

The positioning of concessive adverbial clauses in English¹: Assessing the importance of discourse-pragmatic and processing-based constraints

Daniel Wiechmann and Elma Kerz
RWTH Aachen University

(Received 30 January 2012; revised 20 April 2012)

Abstract

English permits adverbial subordinate clauses to be placed either before or after their associated main clause. Previous research has shown that the positioning is conditioned by various factors from the domains of semantics, discourse pragmatics and language processing. With the exception Diessel (2008), these factors have never been investigated in concert, which makes it difficult to understand their relative importance. Diessel's study, however, discusses only temporal constructions and identifies iconicity of sequence as the strongest predictor of clause position. Since this explanation is, in principle, unavailable for other types of subordinate clauses, the generalizability of Diessel's findings is somewhat limited. The present study offers a multifactorial analysis of 2,000 concessive constructions from the written part of the BNC and assesses the variable importance of six factors for the ordering choice, showing that semantic and discourse-pragmatic factors are much stronger predictors of clause position than processing-based, weight-related ones. On a methodological note, the study proposes that random forests using conditional inference trees constitute the preferred tool for the general type of problem investigated here.

1 INTRODUCTION

When language users of English produce a complex sentence with an adverbial subordinate clause, they have a choice as to where to place the dependent clause relative to the main clause. For all present intents and purposes, they can put it either before or after the main clause.² The examples in (1) to (4) illustrate this variation using pairs of constructions that incorporate adverbial clauses of different semantic types:

- (1) Conditional clause
 - (a) *If it doesn't rain*, we'll go to the party.
 - (b) We'll go to the party *if it doesn't rain*.
- (2) Causal clause
 - (a) *Because he didn't sleep much*, Peter is tired.
 - (b) Peter is tired, *because he didn't sleep much*.
- (3) Temporal clause
 - (a) *When we arrived in Berlin*, it was dark.

¹ We thank two anonymous reviewers and the editor Bas Aarts for their helpful comments on earlier drafts. All remaining errors are of course our own.

² There actually is another third option, namely to place the adverbial clause in medial position, as in: *Bankruptcy regulations, although an essential plank of reform policy, had been repeatedly rejected by the Supreme Soviet, largely because of the fear of mass unemployment.* (BNC HLL) (cf. Quirk *et al.* 1985: 1037). This option is comparatively infrequent though and is not investigated in this study.

- (b) It was dark, *when we arrived in Berlin*. (all from Diessel 2005: 450)
- (4) Concessive clause
- (a) *Although he cannot afford it*, he bought her a lovely gift.
- (b) He bought her a lovely gift, *although he cannot afford it*.

Numerous studies have shown that the preferred positioning of a subordinate clause depends on its semantics. That is to say, the different semantic sub-types of adverbial clauses shown in (1) – (4) apparently prefer different positions within a complex sentence construction (Quirk *et al.* 1985; Biber *et al.* 1999; Diessel 2005). Based on an analysis of spoken and written corpus data, Diessel (2005) shows that conditional clauses are mainly sentence-initial, causal clauses are mainly sentence-final, and temporal clauses are fairly evenly divided between the two.³ Concessive adverbial clauses are found to predominantly follow the associated main clause (see, for instance, Biber *et al.* 1999: 834). In the attempt to explain this variation, prior research has offered various accounts from different areas of linguistic inquiry including discourse-pragmatics and language processing (cf. Chafe 1984, Givon 1985, Thompson and Longacre 1985, Hawkins 1994, Verstraete 2004, *inter alia*). While early studies have offered monocausal explanations for the phenomenon, more recent studies have highlighted the multifaceted nature of the phenomenon and have put forth estimations of the relative importance of various, potentially competing factors (Diessel 2005, 2008). The most comprehensive study in this respect that accommodates processing-based, semantic and discourse-functional perspectives is Diessel’s (2008) work on the positioning of temporal clauses (as in (3)). Drawing on corpus data from spoken and written English, Diessel uses logistic regression to investigate several factors, namely length, complexity, pragmatic import and the principle of iconicity, and concludes that the latter is “the strongest predictor for the initial occurrence of a temporal adverbial clause” (Diessel 2008:482). The generalizability of this finding to other semantic types of complex adverbial constructions seems limited though, as the strongest predictor, iconicity of sequence, does not seem to be straightforwardly applicable to all other semantic domains.⁴ Diessel himself recognized this point as he wrote “[...] it seems that the principle of iconicity is not immediately relevant for the positioning of causal and conditional clauses” (Diessel 2008:470). In order to broaden our understanding of clause ordering, we focus on a semantic type of adverbial clause that represents a construction type for which an explanation in terms of iconicity does not recommend itself, namely concessive clauses (as in (4)). Quirk and colleagues characterize concessive clauses as “indicat[ing] that the situation in the matrix clause is contrary to expectation in the light of what is said in the concessive clause.” (Quirk *et al.* 1985: 1098). Such constructions are commonly used in argumentative text types and “allow us to concede, or admit, that an opposing point has merit and, at the same time, to de-emphasize the opposing point’s importance to the reader” (cf. Altman *et al.* 2010: 71). That is to say that they are used to (i) highlight information which supports the position

³ The associations between certain semantic categories of adverbial clauses and their preferred positions in the sentential structure are not only characteristic of the English language but can also be observed cross-linguistically. Drawing on data from a sample of 40 languages, Diessel (2001) reports that temporal clauses are often found to appear both before and after the main clause, while causal clauses typically occupy the final slot and conditional clauses often precede the main clause.

⁴ The principle of iconicity (of sequence) predicts that the relative ordering of main and adverbial subordinate clauses should reflect the sequential ordering of the event they jointly encode.

of a speaker/writer on the topic at hand and, simultaneously, (ii) to extenuate the importance of conflicting information which may not support his/her position (ibid.). Given the above-mentioned tendency for concessive adverbial clauses to follow the main clause, the question arises as to what makes the language user deviate from this “default choice” by placing a concessive adverbial clause in the initial slot?

The present study sets out to contribute to answering this question on the basis of a multifactorial analysis of corpus data from the written part of the British National Corpus. More specifically, the study makes complementary use of two types of statistical classifiers, random forests of conditional inference trees and binomial logistic regression, so as to assess to what degree the ordering is conditioned on a set of factors proposed in the literature. These factors include (1) the proportional LENGTH, (2) the degree of DERANKING of the subordinate clause, (3) the COMPLEXITY of the construction, (4) the presence of a BRIDGING context, and (5) the choice of SUBORDINATOR. An important, albeit secondary goal of this paper is methodological in nature, namely to promote the primary technique employed here, the random forest method utilizing conditional inference trees. This method provides unbiased variable selection in the individual classification trees allowing us to reliably assess the relative importance of variables that are measured on different scales or that differ in terms of the number of distinguished factor levels, a scenario where traditional tree-based models have trouble and the coefficients of logistic regression models are harder to interpret. For these reasons (as well as others to be discussed in section 3), the method is considered an ideal tool to solve the type of problem at hand, namely measuring the importance of a set of investigated variables.

The remainder of this section will present a discussion of prior research and explicate the factors investigated in this study. Section 2 will present the data underlying the study. Section 3 will explicate the employed methodology. Section 4 will discuss the results before we finally conclude the study in section 5.

1.1 *Processing-related and discourse-pragmatic determinants of linear ordering*

Processing-based accounts typically aim to explain the positioning of an adverbial clause in terms of variables such as the relative length of the clausal elements and their complexity, whereas discourse-functional accounts have focused on variables such as the relative availability of the respective information and the discourse-organizing role of the adverbial clause. We will discuss the relevant factors momentarily but would like to point out at this point that they cannot always be fully disentangled. For example, one important factor to influence linear ordering is grammatical weight, i.e. the relative LENGTH of the constituents. It is a long established empirical fact that in languages like English longer constituents tend to follow shorter ones (cf. Behaghel 1932; Quirk *et al.* 1985). One way of accounting for this preference is to say that this short-before-long ordering is more efficient with respect to the online processing of the whole structure (cf. Hawkins 1994, 2004 and Gibson 1998, 2000). According to Hawkins’ performance based theory of constituent ordering (1994, 2004), heavy constituents are expected to appear sentence-finally, as this ordering is cognitively more efficient in head-initial languages making both production and parsing easier. Similarly, Gibson’s *Dependency Locality Theory* (Gibson 1998, 2000) states that the processing complexity of a linguistic configuration is contingent on the length of its syntactic dependencies.

Sentence-initial subordinate clauses introduce longer dependencies and are thus more difficult to parse. Diessel (2005) draws on Hawkins' theory to account for the greater length of final versus initial adverbial clauses. Temperley (2007) provides empirical support for Gibson's *Dependency Locality Theory*. These accounts share the assumption that length differentials play a causal role in the explanation. Alternatively, however, one might offer a pragmatic, information structural explanation for the tendency of "lighter" constituents to precede "heavier" ones with recourse to the "given-new" principle (cf. Arnold *et al.* 2000): new information necessitates more linguistic material to be encoded relative to given information. Not unrelatedly, discourse-pragmatic accounts have also alluded to the principle of end-focus, which describes the increase of informativeness towards the end of each grammatical unit, for both clauses and multi-clause expressions like the ones investigated here. These strategies give the ongoing discourse a forward impetus – what Firbas (1971) described as "communicative dynamism" (Firbas 1971: 136). From this viewpoint, the role of grammatical weight is more indirect, in fact epiphenomenal. Whatever its role in an explanatorily adequate account, it appears that length is an important predictor for adverbial clause positioning even when the abovementioned semantic variable, type of adverbial clause, is controlled for: Temperley (2007) finds that postmodifying clauses are significantly longer than premodifying clauses and concludes that "[t]he greater length of postmodifying clauses does not appear to be an artifact of length differences between different semantic types of adverbial clauses" (Temperley 2007: 312).

A second variable to influence linear ordering is the COMPLEXITY of the construction. Adverbial clauses may display various degrees of complexity and, given the processing-related considerations discussed earlier, one might expect to find sentence-initial adverbial clauses to be structurally less complex. Of course, what exactly it means for a structure to count as complex is a non-trivial issue in and of itself (cf., e.g., Szmrecsányi 2004 and references therein). We have taken what we consider to be a conservative route and, following Diessel (2008), counted as complex only those adverbial clauses that contained another subordinate clause of any type.

Following Cristofaro (2003), the third variable incorporated into this study is referred to as DERANKING and distinguishes 'balanced' and 'deranked' adverbial clauses (cf. also Diessel and Hetterle 2011). An adverbial clause in English is 'balanced' if it is tensed, and it is considered 'deranked' if it is not tensed but reduced in some way, i.e. if it either comprises a non-finite verb form or is expressed as a verbless construction. Consider the example in (5). The finite (=balanced) adverbial clause in the sentence

- (5) (a) *Although it was specific to attacking the GSM mobile phone network, DePetrillo and Bailey's talk showed how IMSIs could help to reveal the identity of the owner and other information. (PDF at find.pcworld.com/70348)*

can be 'deranked' or reduced to a non-finite clause:

- (b) *Although being specific to attacking [...]*

or even further compressed to:

- (c) *Although specific to attacking [...]*

Of course, both COMPLEXITY and DERANKING correlate with LENGTH: a more complex clause is likely to be longer than a less complex one, and a balanced subordinate clause is likely to be longer than a deranked one. However, these three factors should not be reduced to just a single factor, because (a) their intercorrelations are not perfect (a deranked adverbial clause is shorter than its balanced counterpart but a short adverbial clause is not necessarily deranked) and (b) some variables may be more directly associated with one type of explanation than another one: for example, proportional length is often used to operationalize differences in cognitive efficiency (cf. Hawkins 2004), so above all we might consider length a processing-related variable. Deranking, by contrast, may also be conceived of as expressing a discourse-pragmatic variable (with less pronounced implications for processing-related explanations): the reason for using a reduced form may very well be the communicative need to iconically reflect the degree of clause integration (as non-finite constructions signal a higher degree of clause integration (Givón 1995b: 58). Similarly, a deranked version might have been chosen to compress information (cf. Altenberg 1984; Greenbaum 1988; Granger 1997; Biber and Gray 2010). Compared to their finite counterparts, non-finite or verbless adverbial clauses allow denser information packaging, i.e. they represent greater syntactic integration and more informational compactness. In short, the three correlating factors figure differently in different conceivable types of explanation.

It has also been suggested that the ordering of adverbial clauses is co-determined by the need to support information structuring in discourse. Two discourse pragmatic functions are found to motivate language users to place adverbial clauses in initial positions; the “bridging” function and the “setting the stage” function. As pointed out by Verstraete (2004: 819), initial adverbial clauses are seen as distinct from their counterparts in final position in terms of their “discourse-organizing rather than local functions, such as linking back to the preceding discourse or introducing new frames for upcoming discourse” (see also Thompson 1985; Ramsay 1987; Givón 1990; Ford 1993). In the current paper, we concentrate on one discourse-pragmatic factor only, that of BRIDGING, which describes a scenario where an initial adverbial clause serves as a bridge between the preceding and the following discourse, as in (6):

- (6) In order to predict multisolute solution behaviour, all of the previous solution theories describing non-ideal solutions require empirical parameters obtained by fitting the multisolute data of the solution of interest.

Although *these solution theories* are accurate for the particular subset of solutions for which the empirical parameters can be determined, they cannot be applied to solutions for which there are no multisolute solution data.

Note that bridging is also intimately tied to the notion of (derived) ‘accessibility’, which also has been shown to influence constituent ordering (cf. Arnold *et al.* 2000): language users tend to produce given, accessible information earlier in the linguistic construction than new, inaccessible information.

Finally, the choice of SUBORDINATOR, i.e. the specific lexical item encoding the concessive meaning, is considered as a further predictor to affect the relative positioning

of the adverbial clause. The reason for its inclusion is the idea that there are subtle meaning differences between concessive clauses headed by different subordinators (combined with the more general observation discussed above that different semantic types of adverbial clauses prefer different positions). In their chapter on syntactic and semantic functions of subordinate clauses, Quirk et al. (1985: 1098f.) discuss the list of subordinators that can serve as heads of concessive clauses in English (*although*, *though*, *while*, *whilst*, *whereas*, and the more emphatic forms *even though*, *even when* and *even if*). They suggest that different subordinators can be distinguished on the basis of their degree of versatility/restrictiveness of application (see also Aarts 1988 for a discussion of usage-conditions of concessive subordinators). Specifically, they propose that “[a]lthough and the more informal *though* are the most versatile of the subordinators, since they may in fact relate clauses in which the situations are similar” and that *whereas* is “the most restricted of the four subordinators, requiring antithesis between two situations.” This characterization is echoed in Cowan (2008: 555), who states that the concessive subordinators *while* and *whereas* differ from *although* and (*even*) *though* as “they do not suggest that the main clause might be expected to be false; they simply express a contrast”. The example in (7) demonstrates that *whereas* cannot always be used in place of *although*.

- (7) (a) Although she grew up in Tehran, she doesn't speak Farse.
 (b) *Whereas she grew up in Tehran, she doesn't speak Farse.

Based on these characterizations and observations, we investigate differences in preferred positioning between adverbial clauses headed by the semantically most dissimilar subordinators, viz. *although* and *whereas*. With these motivations for the inclusion of five determinants in place, we may now turn to a description of the data used in this study.

2 DATA & ANNOTATION

We randomly extracted from the written part of the British National Corpus (BNC, Burnard 1995) 1,000 subordinate clauses headed by *although* plus 1000 clauses headed by *whereas* for a total 2,000 data points, which were manually annotated with information pertaining to the variables outlined above. The data points are extracted from the written part of the BNC only as concessive clauses occur predominantly in written registers. Biber and colleagues note that “[...] concessive clauses are notably more common in the written registers (especially news and academic prose) [...]” (Biber *et al.* 1999: 821). The relevant information was captured as follows:

- Dependent variable:
 - ADVCL.FINAL (position): measured as a binary factor; levels: final vs. non-final (i.e. initial and middle)
- Predictor variables:

1. LENGTH: measured as a continuous variable; values expressed as the proportion of the whole construction that belongs to the adverbial clause⁵
2. COMPLEXITY: measured as a binary factor; levels: simple vs. complex
3. DERANKING: measured as a binary factor; levels: balanced vs. deranked
4. BRIDGING: measured as a binary factor: levels: anaphoric vs. non-anaphoric (=presence vs. absence of anaphoric items within the adverbial clause)
5. SUBORDINATOR: measured as a binary factor; levels: *although* vs. *whereas*

Table 1 presents some descriptive statistics of the data.

Table 1: *Descriptive statistics of data set*

ADV.CL. - FINAL	DERANKING	COHERENCE	COMPLEXITY	LENGTH	SUB- ORDINATOR
Final: 1,127	Balanced: 1,883	Anaphoric: 331	Complex : 334	Min.: 0.05	<i>although</i> : 1,000
Non- final: 873	Deranked: 117	Not anaphoric: 1,669	Simple: 1,666	1st Qu.: 0.32	<i>whereas</i> : 1,000
				Mean: 0.44	
				3rd Qu.: 0.55	
				Max.: 0.90	

We observe that a considerable proportion of the concessive adverbial clauses (~ 44%) are in non-final positions. The majority of adverbial clauses (AdvCl) are balanced, contain no anaphoric device that would indicate a bridging context, and are simple, i.e. they do not contain any additional subordinate clause. Their proportional size relative to the whole pattern can be a little as 0.05 and as much as 0.90 and averaging around a value of 0.44. Recall that the exact same amount of ADVCL headed by *although* and *whereas* respectively was not coincidental but was due to our sampling. Figure 1 presents the distribution of these data across positions, i.e. sentence-initial versus sentence-final.

⁵ For example, if the concessive adverbial clause consists of 6 words and the complex sentence of 13 words, the relative length of the adverbial clause is $6/13 = 0.46$.

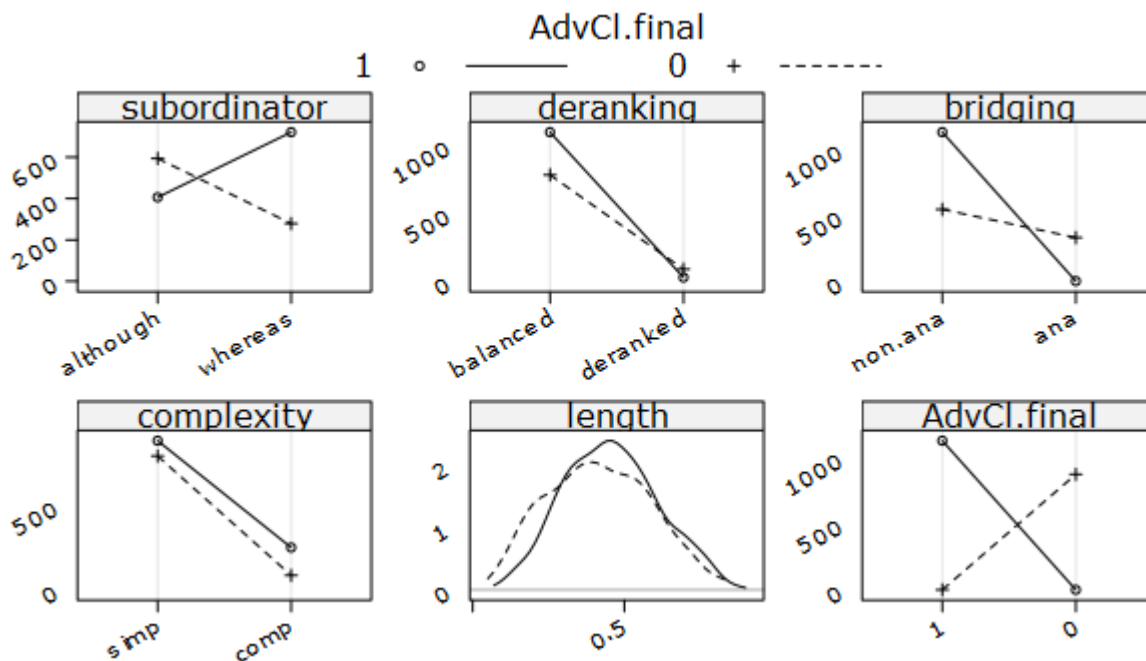


Figure 1: *Distributions across clause positions*

Figure 1 already gives us some indication of notable distributional differences across the two clause positions: the more similar is the shape of the two lines in a given graph, the more similar is the distribution of that variable across the two positions. We can immediately see that the variable COMPLEXITY does not seem to discriminate between sentence final and sentence-initial adverbial clause, whereas SUBORDINATOR appears to do so. The next section is dedicated to providing statistical analyses of these data.

3 METHOD & RESULTS

In the attempt to measure the relative importance of the variables investigated here, we have made use of two statistical classifiers, i.e. two statistical procedures that seek to predict the placement of the adverbial clause, given a set of predictors. These techniques are random forests of conditional inference trees and binomial logistic regression.⁶ We assume some familiarity with the general workings of the latter technique (for an introduction cf. e.g., Baayen 2008), but shall provide a short introduction into the rationale underlying the latter, since the random forest method has not been applied very often in linguistic contexts (but cf., e.g., Wiechmann 2011, Tagliamonte and Baayen (to appear)). This short introduction will, however, focus on some of the most central conceptual ideas regarding the technique (for a general introduction to classification and regression trees cf. Breiman *et al.* 1984; for random forests cf. Breiman 2001; for an introduction into conditional inference framework cf. Hothorn and Zeileis 2006; for an implementation that employs conditional inference trees cf. Strobl *et al.* 2008, 2009). The principal reason why the method of *random forests of conditional inference trees* was employed here is the fact that the method recommends itself whenever we want to assess the importance of correlating variables as its estimation is less prone to error than

⁶The machine learning and pattern recognition literature presents many alternative approaches to the present task (cf., e.g. Baldi and Brunak 2002 or Bishop 2006 for overviews).

that of more commonly employed logistic regression techniques in such situations. When predictors correlate, the effect of a variable can be overshadowed by more significant competitors in simple or mixed effect regression models (Strobl *et al.* 2009: 337).⁷ Nevertheless, it can be helpful to build additional models from different frameworks to attain a more comprehensive understanding of the data. For this reason, we have also fitted a logistic regression model to the data. Not only can building additional models serve as a sanity check, i.e. as a test to quickly evaluate whether a claim or the result of a calculation can possibly be true, it can also offer complementary perspectives on the data. Finally, it also serves a third purpose of providing the reader with a description of the data that they are likely to be more familiar with.

3.1 Using (a forest of) classification trees

We will first have a look at the logic underlying conditional inference trees. We shall then motivate why we would want to use ensembles of such trees (a forest) and finally present a measure of variable importance based on this model.⁸

Like many other types of statistical models used in contexts of data mining (e.g. logistic regression models), the purpose of classification trees in general is to predict a typically binary outcome on the basis of a set of predictor variables. Quite generally, such algorithms work through the data and determine a set of *if-then* logical (split) conditions that permit accurate classification of cases. The particular procedure employed here, so called *Conditional Inference Trees*, involves a recursive partitioning algorithm that works as follows:

1. Test the global null hypothesis of independence between any of the input variables and the response (here: ADVCL.FINAL). Stop if this hypothesis cannot be rejected. Otherwise select the input variable with the strongest association to the response. This association is measured by a p-value corresponding to a test for the partial null hypothesis of a single input variable and the response.⁹
2. Implement a binary split in the selected input variable.
3. Recursively repeat steps (1) and (2).

In other words, the algorithm will split the dataset first on the basis of the most important variable and will then iteratively try to split each resulting sub-set of the data until it can no longer detect statistically significant associations between any of the tested variables and the response. Figure 2 presents a graphical representation of the output of the algorithm when applied to the present data.

⁷ Decision tree algorithms in general handle well mixed types of variables and missing values. They are robust to outliers and monotonic transformations of the input, and robust to irrelevant inputs. The advantages of tree-based models include for example that they are applicable in “small n large p”-situation, i.e. in scenarios where we are dealing with a large number of predictors and a small number of data points. It avoids problems of collinearity that can destabilize the model (cf. Tagliamonte and Baayen (to appear) for a discussion).

⁸ We have used R (R Development Core Team, 2011) for all computations and statistical analyses.

⁹ Continuous variables are usually discretized by defining new discrete-valued attributes that partition the continuous attribute value into a discrete set of intervals. A given partition is then split at a cut point that exhibits the most pronounced differences between the resulting ranges (cf. Breiman *et al.* 1984).

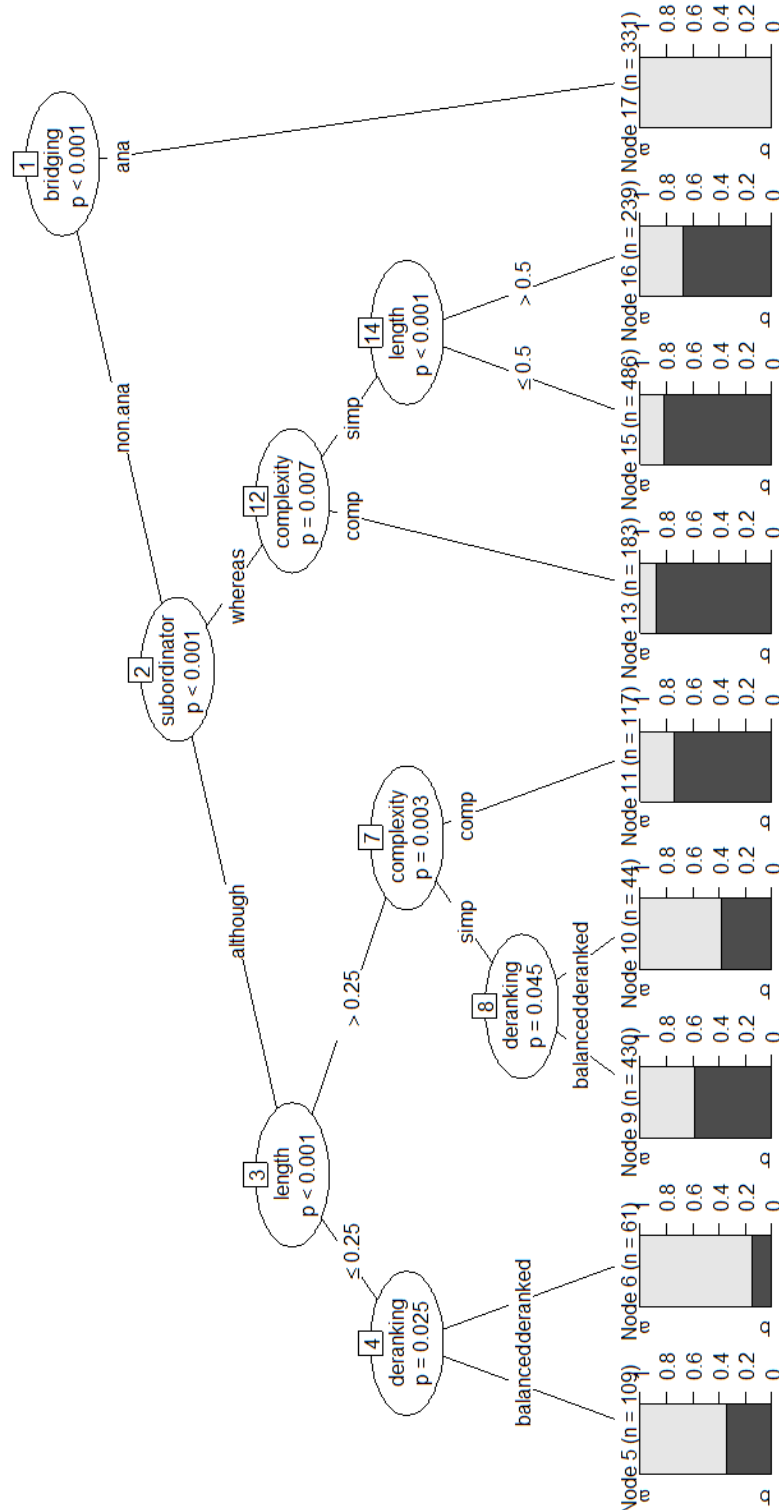


Figure 2: *Conditional inference tree*

The first thing to note is that all five predictors figure in the tree meaning that all predictors are relevant for the classification task at hand. The boxes at the bottom show the proportions of initial and final adverbial clauses of a given subset, which for expository reasons are labeled as “a” and “b” respectively. The best way to interpret the tree is to investigate it from top to bottom: we observe that the variable BRIDGING shows

the strongest association to the response, ADVCL.FINAL, which tells us that the positioning of the adverbial clause can be predicted best when we know whether or not it contains an anaphoric item. If it does, i.e. if BRIDGING=ana(phoric), the adverbial clause invariably assumes the non-final position (the proportion of final positions is 0 in this case). As indicated by Node 17, this is true for 331 cases. The tree fragment branching to the left hand side then represents the subset of cases where there is no such anaphoric item. As indicated by Node 2, the variable with the strongest association to the response now is the predictor SUBORDINATOR, so the next split distinguishes all clauses introduced by *whereas* from those introduced by *although*. Notice that the resulting subsets are further split on the basis of two different variables, namely COMPLEXITY (leading to Node 3) and LENGTH (leading to Node 8). This suggests that there is an interaction between SUBORDINATOR on the one hand and LENGTH and/or COMPLEXITY on the other. We will come back to these interactions in section 3.2 when we apply the logistic regression model. The full interpretation of the tree follows the same logic.

An important thing to note at this point is that a single tree can produce problematic results. Most importantly, it can produce results that may adequately describe structural properties of the data at hand but small changes in the data may affect the decision as to whether or not a variable is included into the tree. Hence we are well advised to be careful with our interpretation at this point. One way to remedy this situation is to build not just one tree (based on the complete data set) but many trees (based on random subsets of the data) and focus on the patterns they “agree on”. In other words, to achieve more robust results and greater generalizability, we may grow a whole “forest of trees”, rather than just a single one (hence the name of the procedure). Specifically, for the present study, we grew a total set of 500 trees using a bootstrapping technique, in which 500 different random subsamples were taken from the original data.¹⁰ On the basis of the resulting model, we can now answer our principal question, namely given that all these variables influence the ordering choice, what are their relative strengths? We can obtain an estimation of the relative importance of the five variables using a *permutation variable importance measure*. The general logic of such measures is as follows. First, the original statistical association of a predictor and the response is broken by permuting the original values of the predictor. Having broken the association between predictor and response, it is then assessed how much the overall classification accuracy of the model drops: the greater the loss in accuracy, the more important the predictor. This measure is preferable over possible alternatives (e.g. Gini importance, Breiman 2001) because it is not biased when predictor variables vary in their number of categories or scale of measurement (cf. Strobl *et al.* 2008 for details).¹¹ Figure 3 presents the results.

¹⁰ The ensemble of conditional inference trees (random forest) was grown using the function *cforest* in the *party* library. The random forest tuning parameter “mtry”, which determines the number of input variables randomly sampled as candidates at each node, was set to the number of input variables to enable bootstrap aggregating (bagging), which reduces variance and helps to avoid overfitting. A forest of 500 trees is in fact a lot larger than what would have been required as the error rate approaches its minimum after about 250 trees.

¹¹ The conditional version of the “varimp” function from the R package “party”, which was used here, adjusts for correlations between predictor variables.

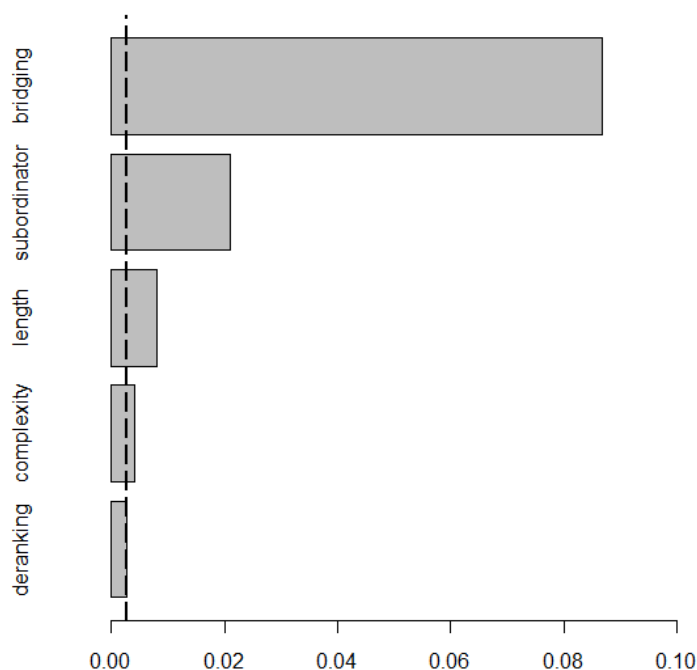


Figure 3: Variable importance of all predictors estimated from random forest model. The dashed line marks the value of the weakest predictor, DERANKING, against which the importance of the remaining predictors can be judged.

Variable importance should be interpreted and reported as a relative ranking of significant predictors. We observe that BRIDGING is judged to be by far the most important predictor followed by SUBORDINATOR and LENGTH. COMPLEXITY and DERANKING apparently play only comparatively minor roles for the ordering of concessive adverbial clauses.

3.2 A complementary view: Logistic regression

The logistic regression model was built by way of stepwise model simplification. That is, we started by fitting a maximal model that included all predictors and all their interactions and then, in a stepwise fashion, deleted all terms whose inclusion did not yield a statistically significant improvement over a model without that term. The resulting minimal adequate model comprised all main effects and one two-way interaction mentioned earlier between SUBORDINATOR and LENGTH.¹² The model shows a good performance in predicting clause position (Nagelkerke's $R^2 = 0.47$, index of concordance $C = 0.82$) and overall performed at level comparable to that of the tree-based models (single tree and random forest models).¹³ Figure 3 presents the *Receiver Operating Characteristic* (ROC) curve where the true positive rate is plotted against the false positive rate (rate of Type I error). In such graphs, the better model is the one with a larger area under the curve.

¹² The suggested interaction between SUBORDINATOR and COMPLEXITY was not judged to be statistically significant ($p > 0.15$).

¹³ The index of concordance C is a generalization of the area under the Receiver Operating Characteristic (ROC) curve. Values > 0.8 indicate a good performance (cf., e.g., Baayen 2008 for details).

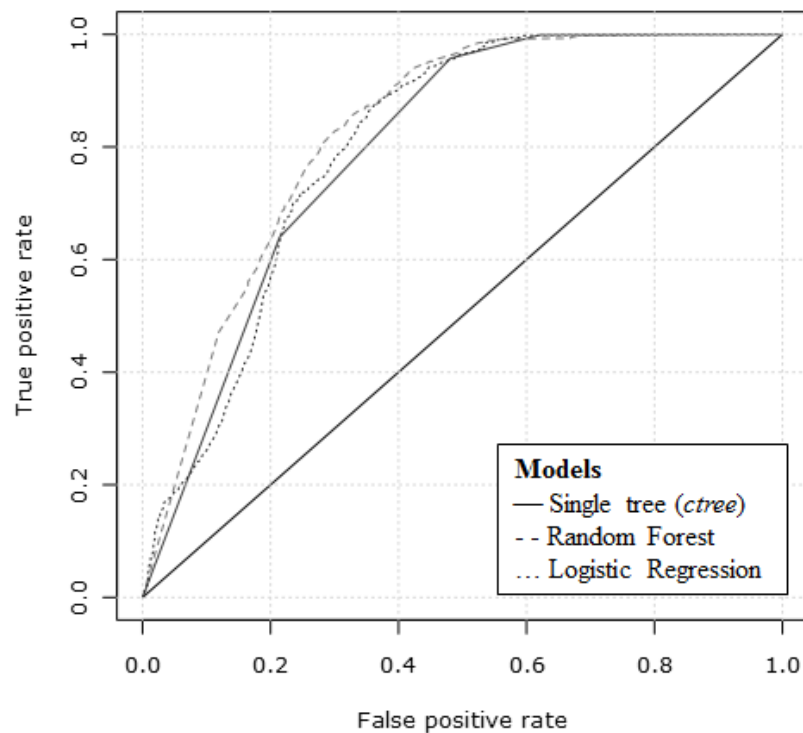


Figure 4: Comparison of random forest and logistic regression model via ROC curves (the better model is the one with a larger area under the curve). The random forest model perform best; the single tree model perform slightly better than the logistic regression model ($C_{\text{Random Forest}}: 0.87 > C_{\text{Single Tree}}: 0.83 > C_{\text{Logistic Regression}}: 0.82$).

Without going into the details of classification accuracy, we can see that (a) the random forest model performs best and (b) that the single tree model performs slightly better than the logistic regression model. These results are corroborated by a comparison of the error matrices of the three models (Table 2), which shows that the random forest method exhibits the lowest the error rate.

Table 2: Error matrices for all models

Random Forest model – overall error: 0.22

		Predicted	
		initial	final
Actual	initial	480	393
	final	54	1073

Single tree model – overall error: 0.23

		Predicted	
		initial	final
Actual	initial	453	420
	final	48	1079

Logistic regression model – overall error: 0.26

		Predicted	
		initial	final
Actual	initial	411	462
	final	29	1098

We may now turn to the estimated coefficients of the logistic regression model shown in Table 3.

Table 3: *Estimated regression coefficients*

Coefficients:	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-6.52	1.05	-6.23	4.58E-10	***
SUBORDINATOR: <i>whereas</i>	2.99	0.36	8.41	< 2e-16	***
DERANKING: <i>deranked</i>	-0.98	0.24	-3.99	6.64E-05	***
BRIDGING: <i>non.ana</i>	6.55	1.00	6.52	6.88E-11	***
COMPLEXITY: <i>simple</i>	-0.79	0.17	-4.51	6.52E-06	***
LENGTH	2.26	0.50	4.49	7.22E-06	***
SUBORDINATOR- <i>whereas</i> :LENGTH	-4.40	0.74	-5.99	2.14E-09	***

Like the random forest, the logistic regression model also judges all predictors to be statistically significant for the task of predicting the position of the adverbial clause. We are also led to assume that BRIDGING is the most important predictor and that SUBORDINATOR and LENGTH are more important than COMPLEXITY and DERANKING. We might want to note that the relative importance of the latter two predictors is reversed here, which may be a result of the fact that correlating predictors negatively influence the validity of their corresponding regression coefficients. Random forests utilizing conditional inference trees are not so prone to this problem. However, given the relatively small effect size/variable importance of these factors, this need not bother us too much. Our results suggest that both variables will not assume very prominent roles in a full explanation of clause ordering. Figure 5 presents the logistic regression analysis in terms of odds ratios. It shows the antilogarithm of the regression coefficients instead of the coefficients themselves.

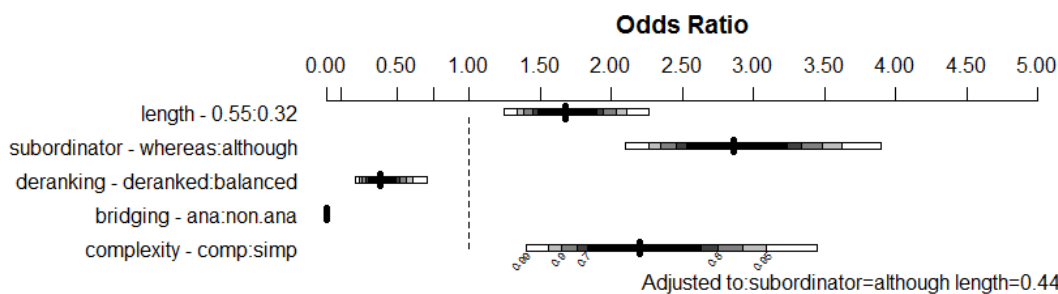


Figure 5: *Presentation of logistic regression analysis in terms of odds ratios*

All odds ratios greater than one, i.e. all values to the right of the dotted line in Figure 5, suggest that the indicated change of value of a given variable increases the odds for final positioning. The chances for final positioning are about 1.6 times greater when the proportional length of the adverbial clause is 0.55 when compared to a length of 0.32.¹⁴ Similarly, the chances for final positioning are about 2.7 times greater when the subordinator is *whereas* if compared to *although*. Jumping to COMPLEXITY, we observe that complex adverbial clauses are about 2 times more likely to occupy a final position than simple ones. Furthermore, we observe that a deranked adverbial clause is judged to be less likely to occur in final position than a balanced one (it is about 0.5 times more likely), and finally, the value approaching zero indicates that it is very highly unlikely to find a adverbial clause serving a bridging function to occur in sentence final position.

These results complement and confirm the output of the random forest method. A convenient property of regression models is that they also allow for an explicit graphical representation of the nature of the strong interaction effect between SUBORDINATOR and LENGTH, which we alluded to in section 3.1. Figure 6 presents an effect plot visualizing the observed interaction, in which, for ease of interpretation, the logit transformed coefficients are transformed back into changes in predicted probability.

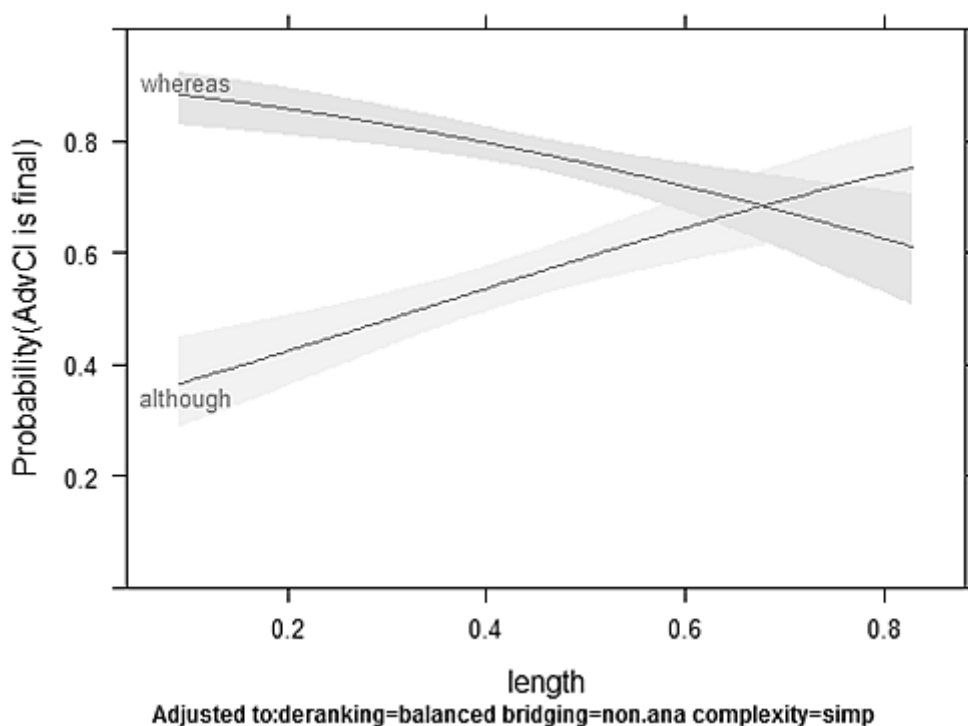


Figure 6: *Effect plot showing the interaction of LENGTH and SUBORDINATOR*
 The graph shows the effect of the variable LENGTH on the relative positioning of the adverbial clause across subordinators as estimated by a binomial logistic regression model (all remaining factors are adjusted to their most frequent levels). For ease of interpretation, the logit transformed coefficients are transformed back into changes in predicted

¹⁴ These values delimit the 1st and 3rd quantile for the variable LENGTH.

probability (the shaded areas surrounding the lines delimit a 95% confidence interval).

Figure 6 shows that the effect of LENGTH depends on the choice of SUBORDINATOR: we observe that the probability of an adverbial clause assuming a sentence final position increases with growing length when the subordinator is *although*, which is very much in line with the processing-related consideration discussed above. Interestingly though, this trend is reversed when the subordinator is *whereas*.

4. DISCUSSION

Our results show that the positioning of concessive adverbial clauses within complex sentence constructions in written English is affected by all five investigated variables (DERANKING, COMPLEXITY, LENGTH, BRIDGING, and SUBORDINATOR). We have found that the positioning is most strongly governed by semantic and discourse-pragmatic factors. The most important variable is the presence of an anaphoric device indicating a bridging context. This finding supports the idea that adverbial clauses are placed in sentence-initial positions when their function is to organize the information flow in the ongoing discourse, i.e. their use is motivated by cohesion and information structuring considerations (cf. Diessel 2005; Givón 1990; Verstraete 2004).

The type of concessive subordinator emerges as the second most important predictor, which reflects that the semantic difference between the two sub-groups of concessive constructions represented by *although* and *whereas* co-determines clause positioning.¹⁵ The *whereas* group has been described as more restricted in usage to the effect that it requires antithesis between the contents of the two clausal constituents. The preferred positioning of *whereas* clauses is sentence final: 721 out of 1000 cases exhibit that position. The contrastive function can also be served by fronted adverbial clauses though, which leads us to the observed interaction between LENGTH and SUBORDINATOR. We observed that the effect of length depends on the type of subordinator heading the adverbial clause. *Although*-clauses exhibit the expected length effect, where the likelihood of final positioning increases with growing proportional size of the adverbial clause. In contrast, *whereas*-clauses showed the opposite tendency meaning that there are many cases in our data where lengthy adverbial clauses are in sentence-initial position. We can explain this interaction and, in particular, the relative heavy grammatical weight of initial *whereas*-clauses in recourse to the contrastive function of such constructions: when the function of a complex sentential construction is to highlight a perceived contrast between two propositions, information structural considerations instruct us to establish in the first of these clauses what might be called the *shared conceptual array*, which is necessary for any kind of comparison. This array is meant to comprise all conceptual building blocks (entities, predications, etc.) that are employed in the contrasted propositions. The verbalization of this shared conceptual background typically is more elaborate in the first of the two clauses as the information

¹⁵ Rather than talking about semantic sub-groups, it appears also possible to interpret the results as presenting evidence for lexically specific knowledge (Dabrowska 2009). We consider the sub-group interpretation to be more conservative as it includes the possibility that a sub-group comprises just a single member but also permits the contingency of additional subordinator not investigated here to join one or the other group (*even though* and *though* are likely to fall into the same group as *although* for example).

structural default seems to be to explicate the shared background first and ‘pick it up’ again in the second clause (rather than the other way around).¹⁶ For instance, the first clause can be longer because it expresses shared discourse referents lexically (viz. relatively heavy constituents), while speakers can refer to them pronominally in the second clause. The verbalization of the first proposition can also be heavier as more lexical material is employed to evoke the necessary semantic frames, which need not be re-invoked in the second clause. Consider the example in (8) for illustration.

- (8) Whereas for Marxist-Leninists, the nationalization of property in the hands of a workers’ State ensured that through the institutions which represented them -- the State, the party, the soviets and the trade unions -- the working masses were now in command, for libertarians this represented little more than a “change of guard”. [BNC FB1]

In (8) the contents expressed by “the nationalization of property in the hands of a workers’ State” are referred to by an anaphoric *this*, which, of course, considerably reduces the amount of linguistic material needed to communicate the intended content. In short, the observed interaction can be explained by the usage conditions on *whereas*-clauses: they are employed to express contrasts between propositions and when they are preposed, they need to introduce all the information shared by the contrasted propositions, i.e. what we have termed here the shared conceptual array. For this reason, they are not in line with the predictions made by the ‘end weight’ principle, which asserts that the likelihood of final positioning should increase with growing proportional size of the adverbial clause.

The variable that is arguably most closely related to processing-based explanations is LENGTH, which occupies only the third rank suggesting that the positioning is first and foremost driven by discourse pragmatic motivations. This conclusion is strengthened by the observed interaction between LENGTH and SUBORDINATOR, which showed that adverbial clauses headed by *whereas* systematically run counter the length constraint. While the effects of the remaining factors that may be associated with processing-based explanations, COMPLEXITY and DERANKING, are statistically significant, it must be acknowledged that their respective effect sizes are comparatively small, which is interpreted as additional evidence for the idea that adverbial clause ordering is most strongly driven by discourse pragmatic considerations, rather than processing related ones.

At this point, a caveat is in order: as we have investigated only written language it seems reasonable to cast into question how far the present results can be generalized so as to also adequately describe the dynamics observable in spoken language. It certainly seems plausible to assume that processing-related variables are most relevant under real-time constraints. Hence, the fact that they do play a comparatively minor (yet statistically significant) role in the present data might just be a consequence of our choosing to focus on written language. While being intuitively plausible, this reasoning goes against several previous word order studies. For instance, in his study on PP ordering, Hawkins 2000 finds strong support from written corpus data for an efficiency-based account of constituent order. Investigating the same phenomenon on the basis of an exhaustive analysis of the ICE-GB corpus, Wiechmann and Lohmann (submitted)

¹⁶ There are exceptions to this tendency, for example in cataphoric constructions, but the anaphoric strategy is a lot more frequent.

report comparable effect sizes of (cognitive) efficiency in spoken and written language. These findings suggest that constituent ordering is not too dissimilar across modalities, provided that register and genre effects are controlled for. This is to be expected if we assume a causal connection between spoken and written language along the following lines: processing biases act on spoken language production and lead to skewed frequency distributions. Skewed distributions “fossilize” and become conventional orders (soft constraints). Conventional orders are adhered to in writing as well (as they are perceived as most natural and idiomatic). Therefore, written language should mirror spoken language. In light of these considerations, the more conservative hypothesis seems to be that a study investigating spoken language would find comparable results.¹⁷

Finally, we would like to come back to the issue alluded to in section 1.1, namely the difficulty to fully disentangle processing- and discourse related effects. To further illustrate the intricate relationship of the two, it is helpful to return to Diessel’s evaluation of the factors influencing the positioning of temporal clauses. Recall that Diessel found that the principle of iconicity (of sequence) was the strongest predictor of clause position. It does not seem to be entirely clear though at what level of description (and/or explanation) the principle of iconicity is situated. In other words: the boundaries between functional, cognitive and psycholinguistic types of explanation are not always clearly demarcated. Following Givón (1985), Diessel claims that “iconicity of sequence, which is commonly characterized as a semantic principle, can be interpreted as a processing principle that contributes to the overall processing load of a complex sentence construction because a non-iconic clause order is difficult to plan and interpret” (Diessel 2088: 484). This strikes us as plausible. The general point that we would like to make on the basis of this example is that, in some sense, many semantic and (discourse) functional principles can be related to and ultimately reduced to processing-related ones (cf. Givon 1995a: Chapter 8 for a discussion of coherence in text and mind).¹⁸ All talk of information structuring since Halliday (1967) and pragmatic inferencing since Grice (1975) can, in principle, be recast in terms of processing principles and biases, which is not problematic per se. It only becomes a problem when one wishes to allocate a certain factor F to either one domain of explanation or the other. We acknowledge this general problem but believe that it is still reasonable to distinguish variables that are traditionally treated as discourse organizational (such as BRIDGING) from those that are more direct expressions of processing biases (such as LENGTH), if only to be able to better situate the factors that drive the ordering.

5. CONCLUSION

Based on a sample of 2000 instance from the written part of the BNC, the present study presents a multifactorial analysis of the positioning of English concessive clauses and sets out to assess the variable importance of five variables that have been proposed to

¹⁷ It is also conceivable that the symmetry can be established from the other direction as well. Concessive constructions are much more frequently employed in written language. It seems plausible to assume that those language users that make use of concessive constructions altogether make use of their experience with written language to derive models for spoken discourse.

¹⁸ There may very well be functional explanations that are not cognitive in some sense. For instance, explanations that make reference to dynamics of groups like Keller’s invisible hand theory of language change (Keller 1994).

affect the relative ordering of main and subordinate clauses. Adhering to what we believe is a conservative classification of the investigated variables, we observed that the presence or absence of an anaphoric item and the type of SUBORDINATOR, the two strongest predictors in our study, are semantic or discourse organizational in nature. The likelihood of fronting is greatest when it includes an anaphoric that takes up a nominal concept of the preceding clause as in (9):

- (9) ... indicate that *the typical temperature differences driving convection in a magma chamber* are 0.1-10°C. Although *these differences* seem small ...

The analysis revealed that the variables more directly associated with processing, i.e. LENGTH, COMPLEXITY and arguably DERANKING, are statistically significant (long, complex and balanced concessive clauses tend to follow their associated main clauses, as (10) illustrates:

- (10) He lived in 1935 (when I last saw him) in the utmost simplicity, although if he had been a little more conciliatory he could always have earned enough for his comfort -- and his wife's.

However, when compared to the functional variables, these factors play only subsidiary roles. Our findings suggest that in order to attain greater explanatory adequacy theories that assign all causal roles to principles of cognitive efficiency (e.g. Hawkins 2004) should broaden their notions of what it means for a structure to be cognitively efficient. More specifically, our findings suggest that this broadening should involve the re-interpretation and incorporation of quantities that are typically situated at the level of discourse-pragmatics.¹⁹ Furthermore, on a methodological note we have shown that random forests of conditional inference trees make for a valuable tool to be employed instead of or in combination with more familiar regression models as they handled correlating variables better and performed slightly better than the regression model in the presented scenario here.

Authors' address:

Department of English, American and Romance Language Studies

RWTH Aachen University

Kármánstr. 17/19

D-52062 Aachen

Germany

wiechmann@anglistik.rwth-aachen.de

kerz@anglistik.rwth-aachen.de

References

Aarts, Bas. 1988. Clauses of concession in written present-day British English. *Journal of English Linguistics* 21, 39-58.

¹⁹ For example, our bridging can be re-phrased in terms of processing efficiency. We thank one anonymous reviewer for pointing out the fact that “[a]naphora resolution is harder in contexts where the antecedent is far away. Hence, the human processor would prefer texts in which an adverbial clause with an anaphoric pronoun is relatively closer to its antecedent in the previous discourse”.

- Altenberg, Bengt. 1984. Causal linking in spoken and written English. *Studia Linguistica* 38, 20-69.
- Altman, Pam, Mari Caro, Lisa Metge-Egan & Leslie Roberts. 2010. *Sentence-combining workbook*. 3rd edn. Belmont, CA: Wadsworth Publishing.
- Arnold, Jennifer E., Thomas Wasow, Anthony Losongco & Ryan Ginstrom. 2000. Heaviness vs. newness: the effects of structural complexity and discourse status on constituent ordering. *Language* 76(1), 28-55.
- Baayen, R. Harald. 2008. *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press.
- Baldi, Pierre & Soren Brunak. 2002. *Bioinformatics: A Machine Learning Approach*. Cambridge, MA: MIT Press
- Bishop, Christopher. 2006. *Pattern Recognition and Machine Learning*. Berlin: Springer-Verlag.
- Behaghel, Otto. 1932. *Deutsche Syntax. Eine geschichtliche Darstellung*. Vol. 4: *Wortstellung, Periodenbau*. Heidelberg: Winter.
- Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad & Edward Finegan. 1999. *Longman grammar of spoken and written English*. London: Longman.
- Biber, Douglas & Bethany Gray. 2010. Challenging stereotypes about academic writing: complexity, elaboration, explicitness. *Journal of English Academic Purposes* 9(1), 2-20.
- Breiman, Leo., Jerome Friedman, Charles J. Stone & R.A. Olshen. 1984. *Classification and Regression Trees*. Chapman and Hall, New York.
- Breiman, Leo. 2001. Random Forests. *Machine Learning*, 45(1), 5-32.
- Chafe, Wallace. 1984. How people use adverbial clauses. *Berkeley Linguistics Society* 10, 437-49.
- Cowan, Ron. 2008. *The teacher's grammar of English*. Cambridge: Cambridge University Press.
- Cristofaro, Sonia. 2003. *Subordination strategies: A typological study*. Oxford: Oxford University Press.
- Dąbrowska, Ewa. 2009. Words as constructions. In Vyvyan Evans & Stéphanie Pourcel (eds.), *New directions in Cognitive Linguistics*, 201-23. Amsterdam: John Benjamins.
- Diessel, Holger. 2001. The ordering distribution of main and adverbial clauses: A typological study. *Language* 77, 345-65.
- Diessel, Holger. 2005. Competing motivations for the ordering of main and adverbial clauses. *Linguistics* 43, 449-70.
- Diessel, Holger. 2008. Iconicity of sequence. A corpus-based analysis of the positioning of temporal adverbial clauses in English. *Cognitive Linguistics* 19, 457-82.
- Diessel, Holger & Katja Hetterle. (2011). Causal clauses: A cross-linguistic investigation of their structure, meaning, and use. In Peter Siemund (ed.), *Linguistic universals and language variation*, 21-52. Berlin: Mouton de Gruyter.
- Firbas, Jan. 1971. On the concept of communicative dynamism in the theory of functional sentence perspective. *Philologica Pragensia* 8, 135-144.
- Ford, Cecilia E. 1993. *Grammar in interaction: Adverbial clauses in American English conversations*. Cambridge: Cambridge University Press.
- Gibson, Edward. 1998. Linguistic complexity: locality of syntactic dependencies. *Cognition* 68, 1-76.

- Gibson, Edward. 2000. The dependency locality theory: a distance-based theory of linguistic complexity. In Alec Marantz, Yasushi Miyashita & Wayne O'Neil (eds.), *Image, language, brain*, 95–126. Cambridge, MA: MIT Press.
- Givón, Talmy. 1985. Iconicity, isomorphism and nonarbitrary coding in syntax. In John Haiman (ed.), *Natural Syntax*, 187–220. John Benjamins: Amsterdam.
- Givón, Talmy. 1990. *Syntax: A functional-typological introduction*, vol. 2. Amsterdam: John Benjamins.
- Givón, Talmy. 1995a. *Functionalism and Grammar*. John Benjamins: Amsterdam
- Givón, Talmy. 1995b. Isomorphism in the grammatical code: cognitive and biological considerations. In Raffaele Simone (ed.), *Iconicity in Language*. 47-77. Amsterdam: John Benjamins.
- Granger, Sylviane. 1997. On identifying the syntactic and discourse features of participle clauses in academic English: native and non-native writers compared. In Jan Arts, Herman Wekker & Inge de Monnik (eds.), *Studies in English language and teaching*, 185-98. Amsterdam & Atlanta: Rodopi.
- Grice, Herbert P. 1975. Logic and conversation. In Peter Cole & John Morgan. (eds.), *Syntax and semantics*, vol 3. 41-58. New York: Academic Press.
- Greenbaum, Sidney. 1988. Syntactic devices for compression in English. In Josef Klegraf & Dietrich Nehls (ed.), *Essays on the English language and applied linguistics on the occasion of Gerhard Nickel's 60th birthday*, 3-10. Heidelberg: Julius Groos.
- Halliday, Michael A.K. (1967). Notes on Transitivity and Theme in English, PartII. *Journal of Linguistics* 3, 199–244.
- Hawkins, John. 1994. *A performance theory of order and constituency*. Cambridge: Cambridge University Press.
- Hawkins, John. 2000. The relative ordering of prepositional phrases in English: Going beyond manner-place-time. *Language Variation and Change* 11, 231-266.
- Hawkins, John. 2004. *Efficiency and complexity in grammars*. Oxford: Oxford University Press.
- Hothorn, Torsten, Kurt Hornik & Achim Zeileis. 2006. Unbiased recursive partitioning: A conditional inference framework. *Journal of Computational and Graphical Statistics* 15(3), 651–674.
- Keller, Rudi. 1994. *Sprachwandel - Von der unsichtbaren Hand in der Sprache*. Tübingen. Francke
- Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech & Jan Svartvik. 1985. *A comprehensive grammar of the English language*. London: Longman.
- R Development Core Team. 2011. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN3-900051-07-0, URL <http://www.R-project.org/>.
- Ramsay, Violetta. 1987. The functional distribution of preposed and postposed *if* and *when* clauses in written discourse. In Russell Tomlin (ed.), *Coherence and grounding in discourse*, 383-408. Amsterdam: John Benjamins.
- Szmrecsanyi, Benedikt. 2004. On operationalizing syntactic complexity. In Gérard Purnelle, Cédric Fairon & Anne Dister (eds.), *Le poids des mots. Proceedings of the 7th International Conference on Textual Data Statistical Analysis. Louvain-la-Neuve*, 10-12.

- Strobl, Carolin, Anne-Laure Boulesteix, Achim Zeileis & Torsten Hothorn. 2008. Bias in Random Forest Variable Importance Measures: Illustrations, Sources and a Solution. *BMC Bioinformatics*, 8-25.
- Strobl, Caroline, James Malley & Gerhard Tutz. 2009. An introduction to recursive partitioning: Rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychological Methods* 14(4), 323-348.
- Tagliamonte, Sali & Harald Baayen. to appear. Models, forests and trees of York English: Was/were variation as a case study for statistical practice.
- Temperley, David. 2007. Minimization of dependency length in written English. *Cognition* 105(2), 300-333.
- Thompson, Sandra. 1985. Grammar and written discourse. Initial and final purpose clauses in English. In Talmy Givón (ed.), *Quantified studies in discourse*. Special issue of *Text* 5, 55-84.
- Thompson, Sandra & Robert E. Longacre. 1985. Adverbial clauses. In Timothy Shopen (ed.), *Language typology and syntactic description*, vol. II, 171-234. Cambridge: Cambridge University Press.
- Verstraete, Jean-Christophe. 2004. Initial and final position of adverbial clauses in English: the constructional basis of the discursive and syntactic differences. *Linguistics* 42, 819–853.
- Wiechmann, Daniel. 2011. Exploring probabilistic differences between genetically related languages. *Language in Contrast* 11(2), 193-215.
- Wiechmann, Daniel & Arne Lohmann (submitted) Modeling Domain Minimization: PP ordering revisited.