Computational Linguistics for Linguistic Complexity (CL4LC 2016)

Osaka, Japan
11 December 2016

ISBN: 978-1-5108-3321-0
# Table of Contents

*Could Machine Learning Shed Light on Natural Language Complexity?*
  Maria Dolores Jimenez Lopez and Leonor Becerra-Bonache ........................................... 1

*Towards a Distributional Model of Semantic Complexity*
  Emmanuele Chersoni, Philippe Blache and Alessandro Lenci ........................................... 12

*CoCoGen - Complexity Contour Generator: Automatic Assessment of Linguistic Complexity Using a Sliding-Window Technique*
  Ströbel Marcus, Elma Kerz, Daniel Wiechmann and Stella Neumann ............................... 23

*Addressing surprisal deficiencies in reading time models*
  Marten van Schijndel and William Schuler .................................................................... 32

*Towards grounding computational linguistic approaches to readability: Modeling reader-text interaction for easy and difficult texts*
  Sowmya Vajjala, Detmar Meurers, Alexander Eitel and Katharina Scheiter ...................... 38

*Memory access during incremental sentence processing causes reading time latency*
  Cory Shain, Marten van Schijndel, Richard Futrell, Edward Gibson and William Schuler ...... 49

*Reducing lexical complexity as a tool to increase text accessibility for children with dyslexia*
  Nuria Gala and Johannes Ziegler .................................................................................. 59

*Syntactic and Lexical Complexity in Italian Noncanonical Structures*
  Rodolfo Delmonte ......................................................................................................... 67

*Real Multi-Sense or Pseudo Multi-Sense: An Approach to Improve Word Representation*
  Haoyue Shi, Caihua Li and Junfeng Hu ........................................................................ 79

*A Preliminary Study of Statistically Predictive Syntactic Complexity Features and Manual Simplifications in Basque*
  Itziar Gonzalez-Dios, María Jesús Aranzabe and Arantza Díaz de Ilarraza ........................ 89

*Dynamic pause assessment of keystroke logged data for the detection of complexity in translation and monolingual text production*
  Arndt Heilmann and Stella Neumann ............................................................................. 98

*Implicit readability ranking using the latent variable of a Bayesian Probit model*
  Johan Falkenjack and Arne Jonsson .............................................................................. 104

*CTAP: A Web-Based Tool Supporting Automatic Complexity Analysis*
  Xiaobin Chen and Detmar Meurers ................................................................................ 113

*Coursebook Texts as a Helping Hand for Classifying Linguistic Complexity in Language Learners’ Writings*
  Ildikó Pilán, David Alfier and Elena Volodina .................................................................. 120

*Using Ambiguity Detection to Streamline Linguistic Annotation*
  Wajdi Zaghouni, Abdelati Hawwari, Sawsan Alqahtani, Houda Bouamor, Mahmoud Ghoneim, Mona Diab and Kemal Oflazer ................................................................. 127
Morphological Complexity Influences Verb-Object Order in Swedish Sign Language
   Johannes Bjerva and Carl Börstell ....................................................... 137

A Comparison Between Morphological Complexity Measures: Typological Data vs. Language Corpora
   Christian Bentz, Tatyana Ruzsics, Alexander Koplenig and Tanja Samardzic ................. 142

Similarity-Based Alignment of Monolingual Corpora for Text Simplification Purposes
   Sarah Albertsson, Evelina Rennes and Arne Jonsson ................................................ 154

Automatic Construction of Large Readability Corpora
   Jorge Alberto Wagner Filho, Rodrigo Wilkens and Aline Villavicencio ...................... 164

Testing the Processing Hypothesis of word order variation using a probabilistic language model
   Jelke Bloem .......................................................................................... 174

Temporal Lobes as Combinatory Engines for both Form and Meaning
   Jixing Li, Jonathan Brennan, Adam Mahar and John Hale ........................................ 186

Automatic Speech Recognition Errors as a Predictor of L2 Listening Difficulties
   Maryam Sadat Mirzaei, Kourosh Meshgi and Tatsuya Kawahara ............................... 192

Quantifying sentence complexity based on eye-tracking measures
   Abhinav Deep Singh, Poojan Mehta, Samar Husain and Rajkumar Rajakrishnan .......... 202

Upper Bound of Entropy Rate Revisited —A New Extrapolation of Compressed Large-Scale Corpora—
   Ryosuke Takahira, Kumiko Tanaka-Ishii and Łukasz Dębowski .................................. 213

Learning pressures reduce morphological complexity: Linking corpus, computational and experimental evidence
   Christian Bentz and Aleksandrs Berdicevskis ....................................................... 222