

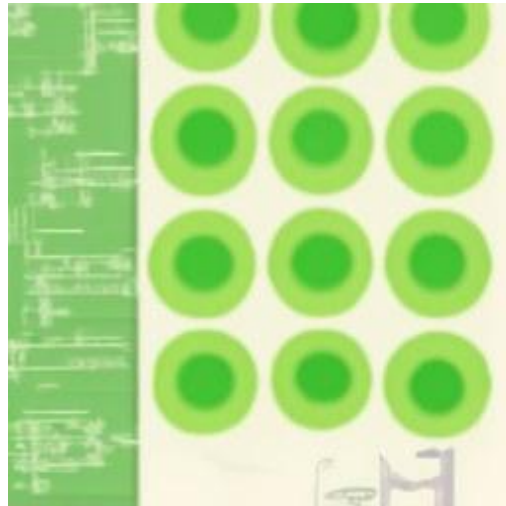
Multi3Generation 2022 Training School

Creative Natural Language Generation

19-23 September 2022

Ca' Foscari University of Venice

Sala Morelli (Calle Contarini, Dorsoduro 3484/D)



"Colorless green ideas sleep furiously",
generated by GLIDE (Nichol et al. 2021)

Monday 19/9

10:00 - 10:30 Training School 2022 registration

10:30 - 10:45 Opening

10:45 - 11:00 EU COST Action: About Multi3Generation (Action Chair Anabela Barreiro, INESC ID, Portugal)

11:00 - 12:30 Invited talk: *Computational Exercises of Creative Language* (Carlo Strapparava, FBK, Italy)

Dealing with creative language and in particular with affective, persuasive and even humorous language has often been considered outside the scope of computational linguistics, and, in general, a challenge for AI systems.

Nonetheless, it is possible to exploit current NLP techniques to address some founding aspects of these linguistic phenomena.

We briefly review some computational experiences about these typical creative genres.

12:30 - 14:00 Lunch

14:30 - 18:00 MC meeting (trainees should not attend this meeting)

Tuesday 20/9

09:00 - 11:00 A gentle introduction to creative NLG (Mika Hämmäläinen and Khalid Alnajjar, University of Helsinki & Rootroo Ltd, Finland)

An introduction to computational creativity (CC) covering theories such as the creative tripod, and H and P creativity. What is considered creative? When is a computer computationally creative? Creativity: mere surface generation or communication of a meaning? These are some of the topics that we will discuss, along with some techniques for Natural Language Generation. In the hands-on session, students will have a task of developing a simple metaphor generator. Some of the code will be provided and the data. Students can experiment with different thresholds and see how they would influence the outcome. Also, they will learn how to construct metaphorical expressions by filling in linguistic templates. Thought exercise: how creative is the metaphor generator according to the different theories covered in the morning session. Needs to be installed: Python, pip

11:00 - 11:30 Coffee break

11:30 - 12:30 A gentle introduction to creative NLG

12:30 - 14:00 Lunch

14:00 - 16:00 hands-on session

16:00 - 16:30 Coffee break

16:30 - 18:00 Invited talk: *Rule-based text generation. How to generate texts with NooJ formalizing morphosyntactic rules* (Mario Monteleone, University of Salerno, Italy)

Automatic Text Generation (ATG) is a Natural Language Processing (NLP) task that aims at writing acceptable and grammatical written text exploiting machine-representation systems, such as for instance knowledge bases, taxonomies and ontologies. In this sense, it is possible to state that an ATG system works like a translator that converts data into a natural-language written representation. The methods to produce the final texts may differ from those used by compilers, due to the inherent expressivity of natural languages.

*ATG is not a recent discipline, even if commercial ATG technology has only recently become widely available. Today, many software environments cope with ATG, as Text Spinner, DKB Lettere, or textOmatic*Composer, to mention just some of them.*

As a discipline strictly connected to NLP, ATG should be based strongly on morph-syntax formalization and semantic predicate use. However, in some cases it seems possible to avoid these steps. A simple example of ATG not involving the use of morpho-syntactic and semantic rules may be the generation of texts using only simple alphabetic letters. This method can prove itself useful when the text to generate is somehow generic in terms of semantics, and fix in terms of syntax. For instance, it can be used to generate a letter to a consumer stating that a credit card spending limit has been reached, or also to generate receipts from an ATM machine, or Social Media notifications.

However, in theory and practice the automatic generation of more complex texts can only be based on a complete system of Natural Language Formalization (NLF), as for instance Maurice Gross' Lexicon-Grammar. Therefore, in order to build an ATG procedure, we will use both Lexicon-Grammar theoretical and practical framework and Max Silberztein's NooJ NLP Environment, which as it is well known are in a strict connection. Starting from Gross' definition of semantic predicates and from the NooJ paraphrase generation routine, our aim will be to write automatically the basic plot of a novel. While achieving our aim, we will take into due account that text writing is difficult to define formally, and is probably one of the most complex challenges NLP routines can choose to tackle.

Finally, in carrying out our research, we will also make extensive references to Text Linguistics (TL) and its theoretical and practical contact points with Formal Linguistics (FL).

Wednesday 21/9

09:00 - 11:00 Developing NooJ linguistic resources for Natural Language Generation (Max Silberztein, Université de Franche-Comté, France and Cristina Mota, INESC-ID Lisboa, Portugal)

11:00 - 11:30 Coffee break

11:30 - 12:30 Developing NooJ linguistic resources for Natural Language Generation

12:30 - 14:00 Lunch

14:00 - 16:00 hands-on session

16:00 - 16:30 Coffee break

16:30 - 18:00 hands-on session

Thursday 22/9

09:00 - 10:00 Annotating empathy in texts (Ioannis Pavlopoulos, Athens University of Economics and Business, Greece and Federico Boschetti, CNR-ILC & VeDPH, Italy)

First part (I. Pavlopoulos)

This talk illustrates techniques to annotate sentiment and emotions in literary texts. The main issues are discussed and the results of experiments conducted with students at the Athens University of Economics and Business are explained.

Second part (F. Boschetti)

This talk illustrates how to create and use Domain-Specific Languages (DSLs) to annotate literary texts. The main benefits of DSLs (such as compactness, familiarity, completeness, etc.) are discussed and the Context-Free Grammar that describes EmpathyDSL is explained step by step.

Third part (I. Pavlopoulos – F. Boschetti)

Students are invited to annotate empathic expressions in two different translations of Iliad 2.211-277 (Thersites' episod) by comparing them through a Domain-Specific Language called EmpathyDSL, created for this task.

10:00 - 10:30 Coffee break

10:30 - 12:30 hands-on session

12:30 - 14:00 Lunch

14:00 - 15:00 CLARIN ERIC, a distributed infrastructure for language resources and technologies and its national node CLARIN-IT (Monica Monachini, ILC-CNR & CLARIN-IT and Francesca Frontini, ILC-CNR & CLARIN ERIC)

Multilingual Language Resources and NLP tools and services are crucial for many downstream tasks. In this presentation we will find out more about CLARIN, the common language resources and technology infrastructure, its network of data centres and its central services.

You will learn how to discover new resources, deposit and preserve newly created ones, find tools that can process and annotate them. In the hands-on session you will test important services such as the CLARIN Virtual Language Observatory and Switchboard. We will also explore the tools and resources offered by ILC4CLARIN, the national centre of the CLARIN-IT consortium.

- 15:00 - 16:00** hands-on session
16:00 - 16:30 Coffee break
16:30 - 18:00 Invited talk: *The many facets of affective expressions in multilingual text* (Alexandra Balahur, Joint Research Centre European Commission, Italy)

Friday 23/9

- 09:00 - 10:30** Small projects ideas for Short-Term Scientific Missions
10:30 - 11:00 Coffee break
11:00 - 12:30 Round table about open questions in creative NLG (moderator: Marco Guerini, FBK, Italy)
12:30 - 14:00 Lunch
14:00 - 15:30 Invited talk: *See You In The Funny Pages: Generating Comics with Meaning, Intent, and Humour* (Tony Veale, University College Dublin, Ireland)

Comic strips, the so-called “9th art“ after film and TV, sequentially integrate words and images to support more possibilities than either a text or an image alone can offer. Their visual expressivity imbues events with strong emotions, lending an intuitive immediacy to the gamut of human concerns. Here I will explore comics as an engaging means of distilling the arguments of a polarizing debate, and as a meaning representation for targeting specific content at particular audiences. I define two XML schemas for communication via comics: one to represent a story and one to represent the comic derived from it. By visualizing the heated arguments of a debate, such as the vaccines debate on Twitter, comics enable interventions that present polarized audiences with more balanced points of view. By focusing on the generation of comics that reflect both sides of a divisive issue, we can produce targeted interventions that can bridge the gulf between the opposing perspectives in an unproductive “echo chamber.”

- 15:30 - 16:00** Closing remarks



MULTI-TASK
MULTI-LINGUAL
MULTI-MODAL



Funded by the
European Union



Ca' Foscari
University
of Venice
Department
of Humanities

[ve]dph

Venice Centre for
Digital and Public
Humanities

DiPText-KC
CLARIN
K CENTRE 