



MORGAN & CLAYPOOL PUBLISHERS

# Automatic Text Simplification

Horacio Saggion

*SYNTHESIS LECTURES ON  
HUMAN LANGUAGE TECHNOLOGIES*

Graeme Hirst, *Series Editor*



# Automatic Text Simplification

# Synthesis Lectures on Human Language Technologies

## Editor

**Graeme Hirst**, *University of Toronto*

Synthesis Lectures on Human Language Technologies is edited by Graeme Hirst of the University of Toronto. The series consists of 50- to 150-page monographs on topics relating to natural language processing, computational linguistics, information retrieval, and spoken language understanding. Emphasis is on important new techniques, on new applications, and on topics that combine two or more HLT subfields.

## Automatic Text Simplification

Horacio Saggion  
2017

## Neural Network Methods for Natural Language Processing

Yoav Goldberg  
2017

## Syntax-based Statistical Machine Translation

Philip Williams, Rico Sennrich, Matt Post, and Philipp Koehn  
2016

## Domain-Sensitive Temporal Tagging

Jannik Strötgen and Michael Gertz  
2016

## Linked Lexical Knowledge Bases: Foundations and Applications

Iryna Gurevych, Judith Eckle-Kohler, and Michael Matuschek  
2016

## Bayesian Analysis in Natural Language Processing

Shay Cohen  
2016

## Metaphor: A Computational Perspective

Tony Veale, Ekaterina Shutova, and Beata Beigman Klebanov  
2016

### Grammatical Inference for Computational Linguistics

Jeffrey Heinz, Colin de la Higuera, and Menno van Zaanen  
2015

### Automatic Detection of Verbal Deception

Eileen Fitzpatrick, Joan Bachenko, and Tommaso Fornaciari  
2015

### Natural Language Processing for Social Media

Atefeh Farzindar and Diana Inkpen  
2015

### Semantic Similarity from Natural Language and Ontology Analysis

Sébastien Harispe, Sylvie Ranwez, Stefan Janaqi, and Jacky Montmain  
2015

### Learning to Rank for Information Retrieval and Natural Language Processing, Second Edition

Hang Li  
2014

### Ontology-Based Interpretation of Natural Language

Philipp Cimiano, Christina Unger, and John McCrae  
2014

### Automated Grammatical Error Detection for Language Learners, Second Edition

Claudia Leacock, Martin Chodorow, Michael Gamon, and Joel Tetreault  
2014

### Web Corpus Construction

Roland Schäfer and Felix Bildhauer  
2013

### Recognizing Textual Entailment: Models and Applications

Ido Dagan, Dan Roth, Mark Sammons, and Fabio Massimo Zanzotto  
2013

### Linguistic Fundamentals for Natural Language Processing: 100 Essentials from Morphology and Syntax

Emily M. Bender  
2013

### Semi-Supervised Learning and Domain Adaptation in Natural Language Processing

Anders Søgaard  
2013

### Semantic Relations Between Nominals

Vivi Nastase, Preslav Nakov, Diarmuid Ó Séaghdha, and Stan Szpakowicz  
2013

### Computational Modeling of Narrative

Inderjeet Mani  
2012

### Natural Language Processing for Historical Texts

Michael Piotrowski  
2012

### Sentiment Analysis and Opinion Mining

Bing Liu  
2012

### Discourse Processing

Manfred Stede  
2011

### Bitext Alignment

Jörg Tiedemann  
2011

### Linguistic Structure Prediction

Noah A. Smith  
2011

### Learning to Rank for Information Retrieval and Natural Language Processing

Hang Li  
2011

### Computational Modeling of Human Language Acquisition

Afra Alishahi  
2010

### Introduction to Arabic Natural Language Processing

Nizar Y. Habash  
2010

### Cross-Language Information Retrieval

Jian-Yun Nie  
2010

### Automated Grammatical Error Detection for Language Learners

Claudia Leacock, Martin Chodorow, Michael Gamon, and Joel Tetreault  
2010

## Data-Intensive Text Processing with MapReduce

Jimmy Lin and Chris Dyer  
2010

## Semantic Role Labeling

Martha Palmer, Daniel Gildea, and Nianwen Xue  
2010

## Spoken Dialogue Systems

Kristiina Jokinen and Michael McTear  
2009

## Introduction to Chinese Natural Language Processing

Kam-Fai Wong, Wenjie Li, Ruifeng Xu, and Zheng-sheng Zhang  
2009

## Introduction to Linguistic Annotation and Text Analytics

Graham Wilcock  
2009

## Dependency Parsing

Sandra Kübler, Ryan McDonald, and Joakim Nivre  
2009

## Statistical Language Models for Information Retrieval

ChengXiang Zhai  
2008

Copyright © 2017 by Morgan & Claypool

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means—electronic, mechanical, photocopy, recording, or any other except for brief quotations in printed reviews, without the prior permission of the publisher.

Automatic Text Simplification

Horacio Saggion

[www.morganclaypool.com](http://www.morganclaypool.com)

ISBN: 9781627058681      paperback

ISBN: 9781627058698      ebook

DOI 10.2200/S00700ED1V01Y201602HLT032

A Publication in the Morgan & Claypool Publishers series

*SYNTHESIS LECTURES ON HUMAN LANGUAGE TECHNOLOGIES*

Lecture #32

Series Editor: Graeme Hirst, *University of Toronto*

Series ISSN

Print 1947-4040    Electronic 1947-4059



# Automatic Text Simplification

Horacio Saggion

Department of Information and Communication Technologies  
Universitat Pompeu Fabra

*SYNTHESIS LECTURES ON HUMAN LANGUAGE TECHNOLOGIES #32*



MORGAN & CLAYPOOL PUBLISHERS

## ABSTRACT

Thanks to the availability of texts on the Web in recent years, increased knowledge and information have been made available to broader audiences. However, the way in which a text is written—its vocabulary, its syntax—can be difficult to read and understand for many people, especially those with poor literacy, cognitive or linguistic impairment, or those with limited knowledge of the language of the text. Texts containing uncommon words or long and complicated sentences can be difficult to read and understand by people as well as difficult to analyze by machines. Automatic text simplification is the process of transforming a text into another text which, ideally conveying the same message, will be easier to read and understand by a broader audience. The process usually involves the replacement of difficult or unknown phrases with simpler equivalents and the transformation of long and syntactically complex sentences into shorter and less complex ones. Automatic text simplification, a research topic which started 20 years ago, now has taken on a central role in natural language processing research not only because of the interesting challenges it possesses but also because of its social implications. This book presents past and current research in text simplification, exploring key issues including automatic readability assessment, lexical simplification, and syntactic simplification. It also provides a detailed account of machine learning techniques currently used in simplification, describes full systems designed for specific languages and target audiences, and offers available resources for research and development together with text simplification evaluation techniques.

## KEYWORDS

syntactic simplification, lexical simplification, readability measures, text simplification systems, text simplification evaluation, text simplification resources