Quantitative Methods in the Humanities and Social Sciences

Dirk Speelman
Kris Heylen
Dirk Geeraerts *Editors*

Mixed-Effects Regression Models in Linguistics

Springer

Quantitative Methods in the Humanities and Social Sciences

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

Dirk Speelman • Kris Heylen • Dirk Geeraerts
Editors

Mixed-Effects Regression Models in Linguistics

Springer

Editors
Dirk Speelman
Faculty of Arts, Research Group QLVL
KU Leuven, Belgium

Kris Heylen
Faculty of Arts, Research Group QLVL
KU Leuven, Belgium

Dirk Geeraerts
Faculty of Arts, Research Group QLVL
KU Leuven, Belgium

ISSN 2199-0956 ISSN 2199-0964 (electronic)
Quantitative Methods in the Humanities and Social Sciences
ISBN 978-3-319-69828-1 ISBN 978-3-319-69830-4 (eBook)
https://doi.org/10.1007/978-3-319-69830-4

Library of Congress Control Number: 2018930011

Printed on acid-free paper

This Springer imprint is published by the registered company Springer International Publishing AG part of Springer Nature.
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

When data consist of grouped observations or clusters, and there is a risk that measurements within the same group are not independent, group-specific random effects can be added to a regression model in order to account for such within-group associations. Regression models that contain such group-specific random effects are called mixed-effects regression models, or simply mixed models. Mixed models are a versatile tool that can handle both balanced and unbalanced datasets and that can also be applied when several layers of grouping are present in the data; these layers can either be nested or crossed.

In linguistics, as in many other fields, the use of mixed models has gained ground rapidly over the last decade. This methodological evolution enables us to build more sophisticated and arguably more realistic models but, due to its technical complexity, also introduces new challenges. This volume brings together a number of promising new evolutions in the use of mixed models in linguistics but also addresses a number of common complications, misunderstandings, and pitfalls. Topics that are covered include the use of huge datasets, dealing with non-linear relations, issues of cross-validation, and issues of model selection and complex random structures. The volume features examples from various subfields in linguistics. The book also provides R code for a wide range of analyses.

The idea for this book first arose at the 2012 Leuven Statistics Days conference, the theme of which was 'Mixed models and modern multivariate methods in linguistics' (http://lstat.kuleuven.be/research/lsd/lsd2012/index.htm). The conference took place at the KU Leuven and was co-organized by LStat (Leuven Statistics Research Centre) and the linguistic research group QLVL. We thank all conference participants for their contributions to the conference. We also thank all authors for contributing to this book, and we thank all anonymous referees for their important criticisms.

Leuven, Belgium
January 2018

Dirk Speelman
Kris Heylen
Dirk Geeraerts

Contents

1 Introduction 1
Dirk Speelman, Kris Heylen, and Dirk Geeraerts

2 Mixed Models with Emphasis on Large Data Sets 11
Geert Verbeke, Geert Molenberghs, Steffen Fieuws, and Samuel Iddi

3 The L2 Impact on Learning L3 Dutch: The L2 Distance Effect 29
Job Schepens, Frans van der Slik, and Roeland van Hout

4 Autocorrelated Errors in Experimental Data in the Language Sciences: Some Solutions Offered by Generalized Additive Mixed Models 49
R. Harald Baayen, Jacolien van Rij, Cecile de Cat and Simon Wood

5 Border Effects Among Catalan Dialects 71
Martijn Wieling, Esteve Valls, R. Harald Baayen, and John Nerbonne

6 Evaluating Logistic Mixed-Effects Models of Corpus-Linguistic Data in Light of Lexical Diffusion 99
Danielle Barth and Vsevolod Kapatsinski

7 (Non)metonymic Expressions for GOVERNMENT in Chinese: A Mixed-Effects Logistic Regression Analysis 117
Weiwei Zhang, Dirk Geeraerts, and Dirk Speelman

Chapter 1
Introduction

Dirk Speelman, Kris Heylen, and Dirk Geeraerts

Abstract As in many other fields, the use of mixed models has recently gained ground rapidly in linguistics. This methodological evolution enables us to build more sophisticated and arguably more realistic models, but, due to its technical complexity, also introduces new challenges. This volume brings together a number of promising new evolutions in the use of mixed models in linguistics, as well as addressing a number of common complications, misunderstandings, and pitfalls. Topics that are covered include the use of huge datasets, non-linear relations, issues of crossvalidation, and issues of model selection and complex random structures. The volume features examples from various linguistic subfields. This introductory chapter succinctly sketches how and why linguistic data often lend themselves to the use of mixed models and introduces the issues raised, and the topics covered, in the chapters of this volume.

1 Mixed Models

When data consist of grouped observations, and there is a risk that measurements within the same group are not independent, group-specific random effects can be added to a regression model in order to account for such within-group associations. Regression models that contain such group-specific random effects are called mixed-effects regression models, or simply mixed models. Mixed models are a versatile tool that can handle both balanced and unbalanced datasets and that can also be applied when several layers of grouping are present in the data; these layers can either be nested or crossed.

As in many other fields, the use of mixed models has recently gained ground rapidly in linguistics. This methodological evolution enables us to build more sophisticated and arguably more realistic models, but, due to its technical

D. Speelman (✉) · K. Heylen · D. Geeraerts
Faculty of Arts, Research Group QLVL, KU Leuven, Belgium
e-mail: dirk.speelman@kuleuven.be; kris.heylen@kuleuven.be; dirk.geeraerts@kuleuven.be

D. Speelman et al. (eds.), *Mixed-Effects Regression Models in Linguistics*,
Quantitative Methods in the Humanities and Social Sciences,
https://doi.org/10.1007/978-3-319-69830-4_1

complexity, also introduces new challenges. This volume brings together a number of promising new evolutions in the use of mixed models in linguistics, as well as addressing a number of common complications, misunderstandings, and pitfalls. Topics that are covered include the use of huge datasets, non-linear relations, issues of cross-validation, and issues of model selection and complex random structures. The volume features examples from various linguistic subfields.

2 Mixed Models in Linguistics

Examples of early adoptions of mixed models in linguistics can be found in different subfields of linguistics. Some examples are [5] in corpus linguistics, [2, 6, 7] in psycholinguistics, [8] in sociolinguistics, [10] in dialectometry/dialectology, etc. One publication that deserves special mention is the 2008 textbook [1], which offers a comprehensive coverage of mixed models in linguistics and has been very instrumental in the wider adoption of this technique in linguistics.

Over the last decade, mixed models have become increasingly popular in linguistics. This has happened for good reasons, since several types of grouping are very common in linguistic data. The following paragraphs list but a few examples, and certainly do not exhaust all possible types of grouping in linguistic data.

– In corpus data, when we study some linguistic variable, sometimes several attestations of that variable were produced by the same *speaker/writer*. In that case, it is possible that instances produced by the same speaker/writer are not independent. Additionally, corpus data are often sampled from a mixture of *genres/text types*, and the utterances in the corpus touch upon different *topics*. Here, again, it is possible that observations within the same genre/text type or observations that share the same topic are not independent. Unfortunately, it is not always easy to 'tag' each observation for speaker, genre or topics information; this can be difficult or even impossible, either because metadata such as speaker information are missing or because it is hard to come up with a proper, or useful, classification of things such as genre or topic. A classification that is often used as a proxy for sources of grouping that are hard to detect directly, is that of the individual *texts/documents* in a corpus. The rationale is that for a number of reasons (which themselves are often hard to identify or disentangle), it can be the case that observations from the same text/document (e.g. the same conversation) are not independent, and that it therefore makes sense to treat individual texts/documents as grouping level. Also, there is the possibility that linguistic variables behave differently depending on their immediate linguistic context. For instance, this may depend on the specific *words* the variables occur with. For instance, a syntactic variation pattern (e.g. a word order alternation pattern) may behave differently, depending on what is the specific main verb in the pattern or in the syntactic context of the pattern.

- In experimental data, often several measurements apply to the same *participant* or to the same *linguistic stimulus*, or to the same combination of both (as is often the case in longitudinal studies).
- In survey data, the *location of residence* of an informant, or his/her *location of origin*, his/her *mother tongue*, etc. can also be possible sources of observation grouping. Also, when asked several similar questions, the *informant* himself/herself can be the source of non-independence of observations.
- In the context of language and education and language acquisition, *students/pupils*, *classes*, *schools*, *cities*, etc. can all be possible sources of observation grouping.

As these examples make clear, linguistic data often lend themselves to the use of mixed models. That being said, the application of mixed models to linguistic data often is far from a trivial matter, for a number of reasons. The following list contains some of these reasons. None of these are unique to linguistics, but together they sometimes make the application of mixed models to linguistics a complicated matter.

- It can be hard to distinguish between random-effect and fixed-effect factors. The distinction is clear in prototypical cases. In a prototypical fixed-effect factor, the variable has a rather limited set of levels, both in the sample and in the population, and all the levels that occur in the population also occur in the sample. In a replication study, the levels that occur in the new sample would be the same as in the original study. In a prototypical random-effect factor, the variable typically has a very large set of levels, certainly in the population, and the levels that occur in the sample are a random subset of the levels that occur in the population. In a replication study, the levels that occur in the new sample would typically differ (to a large extent, if not completely) from those in the original study. Treating such variables as random-effect factors allows us to build models the merits of which are not confined to the specific set of levels that were attested in our sample.

  Whereas variables such as *speaker/writer*, *participant*, *informant* or *word* in many studies approximate the prototypical case of random-effect factors, the situation is rather different for things such as *genre/text type* and *topic*. For those types of variables, two competing approaches both have their merits. One approach would be to opt for coarse-grained classifications (of genres/text type or topics) that can be used as fixed-effect factors. They would have to be such that all levels are attested in the sample, and that the levels jointly cover all situations we want to model. This may imply that we can have no ambition to extrapolate our findings beyond a somewhat restricted, but still rather broad, set of contexts (e.g. three or four broadly defined genres). The alternative approach is to work with much more fine-grained classifications (or even treat each individual text as a separate 'level'), which then are treated as random-effect factors. Of course, to some extent the choice between both approaches is related to our research goals. In some studies, difference between specific genres will be at the heart of the study. In other studies, genre could be considered a nuisance variable.
- Many variables in linguistics, including response variables, are categorical, especially in corpus linguistics. Therefore, the most often used type of mixed

models is that of mixed-effects logistic regression models. From a mathematical point of view, as far as modeling random effect structures is concerned, this is not the most favorable case. Moreover, the Zipfian distribution of word frequencies (with a few very high frequency words and many very low frequency words), as well as the typically very skewed distribution of the amount of utterances per speaker/writer, tend to lead to observation groups of very different sizes. Typically, there are a few very large groups, and many very small groups (often singletons), which again is challenging from a mathematical point of view.

– A considerable number of numerical variables in linguistics are related to each other in non-linear ways (as will also be illustrated in several studies in this book). Also, measurements can show autocorrelation patterns. This is a particular concern in psycholinguistics experiments in which measurements constitute time series, but it can also be an issue in corpus data where, for instance, what happened earlier in a conversation can affect what happens later. Therefore, techniques are needed that can deal with non-linear relations and with autocorrelated patterns.
– The ever-increasing size of linguistic corpora, some of which now have clearly entered the era of big data, as well as the vast amounts of data generated in modern psycholinguistics labs, can lead to huge data sets. Applying regression techniques to such huge data sets, especially when the models are complex, can introduce computational issues.

These are some of the issues addressed in this book.

3 Mixed Models in This Book

This book offers a broad window on the use of mixed models in linguistics, with different chapters zooming in on different subfields of linguistics. Chapter 2, while not specifically discussing linguistic examples, zooms in on the analysis of huge data sets, discussing solutions that can be of great value for the analysis of the type of huge data sets that are often encountered in modern corpus linguistics. Chapter 3 zooms in on applications in second language acquisition and language and education. In Chap. 4, we look at examples from phonology, psycholinguistics and neurolinguistics. In Chap. 5, dialectometric and sociolinguistic data are discussed. In Chaps. 6 and 7, finally, examples from corpus linguistics are discussed.

Throughout the chapters, we encounter very different types of random effect structures. In Chap. 2, we look at the typical type of hierarchical structures of ‘individuals within groups’ that is often encountered in studies on language and education (e.g. pupils within classes within schools). In Chap. 3, the random-effect factors are the mother tongue and the second language of individuals who study Dutch as a third language. In Chap. 4, random-effect factors are participants and words. In Chap. 5, random-effect factors are speakers, words, and locations. In Chaps. 6 and 7, finally, random-effect factors are words.

The different chapters address different important issues related to the use of mixed models in linguistics. For instance, if we briefly revisit the issues that were listed in the previous section, we see that the issue of the sometimes difficult borderline between fixed-effect and random-effect factors plays a role both in Chap. 3 (mother tongue and second language) and Chap. 7 (genre). The specific difficulties of using and interpreting mixed-effect logistic regression analysis are specifically addressed in Chaps. 1 and 6. Non-linear relations are discussed in Chaps. 4 and 5, and autocorrelation patterns are addressed in Chap. 4. Huge data sets, finally, are specifically addressed in Chap. 2.

4 Software Used in the Book

The goal of this book is to not only discuss mixed models at a conceptual level, but to also discuss the practical usage of the technique. Nowadays, many different statistical packages, including all major commercial tools, offers good support for mixed models. In linguistics too, many different tools are being used for running mixed models. That being said, it is probably fair to say that at present, for many linguists the statistical software environment R is the tool of choice for conducting mixed-effects regression analyses. This tendency is also reflected in this book. Most chapters in this book provide R code for the types of analyses that are being discussed. The authors provide this code either by including the most important pieces of R code in the text, or by providing a URL where an R script or an R paper package can be downloaded.

5 Chapters in This Book

To conclude this first chapter, we will introduce the chapters in this book in a bit more detail.

In Chap. 2, Verbeke et al. present an introduction to mixed models (using the alternative term *clusters* to refer to grouped observations) that specifically focuses on the correct interpretation of the parameters in the models, and on possible pitfalls and misunderstandings. For instance, they illustrate that in a logistic mixed model, a type of model that is very often used in linguistics, fixed effects no longer have a population-average interpretation (as they do in linear mixed models). Instead of describing average trends in the population (across clusters), they describe trends in average clusters. Next, they illustrate that Wald tests, likelihood ratio tests, and score test statistics cannot straightforwardly be used to test whether between-cluster variability is significant (i.e. to test whether a certain random effect is needed in a model), but that instead corrections are needed (often using mixtures of χ^2 distributions). Also, they show that the distribution of empirical Bayes predictions for random effects (the so-called BLUPS) should not be used to test distributional

assumptions made in the model. They also discuss pseudo-likelihood techniques, which enable the researcher to analyze data sets that are so large that standard likelihood based inference is no longer feasible, and which therefore, in light of the ever growing size of linguistic corpora and other linguistic data collections, offer a most welcome addition to the linguist's toolset. Most examples in Chap. 2 illustrate cases with nested random effects. In the next chapter, Chap. 3, focus shifts to crossed random effects.

In Chap. 3, Schepens et al. illustrate that crossed random effects may have more complex interrelationships than is often assumed. Focus is on model selection (specifically the selection of the most appropriate random effects structure), where the authors specifically recommend that researchers compare the fit of their crossed random effects models with the fit of models that also include the respective random interaction effects. The chapter reports on a study, based on a large state examination database, of the effect of L1 (mother tongue) and L2 (second language) on the proficiency in Dutch as an L3. L1 and L2 are treated as two crossed sources of random variation. The authors want to inspect whether and how the variation across the levels of one random-effect factor (e.g. L2) depends on the levels of another random-effect factor (e.g. L1). For example: could it be that L1 Spanish learners benefit more from L2 English than L1 German learners do? One way of investigating this type of interrelatedness between random effects, which the authors claim to often be an issue in observational studies, is to incorporate an x-by-y random interaction effect, where x and y are the crossed random effects. The sample used for this study has data for 73 L1s, 44 L2s (one of which is the value 'none'), and 759 L1–L2 combinations. Model selection, as far as the random effect structure is concerned, consisted of the comparison of four models. The first model is a model with a random intercept for L1–L2 (Model 1). The second model is a model with crossed random intercepts for L1 and L2 (Model 2). Next, a model with random intercepts for both L1 and L1–L2 is inspected (Model 3). Finally, a model with the crossed random intercepts for L1 and L2, as well as an additional random intercept for the interaction effect L1–L2 is inspected (Model 4). Additionally, all these models contain the same range of fixed effects, as well as an additional random intercept for 'country of birth'. The authors argue that likelihood ratio tests and inspection of the estimated parameters indicate that Model 4 explains the data significantly better than the other models, with a larger proportion of variance being attributed to L1 factor than to L2 factors. The authors offer a detailed illustration of carefully executed model selection.

In Chap. 4, Baayen et al. address the case of responses constituting time series, which is quite common in experimental data in the field of linguistics. This situation may raise the problem of autocorrelated errors, a problem which in turn can potentially lead to anti-conservatism of p-values as well as to a more blurred window on the quantitative structure of the data. The paper illustrates two tools offered by generalized additive mixed models (gamms), as implemented by the `R` package `mgcv`, for dealing with autocorrelated errors. Generalized additive mixed models extend the generalized linear mixed model with a large array of tools for modeling nonlinear dependencies between a response variable and one or more numeric

predictors. The first tool in `mgcv` that can help us account for autocorrelated errors is the incorporation in the model of a first-order autoregressive process for the errors, which uses an autocorrelation parameter ρ. The second tool is the use of factor smooths for random-effect factors. These smooths are set up (by means of penalization) to yield the non-linear equivalents of random intercepts and random slopes in the classical linear framework.

Three cases studies are discussed. The first case study is on a word naming task; in this kind of task, participants are asked to respond to stimuli that are presented sequentially, so measurements for each participant result in a time series, and possibly a participant's later responses are not independent from his earlier responses. A model with a random intercept for `verb` (i.e. the word) and by-subject wiggly penalized curves for `trial` (i.e. position in the time series), in combination with a rather modest autoregressive parameter ρ of 0.3, is shown to almost completely account for the autocorrelation in the residuals. The second case study is on the pitch contour in the pronunciation of English three-constituent compounds. In this study, there are $12 \times 40 = 480$ elementary time series (viz. all combinations of 12 speaker and 40 English three-constituent compounds); in each elementary time series, pitch is measured at 100 moments in normalized time. The autocorrelation patterns that are much stronger than those in the first case study. Out of the three models the authors compare for this second case study, a model with by-compound and by-participant random wiggly curves as well as a high autoregressive parameter ρ of 0.98, is found to offer the best fit for the data and to remove most of the autocorrelation from the model residuals. The third case study models amplitude over time of the brain's electrophysiological response to visually presented compound words (EEG data). Focus is again on the complex random structure of the data and on the autocorrelation structure in the model residuals.

Throughout the three cases studies, the authors discuss model selection in detail. They illustrate the different ways in which the introduction of random curves and of an autoregressive parameter ρ can impact the models. Regarding the latter, they more specifically argue and illustrate that, when residuals reveal autocorrelational structure, ρ should be chosen high enough to remove substantial autocorrelational structure, but not so high that new, artificial autocorrelational structure is artefactually forced onto the data.

In Chap. 5, Wieling et al. investigate which factors influence the linguistic distances of Catalan dialectal pronunciations from standard Catalan. Using a large data set of catalan dialect pronunciation of 357 words by 320 speakers of varying age coming from 40 locations, the authors show that the speakers of Catalan in Catalonia and Andorra use a variety of Catalan that is closer to the standard than the variety spoken by speakers from Aragon. Because this tendency is particularly strong among younger speakers, they argue that the difference is at least in part due to the introduction of Catalan as an official language in the 1980s in Catalonia and Andorra but not in Aragon. As far as design is concerned, their study adopts a dialectometric approach that is enriched with social factors. More specifically, their study is dialectometric, in the sense that they aggregate over many linguistic variables, but unlike many dialectometric studies, they do

incorporate age group (which gives them a window on linguistic change) and several other social factors. The response variable in their model is pronunciation distance from the standard pronunciation, operationalized as (log-transformed and centered) normalized PMI-based Levenshtein distance. They use a generalized additive mixed-effects regression model in which geography is modeled by a non-linear interaction of longitude and latitude, and in which additionally location is included as a random-effect factor to capture location-based effects not captured by geographic coordinates. Fixed-effect predictors include word-specific variables, as well as location-specific and speaker-related social variables. Model selection is discussed in detail. The authors argue in favor of using both random intercepts and random slopes, in order to avoid anti-conservative p-values. They also argue in favor standardization of predictors. They also touch upon the issue of whether or not to use a maximally complex random-effects structure (see [3, 4]).

In Chap. 6, Barth and Kapatsinski address an issue that is very common in corpus data. If we use *word* as a random factor in the specific context of corpus data, we are faced with two specific properties of natural language: (1) word frequency distributions are such that in observational data such as corpus data, a small number of words will have exceptionally high frequencies; (2) at least according to some linguistics theories (most notably lexical diffusion theory), high frequency words tend to behave differently from other words (e.g. more articulatory reduction and semantic bleaching, and more retention of grammatical patterns that are no longer productive). The authors use simulation and cross-validation tests to investigate what are the implications of this situation for the role of random factors such as *word* in quantitative corpus linguistics research, and how they should be dealt with for the purpose of gauging fixed effects, selecting models and establishing model quality. First of all, they argue that specifically in the case of corpus-like sampling (which is bound to be unbalanced with respect to word frequencies), the inclusion of random effects is needed to obtain accurate fixed-effect coefficients. They show that corpus-like (unbalanced) sampling greatly diminishes the predictive power of fixed-effects-only models; it also hurts mixed-effects models, but to a much lesser extent. Second, they argue that whereas random factors need to be included in the models in order to more accurately capture the fixed effects in the model, at the same time it is the predictiveness of the fixed effects only that should guide model selection. In other words, they argue against using the fit (in their case, the concordance index C) of the complete model (including random effects) to evaluate mixed models, as is often done in linguistic research. Instead, evaluation of the fit of the model should be done by examining how much variance is captured by the fixed effects alone. Extrapolating their findings to the measures introduced in [9], they advocate using marginal R^2 rather than conditional R^2. This chapter can also be read as an argument in favor of cross-validation of regression models, which is a practice that is known in linguistics, but definitely is not common.

In Chap. 7, Zhang et al. present a mixed-effects logistic regression analysis by means of which they model how, in newspaper data and online forum data in Mainland Chinese and Taiwan Chinese, people choose between using either a literal or a metonymic expression when they refer to a government. Literal expressions

included in the study are two Chinese words that are the Chinese counterparts of the English words *government* and *authorities* respectively. Metonymic expressions included in the study are all usages of a PLACENAME FOR GOVERNMENT metonymy encountered in the data (typically country names, or names of capitals or official residences of state leaders or governments). The source of random variation in this study is the main verb of the sentence in which the reference to a government occurs (often, but not always, with this government being the subject of this verb). Fixed-effect predictors in the model include conceptual, grammatical/discursive and lectal variables. The study illustrates that the choice of literal vs. metonymic expressions is the result of a complex interplay of these three types of variables. Most notably, contexts favoring the use of a PLACENAME FOR GOVERNMENT metonymy are the discussion of 'general topics of global importance' (as opposed to domain-specific topics and/or topics of only local importance), the syntactic role of the government being the subject of the sentence, and the situation of the reference to the government featuring in the title of the text (rather than in the body). Two models are being presented and discussed. First, a 'global' model is fit for all data points (including references to many different governments). Then, a second model is fit for the subset of only those observations that refer to the MAINLAND CHINESE GOVERNMENT. An interesting issue regarding the random structure in the data of this study, is the question where to draw the line between what are random-effect and what are fixed-effect factor. Two factors in this paper, viz. lect/variety/style and topic, are factors that in linguistics often are treated as random-effect factors, but that here, in light of the specific research goals (specifically, the lack of desire to extrapolate beyond the topics and lects studied here) are treated as fixed-effect factors.

Acknowledgements The idea for this book arose at the 2012 Leuven Statistics Days conference, the theme of which was "Mixed models and modern multivariate methods in linguistics" (http://lstat.kuleuven.be/research/lsd/lsd2012/index.htm). The conference took place at the KU Leuven and was co-organized by LStat (Leuven Statistics Research Centre) and the linguistic research group QLVL. We thank all conference participants for their contributions to the conference. We also thank all authors for contributing to this book, and we thank all anonymous referees for their important criticisms.

References

1. Baayen RH (2008) Analyzing linguistic data: a practical introduction to Statistics using R. Cambridge University Press, Cambridge
2. Baayen RH, Davidson DJ, Bates DM (2008) Mixed-effects modeling with crossed random effects for subjects and items. J Mem Lang 59(4):390–412
3. Baayen RH, Vasishth S, Bates D, Kliegl R (2015) Out of the cage of shadows. arxiv.org. http://arxiv.org/abs/1511.03120
4. Barr DJ, Levy R, Scheepers C, Tily HJ (2013) Random effects structure for confirmatory hypothesis testing: keep it maximal. J Mem Lang 68(3):255–278