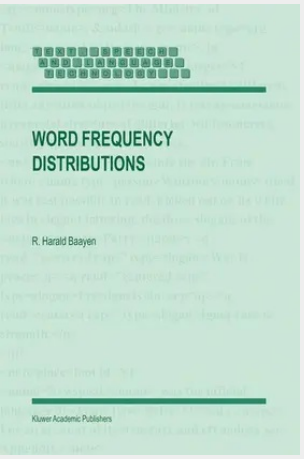
Baayen

Word Frequency Distributions

This book is an introduction to the statistical analysis of word frequency distributions, intended for linguists, psycholinguistics, and researchers work ing in the field of quantitative stylistics and anyone interested in quantitative aspects of lexical structure. Word frequency distributions are characterized by very large numbers of rare words. This property leads to strange statisti cal phenomena such as mean frequencies that systematically keep changing as the number of observations is increased, relative frequencies that even in large samples are not fully reliable estimators ofpopulationprobabilities, and model parameters that emerge as functions of the text size. Special statistical techniques for the analysis of distributions with large numbers of rare events can be found in various technical journals. The aim of this book is to make these techniques more accessible for non-specialists. Chapter 1 introduces some basic concepts and notation. Chapter 2 describes non-parametricmethods for the analysis ofword frequency distributions. The next chapterdescribes in detail three parametricmodels, the lognormal model, the Yule-Simon Zipfian model, and the generalized inverse Gauss-Poisson model. Chapter 4 introduces the concept of mixture distributions. Chapter 5 explores the effectofnon-randomness inword use on the accuracy of the non parametric and parametric models, all of which are based on the assumption that words occur independently and randomly in texts. Chapter 6 presents examples of applications.

This book is an introduction to the statistical analysis of word frequency distributions, intended for linguists, psycholinguistics, and researchers work ing in the field of quantitative stylistics and anyone interested in quantitative aspects of lexical structure. Word frequency distributions are characterized by very large numbers of rare words. This property leads to strange statisti cal phenomena such as mean frequencies that systematically keep changing as the number of observations is increased, relative frequencies that even in large samples are not fully reliable estimators ofpopulationprobabilities, and model parameters that emerge as functions of the text size. Special statistical techniques for the analysis of distributions with large numbers of rare events can be found in various technical journals. The aim of this book is to make these techniques more accessible for non-specialists. Chapter 1 introduces some basic concepts and notation. Chapter 2 describes non-parametricmethods for the analysis ofword frequency distributions. The next chapterdescribes in detail three parametricmodels, the lognormal model, the Yule-Simon Zipfian model, and the generalized inverse Gauss-Poisson model. Chapter 4 introduces the concept of mixture distributions. Chapter 5 explores the effectofnon-randomness inword use on the accuracy of the non parametric and parametric models, all of which are based on the assumption that words occur independently and randomly in texts. Chapter 6 presents examples of applications.



160,49 €
149,99 € (zzgl. MwSt.)

Lieferfrist: bis zu 10 Tage

Artikelnummer: 9780792370178
Medium: Buch
ISBN: 978-0-7923-7017-8
Verlag: Springer
Erscheinungstermin: 31.07.2001
Sprache(n): Englisch
Auflage: 2001
Serie: Text, Speech and Language Technology
Produktform: Gebunden
Gewicht: 1520 g
Seiten: 335
Format (B x H): 160 x 241 mm

Kundenservice Verlag Dr. Otto Schmidt KG
Neumannstraße 10, 40235 Düsseldorf | kundenservice@otto-schmidt.de | 0221-93738-997 (Inland)