Data-driven models for Natural Language Generation

Dimitra Gkatzia (Edinburgh Napier University)

The main means of human-human communication is natural language, however, for machines both understanding and producing natural language is a challenge. Natural language generation is essentially the task of producing natural language by transforming a non-linguistic representation, e.g. from an ontology, database, time-series data etc. It is often seen as the reverse of natural language understanding. Natural language generation is used in various data-to-text scenarios, where the goal is to automatically express data in a way that is understandable and useful to human users. As such, there is a lot of commercial and societal impact potential in this task as it can make data and information accessible to people that would otherwise be impossible. This could apply to users with limited expert knowledge, users with disabilities or hands-free scenarios such as driving. Sometimes data can also be too large and complex to easily access and interpret. All these scenarios benefit from natural language generation.

In this talk, I will initially introduce the field of Natural Language Generation (NLG) and two subfields: (1) data-to-text generation, the summarisation of numerical data using natural language and (2) Referring Expressions Generation, the automatic description of objects in a way that it is easy for the hearer to recognise them. I will then describe my current and past work in the field and I will introduce different models to address various challenges in NLG.