Attractor states and diachronic change in Hawkins's "Processing Typology"[1]

Karsten Schmidtke-Bode
Leipzig University and Friedrich Schiller University Jena

SUPPLEMENTARY MATERIALS (SM)

Contents

SM1 The Family Bias Method 1
SM2 Log-linear modelling of family biases 6
SM3 Mixed-effects regression modelling of the data 6
  SM3.1 The model reported in the paper 7
  SM3.2 Alternative models 8
SM4 Articles in VO languages 11
  SM4.1 Definiteness words 12
  SM4.2 Article words 13
SM5 Other factors in the diachrony of articles 14
  SM5.1 Case marking on NPs 14
  SM5.2 Verbal aspect 18
References 20

SM1: The Family Bias Method

In what follows, I provide a relatively non-technical introduction to Bickel's (2011; 2013) Family Bias Method (henceforth **FBM**) for uninitiated readers, in order to explicate how the data in Section 3 of the paper were generated.

The FBM proceeds from the idea that typological distributions are products of historical developments; so if we postulate that certain functional forces are responsible for the synchronic typological picture, we should expect them to guide the diachronic development of languages around the world. The FBM first probes the developments *within* genetic **taxa** (families, genera, etc.) and then compares whether the same developmental trends

---

[1] The present document, published online in 2019, supplements an article that I contributed to: Schmidtke-Bode, Karsten, Natalia Levshina, Susanne Maria Michaelis & Ilja A. Seržant (eds.). 2019. *Explanation in typology: Diachronic sources, functional motivations and the nature of the evidence*. [Conceptual Foundations of Language Science] Berlin: Language Science Press.

hold *across* the world’s taxa and independently of areal affiliations (e.g. across the commonly distinguished macro areas, see Dryer 1989a). The reasoning behind examining **intragenealogical** developments is the following: When protolanguages split off into various generations of offspring, each daughter is taken to represent an evolutionary trial, as it were, which may either keep or alter a certain structure of interest. If viable universal pressures on this structure are operative, one expects that the daughters either retain the preferred pattern if this was already present in their respective protolanguage or else change it to match the universal, rather than going against it. For example, if we hypothesize that there is a universal functional pressure for A-arguments to be placed before P-arguments in transitive clauses, we would expect that daughter languages either retain A>P order from the protolanguage or else likely change the inherited P>A order to A>P; conversely, the FBM predicts that long-term retention of P>A or a change towards P>A are significantly less likely to occur if such a pressure is indeed at work.

Now, in the absence of detailed diachronic records for most of the world’s language families, developmental patterns are inferred from inspecting the **synchronic internal composition** of the taxon in question: If, among its extant members, the order A>P significantly outnumbers the order P>A, the taxon is said to be “**biased**” towards A>P order; if P>A significantly outnumbers A>P, the taxon is “biased” in the opposite direction; finally, if no significant majority value can be established, the taxon is said to be internally “**diverse**”.[2] Evidence for linguistic universals can be obtained when these results from individual families are compared across the sample: If it turns out that families tend to be biased in the same direction, and in much the same manner in all macro areas, there is evidence for what Bickel (2013) calls a “**directional family bias**”: For example, if significantly more families are biased towards A>P order than towards P>A order, this is said to reflect a universal pressure at work. By contrast, if most of the sample families are biased, but in different directions (i.e. some families towards A>P, others towards P>A), we have a “**non-directional family bias**” and, Bickel argues, no evidence for universal functional pressures. In this situation, it is more likely that family-internal biases are due to faithful inheritance, without a lot of changes going on. Finally, a scenario in which most of the sample languages turn out to be internally diverse does not yield evidence for universal functional pressures either: In such cases, we have various changes, but no clear direction that these take within and across families.

---

[2] Statistical significance is established by a binomial test with $\alpha = 0.1$. The FBM has been made available for public use as the package `familybias` for *R* (Zakharko & Bickel 2011ff.). It comes with detailed documentation of the statistical procedure.

It should have become clear from this exposition that the FBM presupposes a sampling method that deviates radically from traditional controlled sampling in linguistic typology, precisely by taking multiple data points per taxon – preferably, one actually samples as densely as possible within these taxa. But since many data points in typology come from very **small families** and especially **isolates**, the question is how the FBM can incorporate these cases: Because of the limited number of family members (i.e. data points sampled), biases of the above type cannot be calculated but need to be estimated based on what we know from trends in the large families. In other words, we **extrapolate** the results from the larger families to the smaller taxa in the sample. The underlying theoretical argument is that isolates are “loners” only from a synchronic point of view – in fact, however, they are the only remnants of genealogical units that were more numerous at some point in the past but have been reduced, as it were, to a single survivor. Therefore, the synchronic behaviour of the isolate may or may not reflect the situation in the family before it disappeared – since languages do not become extinct due to functional unfitness but for sociolinguistic reasons, it is possible, after all, that an isolate ends up having a linguistic feature that was not representative of the family from which it descended. For this reason, the FBM does not simply take the synchronic data from small families and isolates at face value, but takes into account that they may be deviates from the majority value of their erstwhile family. This is important to understand because it is actually the major point of departure from purely synchronic ways of testing for universals, as Family Bias estimations change the raw data at hand. In order to appreciate this point, we will briefly describe a concrete (but fictitious) example, relating to the order of A and P again.

Suppose that a typological sample includes eight large families, i.e. taxa with at least five representative members in one’s database (by the criterion proposed in Bickel 2011, 2013). For each of those eight families, one can determine how many of its members show the A>P order predicted, for example, by one of Hawkins’s (2004) processing principles, or the alternative P>A order. A binomial test will estimate whether the A>P languages significantly outnumber the P>A languages (or vice versa) or whether there is no significant distributional skewing in the family. In concrete terms: If a family with 12 data points shows 10 languages conforming to the predicted A>P and 2 exhibit P>A, the family is said to be significantly biased (in the predicted direction, $p = 0.04$); if the numbers were swapped, it would also be significantly biased, but in the opposite direction. By contrast, if the same family contained 8 conforming and 4 violating members, no significant bias would emerge ($p = 0.38$), even though the conforming members would be in the majority. This is an important point since it means, for example, that what *looks* like a skewed family (say, 7:2 in favour of one's hypothesis) is not significantly biased under a binomial test ($p = 0.18$) and hence considered an unbiased, i.e. diverse, family. This is because inferential statistical testing of this kind is about the certainty with which we can reject the null hypothesis, and small samples simply do not give one enough confidence to generalize to the underlying population, i.e. to infer the true trend in the family at hand. Therefore, if most of the large (i.e. $n \geqslant 5$) families in one's sample do not contain a substantial number of languages, one will obtain many families that one has to consider unbiased on statistical grounds, even though they may contain seemingly few counterexamples to one's prediction.

The next step in the procedure is to calculate the **proportion** of large families that are biased (as opposed to unbiased/diverse), and to calculate **how strongly** they are biased. These calculations are then taken as the basis for estimating biases in smaller families and isolates. For example, if eight out of ten large families are biased (*in whatever direction!*), the likelihood of being biased is 0.8. Consequently, 80% of the smaller taxa are declared "biased" and 20% are declared "unbiased/diverse"; this allocation is done on a random basis. Thus if our sample contains 130 small families and isolates, 104 of these will be declared "biased". Afterwards, the average strength of the bias in large families is extrapolated to those 104 taxa. For example, if large families tend to be biased to 94% (i.e. across all large biased families, a mean of 6% of the family members do not show the majority value of the family), then 94% of our 104 taxa from above will be declared as "bias-conforming" and 6% as "bias-violating"; this is done randomly again. In this way, 98 small taxa contribute the value that they actually have (e.g. A>P), while the remaining six are assigned the opposite value (i.e. P>A) even though they actually have A>P. Since these random assignments thus invariably entail some misclassifications, the procedure is repeated hundreds or even thousands of times and the average results of all estimations are then taken as the final basis for exploring universal trends. For this reason, the results of the FBM return non-integer numbers, which are then usually rounded for expository purposes.

Against this background, let us return to the actual findings presented in the paper. The design is somewhat more complicated than the A>P example above, since we are no longer dealing with a single variable like a word-order preference but a **correlation** between two variables, viz. the presence of article morphemes (as defined in the paper) and the order of verb and object. In such research designs, the FBM first inspects whether the large taxa in the sample are internally homogeneous with regard to the predictor variable, i.e. the VO/OV distinction. For instance, the Trans New Guinea taxon is represented by 21 languages in the sample, all of which exhibit OV order. We can thus analyse directly how the response variable, i.e. the presence of article morphemes, is distributed in this taxon. By contrast, the Oceanic taxon contains 48 VO and 9 OV languages. In order to probe the hypothesized correlation, we need to separate these two groups and inspect them individually: How did the Oceanic OV languages develop with regard to articles, and how did the Oceanic VO languages do? Bickel (2011, 2013) calls these artificial units “pseudo-groups”, precisely because they do not correspond to actual genetic taxa anymore.

Considering all large taxa that are OV/VO-homogeneous to begin with (such as Trans New Guinea) and all pseudo-groups (such as “Oceanic VO”), our sample contains **29 large taxa** altogether. As can be seen in Table 4 in the paper, replicated here as Table 1, these fall into 18 VO and 11 OV taxa. Within these taxa, we can now examine which ones show significant biases (towards or against!) article morphemes and which ones are unbiased, i.e. internally diverse. For example, the West Chadic taxon is invariably VO, and all 9 members exhibit article morphemes as defined in the paper. This is, of course, a highly significant distribution (p = 0.004) and hence a “bias”. By contrast, the Altaic taxon (in which all 9 data points are OV) shows 5 members with and 8 members without articles; there is no significant bias here (p = 0.58) and hence this taxon is considered “diverse”. The Oceanic VO languages from above, a pseudo-group, has 39 members with article morphemes and 9 without; again, this is a significant “bias” (p < 0.001). And so on.

Table 1. Distribution of biases (for or against) articles among *large* families in the sample (N_{total} = 29 genetic units)

	VO	**OV**	**Totals**
significantly biased	12	0	12
internally diverse	6	11	17
Totals	**18**	**11**	**29**

In a second step, these and further calculations on the large families are extrapolated to smaller families and isolates, in the fashion described for our fictitious example above. After repeating the extrapolation 2,000 times, averaging over the results and rounding them to integers again, we obtain the picture given in Table 3 of the paper, reproduced as Table 2 here:

Table 2. (Rounded) family biases for articles in different word-order types (N_{total} = 217 genetic units, 99 of which are estimated to be “biased” (as opposed to internally “diverse”))

	VO	**OV**	**Totals**
ART morph	50	19	69
No ART morph	15	15	30
Totals	**65**	**34**	**99**

As can be seen, these final results only take the "biased" taxa into account, i.e. all "diverse" taxa have been discarded at this point, and the distribution can now be tested for significance (Fisher exact test, $p = 0.039$, $OR = 2.6$).[3] In other words, the hypothesized correlation is borne out by the distribution of family biases. The final step missing, however, is to test for the areal independence of this correlation. This will be illustrated in the following section.

SM2: Log-linear modelling of family biases

Bickel (2011) proposes log-linear modelling for probing whether a given structural correlation holds across all macro-areas distinguished in one's sample.[4] Figure 2 in the paper shows that Hawkins's hypothesized correlation holds in a similar fashion across all six macro areas: While it is not always *individually* significant (because the number of biases tends to be very small for each area), the relationship between word order and articles is never reversed. Accordingly, a log-linear model in which the presence of articles is regressed against both word order and macro area does not yield a significant interaction between all three parameters. Put differently, there is not enough evidence for claiming that the hypothesized correlation is substantially different in the various macro areas. The only significant two-way interaction that the modelling process yielded was precisely the one between word order and articles ($\chi^2(1) = 4.1$, $p = 0.04$), in keeping with Hawkins's prediction.

SM3: Mixed-effects regression modelling of the data

As outlined in the paper, the Family Bias estimations are complemented here by a statistical technique that is widely applied in other disciplines dealing with variation data and which was proposed for linguistic typology by Cysouw (2010) and Jaeger et al. (2011), viz. mixed-effects regression modelling.[5] The basic procedure is to model the presence of article morphemes as a function of the word-order predictor. But in order to account for the fact that the sample contains multiple data points for many families and multiple data

---

[3] All statistical modelling presented in this document was performed in *R* 3.3.1 (*R* Development Core Team 2016).

[4] For running the log-linear model, I used the `glm()` function in *R*, i.e. a generalized linear model, based on the Poisson distribution.

[5] For creating these models and their visualizations given in the paper, I employed the *R* packages `lme4` (Bates et al. 2015), `effects` (Fox et al. 2016), `rms` (Harrell 2017) and `MuMIn` (Bartón 2016). In all cases, I proceeded by backwards elimination of non-significant terms, beginning with the so-called random effects (Baayen 2008: Ch. 7). In what follows, I will report on the key parameters of each model only (e.g. whether or not the hypothesized effect from word order emerges as significant, and which random effects remained in the model); in case of a *significant* signal of word order (or other predictors), I additionally provide some model diagnostics in the footnotes; otherwise no such information will be given.

points per macro area, the latter two dependencies in the sample are modelled as so-called “**random effects**” in addition to the “**fixed effect**” of word order that we primarily seek to investigate. Specifically, we try to filter out that different families and areas may have different “baseline preferences”, as it were, for developing article morphemes, and we ask if a significant correlation with word order can be established once these baseline preferences are controlled for.[6]

In contrast to the Family Bias Method, the model resulting from a mixed-effects regression procedure is based on all the data in the sample (i.e. it does not discard taxa that the FBM would consider “internally diverse”), and it does not operate with extrapolations (i.e. it does not change the raw data in any way). On the downside, such models are purely synchronic, both in terms of the data as such *and* the inferences drawn from them; the FBM likewise works only with synchronic data, but models them with certain diachronic considerations in mind, so that inferences can be drawn to possible developments within families. As both procedures thus have particular strengths and potential weaknesses, I am using them in conjunction.

SM3.1 The model reported in the paper

An initial model that contained random intercepts for two different genetic groupings of the data (FAMILY and GENUS) as well as a random intercept for AREA failed to converge, as did the same model without a random intercept for FAMILY.[7] A feasible model could be obtained when genetic relationships were taken into account only at the FAMILY level (i.e. highest taxonomic grouping), discarding GENERA-specific variation. This model yielded a significant effect of word order on the presence of article morphemes ($\beta = -0.73$, $p < 0.001$). As can be seen in an effects plot (reproduced as Figure 1 here), the probability of *not* having article morphemes (y-axis) increases significantly as we go from VO to OV (x-axis), just as predicted by Hawkins (2014).

In subsequent ANOVAs, the random intercept for AREA ($sd = 0.56$) turned out to be significant, while the one for FAMILY ($sd = 0.31$) was actually immaterial; accordingly, its presence or absence did not change the fact that the fixed effect of word order was significant. Interestingly, a model additionally

---

[6] Technically, this is done by modelling so-called “random intercepts” for the genetic units (e.g. families and/or genera) and areas in the sample. In addition, large datasets also allow controlling for the possibility that a hypothesized correlation (here: the word-order effect on articles) may be different (e.g. in strength or even directionality) for different families and areas. This amounts to introducing so-called “random slopes” into the regression equation; see Baayen (2008) or Jaeger et al. (2011) for an introduction.

[7] In all reports on regression models, variables modelled as random effects are indicated by SMALL CAPITALS.

containing a by-AREA random slope for the word-order effect on articles did not improve the original model significantly; in other words, the ways in which the predicted relationship between word order type and article morphemes varies across macro areas are negligible.

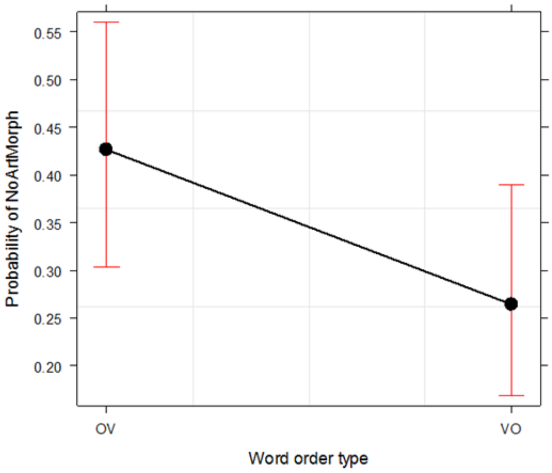


Figure 1. Effect of word order on the probability of (not) having articles in a mixed-effects model (R^2_c = 0.14, C = 0.72)[8]

SM3.2 Alternative models

In footnote 7 of the paper, I briefly mention potential disagreement with regard to the very definition of what should count as an “article morpheme” in the first place. The analysis given in the paper is based on a broad view that also includes languages with incipient grammaticalization of demonstratives into definite articles. But one could say that such cases do not (yet) qualify as sufficiently distinct from demonstratives and hence not as “proper article” morphemes. Accordingly, one may wish to consider these languages along with those that Dryer (2013a) classifies as not having any articles. The raw distribution then looks as follows (Table 3, analogous to Table 2 in the paper):

---

[8] The goodness-of-fit of a mixed-effects model can be expressed in terms of marginal R^2 (R^2_m), which takes only the fixed effects into account, and in terms of conditional R^2 (R^2_c), which additionally takes the random effects into consideration (Nakagawa & Schielzeth 2013). The concordance index C indicates the predictive power of the model. Some model diagnostics: The BLUPs for AREA are normally distributed while those for FAMILY are not. There are no outliers, but some overdispersion (1.18***). The latter was calculated by dividing the residual deviance of the model by its associated degrees of freedom (see Gries 2013: 315); additionally, I also adopted Baayen’s (2008: 199) suggestion for running a significance test for overdispersion, whose result is indicated by the stars above, following the conventional notation for different levels of significance. All in all, our model containing the word-order parameter as the only fixed effect is too simplistic: It is overdispersed, leaves quite a bit of variation unaccounted for and thus likely misses further predictors.

Table 3. Distribution of articles in different word-order types (alternative coding)

	VO	**OV**	**Totals**
ART morph	213	109	322
No ART morph	100	144	244
Totals	**313**	**253**	**566**

Since incipient articles are found in both OV and VO languages, the figures have changed for both language types when compared to the earlier analysis. But is there still a significant difference between VO and OV?

The FBM yields the following results (after 2,000 extrapolation rounds for small families and isolates):

Table 4. (Rounded) family biases for articles in different word-order types (N_{total} = 217 genetic units, 91 of which are estimated to be “biased” (as opposed to internally “diverse”))

	VO	**OV**	**Totals**
ART morph	41	16	57
No ART morph	17	17	34
Totals	**58**	**33**	**91**

As compared to the previous analysis, we now obtain even more taxa that are “diverse” (118 > 126) rather than “biased” (99 > 91). Also, the distribution of the OV taxa has changed from 19:15 in favour of articles to 16:17, i.e. precisely in the predicted direction. However, since the VO taxa also “lost” a few article biases and gained two biases for not having articles, they balance out the change in the OV taxa. As a result, the distribution in Table 4 is similarly skewed as the one in Table 2 above, i.e. just about significant in a Fisher exact test (p = 0.044, OR = 2.53). Subsequent log-linear modelling for areal independence yields some interesting results: On the one hand, it shows again that there is no significant three-way interaction between word order, articles and macro areas, indicating that the word-order effect on articles does not differ depending on areal affiliations. This is illustrated in Figure 2.

On the other hand, however, the log-linear model also reveals that the hypothesized word-effect itself does *not* reach the conventional alpha level (χ^2 (1) = 3.27, p = 0.07), in contrast to the Fisher exact test above, and that, instead, a significant interaction emerges between word order and area ((χ^2 (1) = 12.62, p = 0.027).

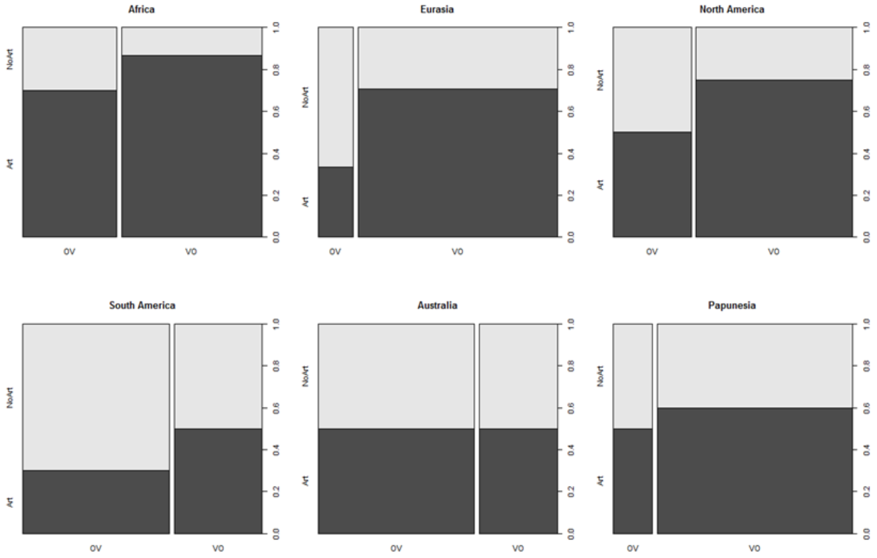


Figure 2. Family biases by macro area (alternative coding of the data)

In this kind of situation, it seems particularly useful to probe a complementary statistical model based on standard regression methods again. A mixed-effects model that contains random intercepts for FAMILY, GENUS and AREA shows that a random effect from the highest taxonomic level is not significant; keeping the remaining random intercepts in the model (sd_{GENUS} = 0.98, sd_{AREA} = 0.53), we obtain a highly significant effect from word order on articles (β = -1.04, p < 0.0001). This is illustrated in Figure 3:

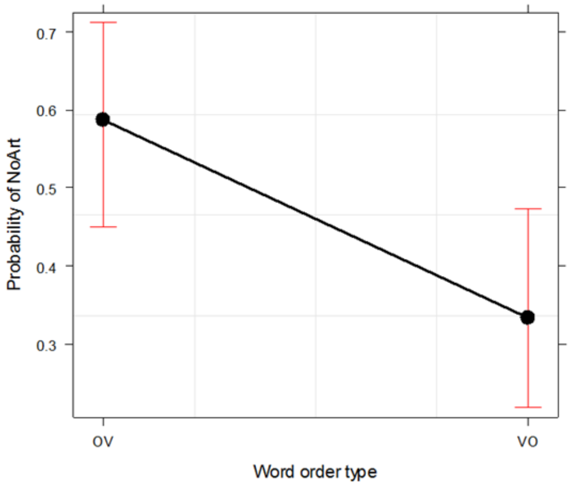


Figure 3. Effect of word order on the probability of (not) having articles in an alternative mixed-effects model (R^2_c = 0.32, C = 0.87)[9]

---

[9] Some model diagnostics: The BLUPs for AREA are normally distributed while those for GENUS are not. There are no outliers, but some overdispersion (1.26***).

Just as in the previous analysis, adding a by-AREA random slope for the word-order effect does not make a significant contribution to the model. Again, then, there is no indication that the word-order effect on articles is substantially different across the macro areas.

In this case, therefore, we have two slightly conflicting analyses: Purely synchronic regression models on *all* data points in the sample yield a highly significant effect in the hypothesized direction, and independently of macro areas. That is, if we count languages with some article-like uses of demonstratives as not having developed “proper” article morphemes, we still find a correlation between word order and articles; in fact, the mixed-effects model is actually a better one than before (compare the goodness-of-fit values given in Figures 1 and 3 above). In other words, if we concentrate on fully grammaticalized (rather than also emerging) instances of articles, the synchronic typological picture is even further in line with Hawkins (2014). On the FBM, by contrast, such a clear correlation does not emerge. Following Bickel’s reasoning, we would thus be forced to conclude that the alleged processing motivation for (not) developing article morphemes is not strong enough to influence the diachronic development of families in a consistent way across the world. But we also have to keep in mind that the FBM works with extrapolations, discards most taxa from the sample because it judges them to be “diverse” according to its own ways of testing family-internal distributions, and so on. I am thus inclined to give the mixed-effects model somewhat more weight in the interpretation of the data.

SM4: Articles in VO languages

In this section, I provide an overview of the mixed-effects models that were run in order to investigate whether VO languages with ADJ-N order (e.g. *an* [*extremely long*] *movie*) are especially prone to develop articles (as predicted by Hawkins (2014: 125)). For the sake of consistency, we stick to the *WALS* data on articles again (Dryer 2013a), narrowed down to the VO languages in the sample (based on Dryer 2013b), and cross-classify these data with Dryer’s (2013c) data on ADJ-N order. The sample for this analysis thus reduces to a subset of the original 566 languages to VO languages that are coded for both article morphemes and ADJ-N order, a total of 278 languages. As with the previous analyses, the various types of article morphemes identified by Dryer (2013a), given in Table 1 of the paper, can be grouped in different ways for analytical purposes. In what follows, I briefly present the results of several possible analyses.

SM4.1 Definiteness words

As a first step, one may wish to further reduce Dryer’s (2013a) article data to free morphemes, for the following reason: Hawkins’s hypothesis for VO languages actually only makes sense for prenominal article morphemes. While the *WALS* data on articles are not coded for their position relative to the noun, I believe we can target such prenominal articles in VO languages better (i) if we leave out affixed articles, as these will tend to be postnominal (i.e. suffixes, by the generalization in Bybee et al. 1990), (ii) if we thus concentrate on free article morphemes, as they tend to be prenominal in VO languages (by the generalization in Dryer 1989b). While this procedure does not yield prenominal articles only, it is the best approximation that I can offer at this point. The following analysis is thus based on the coding indicated in the last column of Table 5:

Table 5. Distribution of articles in VO languages, depending on ADJ-N order (Dryer 2013a,b,c)

	ADJ-N	N-ADJ	Totals	Coding
Distinct DEF word	36	94	130	“DEF Word”
DEF affix	10	29	39	---
DEM used as DEF	4	22	26	“DEF Word”
Only INDEF ART	3	13	16	“No DEF word”
No ART	14	53	67	“No DEF word”
Totals	**67**	**211**	**278**	

Note that languages without a dominant order of adjective and noun were not taken into account here. Based on Table 5, we can now cross-classify the parameter of ADJ-N order with the coding for definiteness words (Table 6):

Table 6. Definiteness words in VO languages depending on ADJ-N order

	ADJ-N	N-ADJ	Totals
DEF word	40	116	156
No DEF word	17	66	83
Totals	**57**	**182**	**239**

These raw data are visualized in the left panel of Figure 3 in the paper, and we run a mixed-effects model on them, we obtain the following result: Random intercepts are significant for GENUS only, but not for FAMILY or AREA; with such a random intercept for GENUS included (*sd* = 1.2), the effect of adjective order is not significant ($\beta = 0.37$, $p = 0.34$), just as the raw data themselves would have suggested. There is thus no evidence that VO languages with

ADJ-N order are significantly more likely to develop definiteness words than those with N-ADJ order.[10]

SM4.2 Article words

In an alternative analysis, we now also include languages with indefinite articles only, but again we remove the affixes from this category (using Dryer’s (2013d) coding of indefinite articles). What we sample then, are languages

- with definiteness words, including incipient uses of demonstratives as articles
- with indefiniteness words, including those that use ‘one’ as indefinite article

and we oppose them to languages without definite and indefinite article words. (Again, languages with affixes only are discarded).

Following Hawkins’s line of thinking, one may also wish to think about adjective order in a more complex fashion: On the one hand, we can simply compare ADJ-N and N-ADJ languages, as we did in Section 4.1 above. On the other hand, one may argue that languages without a dominant order of noun and adjective actually have the critical ADJ-N condition as a frequent option, so that here, too, an increased processing pressure for developing article words may be operative. Therefore, let us run two different models on the data and compare them. Here is what the raw data look like:

Table 7. Article words in VO languages depending on ADJ-N order

	ADJ-N	**N-ADJ**	**flexible**	**Totals**
ART word	34	101	8	143
No ART word	14	53	0	67
Totals	**48**	**154**	**8**	**210**

In the first analysis, we simply compare the ADJ-N and the N-ADJ orders with regard to article words. In a mixed-effects model with a random intercepts for GENUS (*sd* = 1.76), the effect of adjective order does not come out as significant ($\beta = 0.578$, $p = 0.295$). The corresponding visualization is given in the right panel of Figure 3 in the paper. In the second analysis, we subsume the “flexible” languages in Table 7 under the ADJ-N languages. The resulting mixed-effects model on these data includes a random intercept for GENUS (*sd* = 1.42) and, again, the effect of adjective order is not significant ($\beta = 0.84$, $p = 0.13$). On neither analysis, then, is there sufficient evidence that VO lan-

---

[10] The picture does not change at all if the languages with definiteness affixes are included and coded as not having definiteness words.

guages with ADJ-N order are significantly more likely to develop article words than those with N-ADJ order.[11]

SM5: Other factors in the diachrony of articles

Towards the end of the paper (see footnote 14), I point out that research in historical linguistics has actually put forward alternative causal forces in the development of articles, notably information-structural consequences of the loss of **case** and the loss of an **aspectual** system. I do not know of any typological study that has investigated the correlational structure of word order, case systems, aspectual distinctions on the verb and the presence of article morphemes, particularly in a sufficiently large database. At present, the amount of data available in *WALS online* only allows for some preliminary investigations, but I shall include these analyses here for interested readers.

SM5.1 Case marking on NPs

The rise of articles has been linked to the loss of case marking on NPs, either for considerations of online processing (Hawkins 2004) or, more indirectly, as a compensatory functional strategy: The loss of core cases typically engenders a rigidification of constituent order at the clause level, which in turn reduces the former potential of constituent order to be exploited for discourse-pragmatic purposes. In this line of argumentation, article morphemes (both definite and indefinite articles) become grammaticalized precisely in order to mark information-structural distinctions that can no longer be signalled by the syntacticized constituent order (Vennemann 1975, or more recently Hewson & Bubenik 2006, Fischer 2010, Carlier & Lamiroy 2014, among others).

If this is true in general, we would thus ideally also obtain a global, i.e. typological, correlation between the absence of case marking on NPs and the presence of article morphemes. Again, we turn to *WALS* for a preliminary investigation. As for the article data (Dryer 2013a), we stick to the coding introduced in Table 1 of the paper; in other words, we consider all kinds of article morphemes, definite and indefinite, free or bound, incipient or grammat-

---

[11] In response to this finding, John Hawkins (p.c.) referred me to a potential explanation "of why the adjective ordering prediction is weak": In comparison to other types of phrases, APs tend to be shorter on average, often consisting of a single adjective only. Hawkins has suggested elsewhere (notably in Hawkins 2004: 25–26 and Hawkins 2014: 101–102) that especially single-word APs do not have much effect on ordering preferences in corpora and grammatical conventions. This is also why, in his processing account, there is no correlation between the order of nouns and adjectives with the order of verb and object: in contrast to other, usually more complex types of constituents (e.g. PPs or relative clauses), the pressure for minimizing processing domains is very weak for APs, and hence we find more ordering variation in grammars than for these other types of phrases.

icalized. The data for case marking come from Comrie (2013); we can rearrange Comrie’s coding into a basic contrast between his “neutral” type (= languages without case marking on NPs) and all others (e.g. nominative-accusative, ergative-absolutive, tripartite, etc.). If we cross-classify the two variables, we obtain the following raw result:

Table 8. Distribution of articles depending on case marking on full NPs

	+Case	**-Case**	**Totals**
ART morpheme	30	44	74
No ART morpheme	29	23	52
Totals	**59**	**67**	**126**

As can be seen, we are dealing with a rather small sample this time, which is why the results need to be interpreted with some caution. Note also that the sample is more balanced to begin with, i.e. there are rather few data points from the same genetic units this time. Consequently, in a mixed-effects model of the data, the random intercepts for FAMILY and AREA are not significant and show a standard deviation of roughly zero. The fixed effect of case goes into the predicted direction, as shown in Figure 4 below, but it fails to yield a significant signal ($\beta = -0.615$, $p = 0.09$).

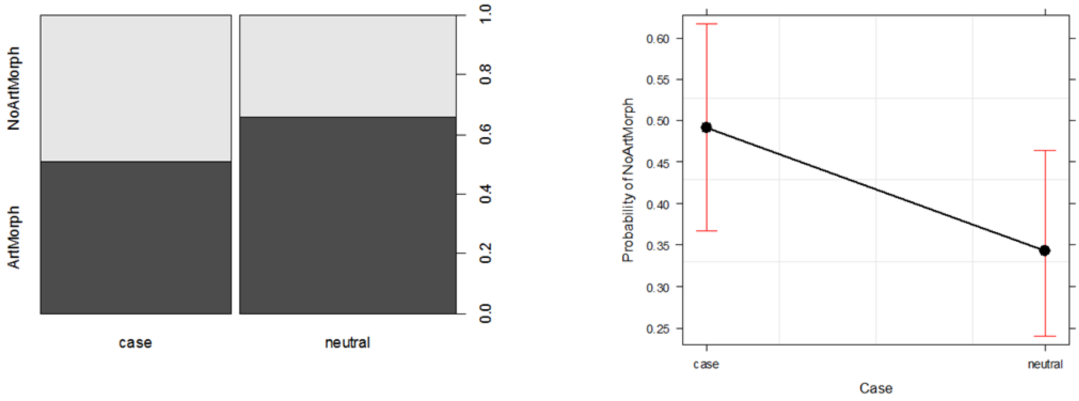


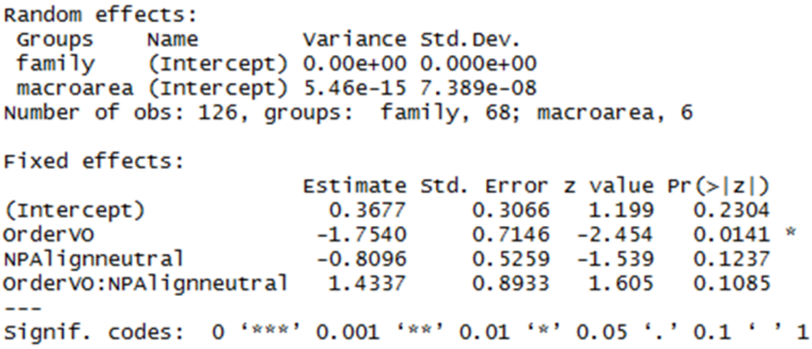
Figure 4. Effect of nominal case marking on the probability of (not) having articles in a mixed-effects model

The same analysis in terms of Bickel’s (2011, 2013) FBM leads to very similar conclusions. Table 9 shows how family biases for or against articles are distributed over the two case conditions. There is no significant difference here between taxa with and without case marking (Fisher exact test, $p = 0.53$, *OR* = 0.6), and subsequent log-linear modelling reveals that no relationships at all are significant.

Table 9. Distribution of articles depending on case marking on full NPs in FBM modelling

	+Case	**-Case**	**Totals**
Bias towards ART	10	15	25
Bias against ART	9	8	17
Unbiased (“diverse”)	26	27	53
Totals	**55**	**50**	**105**

Finally, one may wonder how the postulated effects of word order and case marking turn out if they are modelled simultaneously, i.e. in a multifactorial design. Figure 5 below provides the (partial) output of a mixed-effects regression model that contains word order, case marking and their interaction as fixed effects:



```
Random effects:
 Groups    Name        Variance Std.Dev.
 family    (Intercept) 0.00e+00 0.000e+00
 macroarea (Intercept) 5.46e-15 7.389e-08
Number of obs: 126, groups:  family, 68; macroarea, 6

Fixed effects:
                      Estimate Std. Error z value Pr(>|z|)
(Intercept)             0.3677     0.3066   1.199   0.2304
OrderVO                -1.7540     0.7146  -2.454   0.0141 *
NPAlignneutral         -0.8096     0.5259  -1.539   0.1237
OrderVO:NPAlignneutral  1.4337     0.8933   1.605   0.1085
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 5. Multifactorial mixed-effects model for the influence of case marking and word order on the development of article morphemes

We can see here, first, that there is hardly any variation from the random effects of FAMILY and AREA. Second, while the fixed effect of word order comes out as significant again, that of case marking does not, and neither does the interaction between them.[12] In other words, when word order and case marking are directly pitted against each other, the former appears to be the better, and in fact the only significant, predictor of article development.

The interaction term, while not significant as such, is nevertheless interesting from a linguistic point of view: Its general direction suggests that, when a language has no case marking on NPs (“neutral” alignment in Comrie 2013), its basic word order is quite immaterial for the development of articles, i.e. there is little difference between OV- and VO languages of this type with regard to articles. For case-marking languages, by contrast, the word-order difference matters much more: Case-marked VO languages tend to develop articles, while case-marked OV languages tend not to. This is visualized in Figure 6:

---

[12] There are no signs of multicollinearity between word order and case in this model, judging by the variance inflation factors (Fox & Monette 1992).

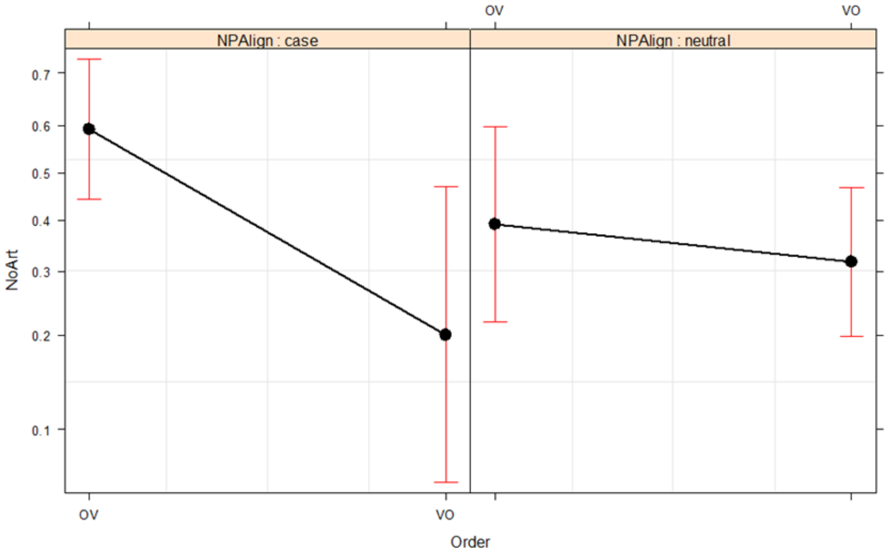


Figure 6. Interaction between case marking and word order in the development of article morphemes ($\beta = 1.43$, $p = 0.11$)

Especially because of these potentially interesting interactions, the relationship between word order, case marking and articles deserves to be investigated in a substantially larger sample than the present one.

Before we leave the topic of case marking, it should be noted that, in the historical linguistic literature, a causal connection between the loss of case and the rise of articles is sometimes seen as untenable or at least quite dubious. Leiss (2000: 225) calls Vennemann’s (1975) hypotheses “Fehlannahmen” [misconceptions] and “widerlegt” [disproved], based on diachronic evidence from the Germanic languages; a similarly skeptical view is held, for example, by McColl Millar (2000) on English. He claims that the definite article in English developed

> “not so much because the language felt an overwhelming need for such a form, but rather because a gap had opened in the semantic fabric of the language due to the specialisation in meaning of *that*. *The* is, in other words, an historical accident. [...] If we look at languages such as Greek, or the Scandinavian dialects, we can see that they developed dedicated determiners before they shifted along the synthetic to analytic continuum. By the same token, there are languages such as Farsi which have also lost all but the smallest vestiges of grammatical gender and case, but have not developed a definite article.” (275–276)

There is thus quite a bit of controversy surrounding the relationship between case and articles.

SM5.2 Verbal aspect

Building on Abraham (1997), Leiss (2000, 2007) has developed the hypothesis that there is a diachronic connection between changes in the aspectual system of a language and the rise or loss of article morphemes. According to Leiss (2007: 73–93), this is due to a certain functional complementarity between the two phenomena:

> “Perfective aspect creates definiteness effects and, vice versa, definite articles create aspectual effects. The same holds for the imperfective aspect, which is able to create indefiniteness effects.” (84) In Gothic, for example, “the perfective aspect [...] blocks the use of the definite article, being strong enough to create definite expressions” (93).
>
> Therefore, “[i]t is assumed that the rise of the definite article is due to changes in the aspectual system of a language. Definiteness and perfective aspect are shown to be just two instantiations of the same grammatical function. [...] No sooner than the verbal part of the pattern, aspect, was subject to erosion, the first occurrences of definite articles can be observed. There is converging evidence from linguistic typology that aspect languages tend to avoid article systems, and article languages tend to avoid aspect.” (73)

It is the last sentence that is particularly interesting here because Leiss does not actually provide large-scale typological evidence for the alleged relationship between articles and aspect, nor does she mention studies furnishing such evidence. In order to make some (again, preliminary) inroads into this issue, we can cross-classify the presence of article morphemes (Dryer 2013a, with the binary coding adopted in the paper) with the presence of a grammatical perfective/imperfective distinction, also drawn from *WALS* (Dahl & Velupillai 2013). The sample naturally reduces to those languages which are coded for both features, amounting to N = 131 overall. The distribution of article morphemes in this sample is shown in Table 10:

Table 10. Distribution of article morphemes in languages with and without a grammatical distinction between perfective and imperfective aspect

	+Asp	**-Asp**	**Totals**
ART morph	38	46	84
No ART morph	23	24	47
Totals	61	70	131

We now proceed in exactly the same way as we did for case marking above. Firstly, we run a mixed-effects regression model on these data. In a model that includes random intercepts for FAMILY (sd = 0.51) and AREA (sd = 0.72), the hypothesized effect from aspect does not emerge as significant ($\beta = 0.358$, $p = 0.39$). This is visualized in Figure 7:

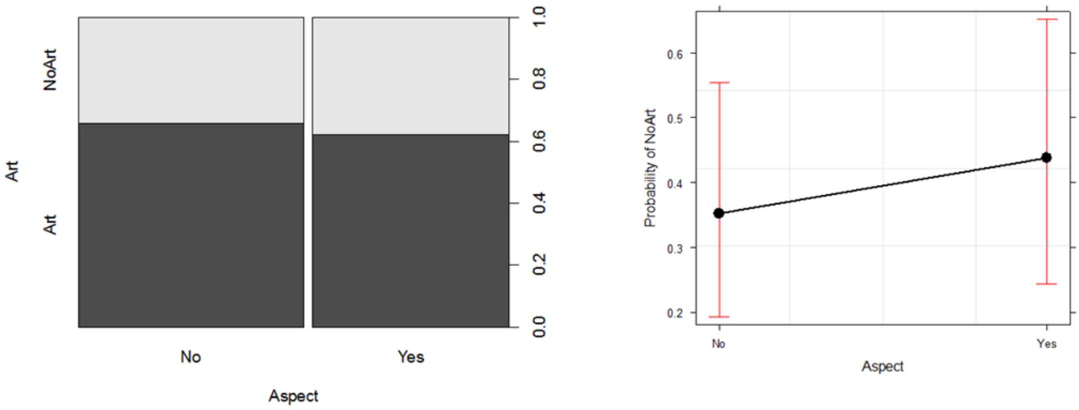


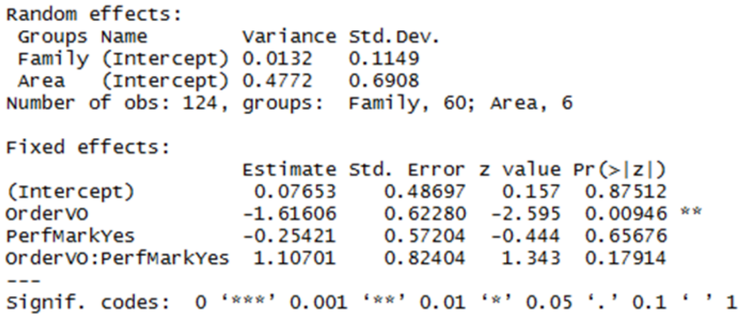
Figure 7. Effect of grammatical aspect marking on the probability of (not) having articles in a mixed-effects model

Secondly, we conduct an FBM analysis. Table 11 shows how family biases for or against articles are distributed over the two aspect conditions. There is no significant difference here between taxa with and without aspect marking (Fisher exact test, $p = 0.71$, *OR* = 0.64), and subsequent log-linear modelling reveals that no relationships at all are significant.

Table 11. Distribution of articles depending on grammatical aspect marking in FBM modelling

	+Asp	**-Asp**	**Totals**
Bias towards ART	6	11	17
Bias against ART	6	7	13
Unbiased (“diverse”)	23	21	44
Totals	**35**	**39**	**74**

Finally, we can pit word order and aspect marking against each other in a multifactorial mixed-effects model. Analogously to Figure 5 above, the following overview shows the (partial) output of such a regression model:



```
Random effects:
 Groups Name        Variance Std.Dev.
 Family (Intercept) 0.0132   0.1149
 Area   (Intercept) 0.4772   0.6908
Number of obs: 124, groups:  Family, 60; Area, 6

Fixed effects:
                    Estimate Std. Error z value Pr(>|z|)
(Intercept)          0.07653    0.48697   0.157  0.87512
OrderVO             -1.61606    0.62280  -2.595  0.00946 **
PerfMarkYes         -0.25421    0.57204  -0.444  0.65676
OrderVO:PerfMarkYes  1.10701    0.82404   1.343  0.17914
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 8. Multifactorial mixed-effects model for the influence of grammatical aspect marking and word order on the development of article morphemes

There is some variation from families and areas this time, but neither of these random effects makes a significant contribution to the model. The picture for the fixed effects looks similar to the one for case marking above: The hypothesized effect from grammatical aspect (coded as PERFMARK here) is not significant, and neither is the interaction between aspect marking and word order; the main effect from word order, however, emerges as significant again.[13] Overall, then, it seems that Hawkins's word-order hypothesis fares better than both aspect and case, at least in the small samples investigated here.

References

Abraham, Werner (1997). The interdependence of case, aspect, and referentiality in the history of German: The case of the verbal genitive. In Ans van Kemenade & Nigel Vincent (eds), *Parameters of morphosyntactic change*, 29–61. Cambridge: Cambridge University Press.

Baayen, R. Harald (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press.

Bartón, Kamil (2016). MuMIn: Multi-model inference. *R* package, version 1.15.6. <https:// CRAN.R-project.org>.

Bates, Douglas, Martin Maechler, Ben Bolker & Steven Walker (2015). lme4: Linear mixed-effects models using 'Eigen' and S4. *R* package, version 1.1-13. <https:// CRAN. R-project.org>.

Bickel, Balthasar (2011). Statistical modelling of language universals. *Linguistic Typology* 15(2). 401–413.

Bickel, Balthasar (2013). Distributional biases in language families. In: Balthasar Bickel, Lenore A. Grenoble, David A. Peterson & Alan Timberlake (eds.), *Language typology and historical contingency,* 415–444. Amsterdam, Philadelphia: John Benjamins.

Bybee, Joan L., William Pagliuca & Revere D. Perkins (1990). On the asymmetries in the affixation of grammatical material. In: William A. Croft, Suzanne Kemmer & Keith Denning (eds.), *Studies in typology and diachrony: Papers presented to Joseph H. Greenberg on his 75th birthday*, 1–42. Amsterdam, Philadelphia: John Benjamins.

Carlier Anne & Béatrice Lamiroy (2014). The grammaticalization of the prepositional partitive in Romance. In: Silvia Luraghi & Tuomas Huumo (eds.), *Partitives and related categories*, 477–523. Berlin: Mouton de Gruyter.

Comrie, Bernard (2013). Alignment of case marking of full noun phrases. In: Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Available online at <http://wals.info/chapter/98>, accessed on 2017-03-18.

---

[13] There are no signs of multicollinearity between word order and aspect in this model, judging by the variance inflation factors (Fox & Monette 1992).

Cysouw, Michael (2010). Dealing with diversity: Towards an explanation of NP-internal word order frequencies. *Linguistic Typology* 14(2–3). 253–286.

Dahl, Östen & Viveka Velupillai (2013). Perfective/imperfective aspect. Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Available online at <http://wals.info/chapter/65>, accessed on 2017-03-22.

Dryer, Matthew S. (1989a). Large linguistic areas and language sampling. *Studies in Language* 13(2): 257–292.

Dryer, Matthew S. (1989b). Article-noun order. *Chicago Linguistic Society* 25. 83–97.

Dryer, Matthew S. (2013a). Definite articles. In: Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online.* Leipzig: Max Planck Institute for Evolutionary Anthropology. Available online at <http://wals.info/chapter/ 37>, accessed on 2016-12-08.

Dryer, Matthew S. (2013b). Order of object and verb. In: Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Available online at <http://wals.info/chapter/83>, accessed on 2016-12-08.

Dryer, Matthew S. (2013c). Order of adjective and noun. In: Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Available online at <http://wals.info/chapter/87>, accessed on 2016-12-08.

Dryer, Matthew S. (2013d). Indefinite articles. In: Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Available online at <http://wals.info/chapter/ 38>, accessed on 2016-12-08.

Fischer, Susann (2010). *Word-order change as a source of grammaticalisation*. Amsterdam, Philadelphia: John Benjamins.

Fox, John & Georges Monette (1992). Generalized collinearity diagnostics. *Journal of the American Statistical Association* 87. 178–183.

Fox, John, Sanford Weisberg, Michael Friendly & Jangman Hong (2016). effects: Effect displays for linear, generalized linear, and other models. *R* package, version 3.1-2. <https://CRAN. R-project.org>.

Gries, Stefan Th. (2013). *Statistics for linguistics with R: A practical introduction.* 2nd ed. Berlin, Boston: de Gruyter Mouton.

Harrell, Frank E. Jr. (2017). rms: Regression modeling strategies. *R* package, version 5.1-1. <https://CRAN. R-project.org>.

Hawkins, John A. (2004). *Efficiency and complexity in grammars*. Oxford: Oxford University Press.

Hawkins, John A. (2014). *Cross-linguistic variation and efficiency*. Oxford: Oxford University Press.

Hewson, John & Vit Bubenik (2006). *From case to adposition: The development of configurational syntax in Indo-European languages*. Amsterdam, Philadelphia: John Benjamins.

Jaeger, T. Florian, Peter Graff, William Croft & Daniel Pontillo (2011). Mixed effect models for genetic and areal dependencies in linguistic typology. *Linguistic Typology* 15(2). 281–320.

Leiss, Elisabeth (2000). *Artikel und Aspekt: Die grammatischen Muster von Definitheit*. Berlin: Mouton de Gruyter.

Leiss, Elisabeth (2007). Covert patterns of definiteness/indefiniteness and aspectuality in Old Icelandic, Gothic, and Old High German. In Elisabeth Stark, Elisabeth Leiss & Werner Abraham (eds.), *Nominal determination: Typology, context constraints and historical emergence*, 73–102. Amsterdam, Philadelphia: John Benjamins.

McColl Millar, Robert (2000). Some suggestions for explaining the origin and development of the definite article in English. In Olga Fischer, Anette Rosenbach & Dieter Stein (eds.), *Pathways of change: Grammaticalization in English*, 275–310. Amsterdam, Philadelphia: John Benjamins.

Nakagawa, Shinichi & Holger Schielzeth (2013). A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution* 4(2). 133–142.

R Development Core Team (2016). *R: A language and environment for statistical computing*. Version 3.3.1. Vienna: R Foundation for Statistical Computing. <http://www.r-project.org>.

Vennemann, Theo (1975). An explanation of drift. In Charles N. Li (ed.), *Word order and word order change*, 269–305. Austin: University of Texas Press.

Zakharko, Taras & Balthasar Bickel (2011ff). familybias: Family bias estimation. *R* package. <https://github.com/IVS-UZH>, accessed on 2016-11-17.