COLING 2018

**The 27th International Conference
on Computational Linguistics**

**Proceedings of the Workshop on Linguistic Complexity and
Natural Language Processing (LC&NLP-2018)**

August 25, 2018
Santa Fe, New Mexico, USA

Copyright of each paper stays with the respective authors (or their employers).

Preface

The workshop on "Linguistic Complexity and Natural Language Processing" focuses on linguistic complexity and its relevance in the field of natural language processing. It is a cross-discipline workshop that foster exchange of ideas between people in the area of artificial intelligence and natural language processing and people dealing with natural language complexity from a cognitive or a theoretical point of view. The main objective of this workshop is to bring together researchers from different areas that have in common their interest on linguistic complexity (from a practical or theoretical perspective) boosting the interchange of knowledge and methods between specialists that have approached complexity from different viewpoints. We want to promote interdisciplinarity among researchers that are dealing with any type of language complexity.

Complexity has become an important concept in several scientific disciplines. There has been a lot of research on complexity and complex systems in the natural sciences, economics, social sciences and, now, also increasingly in linguistics. Moreover, linguistic complexity may be a key point in automatic natural language processing, since results in that field may condition the design of language technologies.

Are all languages equally complex? Does it make sense to compare the complexity of languages? Can languages differ in complexity? Complexity is a controversial concept in linguistics. Until recently, natural language complexity has not been widely researched and still not clear how complexity has to be defined and measured. Twentieth century most theoretical linguists have defended the equi-complexity dogma, which states that the total complexity of a natural language is fixed because sub-complexities in linguistic sub-systems trade off. This idea of equi-complexity, seen for decades as an unquestioned truism of linguistics, has begun to be explicitly questioned in recent years. There have been attempts to apply the concept of complexity used in other disciplines in order to find useful tools to calculate linguistic complexity. Information theory, computational models or the theory of complex systems are examples of areas that provide measures to quantitatively evaluate linguistic complexity.

Many models have been proposed to confirm or refute the hypothesis of linguistic equi-complexity. The tools, criteria and measures to quantify the level of complexity of languages vary and depend on the specific research interests and on the definition of complexity adopted. In fact, there is no agreement in the literature about how to define complexity. Instead, in the literature, we can find a variety of approaches that has led to linguistic complexity taxonomy: absolute complexity vs. relative complexity; global complexity vs. local complexity; system complexity vs. structural complexity, etc. Currently, there is no clear solution to quantify the complexity of languages and each of the proposed models has advantages and disadvantages.

The contributions to the workshop introduce new methods, models, definitions and measures to assess natural languages complexity (in human and automatic processing). They propose computational and formal approaches to linguistic complexity.

We would like to thank everyone who submitted a paper to the workshop, all the authors for their contributions, the members of the programme committee for their help in reviewing papers and, of course, all the people who attended this workshop.

Leonor Becerra-Bonache, M. Dolores Jiménez-López, Carlos Martín-Vide, Adrià Torrens-Urrutia

Organizers

Leonor Becerra-Bonache, Jean Monnet University, France
M. Dolores Jiménez-López, Universitat Rovira i Virgili, Spain
Carlos Martín-Vide, Universitat Rovira i Virgili, Spain
Adrià Torrens-Urrutia, Universitat Rovira i Virgili, Spain

Program Committee:

Leonor Becerra-Bonache, Jean Monnet University, France
Suna Bensch, Umea University, Sweden
Philippe Blache, CNRS, France
Henning Christiansen, Roskilde University, Denmark
Carmen Conti Jiménez, Universidad de Jaén, Spain
Rémi Eyraud, Aix-Marseille Université, France
Benoit Favre, Aix-Marseille Université, France
Ramon Ferrer i Cancho, Universitat Politecnica de Catalunya, Spain
Amaury Habrard, University of Saint-Etienne, France
Jeffrey Heinz, University of Delaware, USA
María del Carmen Horno Chéliz, Universidad de Zaragoza, Spain
Iraide Ibarretxe-Antuñano, Universidad de Zaragoza, Spain
François Jacquenet, University of Saint-Etienne, France
M. Dolores Jiménez-López, Universitat Rovira i Virgili, Spain
Roussanka Loukanova, Stockholm University, Sweden
Carlos Martín-Vide, Universitat Rovira i Virgili, Spain
Larry Moss, Indiana University, USA
Alexis Nasr, Aix-Marseille Université, France
Laurent Prevot, Aix-Marseille Université, France
Adrià Torrens-Urrutia, Universitat Rovira i Virgili, Spain
Michael Zock, Aix-Marseille Université, France

Table of Contents

Workshop Program

Saturday, August 25, 2018

9:30–9:45   *Opening Remarks*

9:45–10:30   **Session 1**

9:45–10:30   *A Gold Standard to Measure Relative Linguistic Complexity with a Grounded Language Learning Model*
Leonor Becerra-Bonache, Henning Christiansen and M. Dolores Jiménez-López

10:30–11:00   *Coffee break*

11:00–12:30   **Session 2**

11:00–11:45   *Computational Complexity of Natural Languages: A Reasoned Overview*
António Branco

11:45–12:30   *Modeling Violations of Selectional Restrictions with Distributional Semantics*
Emmanuele Chersoni, Adrià Torrens Urrutia, Philippe Blache and Alessandro Lenci

12:30–14:00   *Lunch*

Saturday, August 25, 2018 (continued)

14:00–15:30 **Session 3**

14:00–14:45 *Comparing morphological complexity of Spanish, Otomi and Nahuatl*
Ximena Gutierrez-Vasques and Victor Mijangos

14:45–15:30 *Uniform Information Density Effects on Syntactic Choice in Hindi*
Ayush Jain, Vishal Singh, Sidharth Ranjan, Rajakrishnan Rajkumar and Sumeet Agarwal

15:30–16:00 *Coffee break*

16:00–17:30 **Session 4**

16:00–16:45 *Investigating the importance of linguistic complexity features across different datasets related to language learning*
Ildikó Pilán and Elena Volodina

16:45–17:30 *An Approach to Measuring Complexity with a Fuzzy Grammar & Degrees of Grammaticality*
Adrià Torrens Urrutia