ICAME Journal • 48(1) • 2024 • 1–22
DOI: 10.2478/icame-2024-0001

ICAME Journal

How *real* is the quantitative turn? Investigating statistics as the *new normal* in linguistics

Research Article

Sarah Buschfeld[1]*, Sven Leuckert[2], Claus Weihs[1], Andreas Weilinghoff[3]

[1]*TU Dortmund University*
[2]*Technische Universität Dresden*
[3]*University of Koblenz*

Received 6 August 2023; accepted 19 January 2024

Abstract: Statistical approaches in linguistics seem to have gained in importance in recent times, especially in the field of Corpus Linguistics. In particular, the last ten years have seen an upsurge of linguists being dedicated to statistical methods and the improvement of statistical knowledge. This has repeatedly been described as 'the quantitative turn' in linguistics. In the present paper, we assess how real this quantitative turn actually is and whether statistics can be considered the 'new normal' in (corpus) linguistics. To this end, we have analyzed the contributions to six high-impact journals (*Corpora, Corpus Linguistics and Linguistic Theory, ICAME Journal, English World-Wide, Journal of English Linguistics*, and *Language Variation and Change*) for a period of eleven years (January 2011 until December 2021). Our results suggest that, indeed, statistical methods seem to be on the rise in linguistic studies. However, their frequency strongly varies between the journals, and, in general, we have identified some room for improvement in the use of advanced statistical methods, in particular the discussion of true prediction.

Keywords: *quantitative turn • statistical analysis • significance • model evaluation • prediction*

1 Introduction

The discipline of linguistics has not always been as quantitative in orientation and interested in statistical methods as it currently is. In particular over the last 20 years, a growing number of linguistic studies have been published which make use of increasingly sophisticated statistical modeling and techniques. As Joseph (2008: 687) points out, "[l]inguistics has always had a numerical and mathematical side (…) but the use of quantitative methods (…) seems to be ever on the increase; rare is the paper that does not report on some statistical analysis of relevant data or offer some model of the problem at hand" (see also Geeraerts 2019: 190). This development has been labeled the *quantitative turn* in linguistics and seems to be particularly prominent amongst corpus linguists. Further indication of this development are the numerous introductions to statistics for linguists (e.g. Gries 2021; Levshina 2015; Winter 2019), handbooks and textbooks on quantitative corpus linguistics (e.g. Desagulier 2017; Egbert et al. 2020; Gries 2017), as well as a number of existing book series and journals in corpus linguistics (e.g. *International Journal of Corpus Linguistics*; *Routledge Advances in Corpus Linguistics*; *Studies in Corpus Linguistics*). Furthermore, some first empirical evidence for this development can be found, for example, in the studies by Kortmann (2021) and Larsson et al. (2022). The present study expands on these earlier ones in two important respects. First of all, we compare three corpus-linguistic journals (*Corpora*; the *ICAME Journal*; *Corpus Linguistics and Linguistic Theory*) with three linguistic journals of broader thematic and/or methodological foci (*Journal of English Linguistics*; *Language Variation and Change*; *English World-Wide*). We further include an analysis of gradual change by investigating articles

* *Corresponding author: Sarah Buschfeld, E-mail: sarah.buschfeld@tu-dortmund.de*

from all journal volumes between 2011 and 2021 and thus draw on a larger dataset overall. The research questions guiding our analysis are the following:

RQ1: What exact statistical methods are used how frequently in the different journals?
RQ2: How frequently do linguists make use of inferential statistical approaches, i.e. how often do they test for statistical significance?
RQ3: How frequently is a regression/classification model used?
RQ4: How frequently is the fit of regression/classification models assessed?
RQ5: How often is the true predictive power of a model assessed?
RQ6: Do differences in the use of statistical approaches exist between different linguistic journals?
RQ7: Do differences in the use of statistical approaches exist across time, i.e. when comparing an eleven-years time span?

'True prediction' is here understood as the prediction of a target variable for observations outside the sample used for model estimation. We consider this important and more valuable than simply estimating model fit since assessing true prediction allows us to generalize our findings to the wider population, such as speakers not included in an investigated sample.

Our analysis draws on different descriptive and inferential statistical methods (Poisson models and linear as well as logistic time trend analyses). The paper is structured as follows: we first turn towards some historical observations and theoretical considerations concerning the quantitative turn in linguistics (Section 2.1 and Section 2.2) and present earlier empirical findings concerning the quantitative turn (Section 2.3). As a second step, we introduce our data and methods used (Section 3) before we turn to the presentation of our results (Section 4). In the discussion of results (Section 5), we not only discuss our findings in relation to our research questions and with reference to the earlier studies. We also reflect on what these findings mean for the discipline of linguistics in general, i.e. whether we are really confronted with a 'statistical revolution' which comes "at the expense of linguistic description and analysis" (Larsson et al. 2022: 150). We finally present some general conclusions based on our findings (Section 6).

2 The quantitative turn in linguistics

2.1 Historical precursors to the quantitative turn

Over the course of the last two centuries, the study of language in Western Europe has undergone a shift from a culturally oriented philology to what is now commonly defined as *language science* that frequently involves quantitative methods. While describing the details of this complex development goes far beyond the scope of this paper,[1] we provide a brief summary in order to show that the *quantitative turn* analysed in this paper is, in a sense, a logical continuation of a long-standing line of thinking about the *correct* methods of linguistics.

Calls for linguistic analysis to be treated as one of the hard sciences can be found as early as the first half of the 19th century, when linguistics as a discipline in its own right was still in a constitutive stage (Jäger 2003: 67–68). Late 18th- and early 19th-century thinking about language in Europe is often associated with names such as Herder, Humboldt, and Schleicher. However, unlike some of his contemporaries, Schleicher considered language as a *Naturgegenstand* ('a natural object'), meaning that linguistics, according to him, should not merely be inspired by the sciences but instead considered as a full-fledged natural science in its own right (cf. Jäger 2003: 70).

Such discussions about how linguistics can and should be done continued in the 20th century. As Joseph et al. (2001: vii) point out: "[A] characteristic feature of the twentieth century was the attention given to establishing and policing the borders of linguistics as a field of inquiry". Significant changes in the foci and methods of linguistics can be observed particularly after the publication of Saussure's

1 A number of edited collections and handbooks about (parts of) the history of linguistics have been published; examples include Parret (1976), Jacob and Krefeld (2007), and the three volumes by Auroux et al. (2008).

(1916) *Cours de Linguistique Générale* and Bloomfield's (1933) *Language*. While written language had previously been considered as the (only) object worthy of study in linguistics, spoken language became increasingly relevant in the wake of structuralism.[2] The structuralist paradigm emerging from Saussure's and Bloomfield's ideas has shaped Western linguistics significantly and, due to its systematic approach to language, also further prepared the field for the gradual emergence of more quantitative approaches.

2.2 The quantitative turn

As a result and natural consequence of the development of linguistics towards the conceptions and methods of the natural sciences, the so-called *quantitative turn* has emerged as a more recent development. Geeraerts (2019: 190) defines it as "a major shift toward quantitative approaches in the methodology of linguistics" in the early 21st century. Kortmann (2021: 1208) describes the quantitative turn as a shift "both in scale and in quality, a turn concerning the degree (including the degree of sophistication) to which quantitative empirical studies, statistical techniques, and statistical modelling have come to be used and determine linguistic research". While quantitative methods have been common in some areas of linguistics for a long time, such as sociolinguistics and psycholinguistics (Geeraerts 2019: 190), they have become more popular in usage-based approaches only in more recent times (Joseph 2008: 687). In particular, the rise of corpus-linguistic methods has led to an increasing interest in and spread of statistical analyses. This rising popularity of quantitative methods and interest in statistical methodology is reflected in an ever-growing number of handbooks and edited collections (e.g. Köhler et al. 2005; Biber and Reppen 2015; Schützler and Schlüter 2022) as well as textbooks (e.g. Brezina 2018; Stefanowitsch 2020; Gries 2021) on corpus linguistics and statistical analysis in linguistics; many further titles are listed in Sönning and Werner (2021).

In the present paper, our concern is to find empirical evidence of the quantitative turn, comparing the use of statistical methods in six different linguistic journals (cf. Section 3.1 for details). In the next section, we provide overviews of two previous studies with a similar objective, i.e. of Kortmann (2021) and Larsson et al. (2022). In addition, we briefly comment on studies on the quantitative turn in other linguistic disciplines.

2.3 Empirical evidence of the quantitative turn in earlier studies

As mentioned in Section 2.2, our paper is not the first to address the quantitative turn in an empirical fashion. Kortmann (2021) investigates 380 articles that have been published in *English Language and Linguistics* since the journal's foundation in 1997 regarding their use of quantitative methods. He intends to find out for *English Language and Linguistics* how much truth there is to Gries' claim that "10 or 15 years ago it would have been quite difficult to find papers with multifactorial statistical techniques in corpus-linguistic papers – now, monofactorial statistical tests at least are much more frequent, and multifactorial statistical methods are on the rise" (2015: 93).

All articles in Kortmann's analysis are categorized according to four levels, depending on the presence of statistical methods and their complexity: no quantitative methods (i.e. the paper is exclusively qualitative in nature), descriptive statistics (such as relative frequency descriptions), simple statistics (such as chi-squared tests), and advanced statistics (such as regression and mixed-effect models). He finds that the number of purely qualitative articles per volume has been decreasing over time, while simple and advanced statistics have been on the rise; the number of articles solely reporting on the frequency of linguistic features, i.e. taking a descriptive approach, has remained consistent (Kortmann 2021: 1210). Based on his findings, Kortmann (2021) sees the quantitative turn as a positive development, contingent on the careful use of statistical methods.

The next study by Larsson et al. (2022) takes fewer articles overall but more journals into consideration. In total, Larsson et al. (2022) analyse 47 articles published in the journals *Corpora*, *Corpus Linguistics and Linguistic Theory*, the *ICAME Journal*, and the *International Journal of Corpus Linguistics* in 2009

2 In a comparable – if less dramatic – development, computer-mediated communication, which often mixes linguistic features associated with spoken and written language, is now of central interest in linguistics (see, inter alia, Leuckert and Buschfeld 2021).

and 2019. They are interested in the proportion of *statistical reporting* versus *linguistic description*, which they consider relevant because (corpus) linguists have to balance two key considerations:

> (*i*) the advantages of employing sophisticated quantitative and statistical methods for studies that use corpus data, and (*ii*) the need to retain the primary focus on the actual goal of corpus linguistic analysis: describing 'the things linguists are interested in'. (Larsson et al. 2022: 139)

In order to identify the proportions of statistical reporting, they counted the "[n]umber of words devoted to reporting the findings of quantitative or statistical analysis", the "[n]umber of tables/graphs devoted to statistical reporting", and the "[n]umber of distinct statistical tests employed" (Larsson et al. 2022: 142). Linguistic description, on the other hand, is measured by the "[n]umber of words devoted to reporting the findings of qualitative linguistic research or linguistic interpretation", the "[n]umber of tables/graphs devoted to linguistic description", and, finally, the "[n]umber of text excerpts, linguistic examples and word lists" (Larsson et al. 2022: 142).

Comparing the proportions of statistical reporting and linguistic description between 2009 and 2019, Larsson et al. (2022) identify a clear difference: while the proportion of statistical reporting has increased, the proportion of linguistic description has decreased. These two developments go hand in hand, since the "number of text excerpts and linguistic examples in an article negatively correlated with both the proportion of words devoted to statistical reporting […] and, to a lesser extent, the number of distinct statistical techniques" (2022: 152). Furthermore, similar to Kortmann (2021), they find that the use of advanced methods has increased significantly (2022: 146). They conclude that there has been a statistical revolution in corpus linguistics that, despite the advantages of sophisticated statistical analysis, is troubling to them "because we strongly favour the ultimate goal of corpus linguistics: learning about language use with the help of a corpus" (2022: 155). They recommend that, in addition to statistical reporting, linguists "also devote space to interpreting and illustrating the patterns of language use that it represents" (2022: 154), as has traditionally been at the core of linguistic research.

In addition to the (predominantly corpus-linguistic) studies described in some detail above, the quantitative turn has also been discussed in other linguistic disciplines. In an analysis of all studies published in the journal *Cognitive Linguistics* between 1990 and 2012, Janda (2013, 2017) observes a shift from most articles not employing quantitative methods until 2007 to most articles using quantitative methods from 2008 onwards. Her analysis shows that "[o]ver half (75 out of 141 = 53%) of all quantitative articles published in *Cognitive Linguistics* have appeared in 2008–2012" (Janda 2013: 5), with studies reporting on corpus data, experimental data, combinations of both, and language acquisition data.

Focusing on journals in historical linguistics, Jenset and McGillivray (2017) investigate quantitative methodology employed in contributions to *Diachronica*, *Folia Linguistica Historica*, the *Journal of Historical Linguistics*, *Language Dynamics and Change*, *Language Variation and Change*, and *Transactions of the Philological Society*. They find that 60% of studies use qualitative and 40% use quantitative methods; in the contributions that use quantitative methods, linear models and tree-based methods dominate. However, this is due mostly to the inclusion of *Language Variation and Change*, which is more quantitative than the other journals in general and also features more advanced statistical methods. Overall, journals in historical linguistics seem to adopt quantitative methods more slowly than journals with a mainly synchronous focus, such as *Language*, which is a high-ranking journal in general linguistics (see Sampson 2013 for an assessment of quantitative methods in *Language*).

A final paper of note is Palacios Martínez (2020) on methods of data collection in empirical linguistics research. Palacios Martínez (2020: 9) finds that, in his corpus of 1,143 abstracts and 200 papers from 2017, studies prefer experimental methods of data collection (33%), corpus analysis (18%), or are unclear in how data was collected (31%). Surprisingly, combining methods of data collection in his sample is infrequent, suggesting that the assumed increasing popularity of mixing methods and data triangulation are not as prevalent as perhaps expected. Since the focus of Palacios Martínez (2020) is on methods of data collection and less on how the data is handled (qualitatively or quantitatively), our study complements his analysis.

3 Data and method

As mentioned in the introduction to the present article, our aim is to expand on earlier studies, both in terms of number of articles and phenomena investigated. To that end, we compiled the *Meta Studies in Corpora of Linguistic Articles* corpus. The corpus includes articles from six linguistic journals which were analyzed for a period of eleven years. Table 1 summarizes our data set; the abbreviations will be used to refer to the journals in the later presentation of results (cf. Section 4).

Table 1. Setup of the dataset of the present study.

Journal	Abbreviation	Timeframe	Article count
Corpora	Cor	2011-2021	144
Corpus Linguistics and Linguistic Theory	CLLT	2011-2021	153
English World-Wide	EWW	2011-2021	127
ICAME Journal	ICAME	2011-2021	68
Journal of English Linguistics	ENG	2011-2021	133
Language Variation and Change	LVAC	2011-2021	159
Overall			784

The journals were chosen based on their linguistic focus. Three of the journals investigated have a clear corpus-linguistic orientation, i.e. *Corpora*, the *ICAME Journal*, and *Corpus Linguistics and Linguistic Theory*. These three journals are among the leading journals in corpus linguistics and reflect current trends in the field. In addition, from a purely practical perspective, we were able to access all required back issues of these journals for our analysis. As a means of comparison, we also included three thematically and/or methodologically broader journals in our study, i.e. the *Journal of English Linguistics*, *Language Variation and Change*, and *English World-Wide*. While all three of these journals also feature corpus-linguistic studies, they are not limited to corpus linguistics. The *Journal of English Linguistics* explicitly points out its "broad theoretical and methodological scope" on its website; the website for *Language Variation and Change* does not mention any specific methods but that its focus is on "the study of linguistic variation and the capacity to deal with systematic and inherent variation in synchronic and diachronic linguistics". Similarly, *English World-Wide* refrains from listing methods and opts to restrict the journal description to its "focus […] on scholarly discussions of new findings in the dialectology and sociolinguistics of the English-speaking communities (native and second-language speakers)".[3] Overall, the three journals cover a wide array of topics, including cognitive linguistics, historical linguistics, sociophonetics, corpus linguistics, and many others. In addition, the three journals represent leading journals in their respective fields. The overall number of articles investigated from the six journals amounts to 784.

Relating to the first five research questions presented in the introduction (cf. Section 1), we manually browsed through the six journals (i.e. 784 articles; book reviews and editorials were excluded from the analysis) and coded their contents for the following criteria:

1) Which data analysis methods were applied (RQ1; RQ2; RQ3);
2) Whether the results were tested for statistical significance (RQ2);
3) Whether and how the statistical models used were evaluated (RQ4);
4) Whether true prediction (in the sense defined in the introduction to this paper, cf. Section 1) was assessed (RQ5).

3 See https://journals.sagepub.com/description/ENG for the *Journal of English Linguistics*, https://www.cambridge.org/core/journals/language-variation-and-change/information/about-this-journal for *Language Variation and Change*, and https://benjamins.com/catalog/eww for *English World-Wide*.

The results of our analysis, i.e. both descriptive and inferential statistics, will be generated by means of R (R Core Team 2019) and presented and discussed according to (1) the influence the journal has on the method chosen; (2) the frequencies of the use of significance testing, regression/classification models, model evaluation, and true prediction; and (3) the influence of time. For the inferential analyses, we employ Poisson models and linear as well as logistic time trend analyses.

4 Results

4.1 The interaction of journal and method

As a first step, we investigated which specific methods were used in the papers of the six journals and determined their raw frequencies. Whenever a method was used in a paper, it was only counted once, no matter how often it was actually employed. Overall, we were able to identify 125 different data analysis methods/scores used. Table 2 shows the 16 most frequent ones. Amongst these, the chi-square test (chisq-test) is the by far most frequently applied method, followed by the log-likelihood test (log-like-test) and classification/regression methods, i.e. mixed-effects logistic regression (mix-log), logistic regression (log-reg), and mixed-effects linear regression (mix-reg).

Table 2. Raw number of the 16 most frequently used methods in the six journals.

Journal **Method**	**EWW**	**CLLT**	**Cor**	**ENG**	**ICAME**	**LVAC**	**sum**
chisq-test	26	28	13	34	9	8	118
log-like-test	8	12	19	9	4	14	66
mix-log	11	13	0	7	0	25	56
log-reg	8	15	4	7	1	20	55
mix-reg	5	3	2	11	0	34	55
ANOVA	14	5	14	14	1	6	54
corpus description	0	4	22	0	12	0	38
linear regression	2	6	6	2	1	17	34
t-test	4	8	4	8	1	6	31
correlation	2	9	7	6	3	1	28
exact-test	3	6	2	10	1	5	27
trees	6	6	2	2	2	8	26
collocation analysis	0	8	14	4	0	0	26
GLM	1	3	1	5	1	15	26
variable rules analysis	1	1	0	2	0	16	20
cluster analysis	3	7	4	3	2	0	19
No. of papers	127	153	133	144	68	159	

Since the overall number of papers considerably differs between the journals (cf. final row of Table 2), we transformed the raw numbers to percentages in relation to the overall number of papers in the journal and calculated the mean percentage for each method as an indicator of importance, i.e. the higher the mean percentage, the higher the importance of the method (cf. Table 3). Even though the relative importance of the methods differs from the order of the raw frequencies presented in Table 2, the chi-square test (15%) and the log-likelihood test (8.17%) are still the most important methods.

Table 3. Percentages of the 16 most frequently used methods in the six journals.

Journal Method	EWW	CLLT	Cor	ENG	ICAME	LVAC	Mean
chisq-test	20	18	10	24	13	5	15.00%
log-like-test	6	8	14	6	6	9	8.17%
ANOVA	11	3	11	10	1	4	6.67%
corpus description	0	3	17	0	18	0	6.33%
log-reg	6	10	3	5	1	13	6.33%
mix-log	9	8	0	5	0	16	6.33%
mix-reg	4	2	2	8	0	21	6.17%
linear regression	2	4	5	1	1	11	4.00%
correlation	2	6	5	4	4	1	3.67%
t-test	3	5	3	6	1	4	3.67%
trees	5	4	2	1	3	5	3.33%
collocation analysis	0	5	11	3	0	0	3.17%
exact-test	2	4	2	7	1	3	3.17%
GLM	1	2	1	3	1	9	2.83%
cluster analysis	2	5	3	2	3	0	2.50%
variable rules analysis	1	1	0	1	0	10	2.17%

In a next step, we investigated whether any of the journals uses a specific method considerably more often than the other journals. To this end, we determined the z-value for the journal which comes with the highest percentage for a specific method (indicated in the column "highest percentage" in Table 4). The z-value indicates the relative difference between the percentage of the journal which employs a specific method most frequently and the mean percentage of how frequently this method is used across journals (cf. Table 3):

$$z = \frac{(\textit{highest percentage}\text{-}(\textit{mean of percentages}))}{(\textit{standard deviation of percentages})}$$

In a next step, we re-ordered the methods according to these z-values. As illustrated in Table 4, the journals LVAC and COR are characterized by the highest z-values, i.e. these two journals use a number of the methods more frequently than the other journals, in particular variable rules analysis, generalized linear models, mixed-effects regression, log-likelihood test, linear regression, and collocation analysis.

In Table 4, the z-value was used as a descriptive measure to determine the journal which shows the highest percentage for the use of a specific method. In a next step, we applied Poisson regression to determine whether the z-values indicate significant differences between the journals in their representation of specific methods. Poisson regression is preferred over linear regression since the data are percentages from counts (i.e. discrete cardinal) and not measured values (i.e. continuous cardinal). We modeled the influence of method and journal as well as the interaction between method and journal on the percentages presented in Table 4. Significant interactions between method and journal indicate that the frequency of a specific method differs significantly between the journals. The reference levels are ANOVA for "method" and EWW for "journal". Thus, significances relate to these reference levels. Our analysis has yielded a number of significant effects; we only report the most significant ones. As shown in Table 5, significances at the 1% level only appear for interactions with LVAC, namely for GLM, mix-reg, linear regression, and variable rules analysis. This result is similar but not identical to the z-value results (cf. Table 4). One reason for this might be correlations between the use of specific methods. The

Table 4. Percentages and z-values of journals according to the 16 most frequently used methods.

Journal Method	EWW	CLLT	Cor	ENG	ICAME	LVAC	Highest percentage	z-value
variable rules analysis	1	1	0	1	0	10	LVAC	2.025
GLM	1	2	1	3	1	9	LVAC	1.973
mix-reg	4	2	2	8	0	21	LVAC	1.912
log-like-test	6	8	14	6	6	9	Cor	1.867
linear regression	2	4	5	1	1	11	LVAC	1.845
collocation analysis	0	5	11	3	0	0	Cor	1.799
exact-test	2	4	2	7	1	3	ENG	1.794
mix-log	9	8	0	5	0	16	LVAC	1.588
cluster analysis	2	5	3	2	3	0	CLLT	1.521
log-reg	6	10	3	5	1	13	LVAC	1.496
corpus description	0	3	17	0	18	0	ICAME	1.336
t-test	3	5	3	6	1	4	ENG	1.332
chisq-test	20	18	10	24	13	5	ENG	1.288
correlation	2	6	5	4	4	1	CLLT	1.253
trees	5	4	2	1	3	5	EWW/LVAC	1.021
ANOVA	11	3	11	10	1	4	EWW/Cor	0.963

Table 5. Poisson regression – most significant effects only.

	Estimate	p-value	Significance
(Intercept)	2.398	1.82e-15	***
mix-reg:LVAC	2.670	0.000834	***
regression:LVAC	2.716	0.004893	**
varrule:LVAC	3.314	0.005763	**
GLM:LVAC	3.209	0.007746	**
log-reg:LVAC	1.785	0.019569	*
GLM	-2.398	0.021687	*
variable rules analysis	-2.398	0.021687	*
ICAME	-2.398	0.021687	*
correlation:CLLT	2.398	0.021687	*
correlation:ICAME	3.091	0.022716	*
cluster	-1.705	0.026576	*
correlation	-1.705	0.026576	*
linear regression	-1.705	0.026576	*
exact-test	-1.705	0.026576	*

other effects reported in Table 5 are only significant at the 5% level and not at the 1% level. At the 5% significance level, specific interactions, certain methods (such as GLM and variable rules analysis), and the *ICAME Journal* are also significant.

Moreover, in Poisson regression the estimates are logarithmized. Therefore, we interpret exp(estimate), i.e. the transformation of the estimate by the inverse of the logarithm, as a multiplier for the effect of the reference level. A positive estimate, thus, indicates an increase, and a negative estimate indicates a decrease relative to the reference level. The interaction of mix-reg with LVAC, for instance, thus leads to a multiplier of exp(2.67), i.e. to a more than 14 times higher effect than for the reference level. The argument becomes clear when realizing that the effect of mix-reg relative to ANOVA (reference level) for EWW (reference level) is 4 / 11 = 0.3636 (see Table 4), whereas the effect of mix-reg to ANOVA is 21 / 4 = 5.25 for LVAC, and that, indeed, 5.25 = exp(2.67)*0.3636. The fit of the Poisson model is optimal, since $R^2 = 1$.[4]

4.2 Linear time trends

For the time trend analysis, we bundled the different methods according to four categories, i.e. significance tests, regression/classification models, model evaluation, and true prediction, and analyzed the journals in a joint approach, i.e. in one linear regression model. We investigated the relative frequency (percentages) of each category for all six journals in our time period of eleven years as dependent variable. We modeled the influence of the following independent variables on the dependent variable: (1) the categorical variable *journal*, i.e. CLLT, Cor, ENG, EWW, ICAME, LVAC; (2) the continuous variable *time*, i.e. year of publication of the journal. However, instead of using the values 2011, 2012, …, 2021, we employ the values -5, -4, …, 0, 1, 2, …, 5, i.e. we subtract 2016, the mean of the original time values, from the years 2011…2021. This way, the intercept relates to the mean of the percentages of the categories over the years. The reference level for *journal* is CLLT. The journal CLLT is selected because it, in a way, represents the mean development. On this basis, we can calculate the individual intercept for each journal and an individual slope for the time trend of each journal by means of the interaction of *time* and *journal*. This way, the model can identify significant differences between the intercept and slope of the reference journal (CLLT) and the intercepts and slopes of the other five journals. We will report significances at the 5% level.

The means and slopes of the individual linear time trends of the journals can be estimated by means of the following formulas:

1) for journal CLLT:	mean = model intercept
	slope = model coefficient of time
2) for the other journals:	mean = model intercept + model coefficient of *journal*
	slope = model coefficient of *time* + model coefficient of the interaction of *time* with the corresponding journal

Therefore, if we want to test for a significant difference between the slopes for CLLT and another *journal*, we just have to assess the p-value of the interaction of *time* with the corresponding *journal*. For a test on differences in means, p-values of the model coefficients of the *journals* can be considered. Note that we decided to report the results of linear models and not of, for example, Poisson models, even though we modeled percentages from counts. This is because linear models are much easier to interpret and Poisson models lead to nearly identical model fits as linear models.

4.2.1 Results for category significance tests

Turning towards the results for the four categories identified in Section 4.2, Table 6 shows the frequencies for significance tests, relative to the total number of papers per journal and year. In a next step, these data were analyzed by means of the joint regression model introduced in Section 4.2.

4 An alternative method for the estimation of the influence of the journal on the frequency of methods would be negative binomial regression. However, since in this method the estimation of the unknown coefficients starts from the estimates of the Poisson model and this model already has optimal fit ($R^2 = 1$, residuals = 0), negative binomial regression would not change the Poisson model and, thus, would not improve model quality.

Table 6. *Significance tests*: percentages relative to total number of papers per journal and year.

Year / Journal	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	% of papers
CLLT	67	50	63	54	64	71	58	58	80	81	70	67% (103/153)
Cor	44	38	40	40	43	56	41	47	64	40	50	47% (67/144)
ENG	50	36	64	85	83	45	91	31	91	75	56	64% (85/133)
EWW	50	25	58	82	78	91	58	58	67	67	67	63% (80/127)
ICAME	30	00	25	29	20	33	50	71	50	33	33	35% (24/68)
LVAC	93	93	100	100	93	100	93	93	87	93	80	94% (149/159)
						Total						65% (508/784)

Table 7 shows the estimated coefficients of the joint regression model. The means and estimated slopes of the linear time trends for each journal, corresponding to the estimates in Table 7, are shown in Table 8. The estimated model shows that the intercepts of ENG and EWW do not differ significantly from the intercept of CLLT and that no slope differs significantly from the slope of CLLT (cf. Table 7). Furthermore, the slope of ICAME is highest but is characterized by the lowest mean, and the slope of LVAC is negative. Moreover, all means except the mean of LVAC are lower than the mean of CLLT, which shows that the average use of significance tests is highest for LVAC and CLLT. The model fit is acceptable with $R^2 = 0.69$.

Table 7. Estimated model for category *significance tests.*

	Estimate	p-value	Significance
(Intercept)	65.0238	< 2e-16	***
Cor	-19.2801	0.00252	**
ENG	-0.7292	0.90505	
EWW	-1.3415	0.82632	
ICAME	-30.9329	4.76e-06	***
LVAC	28.2662	2.22e-05	***
time	1.7679	0.19930	
Cor:time	-0.6562	0.73437	
ENG:time	-0.2452	0.89905	
EWW:time	0.1283	0.94708	
ICAME:time	1.3295	0.49254	
LVAC:time	-2.8545	0.14373	

The means and estimated slopes of the linear time trends for each journal, corresponding to the estimates in Table 7, are shown in Table 8. Figures 1 and 2 illustrate the corresponding time trends. In LVAC, significance tests appear distinctly most frequently, with a slight decrease over time. In Cor and ICAME they are used least frequently but show an increase over time (see Figure 1). For the journals CLLT, ENG, and EWW, the usage frequencies of significance tests appear to be very similar to each other and lie slightly above the mean development of the overall percentages (see Figure 2 and Figure 3).

For the mean percentages of all journals, we fitted an individual time trend, which is presented in Figure 3, together with a 95% prediction interval of the realizations of the mean and an extrapolation into the future. The uncertainty region of the means increases with the distance to the center time point 2016, in particular for extrapolated, i.e. future, time points. The slope of this time trend is 1.38% per year, corresponding to an increase of approximately 14% over the time span of the study. For the mean percentages, the model fit is low with $R^2 = 0.30$.

Table 8. Means and estimated slopes for the journals in the category *significance tests*.[5]

Journal	Mean	Slope
CLLT	65.02	1.7679
Cor	45.74	1.1117
ENG	64.30	1.5227
EWW	63.68	1.8962
ICAME	34.09	3.0974
LVAC	93.29	-1.0861

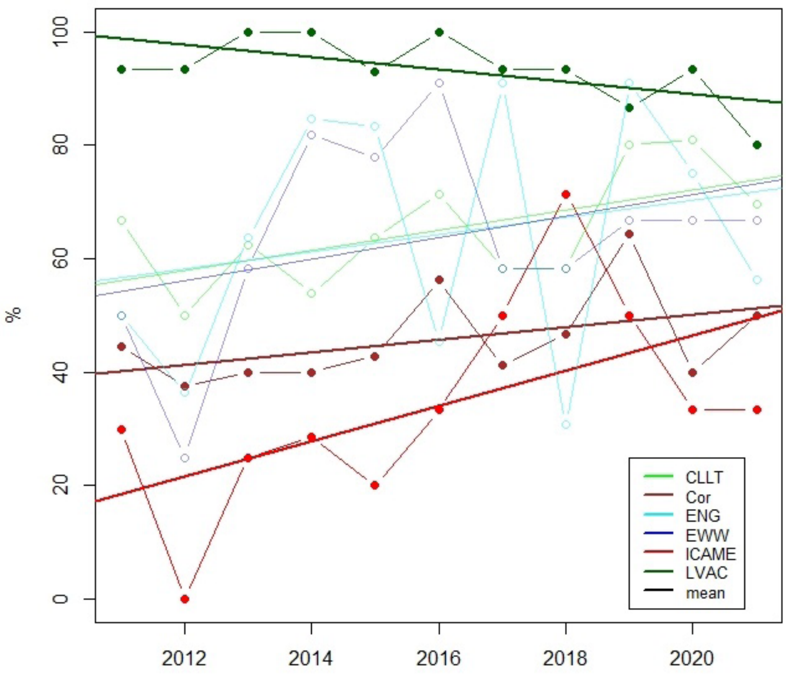


Figure 1. Time trends for the category *significance tests* – high (LVAC) and low frequencies (Cor, ICAME).

5 It has to be noted that the means and slopes were identical for independent modeling of the journals. This suggests that we can actually employ a much simpler model than used here for time trend identification. However, on the basis of the simpler models, differences between means and slopes for the individual journals cannot be assessed concerning their potential significance.

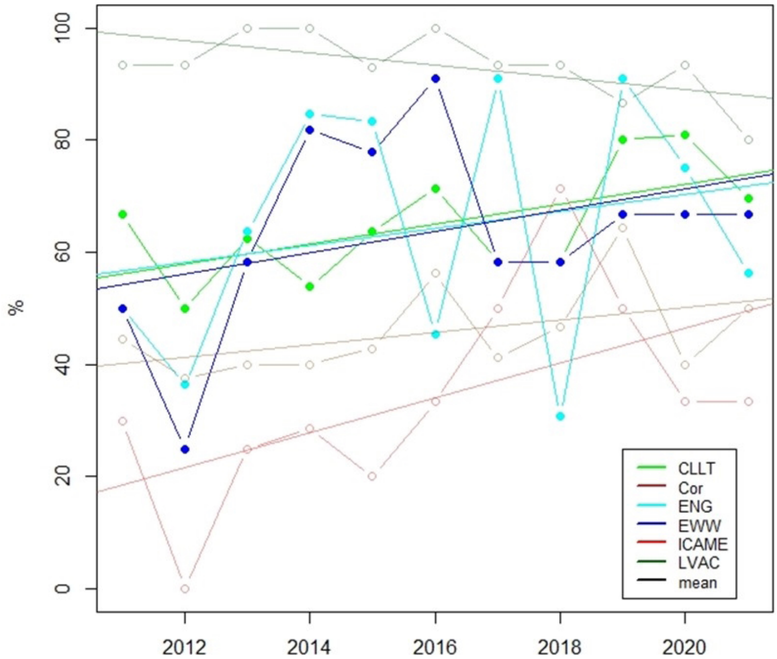


Figure 2. Time trends for the category *significance tests* – medium frequencies (CLLT, ENG, EWW).

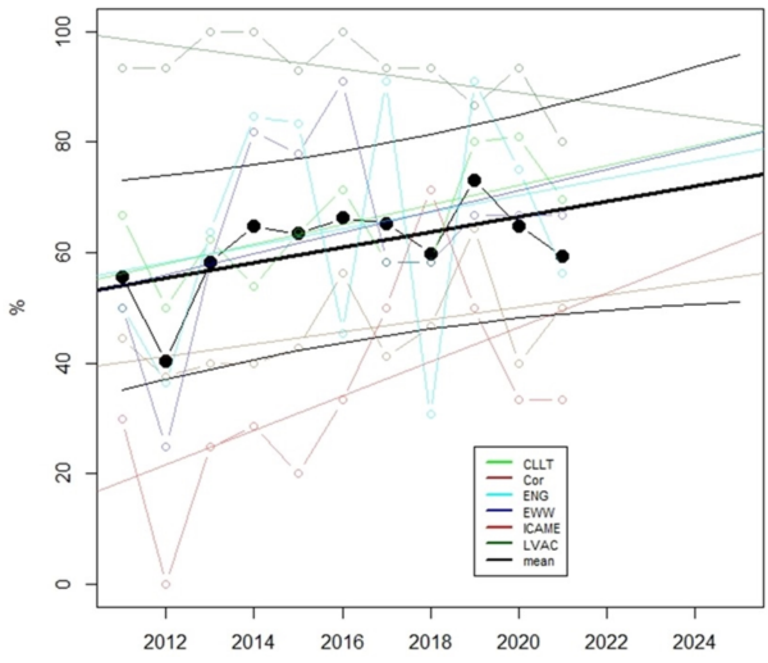


Figure 3. Time trends for the category *significance tests* – mean and 95%-prediction interval.

4.2.2 Results for category regression/classification models

Table 9 shows the percentages of use for the category *regression/classification models* relative to the total number of papers per journal and year. Again, these data were analyzed by means of a joint regression model.

Table 9. *Regression/classification models*: percentages relative to total number of papers per journal and year.

Year / Journal	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	% of papers
CLLT	25	08	50	38	45	21	58	50	40	43	48	39% (60/153)
Cor	56	13	30	10	21	06	06	07	21	20	06	16% (23/144)
ENG	17	18	09	23	33	09	55	31	36	25	38	27% (36/133)
EWW	42	08	25	45	44	36	33	25	33	50	50	35% (45/127)
ICAME	0	0	0	0	0	0	25	29	0	0	0	6% (4/68)
LVAC	73	93	87	87	86	73	87	87	93	87	60	84% (133/159)
						Total						38% (301/784)

As illustrated in Table 10, only the intercept of EWW does not significantly differ from the intercept of CLLT. Furthermore, only the slope of Cor and the slope of LVAC (but only at the 10% level) differ significantly from the slope of CLLT. In addition, our results show that all slopes are estimated to be lower than the slope of CLLT and that all means except the mean of LVAC are estimated to be lower than the mean of CLLT. With $R^2 = 0.84$, the model fit is even better than for the category *significance tests*.

Table 10. Estimated model for the category *regression/classification models*.

	Estimate	p-value	Significance
(Intercept)	38.8813	5.91e-15	***
Cor	-21.0666	0.00014	***
ENG	-12.1887	0.02130	*
EWW	-5.5021	0.28910	
ICAME	-34.0112	1.72e-08	***
LVAC	44.0624	1.18e-11	***
time	2.3471	0.04600	*
Cor:time	-4.7512	0.00505	**
ENG:time	-0.0757	0.96302	
EWW:time	-1.3049	0.42554	
ICAME:time	-1.6003	0.32914	
LVAC:time	-3.0051	0.06991	.

Table 11 presents the means and estimated slopes of the linear time trends for each journal. Figures 4 through 6 show the estimated time trends for the category *regression/classification models* in a similar manner as for the category *significance tests*. Again, LVAC shows the highest and Cor and ICAME the lowest percentages of use (see Figure 4). CLLT, ENG, and EWW, again, are characterized by medium percentages but their general trends are more different from each other than in the analysis of significance tests (cf. Figure 2 and Figure 5).

Table 11. Means and estimated slopes for the journals.

Journal	Mean	Slope
CLLT	38.88	2.3471
Cor	17.82	-2.4041
ENG	26.69	2.2714
EWW	33.38	1.0422
ICAME	4.87	0.7468
LVAC	82.94	-0.6580

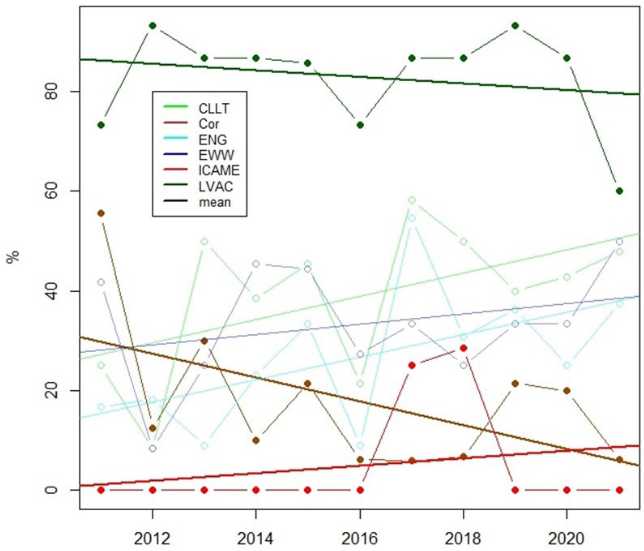


Figure 4. Time trends for the category *regression/classification models* – high (LVAC) and low frequencies (Cor, ICAME).

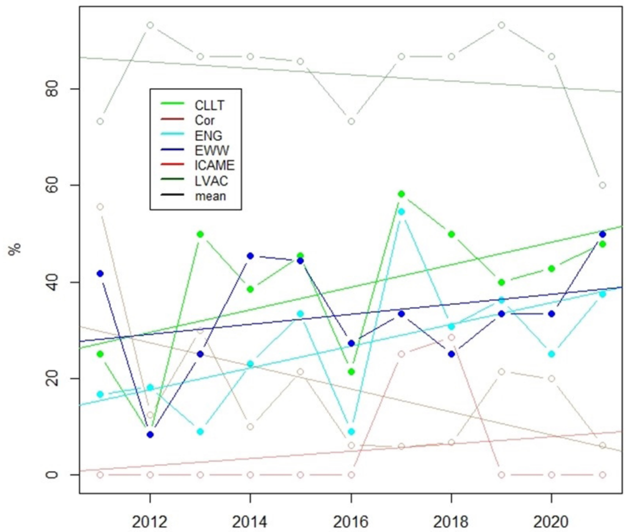
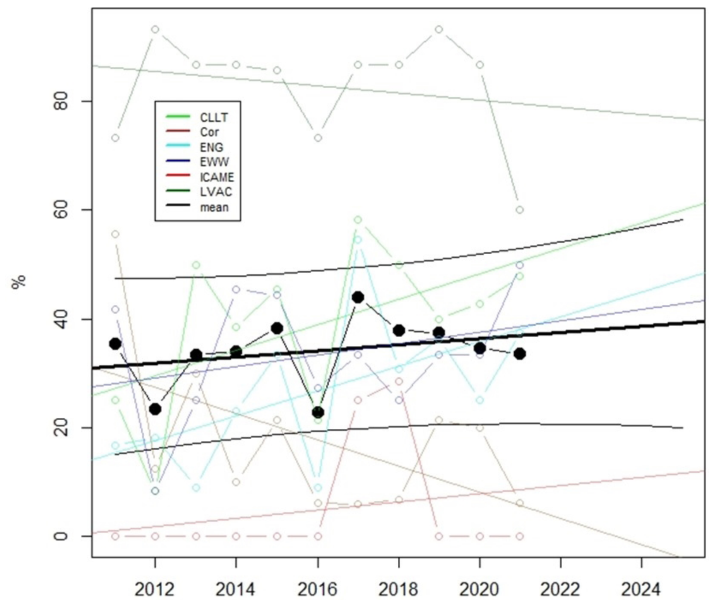


Figure 5. Time trends for the category *regression/classification models* – medium frequencies (CLLT, ENG, EWW).

The slope of the mean percentages is estimated to be 0.56% per year, which corresponds to an increase of around 6% over the 10-year time span of our study (cf. Figure 6). Unfortunately, the fit for this model is poor with $R^2 = 0.09$, which means that the time trend is unreliable.



\

Figure 6. Time trends for the category *regression/classification models* – mean and 95%-prediction interval.

4.2.3 Results for category model evaluation

Table 12 shows the observed percentages of use of the category *model evaluation* relative to the observed number of models per journal and year. It should be noted that we changed the baseline here from the total number of papers to the observed number of models, since model evaluation is only applicable to models. For the analysis of time trends, we excluded the journals Cor and ICAME because of very low frequencies.

Table 12. *Model evaluation*: percentages relative to total number of models per journal and year.

Year Journal	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	% of models
CLLT	67	100	50	60	80	67	71	83	83	56	82	67% (40/60)
Cor	0	0	0	100	67	100	100	0	33	100	100	43% (10/23)
ENG	0	0	100	33	75	100	67	25	50	67	17	44% (16/36)
EWW	20	0	0	20	25	50	0	0	25	17	33	20% (9/45)
ICAME	0	0	0	0	0	0	100	50	0	0	0	75% (3/4)
LVAC	18	43	38	69	25	73	69	85	79	77	83	59% (79/133)
						Total						52% (157/301)

In our model (cf. Table 13), all intercepts but the one for LVAC significantly differ from the intercept of CLLT. For the slope, the opposite is the case, i.e. only the slope for LVAC differs significantly (at 10% level) from the slope of CLLT. In general, all slopes are estimated to be higher than the slope of CLLT, but all means are estimated to be lower than the mean of CLLT. This means that, even though the means of the use of model evaluation are lower in all other journals than in CLLT, their increase over time is higher than in CLLT. With $R^2 = 0.54$, the model fit is the lowest of the three time trend analyses.

Table 13. Estimated model for the category *model evaluation*.

	Estimate	p-value	Significance
(Intercept)	72.618	2.33e-12	***
ENG	-24.134	0.0199	*
EWW	-55.346	2.47e-06	***
LVAC	-12.697	0.2080	
time	0.328	0.8831	
ENG:time	1.2630	0.6891	
EWW:time	0.9751	0.7573	
LVAC:time	5.6480	0.0797	.

Table 14. Means and estimated slopes for the journals and category *model evaluation*.

Journal	Mean	Slope
CLLT	72.62	0.3280
ENG	48.48	1.5910
EWW	17.27	1.3031
LVAC	59.92	5.9760

Table 14 shows the means and the estimated slopes for the journals and the category *model evaluation*. No relevant increase can be observed for CLLT, ENG, and EWW. The estimated time trend for the journal LVAC compared to the trend of the mean percentages of all journals is illustrated in Figure 7. Since the percentages of LVAC show a drastic increase, the uncertainty interval of the mean is much larger than for the other categories. The slope of the mean percentages is estimated to be 2.3% per year leading to an increase of approximately 23% over the 10-year time span. This, however, is mainly driven by the increase for LVAC. The model fit is just acceptable with $R^2 = 0.41$.

4.2.4 True prediction

Turning towards our last category, Table 15 shows the observed numbers for *true prediction* per journal and year relative to the observed number of models per journal and year. Again, prediction is only applicable to models.

Along the lines of our definition of true prediction introduced in the introduction, we searched the full texts of the journals for the following keywords to quantify whether prediction was used: ‘bootstrapping’, ‘cross-validation’, ‘out-of-bag’, ‘predictive accuracy’, and ‘random forest’. As Table 15 illustrates, a relevant number of papers which utilize or discuss true prediction exists only for CLLT. The numbers further suggest that no time trend can be identified. Therefore, we did not employ a trend analysis for this category. Figure 8 visualizes the percentages of true prediction used in the models according to journal, illustrating the high percentage for CLLT and a comparatively high percentage for COR, but for a relatively small number of models. ICAME was excluded from the analysis because of an extremely low number of models used (cf. Table 9).

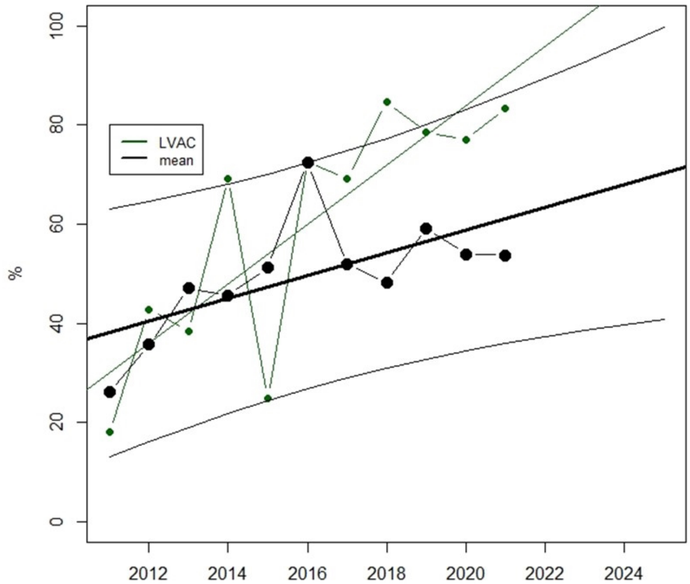


Figure 7. Time trends for the category *model evaluation* – LVAC compared to mean and 95%-prediction interval.

Table 15. *True prediction*: observed numbers per journal and year.

Year Journal	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	all
CLLT	2	0	2	0	2	1	2	0	0	3	2	14/60
Cor	0	0	0	0	0	0	0	0	1	2	0	3/23
ENG	0	0	0	0	1	0	1	0	0	0	0	2/36
EWW	0	0	0	1	0	0	0	0	1	1	0	3/45
ICAME	0	0	0	0	0	0	2	1	0	0	0	3/4
LVAC	0	1	0	0	0	0	0	1	1	0	2	5/133
						Total						30/301

4.3 Prediction of future developments

Figure 7 indicates that using linear models to predict the future development of a specific category might lead to an interpretation problem. As our example illustrates, predictions of percentages into the future might take values higher than 100% if the prediction interval is considered. The reason for this problem lies in the linearity of the model and in the large variation of the mean percentages. Overall, the model does not take into account that our percentages can never be higher than 100%. However, models exist that explicitly include the estimation of boundaries, so-called asymptotes, based on the observed data. In order to demonstrate the effect of such modeling, we applied a so-called S-shaped model (also called logistic curve) to the mean percentages of the category 'model evaluation'. The following formula corresponds to the S-shaped model used in this study:

$$mean = \frac{Asymp}{1+e^{(c-time)/s}} + \text{error}$$

where Asymp refers to the estimated height of the asymptote, c and s are additional unknown parameters that need to be estimated, and 'time' represents the year for which the mean is observed or predicted.

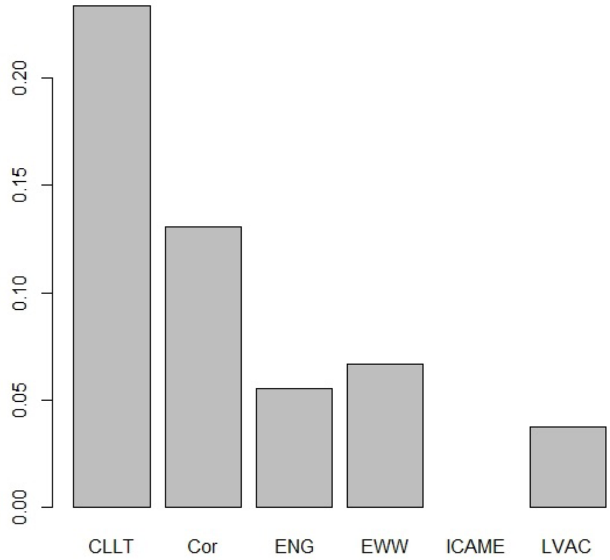


Figure 8. Bar plot of the number of predictive analyses relative to the number of models.

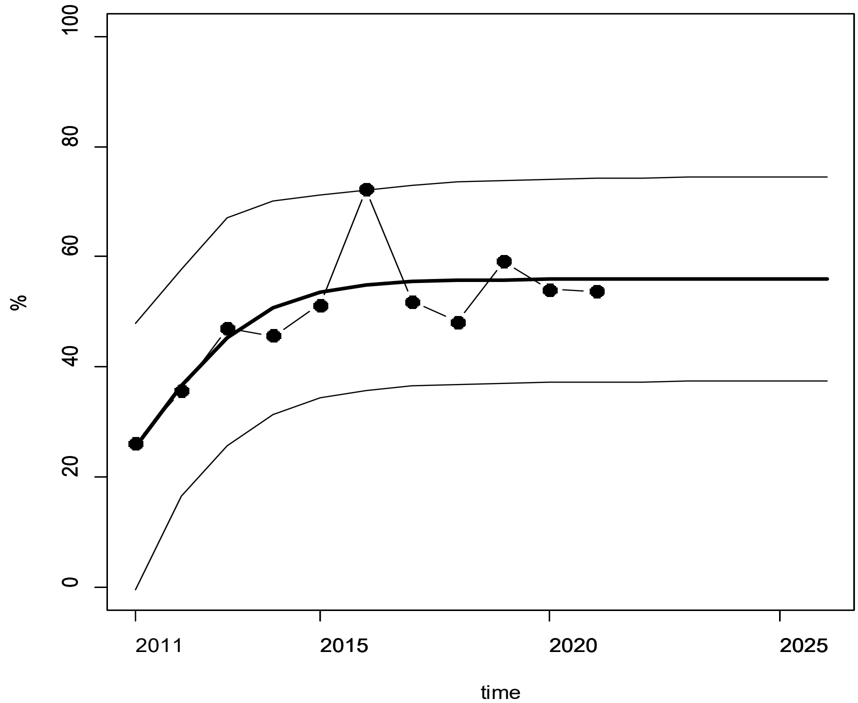


Figure 9. Graph of S-shaped time trend for category *model evaluation* and 95%-prediction interval.

As the formula suggests, the model values are always lower than the value of Asymp and if 'time' takes higher values, the value of the model takes values closer to Asymp.

Figure 9 shows the S-shaped curve for our analysis of the category 'model evaluation' and the corresponding 95% prediction interval. It further illustrates that the S-shaped model estimates an upper limit for the mean percentages, in our case 55.9%. This appears to be a reasonable estimation, since the latest observations of the mean in the time series show similar values. However, as the large prediction interval suggests, the uncertainty of future realizations of mean percentages is still quite large but not as large as for the linear model presented in Figure 7. The R^2 of the S-shaped model is 0.70, which indicates a much better fit than of the linear model in Figure 7. Therefore, the data supports the S-shaped model over the linear model and the upper limit appears to be quite reasonable considering the data.

5 Discussion

Overall, our analysis identified 125 different data analysis methods/scores used throughout the journals (RQ1). Significance testing is the most often employed statistical category across journals and is used in 65% of the papers (508/784) with an increase of 14% over the 11 years investigated (RQ2). Of this category, the chi-square test is the most frequently used individual method of the 125 methods/scores identified overall and is used in 15% of all papers (118/784). As the second strongest category of statistical methods, regression/classification models are used in 38% of the papers (301/784) with a (somewhat uncertain) increase of 6% over the 11 years. In general, LVAC is particularly strong in the use of such advanced modeling methods like (mixed effects) linear and logistic regression (RQ3). Evaluation of models is used for 52% of the models (157/301) and at least LVAC shows a significant increase over the 11 years (RQ4). True prediction was only used for 10% of the models and is most frequently employed in CLLT (RQ5).

Coming back to our overall research question, what do our results suggest about the validity of the quantitative turn? We indeed have identified an increase in the use of advanced quantitative methods overall, i.e. for inferential approaches in general (RQ7) and in particular for model evaluation in LVAC; CLLT is strongest in model evaluation but the numbers for LVAC are clearly on the rise. LVAC is also clearly leading the way in the use of significance testing and regression/classification models. Furthermore, CLLT is the only journal in which, at least occasionally, true prediction is analyzed and discussed. Therefore, LVAC and CLLT appear to be leading the way in the use of (advanced) statistical methods and thus the quantitative turn. Somewhat surprisingly, two of the three corpus linguistic journals, i.e. *Corpora* and the *ICAME Journal*, underperform in the use of significance tests and regression/classification models when compared to the other journals. Therefore, we could identify important differences between the journals when it comes to their use of statistical methods (RQ6). However, to an extent, this may be due to the foci of the journals. Several articles in the *ICAME Journal* introduce a new corpus or software tool and then only provide a small case study to illustrate potential research avenues using the new corpus or tool.[6] These case studies tend to rely on descriptive statistics, since more extensive statistics would usually go beyond the scope of the respective papers. For *Corpora*, a potential explanation may be the focus on interdisciplinarity and multilinguality, both of which invite contributions from various fields and contexts that may be less advanced in statistical methodology or where complex statistical analysis would not be required or appropriate. Another journal, EWW, clearly underperforms in model evaluation. This seems to suggest that the use of advanced statistical methods might not be as advanced and widespread in World Englishes research as in other linguistic subdisciplines – at least not yet.

Our study has therefore shed some new and more detailed light on the development and use of (advanced) statistical methods in linguistics and at the same time also confirms some earlier observations. In this respect, Kortmann (2021) has already pointed out that the somewhat simpler methods are still widely used, in our study chi-square and log-likelihood tests as well as ANOVAs (see also Janda 2013). In general, some kind of quantitative turn seems to be underway in linguistics (see also Kortmann 2021 and Larsson et al. 2022 for similar observations), but our findings clearly suggest that advanced methods are still in the minority and could be expanded in quality and quantity. In particular, true prediction is only

6 See also the description on the journal's website, which points out that it "features original studies on recent advances in the exploitation of corpora, corpus compilation and software applications. It also features a large number of reviews of scholarly work in the discipline" (https://sciendo.com/journal/icame).

used occasionally for model assessment, which is where we see the greatest potential for statistical advancement in linguistic studies.

For predicting future trends in our analyses, we have employed an S-shaped model for the category *model evaluation* as an example. We have discussed that such models are better suited for predicting future developments of percentages since they take into account that percentages can never exceed the 100% threshold. In this respect, the S-shaped model for the category *model evaluation* predicts that model evaluation will not go beyond the 80% threshold (upper limit of prediction interval) in the near future (cf. Figure 9). This appears to be a more realistic assessment of future trends than what is predicted in the linear case, for which the prediction interval already exceeds the 100% threshold in 2025 (cf. Figure 7).

Finally, we would like to reflect on one further claim made by Larsson et al. (2022: 150), namely that a 'statistical revolution' seems to be underway that comes "at the expense of linguistic description and analysis". Our findings do certainly not point towards a statistical revolution. However, it can be observed that, at times, papers making use of highly complex and advanced statistical methods may be difficult to follow and comprehend, in particular since, sometimes, statistical methods seem to be given more prominence than the linguistic interpretation of the results. This is a double-edged sword. On the one hand, we should, of course, strive to enhance our knowledge and understanding of advanced statistical modeling. As our results have shown, this is particularly true if we want to make solid predictions about the behavior of, for example, larger speech groups not part of the actual data sample or about future developments. On the other hand, we need to guarantee sufficient involvement with linguistic analysis and interpretation and make sure that our findings are still accessible, also to linguists who are not experts in statistical modeling.

6 Conclusion

The present study has shed empirical light on the question whether we have been facing a quantitative turn in linguistics in recent times. As our general findings suggest, an important development towards more quantitative and complex statistical methods has been underway throughout the last decade; however, advanced statistical methods, in particular true prediction, are still in the minority. When it comes to our initial question of how *real* the quantitative turn in linguistics really is, we can only partly confirm that statistics should be considered the *new normal* in (Corpus) Linguistics. As our analyses have revealed, this clearly depends on the respective journal, as important differences in the use of advanced statistical methods exist between LVAC and CLLT on the on the one hand and *Corpora*, the *ICAME Journal*, and EWW on the other. For predicting future trends, we have identified S-shaped models to be far better suited than linear models, since S-shaped models avoid predictions of percentages which are higher than 100%.

Future studies investigating the quantitative turn could venture into various directions. First, it would be intriguing to compare our findings and those of Kortmann (2021) and Larsson et al. (2022) with additional journals from different linguistic disciplines. Since different sub-disciplines each have individual research traditions, they may also develop in different directions. Second, it could be insightful to see how linguistics in other philologies (e.g. German linguistics, Romance linguistics, etc.) is developing. This might reveal to what extent English linguistics, as likely the biggest philology, is spearheading methodological changes in linguistic research.

Finally, we would like to conclude that linguists should strive towards more advanced statistical methods but, of course, not at the expense of linguistic analysis and interpretation. As Rome was not built in a day, such a development towards more advanced statistical methods simply may take some time but the discipline of linguistics seems to be on a good way to furthering its understanding and implementation of statistical approaches.

Acknowledgment

We would like to thank Beatrix Weber and an anonymous reviewer for their comments on an earlier version of this paper. Any remaining shortcomings are our own.

References

Auroux, Sylvain, E. F. K. Koerner, Hans-Josef Niederehe and Kees Versteegh (eds.). 2008. *History of the language sciences / Geschichte der Sprachwissenschaften / Histoire des sciences du langage: An international handbook on the evolution of the study of language from the beginnings to the present / Ein internationales Handbuch zur Entwicklung der Sprachforschung von den Anfängen bis zur Gegenwart*. Berlin: Mouton de Gruyter.

Biber, Douglas and Randi Reppen (eds.). 2015. *The Cambridge handbook of English corpus linguistics*. Cambridge: Cambridge University Press.

Bloomfield, Leonard. 1933. *Language*. New York: Holt, Rinehart & Winston.

Brezina, Vaclav. 2018. *Statistics in corpus linguistics: A practical guide*. Cambridge: Cambridge University Press.

Desagulier, Guillaume. 2017. *Corpus linguistics and statistics with R: Introduction to quantitative methods in linguistics*. Cham: Springer.

Egbert, Jesse, Tove Larsson and Douglas Biber. 2020. *Doing linguistics with a corpus: Methodological considerations for the everyday user*. Cambridge: Cambridge University Press.

Geeraerts, Dirk. 2019. Review of *Quantitative historical linguistics: A corpus framework* by Gard B. Jenset and Barbara McGillivray. Oxford: Oxford University Press, 2017. *Language* 95 (1): 190–191. https://doi.org/10.1353/lan.2019.0022.

Gries, Stefan Th. 2015. Some current quantitative problems in corpus linguistics and a sketch of some solutions. *Language and Linguistics* 16 (1): 93–117. https://doi.org/10.1177/1606822X14556606.

Gries, Stefan Th. 2017. *Quantitative corpus linguistics with R: A practical introduction*. London: Routledge.

Gries, Stefan Th. 2021. *Statistics for linguistics with R: A practical introduction*. Berlin: Mouton de Gruyter.

Jacob, Daniel and Thomas Krefeld (eds.). 2007. *Sprachgeschichte und Geschichte der Sprachwissenschaft*. Tübingen: Gunter Narr.

Jäger, Ludwig. 2003. Erkenntnisobjekt Sprache: Probleme der linguistischen Gegenstandskonstitution. In A. Linke, H. Ortner and P. R. Portmann-Tselikas (eds.). *Sprache und mehr: Ansichten einer Linguistik der sprachlichen Praxis*, 67–97. Tübingen: Max Niemeyer. https://doi.org/10.1515/9783110911985.67.

Janda, Laura A. 2013. Quantitative methods in Cognitive Linguistics: An introduction. In L. A. Janda (ed.). *Cognitive linguistics: The quantitative turn. The essential reader*, 1–32. Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110335255.1.

Janda, Laura A. 2017. The quantitative turn. In B. Dancygier (ed.). *The Cambridge handbook of cognitive linguistics*, 498–514. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781316339732.032.

Jenset, Gard B. and Barbara McGillivray. 2017. *Quantitative historical linguistics: A corpus framework*. Oxford: Oxford University Press.

Joseph, Brian D. 2008. The editor's department: Last scene of all… *Language* 84 (4): 686–690. https://doi.org/10.1353/LAN.0.0063.

Joseph, John E., Nigel Love and Talbot J. Taylor. 2001. Landmarks in linguistic thought II: The western tradition in the twentieth century. London: Routledge.

Köhler, Reinhard, Gabriel Altmann and Rajmund G. Piotrowski (eds.). 2005. *Quantitative Linguistik / Quantitative linguistics: Ein internationales Handbuch / An international handbook*. Berlin: Mouton de Gruyter.

Kortmann, Bernd. 2021. Reflecting on the quantitative turn in linguistics. *Linguistics* 59 (5): 1207–1226. https://doi.org/10.1515/ling-2019-0046.

Labov, William. 2004. Quantitative analysis of linguistic variation. In U. Ammon, N. Dittmar, K. J. Mattheier and P. Trudgill (eds.). *Sociolinguistics / Soziolinguistik: An international handbook of the science of language and society / Ein internationales Handbuch zur Wissenschaft von Sprache und Gesellschaft*, 6–21. Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110141894.1.1.6.

Larsson, Tove, Jesse Egbert and Douglas Biber. 2022. On the status of statistical reporting versus linguistic description in corpus linguistics: A ten-year perspective. *Corpora* 17 (1): 137–157. https://doi.org/10.3366/cor.2022.0238.

Leuckert, Sven and Sarah Buschfeld. 2021. Modelling spoken and written English: An introduction. *Anglistik: International Journal of English Studies* 32 (2): 7–14. https://doi.org/10.33675/ANGL/2021/2/4.

Levshina, Natalia. 2015. *How to do linguistics with R: Data exploration and statistical analysis*. Amsterdam: John Benjamins.

Palacios Martínez, Ignacio. 2020. Methods of data collection in English empirical linguistics research: Results of a recent survey. *Language Sciences* 78: 101263. https://doi.org/10.1016/j.langsci.2019.101263.

Parret, Herman (ed.). 1976. *History of linguistic thought and contemporary linguistics*. Berlin: Walter de Gruyter.

R Core Team. 2019. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Sampson, Geoffrey R. 2013. The empirical trend: Ten years on. *International Journal of Corpus Linguistics* 18 (2): 281–289. https://doi.org/10.1075/ijcl.18.2.05sam.

Saussure, Ferdinand de. 1916. *Cours de linguistique générale*. Publié par Charles Bally et Albert Séchehaye avec la collaboration de Albert Riedlinger. Édition critique préparée par Tullio de Mauro. Paris: Payot.

Schützler, Ole and Julia Schlüter (eds.). 2022. *Data and methods in corpus linguistics: Comparative approaches*. Cambridge: Cambridge University Press.

Sönning, Lukas and Valentin Werner. 2021. The replication crisis, scientific revolutions, and linguistics. *Linguistics* 59 (5): 1179–1206. https://doi.org/10.1515/ling-2019-0045.

Stefanowitsch, Anatol. 2020. *Corpus linguistics: A guide to the methodology*. Berlin: Language Science Press.

Winter, Bodo. 2019. *Statistics for linguists: An introduction using R*. London: Routledge.