Applying advanced quantitative methods in bi-/multilingualism

Article

Accepted Version

AUTHORS' ACCEPTED VERSION

Applying Advanced Quantitative Methods in Bi-/Multilingualism

Christos Pliatsikas[1,2], George Pontikas[1], Ian Cunnings[1]

1. School of Psychology and Clinical Language Studies, University of Reading
2. Centro de Investigación Nebrija en Cognición, Universidad Nebrija, Madrid, Spain

Address for correspondence:

Christos Pliatsikas
c.pliatsikas@reading.ac.uk
School of Psychology and Clinical Language Sciences, University of Reading, Harry Pitt Building, Whiteknights, Reading, UK, RG6 6ES.

Advances in any field of scientific enquiry, be them theoretical or applied, rely on robust research methods and statistical inference. Linguistics broadly defined, and second language acquisition and bi-/multilingualism in particular, are no exceptions in this regard. Indeed, over the last decade or so there has been a step-change in the quantitative methods used within second language research and bi-/multilingualism. This special issue comprises six papers that highlight how advanced methods can benefit the bi-/multilingual research community, ranging from advanced statistics to advanced experimental methods and approaches to data handling.

While statistical tests such as t-tests and ANOVA were prevalent in the field only a decade or so ago, use of mixed-effects models has increased considerably over recent years (Cunnings, 2012). This was motivated by the need to overcome long-standing problems in generalising study findings to both study participants and the linguistic materials tested, but also to provide a framework for analysing different types of data, including continuous and binary dependent variables, and for modelling different types of fixed and random effects. Nevertheless, research practices in bi-/multilingualism continue to evolve. For example, a key issue in studying any bi-/multilingual language development is modelling change over time. A limitation of traditional approaches to such longitudinal analysis is the assumption of linearity over time, an assumption that is not appropriate for studying u-shaped curves in development, visual world eye-tracking data or neuroimaging techniques, for example. In this special issue, Coretta and Casillas (2025) propose using General Additive Models to address this problem and provide an introductory tutorial with concrete examples and the R code on how to do so. The practical includes case studies modelling U-curve developmental trajectories in L2 learning and L2 vowel production by simultaneous and sequential bilinguals and demonstrates how to fit both univariable and more complex multivariable hierarchical models.

Additionally, although frequentist statistical analyses remain dominant within the field, some researchers have begun to adopt Bayesian methods, which have the ability to quantify the evidence for a range of possible effect values, given the data, rather than merely classifying an effect as being ‘significant’ or not. Veríssimo (2025) introduces the reader to this approach by reanalysing available datasets and outlining the conceptual differences between a Bayesian approach and frequentist statistics. Veríssimo discusses Bayesian statistics alongside a significant change in statistical practice, the shift from significance testing to estimation which is well served by a Bayesian approach.

He et al. (2025) look at how data-driven approaches can benefit the field. Research in bi-/multilingualism is typically conducted in two, equally important, ways: either by running comparisons between different groups (e.g., between first and second language users, or between monolingual, bilingual and/or multilingual language groups), or by focusing on the heterogeneity of mono-/bi-/multilingual experiences, which are usually devised as continuous factors (e.g., L2 immersion, age of acquisition etc.). However, a key difficulty with analysing multiple individual differences is potential multicollinearity, while groupings may rely on arbitrary cutoffs (e.g., between lower vs. higher proficiency L2 learners). He and colleagues demonstrate how using principal component analysis can help minimise multicollinearity by transforming multiple correlated predictors into uncorrelated principal components, and how non-hierarchical cluster analysis can be used to group participants using data-driven groupings rather than arbitrary cutoffs. In particular, they provide an example of using these techniques in the analysis of eye-tracking during reading data.

In terms of advanced approaches to data handing, Keating (2025) points out that power data transformations are essential when working with timed measures such as reaction times, as they tend to violate the assumptions of typically used linear models. Keating uses eye-tracking data to present an R primer for the use of the Box-Cox coefficient (Box & Cox, 1964),

a powerful and easy to implement transformation that can successfully reduce the positive skewness typically observed in timed measures.

In evaluating new and underused methodologies in the field, Bramlett and Wiener (2025) validate the use of the web-based visual world paradigm by conducting an in-person visual world eye-tracking study with webcams. This method is gaining in popularity and has immense potential to enhance open science and to include diverse groups of hard-to-reach participants. However, there are significant technical challenges which require careful experimental design. Bramlett and Wiener consider these challenges and, in a hands-on tutorial on data (pre)processing, or wrangling, recommend best practices around quality control of data, time adjustments and binning of the data.

Finally, Khoe and Frank (2025) look at cognitive computational models of sentence processing and their potential uses. After reviewing evidence from work using the Bilingual Dual-path model for sentence production (Tsoukala et al., 2021), they use this model to conduct new simulations which show evidence for structural priming between two typologically unrelated languages (Indonesian and Dutch). This finding bears resemblance to effects reported for language pairs that are more related typologically. This demonstrates the predictive value of such models and they opportunities they provide to test hypotheses about linguistic contexts that are not typically tested in lab-based experiments.

In sum, in this special issue we present a collection of papers introducing advanced methods and analyses which will allow bi-/multilingual researchers to examine new research questions and provide novel ways of shedding new light on long-standing debates in the field. In so doing, we hope to engender a positive shift in the use of statistical methods within the field.

References

Box, G. E. P., & Cox, D. R. (1964). An Analysis of Transformations. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *26*(2), 211–243. https://doi.org/10.1111/j.2517-6161.1964.tb00553.x

Bramlett, A. A., & Wiener, S. (2025). The art of wrangling: Working with web-based visual world paradigm eye-tracking data in language research. *Linguistic Approaches to Bilingualism*. https://doi.org/10.1075/lab.23071.bra

Coretta, S., & Casillas, J. V. (2025). A tutorial on generalised additive mixed effects models for bilingualism research. *Linguistic Approaches to Bilingualism*. https://doi.org/10.1075/lab.23076.cor

Cunnings, I. (2012). An overview of mixed-effects statistical models for second language researchers. *Second Language Research*, *28*(3), 369–382. https://doi.org/10.1177/0267658312443651

He, S., Divjak, D., & Milin, P. (2025). Optimising Participant Grouping Methods in Bilingualism Studies: Insights from Eye-Tracking Data. *Linguistic Approaches to Bilingualism*.

Keating, G. D. (2025). Normalization of timed measures in bilingualism research: Make it optimal with the Box-Cox transformation. *Linguistic Approaches to Bilingualism*. https://doi.org/10.1075/lab.24017.kea

Khoe, Y. H., & Frank, S. L. (2025). The Bilingual Dual-path model: Simulating bilingual production, processing, and development. *Linguistic Approaches to Bilingualism*. https://doi.org/10.1075/lab.23072.kho

Tsoukala, C., Broersma, M., van den Bosch, A., & Frank, S. L. (2021). Simulating Code-switching Using a Neural Network Model of Bilingual Sentence Production. *Computational Brain & Behavior*, *4*(1), 87–100. https://doi.org/10.1007/s42113-020-00088-6

Veríssimo, J. (2025). A gentle introduction to Bayesian statistics, with applications to bilingualism research. *Linguistic Approaches to Bilingualism*.