LiRA@NLG 2017

Linguistic Resources for Automatic Natural Language Generation

Santiago de Compostela, Spain

September 4, 2017

Edited by

KRISTINA KOCIJAN

PETER MACHONIS

MAX SILBERZTEIN

LiRA@NLG 2017

Proceedings of the Linguistic Resources for Automatic Natural Language Generation

Workshop

Santiago de Compostela, Spain

September 4, 2017

Edited by Kristina Kocijan, Peter Machonis & Max Silberztein

Preface

The Linguistic Resources for Automatic Natural Language Generation (LiRA@NLG) workshop of the International Natural Language Generation INLG2017 Conference held at Santiago de Compostela, September 4, 2017, brought together participants involved in developing large-coverage linguistic resources and researchers with an interest in expanding real-world Natural Language Generation (NLG) software.

Linguists and developers of NLG software have been working separately for many years: NLG researchers are typically more focused on technical issues specific to text generation—where good performance (e.g. recall and precision) is crucial—whereas linguists tend to focus on problems related to the development of exhaustive and precise resources that are mainly 'neutral' vis-a-vis any NLP application (e.g. parsing or generating sentences), using various grammatical formalisms such as NooJ, TAG or HPSG. However, recent progress in both fields is reducing many of these differences, with large-coverage linguistic resources being more and more used by robust NLP software. For instance, NLG researchers can now use large dictionaries of multiword units and expressions, and several linguistic experiments have shown the feasibility of using large phrase-structure grammars (a priori used for text parsing) in 'generation' mode to automatically produce paraphrases of sentences that are described by these grammars. The eight papers presented at the LiRA@NLG workshop focused on the following questions:

- How do we develop 'neutral' linguistic resources (dictionaries, morphological, phrase-structure and transformational grammars) that can be used both to parse and generate texts automatically?

- Is it possible to generate grammatical sentences by using linguistic data alone, i.e. with no statistical methods to remove ambiguities? What are the limitations of rule-based systems, as opposed to stochastic ones?

The common themes that these articles explore are: how to build large-coverage dictionaries and morphological grammars that can be used by NLG applications, how to integrate a linguistically-based Generation module into a Machine-Translation system, and how to construct a syntactic grammar that can be used by a transformational engine to perform paraphrase generation. Linguists as well as Computational Linguists who work on Automatic Generation based on linguistic methods will find advanced, up-to-the-minute studies on these topics in this volume:

- Max Silberztein's article, "Automatic Generation from FOAF to English: Linguistic Contribution to Web Semantics," presents an automatic system capable of generating a large number of English sentences from *Friend Of A Friend* (FOAF) statements in the RDF Turtle notation using NooJ's transformational engine both in Parse and Generation modes.

- Silvia García Méndez, Milagros Fernández Gavilanes, Enrique Costa Montenegro, Jonathan Juncal Martínez and Francisco Javier González Castaño's paper, "Lexicon for Natural Language Generation in Spanish Adapted to Alternative and Augmentative Communication," introduces *Elsa*, the first lexicon for Spanish automatically generated from a pictogram resource and tailored for Augmentative and Alternative Communication (AAC).

- Essia Bessaies, Slim Mesfar and Henda Ben Ghezala's study, "Generating Answering Patterns from Factoid Arabic Questions," presents the module that generates answers in Arabic for factoid questions, in an automatic Question Answering system.

- Kristina Kocijan, Božo Bekavac and Krešimir Šojat's contribution, "Language Generation from DB Query," demonstrates how to produce short textual summary written in Croatian from the pieces of data found in databases in the domain of airline tickets.

- Peter Machonis' article, "Using Electronic Dictionaries and NooJ to Generate Sentences Containing English Phrasal Verbs," illustrates a method for generating English sentences that contain phrasal verbs; the main difficulty is due to the fact that phrasal verbs are discontinuous.

- Hela Fehri and Sondes Dardour's study, "Generating Text with Correct Verb Conjugation: Proposal for a New Automatic Conjugator with NooJ," presents an automatic system capable of generating verbs conjugated forms in three languages: Arabic, French and English.

- Jouda Ghorbel's paper, "Formalization of Speech Verbs with NooJ for Machine Translation: the French Verb *accuser*," shows how to disambiguate French verbs (taking the verb *accuser* as an example) and generate the appropriate, exact translation into Arabic.

- Ikram Bououd and Rania Fafi's paper, "Using Serious Games to Correct French Dictations: Proposal for a New Unity3D/NooJ Connector," presents a system that generates written text from a vocal input, which can be used to help students learn French as a second language.

It is worth noting that several participants are using the NooJ software to develop the large-coverage linguistic resources needed by their NLG applications. We think that readers will appreciate the importance of this volume, both for the intrinsic value of each linguistic formalization and the underlying methodology, as well as for the potential for developing Automatic Generation applications.

Kristina Kocijan
Peter Machonis
Max Silberztein

Organizing Committee

Kristina Kocijan, University of Zagreb, Croatia

Peter Machonis, Florida International University, USA

Max Silberztein, Université de Franche-Comté, France

Scientific Committee

Héla Fehri, University of Gabes, Tunisia

Yuras Hetsevich, United Institute of Informatic Problems, Belarus

Kristina Kocijan, University of Zagreb, Croatia

Elena Lloret Pastor, Universidad de Alicante, Spain

Peter Machonis, Florida International University, USA

Slim Mesfar, University of Carthage, Tunisia

Simon Mille, Universitat Pompeu Fabra, Spain

Max Silberztein, Université de Franche-Comté, France

Table of Contents