Fabra, Barcelona, Spain
- Diego Moussalem, Paderborn University, Germany
- Anastasia Shimorina, Université de Lorraine/LORIA, Nancy, France

# Table of contents

<b>I</b>	<b>Long papers</b>	<b>1</b>
	<b>A Case Study of NLG from Multimedia Data Sources: Generating Architectural Landmark Descriptions</b>	<b>2</b>
	<i>Simon Mille, Petros Alvanitopoulos, Roberto Carlini Salguero, Jens Grivolla, Montserrat Marimon Felipe, Georgios Meditskos, Maria Rousi, Klearchos Stavrothanasopoulos, Spyridon Symeonidis, Stefanos Vrochidis, Leo Wanner</i>	
	<b>OWLSIZ: An isiZulu CNL for structured knowledge validation</b>	<b>15</b>
	<i>Zola Mahlaza, C. Maria Keet</i>	
<b>II</b>	<b>Short papers</b>	<b>26</b>
	<b>A General Benchmarking Framework for Text Generation</b>	<b>27</b>
	<i>Diego Moussallem, Paramjot Kaur, Thiago Ferreira, Chris van der Lee, Anastasia Shimorina, Felix Conrads, Michael Röder, René Speck, Claire Gardent, Simon Mille, Nikolai Ilinykh, Axel-Cyrille Ngonga Ngomo</i>	
	<b>Controllable Neural Natural Language Generation: comparison of state-of-the-art control strategies</b>	<b>34</b>
	<i>Yuanmin Leng, François Portet, Cyril Labbé, Raheel Qader</i>	
	<b>Enhancing Sequence-to-Sequence Modelling for RDF triples to Natural Text</b>	<b>40</b>
	<i>Oriol Domingo, David Bergés, Roser Cantenys, Roger Creus, José A. R. Fonollosa</i>	
	<b>Utilising Knowledge Graph Embeddings for Data-to-Text Generation</b>	<b>48</b>
	<i>Nivranshu Pasricha, Mihael Arcan, Paul Buitelaar</i>	
<b>III</b>	<b>System descriptions</b>	<b>54</b>
	<b>The 2020 Bilingual, Bi-Directional WebNLG+ Shared Task: Overview and Evaluation Results (WebNLG+ 2020)</b>	<b>55</b>
	<i>Thiago Castro Ferreira, Claire Gardent, Nikolai Ilinykh, Chris van der Lee, Simon Mille, Diego Moussallem, Anastasia Shimorina</i>	
	<b>CycleGT: Unsupervised Graph-to-Text and Text-to-Graph Generation via Cycle Training</b>	<b>77</b>
	<i>Qipeng Guo, Zhijing Jin, Xipeng Qiu, Weinan Zhang, David Wipf, Zheng Zhang</i>	
	<b>Denoising Pre-Training and Data Augmentation Strategies for Enhanced RDF Verbalization with Transformers</b>	<b>89</b>
	<i>Sebastien Montella, Betty Fabre, Tanguy Urvoy, Johannes Heinecke, Lina Rojas-Barahona</i>	
	<b><math>\mathcal{P}^\infty</math> : A Plan-and-Pretrain Approach for Knowledge Graph-to-Text Generation</b>	<b>100</b>
	<i>Qipeng Guo, Zhijing Jin, Ning Dai, Xipeng Qiu, Xiangyang Xue, David Wipf, Zheng Zhang</i>	
	<b>Improving Text-to-Text Pre-trained Models for the Graph-to-Text Task</b>	<b>107</b>
	<i>Zixiaofan Yang, Arash Einolghozati, Hakan Inan, Keith Diedrick, Angela Fan, Pinar Donmez, Sonal</i>	

<b>Leveraging Large Pretrained Models for WebNLG 2020</b>	<b>117</b>
<i>Xintong Li, Aleksandre Maskharashvili, Symon Jory Stevens-Guille, Michael White</i>	
<b>Machine Translation Aided Bilingual Data-to-Text Generation and Semantic Parsing</b>	<b>125</b>
<i>Oshin Agarwal, Mihir Kale, Heming Ge, Siamak Shakeri, Rami Al-Rfou</i>	
<b>NILC at WebNLG+: Pretrained Sequence-to-Sequence Models on RDF-to-Text Generation</b>	<b>131</b>
<i>Marco Antonio Sobrevilla Cabezudo, Thiago A. S. Pardo</i>	
<b>NUIG-DSI at the WebNLG+ challenge: Leveraging Transfer Learning for RDF-to-text generation</b>	<b>137</b>
<i>Nivranshu Pasricha, Mihael Arcan, Paul Buitelaar</i>	
<b>RDFjsRealB: a Symbolic Approach for Generating Text from RDF Triples</b>	<b>144</b>
<i>Guy Lapalme</i>	
<b>Semantic Triples Verbalization with Generative Pre-Training Model</b>	<b>154</b>
<i>Pavel Blinov</i>	
<b>Text-to-Text Pre-Training Model with Plan Selection for RDF-to-Text Generation</b>	<b>159</b>
<i>Natthawut Kertkeidkachorn, Hiroya Takamura</i>	
<b>The UPC RDF-to-Text System at WebNLG Challenge 2020</b>	<b>167</b>
<i>David Bergés, Roser Cantenys, Roger Creus, Oriol Domingo, José A. R. Fonollosa</i>	
<b>Train Hard, Finetune Easy: Multilingual Denoising for RDF-to-Text Generation</b>	<b>171</b>
<i>Zdenek Kasner, Ondrej Dusek</i>	
<b>WebNLG 2020 Challenge: Semantic Template Mining for Generating References from RDF</b>	<b>177</b>
<i>Trung Tran, Dang Tuan Nguyen</i>	
<b>WebNLG Challenge 2020: Language Agnostic Delexicalisation for Multilingual RDF-to-text generation</b>	<b>186</b>
<i>Giulio Zhou, Gerasimos Lampouras</i>	

## Part I

# Long papers