Bilingualism: Language and Cognition

cambridge.org/bil

Nonlinearities in bilingual visual word recognition: An introduction to generalized additive modeling

Koji Miwa[1] and Harald Baayen[2]

[1]Graduate School of Humanities, Nagoya University, Nagoya and [2]Department of Linguistics, Eberhard Karls University Tuebingen, Tuebingen

Review Article

Cite this article: Miwa K, Baayen H (2021). Nonlinearities in bilingual visual word recognition: An introduction to generalized additive modeling. *Bilingualism: Language and Cognition* **24**, 825–832. https://doi.org/10.1017/S1366728921000079

Received: 25 January 2020
Revised: 3 February 2021
Accepted: 3 February 2021
First published online: 17 March 2021

Keywords:
nonlinearity; generalized additive modeling; quantile regression; cognate processing; visual word recognition

Address for correspondence:
Koji Miwa, Bunkeisogokan Room 711, Graduate School of Humanities, Nagoya University, Foro-cho, Chikusa-ku Nagoya, 464-8601, Japan Email: kojimiwa@nagoya-u.jp

Abstract

This paper introduces the generalized additive mixed model (GAMM) and the quantile generalized additive mixed model (QGAMM) through reanalyses of bilinguals' lexical decision data from Dijkstra et al. (2010) and Miwa et al. (2014). We illustrate how regression splines can be used to test for nonlinear effects of cross-language similarity in form as well as for controlling experimental trial effects. We further illustrate the tensor product smooth for a nonlinear interaction between cross-language semantic similarity and word frequency. Finally, we show how the QGAMM helps clarify whether the effect of a particular predictor is constant across distributions of RTs.

Introduction

In bilingual visual word recognition research, a central question has been whether readers activate two languages automatically when reading in one language. To address this question, researchers have investigated whether and, if so, how cross-language similarities in orthography, phonology, and semantics co-determine word processing speed. Significant contributions of cross-language similarity measures to response times provides evidence that both languages are automatically co-activated in reading.

Up until a decade ago, cross-language similarities were typically coded as categorical predictors and tested with by-participant (F_1) and by-item (F_2) ANOVAs (e.g., high vs. low cross-language orthographic similarity, as in Dijkstra, Grainger & van Heuven, 1999). More recently, linear mixed-effects modeling (LMM, Baayen, Davidson & Bates, 2008) replaced the classical ANOVA, with the advantage that a single statistical model can now include both participants and items as (crossed) random effects. The LMM also enabled researchers to move away from unnecessary dichotomization of continuous predictors (for orthographic similarity as a continuous rather than a factorial predictor, see Dijkstra, Miwa, Brummelhuis, Sappeli & Baayen, 2010) and further allowed them to include covariates in the regression model, as opposed to seeking to match items a priori on all dimensions that might potentially be confounded with the factorial manipulation of interest (see, e.g., Baayen, 2010). In this tutorial introduction, we outline more recent mathematical models that offer further precision for understanding experimental data in psycholinguistics: the generalized additive mixed model (GAMM) and the quantile generalized additive mixed model (QGAMM). Whereas the LMM provided a step forward compared to the classical ANOVA, it is restricted in the sense that it assumes that the effects of continuous regressors are linear.

The GAMM relaxes this linearity assumption and offers researchers further flexibility to detect nonlinear trends in their data. As will be shown, nonlinearity is ubiquitous in language studies. The GAMM is capable of handling an interaction between a nonlinear predictor and a factor, as well as an interaction between two nonlinear predictors. The QGAMM is a recent extension of the GAMM that enables constructing models for any desired quantile of the response variable's distribution.

Throughout this paper, familiarity with the LMM and basic knowledge in R are assumed (see Pinheiro & Bates, 2000 and Baayen, Davidson & Bates, 2008 for introduction of the LMM with R). In what follows, we make use of treatment coding for factors. We first demonstrate how the GAMM can be applied in bilingual processing research, reanalyzing lexical decision data from Dijkstra et al. (2010, Experiment 1) and Miwa, Dijkstra, Bolger, and Baayen (2014, Experiments 1 and 2). The former study tested 21 Dutch-English bilinguals reading 194 English words and 194 nonwords in a lexical decision experiment. We reanalyzed data for 189 target words. The latter study tested 19 Japanese-English bilinguals and 19 English monolinguals reading English words in a lexical decision experiment with 250 words and 200 nonwords. This time we reanalyzed data for 228 target words unless otherwise noted. All

© The Author(s), 2021. Published by Cambridge University Press

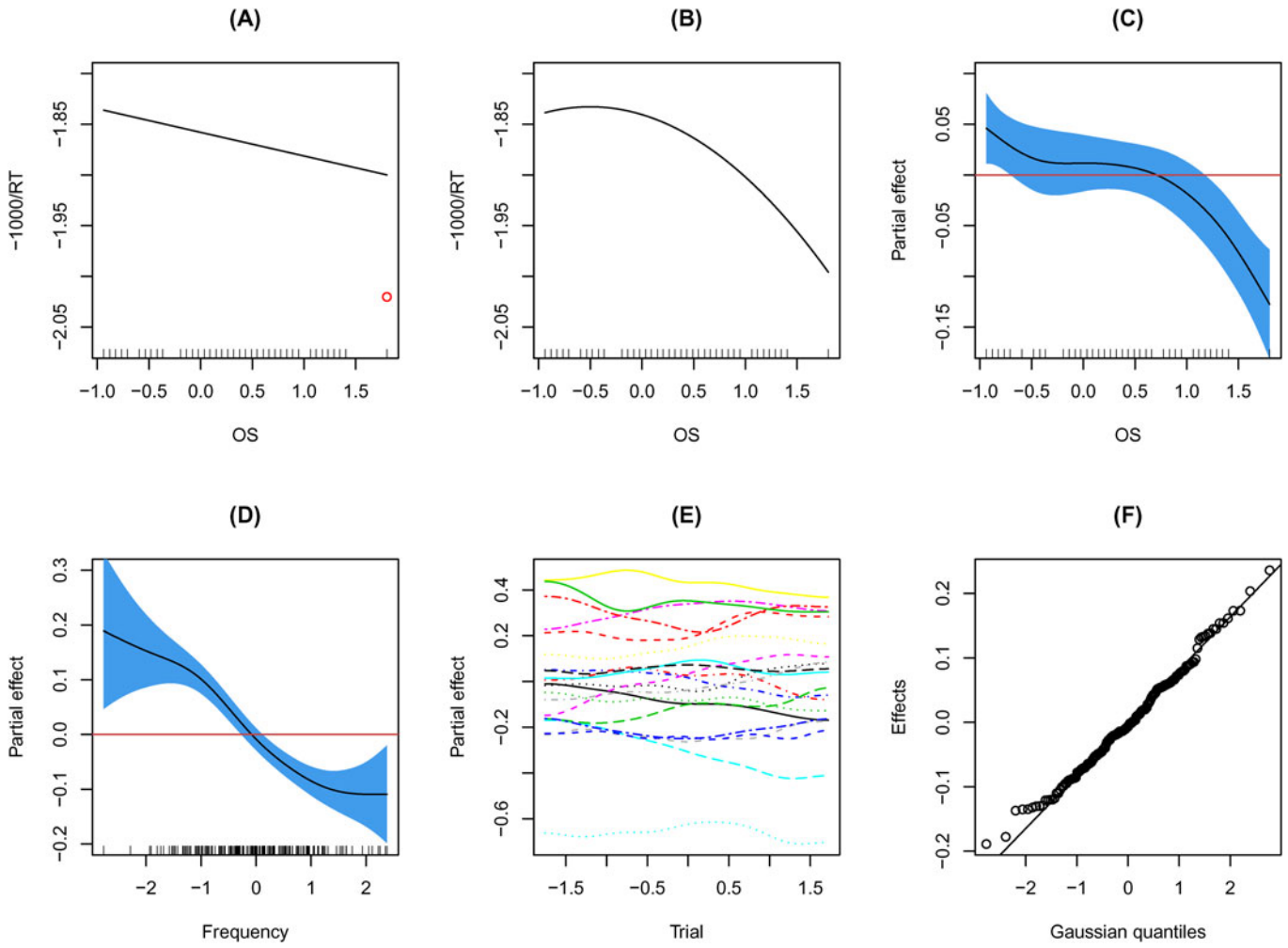


Fig. 1. LMMs and a GAMM fitted to Dutch-English bilinguals' RTs. (A) the effects of *OS* and *Ident* (red circle) in a LMM, (B) the quadratic effect of *OS* in a LMM, (C) the nonlinear effect of *OS* in a GAMM, (D) the nonlinear effect of Frequency in a GAMM, (E) by-Participant random wiggly curves for Trials in a GAMM, (F) the distribution of the random intercepts. Note that all the predictors are standardized and that the presented effects are partial effects (i.e., the contribution of the predictor to the fitted values, independently of the contributions of the other predictors in the model)

numerical predictors were standardized. R code for the analyses presented below is available in the supplementary material (https://osf.io/g5ax4/).

Nonlinearity in cross-language form similarity effects

In the LMM analysis of Dijkstra et al. (2010), a rated orthographic similarity (rated on a 7-point Likert scale, henceforth *OS*) between target English words and their Dutch translation equivalents (e.g., tomate – tomaat) was included as a covariate, together with a factor (*Ident*) coding for whether an English word had an identical cognate in Dutch. Identical cognates revealed substantial facilitation on top of a facilitatory effect of *OS*, as shown in Figure 1, Panel A. From a modeling perspective, the factor *Ident* is somewhat awkward, as it marks the maximum value of *OS*. Within the context of the LMM, the effect of *OS* can be modeled as nonlinear by regressing the response on powers of *OS* (e.g., *OS* squared, cubed etc.). Panel B of Figure 1 shows the effect of *OS* when its functional form is given by a second-degree polynomial. In this model, *Ident* is not included as predictor. Model comparison clarifies that the polynomial model lacks precision when pitted against the model with a linear effect of *OS* in combination with *Ident*; the substantial facilitation for *Ident* is no longer visible.

The generalized additive mixed model (GAMM) offers a toolkit for building a more precise statistical model for the present dataset. In what follows, we make use of the `mgcv` package (Wood, 2003; 2004; 2017). An introduction to the main concepts underlying GAMMs is available in Baayen, Vasishth, Kliegl, and Bates (2017). Central to GAMMs is the concept of a spline, a function that approximates a wiggly curve by means of a weighted sum of simple nonlinear functions known as basis functions. The more wiggly a curve is, the more basis functions are required to approximate it. In order to avoid overfitting, GAMMs implement a penalty for wiggliness, the assumption being that the truth is more likely to be simple (less wiggly) than complex (highly wiggly). Penalization reduces the weights of the basis functions. When the weight of a basis function is reduced to zero, it no longer contributes to the model. Often, penalized weights are substantially reduced, but not zero, compared to unpenalized weights, which would provide the closest fit to the data, but at the price of overfitting. The effective degrees of freedom (edf) of a spline function, which is used to evaluate significance, will be larger when more basis functions are required and when these basis functions have larger weights. When the edf of a spline is close to 1, the

https://doi.org/10.1017/S1366728921000079 Published online by Cambridge University Press

functional shape of the spline will be very similar to a straight line. Importantly, the penalization method implemented in GAMMs ensures that when the functional relation between a response and a predictor is in fact linear, the model will detect linearity and will eliminate all wiggly basis functions by driving their weights to zero.

Like the LMM, the GAMM allows researchers to include participants and items as crossed random effects. Within the context of the LMM, one can allow a regression line to have intercepts and slopes that differ by subject (or item) by setting up a population intercept and slope, and adding by participant (or by-item) adjustments for both intercept and slope. Within the context of the GAMM, it is possible to set up the nonlinear equivalent of random intercepts and random slopes by means of special splines known as factor smooths. Factor smooths implement shrinkage for wiggly curves, just as the LMM implements shrinkage for random intercepts and random slopes.

As a first illustration, we fit a GAMM to the data of Dijkstra et al. (2010). We set the model up in such a way that the effects of *OS* and word frequency (here, we use the standardized log English subtitle frequency of Brysbaert & New (2009), henceforth *Frequency*) can be nonlinear. If it turns out that these effects are actually linear, the edf values for these parameters should become 1. As participants often show nonlinear trends over experimental time, we set up a by-participant factor smooth for trial number (*Trial*), in order to incorporate ups and downs in attention and motivation as the experiment unfolds (see Baayen et al., 2017 for further discussion)[1]. This factor smooth also estimates the "intercepts" (the points where the wiggly curves cross the Y-axis), so it is not necessary to request separate by-subject random intercepts. We do request by-word random intercepts. The R code for this model,

```
> dijkstra.gam = bam(invRT ~ s(OS) +
                       s(Frequency) +
                       s(Trial, Participant, bs="fs", m = 1) +
                       s(Word, bs="re"),
                     data = dijkstra, discrete = TRUE)
```

requests regression splines with the `s()` directive. For factor smooths, additional arguments are required: in addition to the continuous variable (*Trial*), we specify the grouping factor (*Participant*), ask for a factor smooth with `bs="fs"`, and request shrinkage with `m = 1`. By-word random intercepts are also set up with the `s()` directive, with the basis function parameter set to "`re`". The directive "`discrete = TRUE`" requests an algorithm that can substantially reduce computation time with hardly any loss of accuracy. The model summary, obtained with `summary(dijkstra.gam)`, is presented in Table 1.

Unlike the LMM, a GAMM summary has two parts: a parametric part for linear terms and a nonparametric part for smooth terms. Note that random effects are listed together with fixed effects. An F-test (detailed in Wood, 2013) is reported that clarifies whether a smooth term provides a noteworthy contribution to the model fit. Comparison of models with and without a given smooth term typically leads to the same conclusions as this F-test. For this model, the only parametric parameter is the intercept, estimated at -1.88. In the non-parametric part of the summary, we see that the effective degrees of freedom (edf) for the two covariates, *OS* and *Frequency*, are close to 3, indicating some wiggliness. To interpret the smooth terms of the model visualization is essential. Figure 1, Panel C presents the effect of *OS*. Where the confidence interval does not include zero (the red line), the effect is significant. Thus, for the smallest values of orthographic similarity, we find longer RTs, for most of the range of OS, there is no effect, and then for the higher values of *OS*, we find shorter RTs. The effect shown is the partial effect of the predictor. Panel D visualizes the effect of word frequency, which is strong in the middle range of log frequency, and tapers off at both tails of the distribution. Panel E presents the by-participant random curves for *Trial*. The main difference between participants that strikes the eye is their intercept, which differentiates between slower and faster subjects. But there is considerable variability between participants, with some showing stable behavior, with other showing nearly linear trends up or down, and some others showing undulating patterns suggestive of fluctuations in attention. Finally, Panel E presents a quantile-quantile plot for the model residuals, which roughly follow a normal distribution, as required.[2]

Table 1. Summary of the GAMM fitted to lexical decision RTs of Dijkstra et al. (2010). *Note*: s(OS): spline for OS, s(Frequency): spline for *Frequency*, s(Trial, Participant): factor smooths for *Trial* by *Participant*, s(Word): by-word random intercepts.

Parametric terms	Estimate	SE	t-value	p-value
(Intercept)	−1.88	0.06	−32.60	<.001
Smooth terms	**edf**	**Ref.df**	**F**	**p-value**
s(OS)	2.75	2.94	11.63	<.001
s(Frequency)	3.17	3.44	23.24	<.001
s(Trial, Participant)	65.67	189.00	12.76	<.001
s(Word)	113.60	175.00	2.19	<.001

Model fit can be improved by including random slopes for *OS* and *Frequency*. The effect of these random slopes is that they will tilt, somewhat differently for each subject, the orientation of the regression curve. The terms to add to the model specified above for by-subject random slopes for frequency and *OS* are, respectively,

[1]Factor smooths for participant by trial typically account for a large part of trial-to-trial dependencies in reaction time experiments. Further control over these dependencies can be gained by incorporating in the model an AR(1) process in the residual error, see Baayen, Vasishth et al. (2017) and Baayen, van Rij, de Cat & Wood (2018).

[2]GAMs, as implemented in the mgcv package, provide many tools for dealing with datasets that violate the iid assumptions of the standard Gaussian modeling framework. When the data are not homoskedastic, the gaulss family argument enables fitting a model in which not only the mean, but also the variance, varies (potentially nonlinearly) with a predictor. When residuals follow a t-distribution rather than a normal distribution, this can be accommodated with the scat family argument. Autocorrelations in the residuals can be modeled to some extent by positing an AR(1) process in the errors. For examples, see Chuang et al. (2020) for worked examples.

https://doi.org/10.1017/S1366728921000079 Published online by Cambridge University Press

```
s(Participant, Frequency, bs = "re")
```

and

```
s(Participant, OS, bs = "re")
```

Addition of these two terms decreases the fREML scores of the model from 1064 to 1056, which is a substantial improvement, as indicated by model comparison using the `compareML()` function from the `itsadug` package (van Rij, Wieling, Baayen & van Rijn, 2017). Like the `anova()` function for the LMM, `compareML()` is used to compare nested GAMMs.

It is noteworthy that once the factor *Ident* is taken in to account, the effect of *OS* becomes linear, with a substantially increased p-value. Since model comparison, using `compareML()`, indicates that the inclusion of *Ident* provides only a modest improvement in goodness of fit (the ML score of the model decreased from 1049 to 1045), we are still left with two theoretical possibilities for future research. The first possibility is that *Ident* is not necessary as predictor, as it is confounded with the endpoint of the scale of orthographic similarity. In this case, we let the spline function do the work for us, with the implication that cognates that are spelled almost but not completely the same also have a processing advantage, albeit a smaller one than that of identical cognates. The second possibility is that *Ident* is a theoretically important predictor, the idea being that under identity there is substantial facilitation for cognates, the conclusion reached by Dijkstra et al. (2010). Statistics cannot decide which possibility is to be preferred.

Using Miwa et al.'s (2014) data obtained from Japanese-English bilinguals, we similarly fitted a GAMM to RTs to investigate an effect of cross-language phonological similarity. There was no sign of nonlinearity (see the supplementary material).

Nonlinearity in responses to words and nonwords

With the GAMM, it is possible to test for an interaction between a nonlinear predictor and a factor. In this example, using the `miwacomp` dataset (Miwa et al., 2014, 19 Japanese in Experiment 1 and 19 English monolinguals in Experiment 2, 250 words, 200 nonwords), we examined how response patterns changed throughout the experiment for words and nonwords. To understand how participants responded to words and nonwords throughout the experiment in detail, we tested how RTs changed as the experiment went by (*Trial*) for different levels of the factor *StimulusType* (levels: *Word*, *Nonword*). This was achieved with the `by`-directive within the `s()` function, which requests two wiggly curves, one for each factor level. The main effect of *StimulusType* should be included so that two wiggly curves can be set at appropriate places on the y-axis.

A comparison of the model with and without the interaction indicated that the two different wiggly curves for *Word* and *Nonword* were indeed necessary. In addition to RTs to words being overall shorter than to nonwords, participants showed different learning effects for words and nonwords throughout the experiment: nonlinear for words and linear for nonwords (Figure 2, Panels A and B, see also Table 2). To see how the two curves differ, we compute a difference curve for *Word* and *Nonword* (see the supplementary material for the procedure). Panel C indicates that the bilingual participants' RTs to *Word* and *Nonword* were equally slow at the beginning. RTs to words became rapidly shorter until about a third into the experiment. After that, the RTs to words remained faster than RTs to nonwords, but with the difference staying the same, until the end of the experiment. Possibly, bilingual participants initially made either-word-or-nonword decisions and then optimized their response criteria to make if-not-word-then-nonword decisions.

Using the same procedure, we computed the difference curve for monolingual participants (Panel D). They clearly distinguished words from nonwords from the very beginning – the difference curve is situated well above the zero-line across the whole experiment – likely because the processing of real words was highly automatized.

Nonlinear interaction involving cross-language similarity

With GAMMs, we can model interactions between two nonlinear predictors. Under the assumption that languages have semantic representations in common (see, e.g., Bilingual Interactive Activation (BIA+) model, a localist-connectionist model of bilingual visual word recognition with orthographic, phonological, and semantic representations, Dijkstra & van Heuven, 2002), cross-language semantic similarity (henceforth *SS*) is expected to facilitate word recognition. Furthermore, because bottom-up processing to the semantic level should proceed faster for higher frequency words, there may be more opportunity for *SS* to contribute for these words. Testing for an interaction of *SS* by *Frequency* requires a tensor product smooth with the `te()` function. Since *SS* is a meaningful measure only for bilinguals, we investigate a three-way interaction by requesting two wiggly regression surfaces, one for each of the two levels of *FirstLanguage*. The main effect of *FirstLanguage* should be included so that two wiggly surfaces can be set at appropriate places on the y-axis.

```
> jpn.gam1 = bam(invRT ~ StimulusType +
                         s(Trial, by = StimulusType) +
                         s(Trial, Participant, bs = "fs", m = 1) +
                         s(Word, bs = "re"),
                 data = miwacomp[miwacomp$FirstLanguage == "Japanese", ],
                 discrete = TRUE)
```

https://doi.org/10.1017/S1366728921000079 Published online by Cambridge University Press

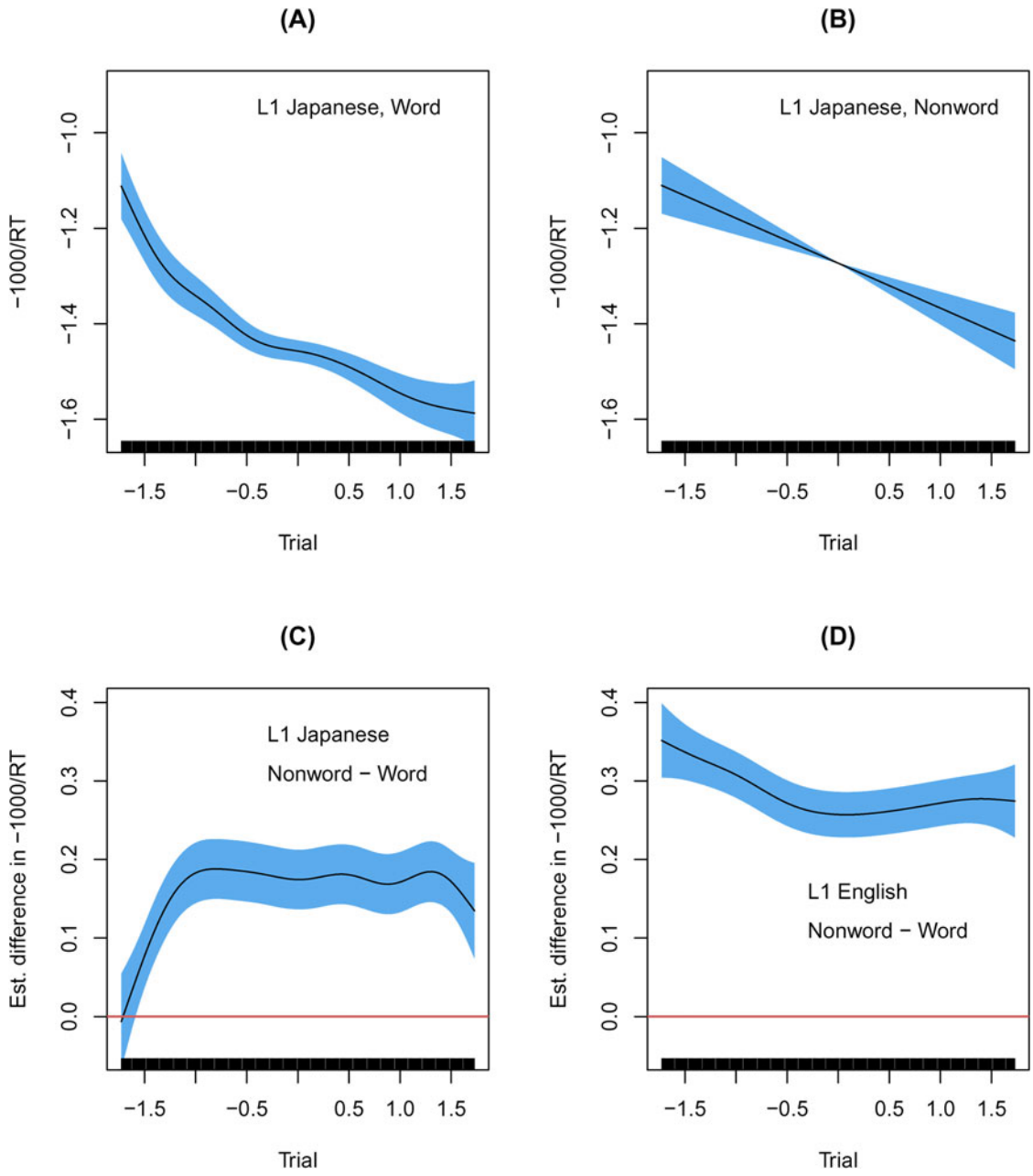


Fig. 2. Trial effects for words and nonwords observed through GAMMs. (A) the nonlinear trial effect seen in Japanese participants' responses to words, (B) the linear trial effect seen in Japanese participants' responses to nonwords, (C) the difference curve computed between the panels A and B, (D) the difference curve computed for non-native speakers of Japanese. Note that *Trial* is standardized and that the presented effects are partial effects (i.e., the contribution of the predictor to the fitted values, independently of the contributions of the other predictors in the model)

```
> miwa.gam3 = bam(invRT ~ FirstLanguage +
                          te(SS, Frequency, by = FirstLanguage) +
                          s(Trial, by = FirstLanguage) +
                          s(Trial, Participant, bs = "fs", m = 1) +
                          s(Participant, Frequency, bs = "re") +
                          s(Word, bs = "re"),
                  data = miwa, discrete = TRUE)
```

https://doi.org/10.1017/S1366728921000079 Published online by Cambridge University Press

Table 2. Summary of the GAMM fitted to lexical decision RTs of Miwa et al. (2014)

Parametric terms	Estimate	SE	t-value	p-value
(Intercept)	−1.27	0.04	−29.63	<.001
StimulusType Word	−0.16	0.01	−11.39	<.001
Smooth terms	**edf**	**Ref.df**	**F**	**p-value**
s(Trial):StimulusType Nonword	1.00	1.00	30.48	<.001
s(Trial):StimulusType Word	5.93	6.96	13.94	<.001
s(Trial, Participant)	114.92	170.00	25.65	<.001
s(Word)	374.28	448.00	5.06	<.001

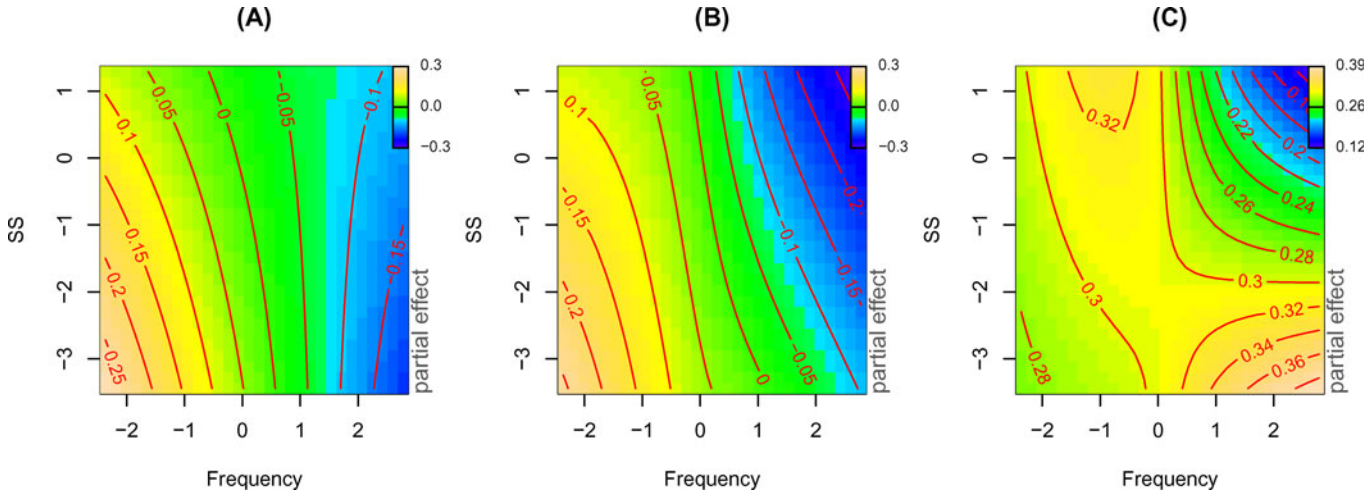


Fig. 3. Nonlinear interactions between semantic similarity and target word frequency observed through GAMMs. (A) the nonlinear interaction between SS and Frequency for non-native speakers of Japanese, (B) the nonlinear interaction between SS and Frequency for native speakers of Japanese, (C) the difference surface computed between the panels A and B. Note: colder color indicates shorter RTs.

Both wiggly surfaces were evaluated as significant, and are visualized in Figure 3, Panel A for monolinguals and Panel B for bilinguals. These figures should be interpreted like a topographic contour map, with colder colors indicating shorter RTs. Unexpectedly, monolingual participants were also sensitive to some aspect of *SS*, possibly because when words share semantics across languages, they have more prototypical meanings. Here, we focus on how bilinguals made use of *SS* and *Frequency*, with monolinguals as the baseline of comparison. We therefore created a difference surface by subtracting the monolinguals' surface from the bilinguals' (see the supplementary material for the procedure). The difference surface was significant and is visualized in Panel C in Figure 3. For bilingual participants, as can be seen in the right half of the difference surface, *SS* afforded more facilitation compared to monolinguals, specifically for high-frequency words.

Response times observed through the additive quantile regression

Just like LMMs, GAMMs estimate the mean, or more precisely, the expectation of the mean of the response. However, effects of predictors are not necessarily uniform across the distribution of response times. Effects may primarily affect early reaction times, or conversely, play a role primarily during late decision stages (see, e.g., Ratcliff, 1979; Schmidtke, Matsuki & Kuperman, 2017). Quantile regression makes it possible to predict any desired quantile of a response distribution. Thus, at the 0.1 decile, one can study early effects, at the median one can study the central tendency, and at the 0.9 decile, late effects. When fitting a model for a specific quantile, all datapoints in the distribution are taken into account, but datapoints far away from a given quantile are given less weight. By drawing multiple regression curves at different quantiles of the dependent variable, the effects of a predictor across the distribution of RTs can be gauged.

In the following example, we illustrate QGAMMs (Fasiolo, Goude, Nedellec & Wood, 2017). Importantly, QGAMMs do not make any distributional assumptions about the residuals. Thus, no transformations of the response are required. The following model, fitted with `mqgam()` function of the `qgam` package (Fasiolo et al., 2017) predicts the specified quantiles of the RT distribution of the experiment reported by Dijkstra et al. (2010). For details on how to summarize and plot the resulting `mqgam` object, see the supplementary materials.

```
> qntls = seq(0.1, 0.9, by = 0.2)
> ldt.qgam = mqgam(RT ~ s(OS, k = 5) +
                         s(Frequency) +
                         s(Trial) +
                         s(Participant, bs = "re"),
                   data = ldt, qu = qntls)
```

https://doi.org/10.1017/S1366728921000079 Published online by Cambridge University Press

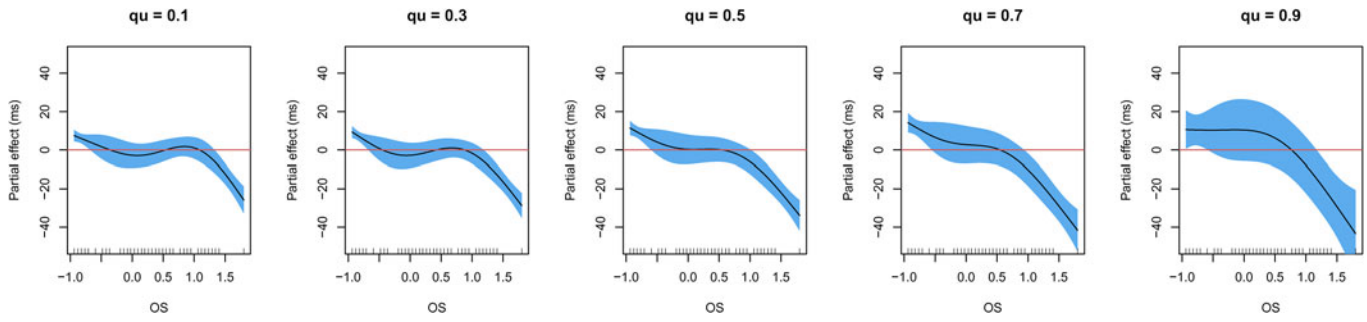


Fig. 4. Effects of orthographic similarity at different deciles of the RT distribution obtained with QGAMMs.

The effect of *OS* is visualized in Figure 4. From left to right, the panels show the effect of *OS* at the 1st, 3rd, 5th, 7th, and 9th deciles. Its effect is present across all deciles. The nonlinear effect observed in the GAMM analysis (Figure 1, Panel C) is well-replicated at lower deciles (note that a QGAMM does not assess the difference between the quantiles statistically). From this analysis, we can infer that OS indeed contributes already at the earliest stages of word recognition.

A QGAMM can also be used to track the time-course of an interactive effect of two (or more) predictors (for the interaction between *OS* and *Frequency*, see the supplementary material).

Conclusion

GAMMs and QGAMMs provide the analyst with tools that improve substantially on what the LMM makes available. Regression splines allowed us to address in more detail whether the facilitation for identical cognates reported by Dijkstra e al. (2010) is indeed substantial, or can instead be understood as the endpoint of a nonlinear trend. A reanalysis of Miwa et al.'s (2014) data clarified that bilinguals and monolinguals responded differently to words and nonwords in the course of the experiment. We also illustrated the tensor product smooth for the nonlinear interaction of semantic similarity and word frequency, resulting in a wiggly regression surface. Using the QGAMM, we illustrated nonlinear orthographic similarity effects for both short and long response times.

GAMMs and QGAMMs are designed in such a way that if the true effect is linear, the model will discover this and report the effect as such. However, many effects in lexical processing turn out to be nonlinear, including not only the effect of word frequency, but also subject-specific effects of learning, fatigue, or fluctuations in attention, which can be modeled using factor smooths. The examples presented in this introduction cover only a small part of what is possible, and these statistical techniques are undergoing rapid further development. For interested readers, there are various other applications of GAMMs in language studies (see Chuang, Fon, Papakyritsis & Baayen, 2020; Hendrix, Bolger & Baayen, 2017; Murakami, 2016; van Rij et al., 2019; Wieling, 2018). For understanding the quantitative structure of experiment datasets and how a response variable is shaped by both linguistic predictors, participant properties, and the experimental task, GAMMs and QGAMMs have much to offer for bilingualism research.

Supplementary Material. For supplementary material accompanying this paper, visit http://dx.doi.org/10.1017/S1366728921000079

References

Baayen RH (2010) A real experiment is a factorial experiment? *The Mental Lexicon* **5**, 149–157.

Baayen RH, Davidson DJ and Bates DM (2008) Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* **59**, 390–412.

Baayen RH, van Rij J, De Cat C and Wood SN (2018) Autocorrelated errors in experimental data in the language sciences: Some solutions offered by Generalized Additive Mixed Models. In Speelman D, Heylen K and Geeraerts D (eds), *Mixed Effects Regression Models in Linguistics*, (pages 49–69). Berlin: Springer.

Baayen RH, Vasishth S, Kliegl R and Bates D (2017) The cave of shadows: Addressing the huma factor with generalized additive mixed models. *Journal of Memory and Language* **94**, 206–234.

Brysbaert M and New B (2009) Moving beyond Kucera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods* **41**, 977–990.

Chuang YY, Fon J, Papakyritsis I and Baayen RH (2020) Analyzing phonetic data with generalized additive mixed models. In Ball MJ (Ed.), *Handbook of clinical phonetics*. London: Routledge.

Dijkstra T, Grainger J and van Heuven WJB (1999) Recognition of cognates and interlingual homographs: The neglected role of phonology. *Journal of Memory and language* **41**, 496–518.

Dijkstra T, Miwa K, Brummelhuis B, Sappeli M and Baayen RH (2010) How cross-language similarity and task demands affect cognate recognition. *Journal of Memory and Language* **62**, 284–301.

Dijkstra T and van Heuven WJB (2002) The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition* **5**, 175–197.

Fasiolo M, Goude Y, Nedellec R and Wood SN (2017) *Fast calibrated additive quantile regression*. URL: https://arxiv.org/abs/1707.03307

Hendrix P, Bolger P and Baayen H (2017) Distinct ERP signatures of word frequency, phrase frequency, and prototypicality in speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **43**, 128–149.

Miwa K, Dijkstra T, Bolger P and Baayen RH (2014) Reading English with Japanese in mind: Effects of frequency, phonology, and meaning in different-script bilinguals. *Bilingualism: Language and Cognition* **17**, 445–463.

Murakami A (2016) Modeling systematicity and individuality in nonlinear second language development: The case of English grammatical morphemes. *Language Learning* **66**, 834–871.

Pinheiro JC and Bates DM (2000) *Mixed-effects models in S and S-PLUS*. New York: Springer.

Ratcliff R (1979) Group reaction time distributions and an analysis of distribution statistics. *Psychological Bulletin* **86**, 446–461.

Schmidtke D, Matsuki K and Kuperman V (2017) Surviving blind decomposition: A distributional analysis of the time-course of complex word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **43**, 1793–1820.

van Rij J, Hendriks P, van Rijn H, Baayen RH and Wood SN (2019) Analyzing the time course of pupilometric data. *Trends in Hearing* **23**, 1–22.

van Rij J, Wieling M, Baayen R and van Rijn H (2017) "itsadug: Interpreting time series and autocorrelated data using GAMMs." R package version 2.3.

Wieling M (2018) Analyzing dynamic phonetic data using generalized additive mixed modeling: a tutorial focusing on articulatory differences between L1 and L2 speakers of English. *Journal of Phonetics* **70**, 86–116.

Wood SN (2003) Thin-plate regression splines. *Journal of the Royal Statistical Society (B)* **65**, 95–114.

Wood SN (2004) Stable and efficient multiple smoothing parameter estimation for generalized additive models. *Journal of the American Statistical Association* **99**, 673–686.

Wood SN (2013) On p-values for smooth components of an extended generalized additive model. *Biometrika* **100**, 221–228.

Wood SN (2017) *Generalized additive models: An introduction with R* (2nd edition). Chapman and Hall/CRC.

https://doi.org/10.1017/S1366728921000079 Published online by Cambridge University Press