Automated Grammatical Error Detection for Language Learners

Second Edition

Synthesis Lectures on Human Language Technologies

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.

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

Automated Grammatical Error Detection for Language Learners, Second Edition

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

ISBN: 978-3-031-01025-5 paperback
ISBN: 978-3-031-02153-4 ebook

DOI 10.1007/978-3-031-02153-4

A Publication in the Springer series
SYNTHESIS LECTURES ON HUMAN LANGUAGE TECHNOLOGIES

Lecture #25
Series Editor: Graeme Hirst, *University of Toronto*
Series ISSN
Synthesis Lectures on Human Language Technologies
Print 1947-4040 Electronic 1947-4059

Automated Grammatical Error Detection for Language Learners

Second Edition

Claudia Leacock
CTB McGraw-Hill

Martin Chodorow
Hunter College and the Graduate Center, City University of New York

Michael Gamon
Microsoft Research

Joel Tetreault
Yahoo! Labs

SYNTHESIS LECTURES ON HUMAN LANGUAGE TECHNOLOGIES #25

ABSTRACT

It has been estimated that over a billion people are using or learning English as a second or foreign language, and the numbers are growing not only for English but for other languages as well. These language learners provide a burgeoning market for tools that help identify and correct learners' writing errors. Unfortunately, the errors targeted by typical commercial proofreading tools do not include those aspects of a second language that are hardest to learn.

This volume describes the types of constructions English language learners find most difficult—constructions containing prepositions, articles, and collocations. It provides an overview of the automated approaches that have been developed to identify and correct these and other classes of learner errors in a number of languages.

Error annotation and system evaluation are particularly important topics in grammatical error detection because there are no commonly accepted standards. Chapters in the book describe the options available to researchers, recommend best practices for reporting results, and present annotation and evaluation schemes.

The final chapters explore recent innovative work that opens new directions for research. It is the authors' hope that this volume will continue to contribute to the growing interest in grammatical error detection by encouraging researchers to take a closer look at the field and its many challenging problems.

KEYWORDS

grammatical error detection, statistical natural language processing, learner corpora, linguistic annotation

Martin Chodorow: To Mamie,
and to the memory of my parents.

Claudia Leacock: To my daughters,
Tess Elspeth Dougherty and Tracy Duva Dougherty.

Contents

Acknowledgments

We are grateful to Øistein E. Andersen, Aoife Cahill, Robert Dale, Michael Flor, Jennifer Foster, Ross Israel, John Lee, Nitin Madnani, Hwee Tou Ng, and an anonymous reviewer for their feedback and helpful suggestions.

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