



## Book Review

**David Correia Saavedra**, *Measurements of grammaticalization: Developing a quantitative index for the study of grammatical change*, 2021. Berlin: De Gruyter Mouton, pp. xii + 246. €114.95, ISBN: 9783110752946.

Reviewed by **David Lorenz**, Centre for Languages and Literature, Lund University, Lund, Sweden,  
E-mail: david.lorenz@englund.lu.se

<https://doi.org/10.1515/flin-2023-2036>

*Grammaticalization* has been a major keyword in linguistics for several decades now – as an empirical phenomenon or a theoretical framework, but perhaps best seen as a generalization that captures a set of important patterns of language change. Yet, some questions have never been fully settled. Exactly what is within or without the scope of what we classify as grammaticalization? The boundaries seem to be fuzzy. What are the features, necessary or sufficient, that define it? The common parameters (most famously in Lehmann 1982) do not provide a definite diagnostic, they rather make for a kind of family resemblance: the similarities between instances of grammaticalization can be described by these parameters, but there is none that is shared by all instances.

In light of these issues it is clear that grammaticalization as a concept involves a lot of nuances and gradience. It is also clear that a thorough quantitative approach is useful and, indeed, needed. David Correia Saavedra's book *Measurements of grammaticalization* is all about just that. It proposes statistical models to quantify the gradience and gradualness of grammaticalization. The author outlines the endeavor with a set of research questions: “Can the gradualness of grammaticalization be quantified?” (p. 2); “What are the factors that lie behind the gradualness of grammaticalization?” (p. 6); and, thirdly, whether a quantitative index can serve to test the unidirectionality of grammaticalization (p. 8). Developing this quantitative index is the main project that the book reports on. In what follows I will review this project, its presentation and what it might entail for grammaticalization theories. It should be noted from the outset that the book presents an approach that is innovative in many ways and has potential to be refined and developed further; that is, it provides a blueprint rather than ultimate solutions. There is and there should be room for debate and for improvement on the details.

So how does the machinery work and what is it based on? Correia Saavedra endorses Hopper and Traugott's (2003: 18) widely accepted definition of grammaticalization as “the change whereby lexical items and constructions come in certain linguistic contexts to serve grammatical functions and, once grammaticalized, continue to develop new grammatical functions”. Degrees of grammaticalization are

to be measured by a set of factors which are mostly operationalizations of Lehmann's (1982) parameters, however not mapped one-to-one from these parameters, as we will see. As the data are to be read semi-automatically from big corpora, the selection is to a large extent technically motivated and, inevitably, technical limitations occur. For example, the study only considers full single words (or rather, corpus tags), bound morphemes are not included, and multi-word units are captured indirectly by their 'main word' (e.g. *long* in *as long as*). Phonetic erosion is approximated by letter count, which clearly cannot capture all aspects of reduction, especially sub-phonemic reduction. What corpus tools are good at capturing is co-occurrence of elements in text, which no less than five of Correia Saavedra's variables are based on: the diversity of collocates, i.e. neighboring elements (one with a wide window of four words each to the left and right, and two narrow windows for the immediate neighbor to the left and to the right, respectively), and the diversity of colligates, i.e. neighboring parts-of-speech (one each for the left and right neighbor).

For readers familiar with the subject matter, I have listed the parameters and the corresponding operationalizations, based on Correia Saavedra's explanations, in Table 1.

As can be seen, the parameters are unevenly represented, some not at all, and two variables in the model do not directly reflect any of them. Now, it is well known that not all parameters will apply to a given individual case, and as far as I know there is no claim that they should be distributed evenly across cases, or that there can be no other diagnostics. However, their relative weighting is at issue when, for example, Wiemer (2014) proposes that 'syntactic tightening', i.e. a decrease of syntagmatic variability, might be the most applicable parameter cross-linguistically. This one is covered here by collocate and colligate diversity. But we might as well say

**Table 1:** Grammaticalization parameters and their operationalization.

Lehmann's parameter	Correia Saavedra's operationalization
Paradigmatic weight (integrity)	Letter count (phonological integrity) Collocate diversity (wide window: semantic integrity)
Paradigmatic cohesion (paradigmaticity)	–
Paradigmatic variability	–
Syntagmatic weight (structural scope)	–
Syntagmatic cohesion (bondedness)	Collocate diversity (narrow windows) Colligate diversity
Syntagmatic variability	Collocate diversity (narrow windows) Colligate diversity Token frequency Dispersion

that at least in English, a crucial feature is paradigmaticity as in the shift from an open to a closed class (e.g. NOUN > PREPOSITION), which cannot be quantified with the approach taken because part-of-speech tags are used for classification. These are important issues as they have a direct impact on the results and ultimately bear on the question of what insights we can or should gain from quantitative measurements of grammaticalization, especially when it comes to a summary score.

On a similar note, one could also turn to the more succinct model of four parameters developed by Heine and Kuteva (2007) and Narrog and Heine (2021). These are chronologically ordered yet overlapping: context extension, desemantization, decategorialization and erosion. Most of Correia Saavedra's variables could probably be subsumed under 'decategorialization', which seems appropriate as this is where the structural consequences of grammaticalization come to the fore. The first two parameters, however, might be captured by collocate diversity (wide window) but probably nothing else, raising questions about the early stages of grammaticalization (a point I will return to below); and erosion, again, is only approximated by letter count. Correia Saavedra generally acknowledges the limitations with regard to the operationalization of parameters, and he repeatedly resorts to the position that his selection of variables is based on technical feasibility but need not be the final solution: "Adding new parameters or finding better alternatives to the current ones is therefore certainly a possibility for future research" (p. 40).

The main engine of the machine consists of logistic regression and principal components analysis. The way it works from here is as clever as it is straightforward. Data are annotated for the eight variables, which are then used as predictors in the statistical models. Linguistic elements (rather than individual tokens) serve as data points. The fitted values of the regression can then be read as a 'grammaticalization score' for each element. The quantitative scale of the study becomes apparent here: with 528 items selected from the BNC, each with a token frequency of >1,000, the machine is fueled by well over half a million tokens (the exact number is not reported). This is for the written BNC; the replication on the much smaller spoken part of the corpus (Chapter 5.5) suggests that the method also works on with substantially smaller data sets.

Then, what do the models show? I will discuss only a few aspects here. Firstly, when ranking elements by their grammaticalization score (as resulting from the model), we get largely plausible results. Articles (*a, the*) show up at the top, verb forms and nouns (e.g. *say, knew, year*) at the bottom. Ranking parts-of-speech by their average score confirms this, the top scorers are negation markers, conjunctions and prepositions. This is reassuring as it means that, by and large, the measures match up with the general notion of what is 'grammaticalized', but for the same reason it is also not terribly exciting. The quantification gets more interesting in the detail, as Correia Saavedra shows with a ranking of different modal verbs (pp. 107–108). Still, the

question remains what such a grammaticalization score really means. Is the modal *would* (score: 0.991) more grammaticalized than *might* (0.615)? If so, in what way? Correia Saavedra's strategy seems to be to look back into the individual factors. For example, *would* is more frequent than *might* and “[t]oken frequency is the most prominent parameter in the model” (p. 109). This answer is, of course, satisfactory with regard to understanding the performance of the method, but less so when it comes to its explanatory value. Similarly, the model's misclassification of some forms of auxiliary BE ('*m*, '*re* and '*am* receiving very low scores) is explained by the low diversity of their left collocates. These issues show how the summary scores themselves do not give us handy take-away results, but they can produce real findings when their input is taken into account. Given the debate over scope expansion versus reduction in grammaticalization, the model reveals a general trend for English: expansion to the left, reduction to the right of the grammaticalizing element. It also shows that this effect is carried by some types of elements (e.g. articles, prepositions) while others deviate (e.g. the above auxiliary forms).

Such findings reveal the limits and the potential of the approach at the same time. On the one hand, they cast doubt on the generality and representativeness of the model – how would it change, for example, if bound morphemes were included? On the other hand, they point to a data-driven way of capturing differences between types of grammatical elements. This is reminiscent of Norde and Beijering's (2014) 'clustering approach' (which receives surprisingly little attention in the book). Norde and Beijering use categorical parameters, but for all I can see, Correia Saavedra's quantitative measures could well be fed into a clustering/mapping algorithm (e.g. Discriminant Analysis or Support Vector Machine) that would map out the elements by their quantified grammaticalization features. One would expect to see different clusters of parts-of-speech (articles in one corner, nouns in the other), but also groups of grammatical types – e.g. how similar are modals and auxiliaries to each other, and to, say, adverbs? This is but one direction in which the approach could be developed. In my view, it is such considerations and outlooks that make this work most inspiring and thought-provoking, and the quantitative indices provide plenty of food for this. Much of the thought-work on this plane is, however, left to the reader. The author's focus is clearly on the feasibility and technical implementation of a quantitative index from corpus data. The style of presentation is methodological and at times instructional, making the empirical parts a dry read if one is looking for theory-building discussions and concepts. Yet what Correia Saavedra achieves is to provide a methodological template – basic recipes that other researchers can replicate and adapt. For example, the research setup described with the BNC and COHA in Chapters 4–6 could well be reproduced using other corpora, and with other languages than English. A researcher could follow the steps laid out here and just make the adaptations that the different data set requires.

In Chapter 6, Correia Saavedra presents a diachronic study. It comes with a slightly different approach, designed to test the unidirectionality hypothesis, which in this framework would mean an increase in grammaticalization scores. The trick here is that lexical and grammatical homonyms are compared, e.g. main verb and auxiliary *HAVE* or then noun *event* and the complex preposition *in the event of*. Grammaticalization scores are measured for each item in each decade of the COHA corpus. The expectation seems to be that they should increase over time for the grammatical counterparts, at least in cases of on-going grammaticalization. As before, the idea and the technicalities are very transparent and carefully explained. The design is clever, to be sure, but it has a major drawback which soon becomes apparent. The researcher has to classify each token as either 'lexical' or 'grammatical' (by POS tag and manual correction), while in reality, grammaticalizing items are often ambiguous. Correia Saavedra mentions this as a problem (with the temporal or conditional reading of *as long as* in *You can stay as long as you want to*), but seemingly shrugs it off by stating that "such cases are rare and go beyond the scope of the present study" (p. 145). This may be so with respect to the study design, but I have some doubts that this dismissiveness is generally warranted, since such ambiguities really are a feature, not a side-effect of grammaticalization, arising in 'bridging' or 'critical' contexts (cf. Heine 2002; Diewald 2006).

The results are rather mixed. The grammatical elements do have higher scores overall (though with exceptions) and many show an increase over time, but other findings are unexpected or problematic, and as so often, these are the more interesting ones. Some items show stability or fluctuation rather than directional change; none of the measured variables is consistently associated with increasing values. Sure, some items already start out fully grammaticalized and do not develop any further (especially since phonetic erosion is not in the equation). Other cases of decreasing scores (e.g. *going to* V, *keep* V<sub>ing</sub>, *used to* V) are chalked up to their auxiliary usage, as auxiliaries are underrepresented in the model for calculating the scores. So again, a result is explained by an aspect of method, but this still leaves us with the question what makes auxiliaries so different, and we are back at the issue raised above, that different types may grammaticalize by different features. Finally, some of the lexical counterparts also show high or increasing scores. Correia Saavedra wonders if this is "because they have grammatical qualities and/or features that are similar to that [sic] of grammatical elements, which made them ripe targets for grammaticalization in the first place?" (p. 179). This is a good question, because grammaticalization emerges in specific contexts, and if the answer is 'yes', it would suggest that the development sets in well before we would actually identify 'grammaticalized' uses of an item. The findings here may well be influenced by early 'bridging contexts' that enable reanalysis, or even by 'untypical contexts' (Diewald 2006) where a grammatical meaning comes in by conversational implicature. It is

just not clear how well Correia Saavedra's grammaticalization index captures these early stages of the process. It is also not clear if or how the index can distinguish 'retraction', i.e. a decline in grammaticalized uses before they have reached conventionalization (cf. Narrog and Heine 2021: 65–66). Perhaps more fine-tuned instruments are needed to deal with such gradual micro-changes, such as Petré and Van de Velde's (2018) token-based quantification of grammaticalization features. That said, on macro-level the application of the quantitative index to diachronic data seems to work well when it catches the differentiation of grammaticalizing items from their lexical source (e.g. *in order to* and *quit V<sub>ing</sub>* in this case study). It is certainly a promising route that warrants further attention.

The book never fully shifts its focus from method to theory building, neither in the section entitled "Theoretical relevance and methodological concerns" (§6.5) nor in the concluding chapter. This is not necessarily a shortcoming as the methodological proposal is indeed innovative, relevant and inspiring. Still, as a reader, I for one would have hoped for deeper conceptual insights into the phenomenon of grammaticalization. The main conclusion from the empirical studies seems to be that they "support the idea that there are actually shared characteristics between very different processes" (p. 181 and similarly p. 193), so that we might distinguish between common and case-specific features.

What follows? Are these shared characteristics useful as defining features of grammaticalization? Perhaps not, as Correia Saavedra also concedes that "it is difficult to find universal characteristics of grammaticalization that can be transcribed into specific quantitative variables and that work with all possible cases of grammaticalization" (p. 180). Moreover, following the results of Chapters 5 and 6, we would basically rely on frequency and collocate diversity to establish what is or is not grammaticalization. Frequency in particular would be problematic as a diagnostic feature because, logically, any new grammatical use of a form will start out at low frequency; moreover, low-frequency source constructions are attested (e.g. Hoffmann 2004) and grammatical elements themselves are not immune to frequency decline. What we learn from its prominence in the models, then, is that frequency is nonetheless a highly relevant concomitant of grammaticalization. Still, we do need an independent working definition after all, one that refers directly to function. If what can be measured are merely the symptoms rather than the thing itself, then so be it. The quantitative measures still prove their worth in revealing common patterns, and also, prospectively, in systematizing the subtypes of the process.

Another open question is how widely the measurements can be generalized. Assume that some kind of cline between 'lexical' and 'grammatical' elements is a linguistic universal – would the quantifications then be directly applicable to other languages? It rather seems that typological differences set different preconditions, for example the positioning of modifier and head in a phrase. At the very end of the

book, Correia Saavedra mentions an application to Chinese, where indeed the left/right asymmetry of colligate diversity seems to be reversed compared to English (Sun and Correia Saavedra 2020). This shows that a transfer of the approach to other settings is possible (with adaptations) and, again, this holds a vast potential for further development.

One thing that makes grammaticalization so fascinating is that it is realized by tiny steps but plays out on a grand scale. Any operationalization of the concept, quantitative or otherwise, faces the challenge of straddling the range from detailed features to broad trends. *Measurements of grammaticalization* offers a valuable macro-perspective on grammatical change. What is more, it offers a way forward to study such changes with quantitative data analysis and empirical rigor. This is no small achievement. It is to be hoped that the linguistic research community will take up the approach and develop it further. Then we will see how far it takes us in mapping out the tiny steps as well as the grand scale of the phenomenon.

## References

Diewald, Gabriele. 2006. Context types in grammaticalization as constructions. *Constructions* Special Volume 1:9. 1–29. <https://doi.org/10.24338/cons-443>.

Heine, Bernd. 2002. On the role of context in grammaticalization. In Ilse Wischer & Gabriele Diewald (eds.). *New reflections on grammaticalization*, 83–101. Amsterdam: John Benjamins.

Heine, Bernd & Tania Kuteva. 2007. *The genesis of grammar: A reconstruction*. Oxford: Oxford University Press.

Hoffmann, Sebastian. 2004. Are low-frequency complex prepositions grammaticalized? On the limits of corpus data – and the importance of intuition. In Hans Lindquist & Christian Mair (eds.). *Corpus approaches to grammaticalization in English*, 171–210. Amsterdam: John Benjamins.

Hopper, Paul & Elizabeth C. Traugott. 2003. *Grammaticalization*, 2nd edn. Cambridge: Cambridge University Press.

Lehmann, Christian. 2015 [1982]. *Thoughts on grammaticalization*, 3rd edn. Berlin: Language Science Press.

Narrog, Heiko & Bernd Heine. 2021. *Grammaticalization*. Oxford: Oxford University Press.

Norde, Muriel & Karin Beijering. 2014. Facing interfaces: A clustering approach to grammaticalization and related changes. *Folia Linguistica* 48(2). 385–424.

Pétré, Peter & Freek Van de Velde. 2018. The real-time dynamics of the individual and the community in grammaticalization. *Language* 94(4). 867–901.

Sun, Linlin & David Correia Saavedra. 2020. Measuring grammatical status in Chinese through quantitative corpus analysis. *Corpora* 15(3). 317–342.

Wiemer, Björn. 2014. Quo vadis grammaticalization theory?, or: Why complex language change is like words. *Folia Linguistica* 48(2). 425–467.