

Quantitative Methods in the Humanities and Social Sciences

Editorial Board

Thomas DeFanti, Anthony Grafton, Thomas E. Levy, Lev Manovich,
Alyn Rockwood

Quantitative Methods in the Humanities and Social Sciences is a book series designed to foster research-based conversation with all parts of the university campus – from buildings of ivy-covered stone to technologically savvy walls of glass. Scholarship from international researchers and the esteemed editorial board represents the far-reaching applications of computational analysis, statistical models, computer-based programs, and other quantitative methods. Methods are integrated in a dialogue that is sensitive to the broader context of humanistic study and social science research. Scholars, including among others historians, archaeologists, new media specialists, classicists and linguists, promote this interdisciplinary approach. These texts teach new methodological approaches for contemporary research. Each volume exposes readers to a particular research method. Researchers and students then benefit from exposure to subtleties of the larger project or corpus of work in which the quantitative methods come to fruition.

More information about this series at <http://www.springer.com/series/11748>

Pascual Cantos-Gómez • Moisés Almela-Sánchez
Editors

Lexical Collocation Analysis

Advances and Applications

 Springer

Editors

Pascual Cantos-Gómez
Department of English
University of Murcia
Murcia, Spain

Moisés Almela-Sánchez
Department of English
University of Murcia
Murcia, Spain

ISSN 2199-0956

ISSN 2199-0964 (electronic)

Quantitative Methods in the Humanities and Social Sciences

ISBN 978-3-319-92581-3

ISBN 978-3-319-92582-0 (eBook)

<https://doi.org/10.1007/978-3-319-92582-0>

Library of Congress Control Number: 2018950661

© Springer International Publishing AG, part of Springer Nature 2018

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG

The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Introduction

Borderline phenomena are a fertile ground for scientific inquiry. They stimulate theoretical controversy and open up new opportunities for exploring innovative methodologies. The concept of collocation is illustrative of these possibilities. The special character of collocation, particularly its intermediate position between lexical and grammatical patterning, has favored an integration of perspectives of analysis that in previous stages of linguistics had belonged to separate areas of study. This integration of perspectives is proving fruitful. Six decades after the concept of collocation was introduced – it is attributed to the writings of J. Firth published in the 1950s – the range of topics explored in the literature on collocation and the sophistication of the methods proposed in this field are still far from being exhausted.

Collocational studies are, we dare say, one of the most productive areas of research over the last five decades, judging by the abundance of literature dealing with the topic and by the multiplicity of theoretical insights, methodological frameworks, and practical applications that have resulted from this field of research. The results obtained from collocational research have played a central role in the *lexicalist turn* of the last decades and in the reformulation of the boundaries between vocabulary and grammar. Concepts such as the Sinclairian *idiom principle* or Hoey's *lexical priming* are good epitomes of this tendency. So is the integration of corpus collocation studies and construction grammar, famously initiated by Gries and Stefanowitsch. The fruitfulness of collocational research is further illustrated by the diversity and the effectiveness of practical applications derived from advances in this field. Applied collocational research has produced promising results in various disciplines, including lexicography, second language teaching/learning, and computational linguistics, among others.

It is today beyond question that one of the key factors in the boosting of collocational research has been the incorporation of the new technologies into the tools of linguistic description. As Sinclair envisioned four decades ago, the use of computers and electronic corpora has facilitated the creation of ever more powerful methods of description that, in turn, have made it possible to lay bare forms

of lexico-grammatical organization that had remained unnoticed to the unaided observer. This volume lays special emphasis on the coupling of collocational research and computational corpus tools. The common denominator of the papers presented here is the use of computational corpora and quantitative techniques as a means to explore aspects of language patterning that overlap the boundaries between lexis and grammar.

The book opens with a proposal for integrating both collocational and valency phenomena within the overarching theoretical framework of construction grammar. This first chapter, by Thomas Herbst, combines insights from Bybee's usage-based approach to language, from Goldberg's construction grammar, and from Gries and Stefanowitsch's collocation analysis as a way to account for properties of both collocational patterns and valency patterns.

In Chap. 2, Violeta Seretan makes the case for integrating advances in syntactic parsing and in collocational analysis. After observing that parsing technologies and collocational research have often followed separate paths, Seretan contends that these two areas would benefit mutually from a joint approach to syntactic analysis and to collocation extraction.

Chapter 3 submits an interesting and innovative proposal for complementing corpus data and dictionaries in the identification of specific types of collocations consisting of restricted predicate-argument combinations (*collocates* and *bases*, in Hausmann's terminology). The chapter is authored by Isabel Sánchez-Berriel, Octavio Santana Suárez, Virginia Gutiérrez Rodríguez, and José Pérez Aguiar. As the authors explain, association measures face serious limitations as methods for extracting this type of collocations, which are structurally and semantically more restricted than the Sinclairian node-collocate pair. The strategy proposed by the authors of this chapter for solving this problem is to complement corpus collocational data with network analysis techniques applied to dictionary entries.

In Chap. 4, Vaclav Brezina explains the potential of collocational graphs and networks both as a visualization tool and as an analytical technique. Brezina provides three case studies showing the use of this technique in several areas of descriptive and applied linguistics, particularly in discourse analysis, language learning research, and lexicography.

In Chap. 5, Alexander Wahl and Stefan Gries propose a new, data-driven approach to the identification and extraction of multi-word expressions from corpora. The approach, termed by the acronym MERGE (Multi-word Expressions from the Recursive Grouping of Elements), is based on the selection of bigrams using log-likelihood and their successive combination into larger sequences. The results are validated via human ratings.

Finally, in Chap. 6, Peter Uhrig, Stefan Evert, and Thomas Proisl undertake a thorough analysis and evaluation of factors influencing the performance of collocation extraction methods in parsed corpora. The authors compare the impact of several factors, including parsing scheme, association measure, frequency threshold, type of corpus, and type of collocation. The results of this profound study offer valuable criteria for methodological decisions on collocation extraction.

We would like to conclude this introduction by expressing our gratitude to all the contributors to this volume for having joined us in this project and for helping to make it a reality. A word of gratitude goes also to the referees who have kindly agreed to assist us in the review process, supplying valuable feedback and advice to the authors.

Thanks are also due to Springer's staff Matthew Amboy, Editor Operations Research, for believing in this project and for his assistance and support throughout the preparation of this book, and to Faith Su, Assistant Editor, for her guidance during the production of this volume.

We are confident that this collection can contribute to the development of collocation analysis by providing an interesting illustration of the current trends in this field of research.

Universidad de Murcia, Murcia, Spain

Moisés Almela
Pascual Cantos

Contents

1 Is Language a Collostruction? A Proposal for Looking at Collocations, Valency, Argument Structure and Other Constructions	1
Thomas Herbst	
2 Bridging Collocational and Syntactic Analysis	23
Violeta Seretan	
3 Network Analysis Techniques Applied to Dictionaries for Identifying Semantics in Lexical Spanish Collocations	39
Isabel Sánchez-Berriel, Octavio Santana Suárez, Virginia Gutiérrez Rodríguez, and José Pérez Aguiar	
4 Collocation Graphs and Networks: Selected Applications	59
Vaclav Brezina	
5 Multi-word Expressions: A Novel Computational Approach to Their Bottom-Up Statistical Extraction	85
Alexander Wahl and Stefan Th. Gries	
6 Collocation Candidate Extraction from Dependency-Annotated Corpora: Exploring Differences across Parsers and Dependency Annotation Schemes	111
Peter Uhrig, Stefan Evert, and Thomas Proisl	