1. Record Nr.	UNINA9910787416903321
Autore	Moisl Hermann <1949->
Titolo	Cluster analysis for corpus linguistics / / by Hermann Moisl
Pubbl/distr/stampa	Berlin ; ; Boston : , : De Gruyter, , [2015] ©2015
ISBN	3-11-036381-X 3-11-039317-4
Descrizione fisica	1 online resource (398 p.)
Collana	Quantitative linguistics ; ; 66
Classificazione	ES 900
Disciplina	410.1/880151953
Soggetti	Corpora (Linguistics) - Data processing Cluster analysis - Data processing Natural language processing (Computer science) Quantitative linguistics Computational linguistics
Lingua di pubblicazione	Inglese
Formato	Materiale a stampa
Livello bibliografico	Monografia
Note generali	Description based upon print version of record.
Nota di bibliografia	Includes bibliographical references (pages 311-358) and index.
Nota di contenuto	Front matter -- Preface -- Contents -- List of figures -- 1. Introduction -- 2. Motivation -- 3. Data -- 4. Cluster -- 5. Hypothesis generation -- 6. Literature Review -- 7. Conclusion -- 8. Appendix -- References -- Subject index
Sommario/riassunto	The standard scientific methodology in linguistics is empirical testing of falsifiable hypotheses. As such the process of hypothesis generation is central, and involves formulation of a research question about a domain of interest and statement of a hypothesis relative to it. In corpus linguistics the domain is text, and generation involves abstraction of data from text, data analysis, and formulation of a hypothesis based on inference from the results. Traditionally this process has been paper-based, but the advent of electronic text has increasingly rendered it obsolete both because the size of digital corpora is now at or beyond the limit of what can efficiently be used in the traditional way, and because the complexity of data abstracted from them can be impenetrable to understanding. Linguists are increasingly turning to mathematical and statistical computational methods for help, and cluster analysis is such a method. It is used

across the sciences for hypothesis generation by identification of structure in data which are too large or complex, or both, to be interpretable by direct inspection. This book aims to show how cluster analysis can be used for hypothesis generation in corpus linguistics, thereby contributing to a quantitative empirical methodology for the discipline.