

IntelLanG 2020

**Proceedings of the Workshop on Intelligent Information Processing
and Natural Language Generation**

7 September, 2020
Santiago de Compostela
Spain

©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

ISBN 978-1-952148-26-2

Introduction

We are pleased to present the proceedings of the Intelligent Information Processing and Natural Language Generation workshop (IntellLang 2020), which was held as part of the 24th European Conference on Artificial Intelligence (ECAI 2020) organized in Santiago de Compostela as a fully digital conference. The workshop was organized with the cooperation of the Spanish Network of Excellence on Intelligent Data Processing and Natural Language Generation and endorsed by SIGGEN, the ACL Special Interest Group in Natural Language Generation.

Natural Language Generation (NLG) studies systems for the automatic creation of text from non-linguistic information. The transformation of input data into the final output text involves a number of steps, each involving non-trivial choices. This is arguably the case also in neural NLG systems, where such choices are made implicitly in the case of end-to-end models, or can be handled by dedicated modules.

The use of intelligent data and information processing techniques can help in many relevant aspects of the NLG problem, for example in the contribution of formalisms for knowledge modeling and management, KDD, Data Mining and Machine Learning techniques and tools for the analysis of data, or in the development of models for the evaluation of the quality of the proposals, among many others. Artificial Intelligence information processing techniques can also gain a lot from their interaction with the particular area of NLG, as is the case with the explainable artificial intelligence research field.

The aim of this first edition of the IntellLang workshop was to identify challenges and to value current results that arise from the interaction of intelligent information processing techniques and research in natural language generation, both at the level of models and applications. The workshop provided a forum for discussion of these new research directions, with the stage set by invited talk by Professor Kees van Deemter, followed by the presentation of four long papers and four short papers which are collected in these proceedings.

Kees van Deemter's invited talk took as its starting point the notion of "theoretical NLG", that is, the study of NLG as a window into language generation procedures and models which can shed light on broader questions related to language and communication. Drawing on research on referring expression generation and research on generating and on the generation of quantified phrases, the talk highlight multiple areas where collaboration between NLG researchers, theoretical linguists, logicians and cognitive scientists can lead to models whose goal is not only to perform adequately in a given domain, but to generate predictions and further our theoretical understanding of the phenomena under consideration.

In "SportSett:Basketball - A robust and maintainable dataset for Natural Language Generation", by Thomson et al., the authors investigate the data requirements for the difficult real-world problem of generating statistic-focused summaries of basketball games. For this, they introduce the Sport-Sett:Basketball database, an easy-to-use resource that allows for researchers to easily query data, from many different dimensions, for output in a variety of formats for different architectures.

In "Automatic Follow-up Question Generation for Asynchronous Interviews", Pooja Rao S B et al propose a follow-up question generation model capable of generating relevant and diverse follow-up questions. This system is based on a 3D virtual interviewing system, Maya, a virtual agent-based interviewing system equipped with verbal interactivity from follow-up question generation.

In "How are you? Introducing stress-based text tailoring", by Balloccu et al, the authors study the impact of stress on reading and interpretation of text and propose a method for tailoring a document by exploiting complexity reduction and affect enforcement. Their research is framed in the context of project

NeuroFAST, which focuses on the socio-psychological forces that could influence eating behaviour.

In "Neural Language Generation for a Turkish Task-Oriented Dialogue System", by Artun Burak Mecik, Volkan Ozer, Batuhan Bilgin, Tuna Cakar, and Seniz Demiry, the goal is to develop a Turkish task-oriented dialogue system that enables users to navigate over a map to obtain information about dinner venues according to their preferences, and make reservations based on received recommendations. This work is the first that proposes the use of a neural generation model in a Turkish conversational system.

In "Analyzing daily behaviours from wearable trackers using linguistic protoforms and fuzzy clustering", by Martinez-Cruz et al, a methodology is proposed for analyzing common activity patterns on the basis of data provided by wearable devices, based on the use of linguistic protoforms inspired by Zadeh's Computing with Words and Perceptions paradigm, and fuzzy clustering. The methodology has been illustrated by means of a case study conducted for 200 days using the Fitbit device, recording HKIs related to duration of sleep stages and heart rate, in order to analyze restlessness patterns during sleep.

In "FitChat: Conversational AI for active aging", Wiratunga et al introduce the FitChat conversational bot, intended to encourage users to improve their physical activities. The approach uses a co-creation methodology to identify effective conversational skills by means of an iterative refinement process. The system was evaluated with seven users.

The paper "Fuzzy Logic for Vagueness Management in Referring Expression Generation" by Marín et al, provides an overview of some of the contributions regarding the use of Fuzzy Logic to referring expression generation. While fuzzy logic can capture the semantics of linguistic terms and expressions, due to the graduality associated with the fulfilment of such terms and expressions by objects, the fulfilment of referring expressions and consequently the referential success with respect to particular objects or sets becomes a matter of degree. A review of proposals for developing measures of referential success is provided, with special emphasis on those based on specificity measures.

The contribution "Iterative Neural Scoring of Validated Insight Candidates" by Susaiyah et al deals with the problem of providing comparative insights taking the form of comparative statements about the value of a certain measure in different contexts. This problem is relevant in many applications, for instance in health self-management services based on wearable devices, where statements like "On Weekdays you walk less than on Weekends" must be assessed in terms of their statistical significance, interestingness for the user, and validity for achieving the desired goals, among others. The proposal in this work is to use neural networks and transfer learning algorithms in order to assess statistical significance and, at the same time, to learn user preferences and its changes on time using an online-learning scheme.

In summary, we believe that the workshop attracted a broad spectrum of contributions, emphasising either or both of the workshop's main themes - NLG and Information Processing. Our hope is that these contributions will serve to enhance the sharing of ideas among the two communities.

Finally, we would like to thank everyone who contributed to the success of this workshop, especially the authors, the program committee members, the organizers of the ECAI 2020 conference and the ECAI 2020 workshop chairs.

Daniel Sánchez, Raquel Hervás and Albert Gatt

Granada, Madrid & Malta

September 2020

Organisers:

Daniel Sánchez, Raquel Hervás, Albert Gatt

Program Committee:

Jose M. Alonso, Anya Belz, Alberto Bugarín, Josep Carmona, Luka Eciolaza, Claire Gardent, Albert Gatt, Dimitra Gkatzia, Raquel Hervás, Henrik Leopold, Elena Lloret, Nicolás Marín, Simon Mille, Jose Angel Olivas, Ehud Reiter, Mariano Rico, Daniel Sánchez, Leo Wanner, Slawomir Zadrozny, Sina Zarriß

Invited Speakers:

Kees van Deemter, Utrecht University, The Netherlands

Invited Talk

Kees van Deemter: Restoring the link between linguistics and computation: the case of quantified expressions

Various people have observed that large areas of Natural Language Processing have grown further and further apart from the concerns of researchers whose main interest is in language and communication. In this talk I will explore the question of what an optimal collaboration between linguists and computer scientists might look like, in light of the research questions, methods, and tools that both sets of researchers can now offer. The focus of my exploration will be the notion of quantification, which has long been studied by linguists and logicians, but which has not often been the focus of work in modern NLP.

Table of Contents

Analyzing daily behaviours from wearable trackers using linguistic protoforms and fuzzy clustering	1
<i>Carmen Martinez, Javier Medina Quero, Macarena Espinilla and Sergio D. Gramajo</i>	
Automatic Follow-up Question Generation for Asynchronous Interviews	10
<i>Pooja Rao S B, Manish Agnihotri and Dinesh Babu Jayagopi</i>	
FitChat: Conversational AI for Active Ageing	21
<i>Nirmalie Wiratunga, Anjana Wijekoon, Chamath Palihawadana, Kay Cooper and Vanessa Mendham</i>	
SportSett:Basketball - A robust and maintainable data-set for Natural Language Generation	32
<i>Craig Thomson, Ehud Ehud Reiter and Somayajulu Sripada</i>	
Iterative Neural Scoring of Validated Insight Candidates	41
<i>Allmin Susaiyah, Aki Härmä, Ehud Reiter and Milan Petković</i>	
Neural Language Generation for a Turkish Task-Oriented Dialogue System	51
<i>Artun Burak Mecik, Volkan Ozer, Batuhan Bilgin, Tuna Cakar and Seniz Demir</i>	
How are you? Introducing stress-based text tailoring	62
<i>Simone Balloccu, Ehud Reiter, Alexandra Johnstone and Claire Fyfe</i>	
Fuzzy Logic for Vagueness Management in Referring Expression Generation	71
<i>Nicolás Marín, Gustavo Rivas-Gervilla and Daniel Sánchez</i>	