

Full title:

Domain Minimization and beyond: Modeling PP ordering

Short title:

Domain Minimization and beyond: Modeling PP ordering

Authors:

Daniel Wiechmann (RWTH Aachen University)*

Arne Lohmann (University of Vienna)

*Lead author:

Dr. Daniel Wiechmann

RWTH Aachen

Institut für Anglistik, Amerikanistik & Romanistik

Karmanstr. 17-19

52062 Aachen

Germany

wiechmann@anglistik.rwth-aachen.de

Domain Minimization and beyond: Modeling PP ordering

Abstract

An important account of linear ordering in syntax is John A. Hawkins's theory of cognitive efficiency and the principles of domain minimization formulated therein. In its latest formulation the theory postulates syntactic and semantic minimization principles. With regard to the relative strength of these principles, prior research into the dynamics of these constraints has come to differing conclusions. Using the relative ordering of prepositional phrases in English as a test phenomenon, the present study contributes to the further development of a theory of syntactic serialization through the multifactorial analysis of naturalistic data from a corpus of present-day British English. We find that lexical-semantic dependency constitutes the strongest constraint on serialization followed by the weight-related, syntactic one. More specifically, our results show that while syntactic minimization has much greater data coverage – it applies to a much larger proportion of the data, the lexical-semantic factor has a much greater effect size, thus is much more seldomly violated. In addition to assessing the relative importance of the two minimization principles, we also investigate the effects of other potential co-determinants of PP order, namely the MANNER > PLACE > TIME generalization and pragmatic information status. Our results suggest that these play statistically significant but tangential roles in PP ordering.

Acknowledgements:

Preliminary versions of this paper were presented at the workshop on Structural Alternations: Speaker and Hearer Perspectives at the University of Groningen in August 2011 and at the Linguistics Colloquium at RWTH Aachen University in November 2011. We thank the audiences of both events for their helpful questions and comments. All of the data analysis was carried out using the statistical computing language R (R Development Core Team 2011); thanks to the R Development Core Team for providing such a valuable software package.

1. Introduction

Research into syntactic constituent ordering has accumulated increasing evidence for the idea that language users tend to prefer constituent orders that impose fewer demands on (verbal) working memory (Hawkins, 1994, 2004; Gibson, 1998; Wasow, 2002; *inter alia*). One influential proposal in this context is John Hawkins's theory of processing efficiency and the principles of domain minimization formulated therein (cf., e.g., Jaeger & Tily, 2011, for a recent overview of processing complexity and communicative efficiency, which demonstrates the pivotal role Hawkins's proposal plays in bridging theoretical linguistic and psycholinguistic research; for recent illustrations of the typological relevance of the theory, cf. Hawkins, 2009, and Diessel, 2005, who present recent formulations and applications of the hypothesis that the patterns of conventionalized syntactic structures in grammars reflect degrees of preference in performance). Essentially, these principles predict that – given a structural choice – speakers will prefer a structure S over a possible alternative S' in proportion to the overall difference in efficiency between S and S', where efficiency depends on the number of linguistic units that need to be processed to recognize domains of dependent elements. The latest version of the theory considers both syntactic and semantic dependency domains, whose minimization contributes to an increase in efficiency of a structure. One phenomenon that is well suited to serve as a test-bed for the theory is that of clauses containing multiple PPs, in which the verbalization of the to-be-communicated message involves a choice between two alternative orderings. For illustration, consider the examples in (i) – (ii) (example taken from Hawkins, 2004:114):

(i) He _{VP}[_V counted _{PP1}[on his son] _{PP2}[in his old age]].

—————

(ii) He _{VP}[_V counted _{PP2}[in his old age] _{PP1}[on his son]].

—————

The solid lines in the examples indicate the sequence of words that have to be parsed so as to recognize the internal structure of the respective verb phrases, which stretches from the verb to the head of the second PP. The ordering in (i) is a little more efficient as one word less has to be processed to recognize the immediate constituents of the VP. In addition to this syntactic domain, the sentence also exhibits a lexical-semantic dependency, which ranges from counted to on as indicated by the dashed line. Such domains are characterized by the fact that certain semantic properties of the predicate can only be assigned once both of these elements have been processed. Again the theory predicts that speakers prefer (i) over (ii) as the lexical-

semantic dependency domain is much shorter in (i) (since count and on occur in immediate adjacency) thereby imposing fewer demands on working memory.

Given that the theory identifies two potentially competing forces, it is of direct theoretical importance to inquire about their relative importance: without testable statements pertaining to the relative strength of opposing constraints, a theory falls short of being falsifiable in cases where these constraints are in conflict as no conceivable empirical state of affairs could possibly prove it wrong (cf., e.g., Newmeyer, 1998, for a discussion). Hawkins himself (e.g. in Hawkins, 2000:258) argues that syntactic domain minimization is the “strongest single predictor”, relegating semantic dependencies to a secondary role. We believe, however, that Hawkins’s assessment is not unproblematic: first there is evidence from sentence recall experiments coming to a different conclusion. Marblestone 2007 suggests that it is semantic dependencies that constitute the stronger constraint. And second, Hawkins’s notion of strength seems to conflate two logically independent concepts into a single, semantically opaque one: specifically, he ignores the possibility of the coverage of a principle, i.e. how many instances are affected by it, and the magnitude of its effect, i.e. how likely is it that it determines the ordering given its applicability, to yield different results.

The present study sets out to (re-)assess the role of both syntactic and semantic domain minimization through multifactorial analysis, aiming to disentangle the “strength-related” notions by framing the issue in the language of regression modeling. The study also elaborates on Hawkins’s work in that it is based on a much larger and more representative data set, which comprises both spoken and written language allowing for a comparison of results across modalities.

The remainder of this section will provide a brief explanation of Hawkins’s principles of domain minimization and in particular its application to the phenomenon of PP ordering (see Hawkins, 2000). Section 2 will present the corpus data that the present study is based on, discuss the operationalization of the variables and introduce the methodology employed in the analysis of the data. Section 3 comprises the empirical part of the study: Hawkins’s proposed domain minimization principles are framed against a background of the MANNER > PLACE > TIME (MPT) generalization and the information status of the PPs. MPT states that manner information should precede spatial information, which in turn should precede temporal information. By information status we are referring to the relative degree of givenness (Gundel, Hedberg and Zacharski, 1993) or accessibility (Ariel, 1990, 2001) of the referents of the NPs that are contained in the respective PPs. We will introduce these additional variables and present various statistical models that are geared to better understand how PP order is

affected by (a) the two minimization constraints, (b) MPT, and (c) information status. Section 3 will also investigate potential contrasts between spoken and written language. In section 4, we will provide a discussion of our findings and propose an explanation of the relative strengths of the ordering constraints, before we conclude the study in Section 5.

1.1 Theoretical background on Domain Minimization

Before we proceed with our empirical assessment of the relative strength of the syntactically and semantically grounded principles of domain minimization a few words are in order that lay out the theoretical background of the present issue. Building on a line of thinking that dates back at least to Behaghel’s (1932) ideas on phrase ordering, Hawkins’s view holds that language use is rational in the sense that there is a general tendency to maximize the efficiency of the formal means employed in linguistic communication. One of the subsidiary principles that Hawkins proposes to define what exactly it means for a form to be efficient is Minimize Domains:

Minimize Domains

The human processor prefers to minimize the connected sequences of linguistic forms and their conventionally associated syntactic and semantic properties in which relations of combination and/or dependency are processed. (Hawkins, 2004:33)

Drawing on the relations of combination and dependency, we can distinguish a structural domain, the so called Phrasal Combination Domain or PCD, from a semantic domain, the so called Lexical Dependency Domain or LDD. Let us briefly illustrate these domain types and how their presence and magnitude can be assessed on the basis of the examples presented as (i) and (ii) above (cf. Figure 1).

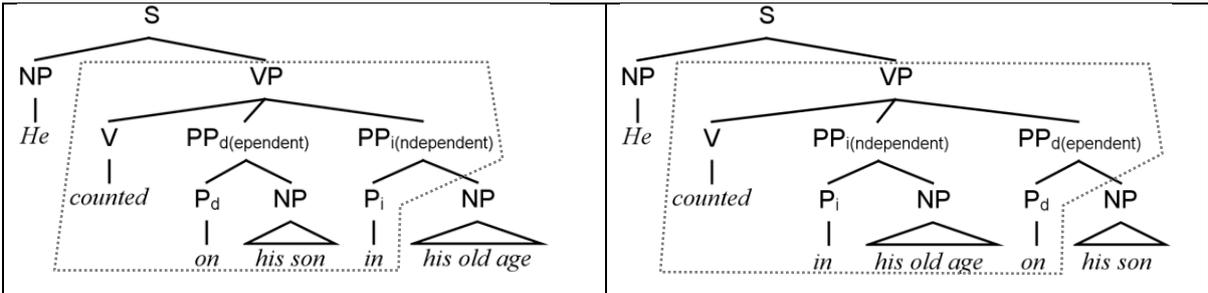


Figure 1: Differences in phrasal combination domain (PCD) length of alternative PP orders

Figure 1 illustrates the relevant properties regarding the efficiency of the two competing structures. The dashed lines around a given tree fragment delineate the respective phrasal combination domains (PCDs) for the VPs, i.e. the smallest connected sequences of terminal elements that must be processed to identify the VP-internal structure. Comparing their sizes (in words), we may conclude that the PP-order on the left is a little more efficient than the one on the right, as it permits that the three immediate constituents of the VP can be recognized on the basis of only five terminal nodes, whereas for the PP-order on the right we need to process six words to be able to identify the three nodes immediately dominated by the VP node. In other words, the PCD-minimization principle predicts that the left-hand variant be preferred (even though this preference is minimal in the example).

Furthermore, the two PPs differ with respect to their semantic relationship to the verb in that only one of the two PP is interdependent with the verb (\rightarrow relation of dependency). A PP is semantically interdependent with the verb if it encodes semantic properties whose processing is necessary to understand the meaning of the main predicate. In the present example, we observe that it is necessary to process the terminal node on to be able to recognize that the overall meaning of the sentence does not describe an actual event of counting—viz. it is not about ascertaining the number of elements of some set—but that it really describes a relation of reliance. Hence, the ordering on the left allows for a faster recognition of the predicate as count and on are put in adjacency, whereas the order on the right requires that one processes all elements of an intervening PP, in his old age, before that (inter)dependency can be recognized. In order to decide on a principled basis whether or not a given PP is to be categorized as semantically dependent or independent, Hawkins proposes two semantic entailment tests: the verb entailment test and the pro-verb entailment test (Hawkins, 2000: 242).

Verb entailment test:

If $[X V PP PP]$ entails $[X V]$, then assign V_i (= verb is independent)

If not, assign V_d (= verb is dependent)

We may illustrate this using the example in (ii), which is repeated here as (iii).

(iii) He counted_{PP1}[in his old age] PP2[on his son].

Applying the test, we find that He counted in his old age on his son does not entail He counted, so we mark the verb as being dependent and assign the label V_d . The second test geared to identify an interdependency of a PP and the verb and assumes the following form:

Pro-verb entailment test:

If [X V PP] entails [X Pro-V PP] or [*something* Pro-V PP] for any pro-verb sentence listed below, then assign P_i . If not, assign P_d

Pro-verb sentences: X did something PP; X was PP; something happened PP;
something was the case PP; something was done (by X) PP.

Applying it to our example in (iii), we find that He counted in his old age entails He did something in his old age, so the first PP is independently processable and marked as P_i . Testing the second PP, however, we get a different result as He counted on his son does not entail He did something on his son. So, this PP is assigned P_d . We considered both tests as potentially providing sufficient conditions for the establishment of the very same type of lexical dependency, which seems to be intended given the definition of a lexical dependency domain.¹

The LDD-minimization principle asserts that language users should prefer those orders in which the distance between semantically interdependent units is shorter, so again the structure on the left-hand side in Figure 1 is considered more efficient and hence more likely to be produced. The structures in (i) and (ii) represent a scenario in which the independently processable PP is longer than the dependent one, meaning that the two motivations pull in the same direction. However, it is also possible for them to compete with each other if the dependent PP happens to be the longer one. Our discussion of the relative importance of PCD- and LDD-minimization will include a closer look at such conflict cases. That is, in addition to fitting regression models to the complete data set, we will also focus on that subset of our data in which the minimization principles make competing predictions.

2. Data and annotation

The data for the present study were extracted from the British component of the International Corpus of English (ICE GB, Nelson, Wallis and Aarts, 2002).² In order to not miss relevant

data due to particular decisions in the annotation of the ICE data, our search pattern was designed for maximal generality and matched all ‘V PP PP’ sequences in the corpus (search string: “((,VP)(,PP)((,PREP))(,PP)((,PREP)))” - number of hits = 2,727). We then manually weeded out all those instances that did not instantiate the target construction or threatened to introduce confounding variables. In particular, we removed all cases in which (a) the two PP were hierarchically ordered, i.e. one PP was contained in the other, or in which (b) any other material occurred in the relevant clause apart from the two PPs.³ Following this procedure, we are left with 1256 data points. These data split up into 719 instances from the spoken and 537 instances from the written medium allowing us to measure and compare the relative strength of the combinatorial/dependency domain across modalities. The data set of the present study constitutes an advancement over prior work (Hawkins, 2000, 2004) as (1) it is more voluminous, (2) it is more balanced in terms of its register make up, (3) it is compiled from a narrower and more recent time span, the 1990s, and, finally, (4) it has been taken from a single variety, namely British English. Hawkins’s data, in contrast, were gathered rather unsystematically from a 500 page corpus of written-only text from novels and (popular) science and amount to only 394 data points. These works were produced in the time span between 1949 and 1994 and comprise both British and American English. Thus, the present results stand on a more solid empirical basis and promise a greater potential for generalization from sample to population. Table 1 presents a comparative overview.

Table 1: Comparison of data sets - Hawkins (2000) versus present study

	Hawkins’s data set	Present data set
What?	500 page corpus of selected text from novels and some (popular) science	Exhaustive extraction from ICE-GB
Time	1949 - 1994	1990s
Variety	British and American	British
Medium	Written	Written and Spoken
Sample size	394	1,256

Once extracted and filtered, the data were annotated with respect to syntactic and lexical-semantic dependencies. The annotation needed for the evaluation of the syntactic constraint, i.e. the PCD-minimization, was straightforward: all that was needed was a measurement of the weight of the two PPs, i.e. the amount of linguistic material. The basic unit in that

measurement was that of a word, a string of characters enclosed in whitespaces. Counting words improves the comparability of our results to Hawkins's work; furthermore it has been shown to be the unit of choice also in other works on weight effects (see Szmrecsanyi, 2004).⁴ We also measured phrasal weight in terms of number of characters (so as to be able to consider word length-effects) but this refinement did not substantially influence the results. The magnitude of domain minimization was then expressed as the difference in PCD-length of the observed order and the (not actualized) alternative order. For example, the sentence The astronomer _{PCD}[gazed into the sky through] his brand-new telescope would receive an observed PCD length of 5 words. Its non-actualized, but possible, alternative ordering, i.e. The astronomer _{PCD}[gazed through his brand-new telescope into] the sky would have a PCD length of 6 words meaning that this data point would receive a value for the difference in PCD length (= Δ PCD) of (6 – 5 =) 1, which would indicate that the observed order is a little more efficient. The next step in annotating the data was to identify potential semantic interdependencies between V and any of the PPs, i.e. to detect the presence of an LDD. This was done by way of applying the semantic entailment tests described earlier, which both authors applied independently.⁵ In most cases the outcome of the tests was uncontroversial, yet there were some problematic cases. In about 18% of the relevant data our coding decisions regarding the outcome of the entailment tests did not match up.⁶ Inspecting the areas of divergence, it turned out that the differences in judgment almost fully reduced to subtle changes in the semantics of polysemous items (verbs and/or prepositions) that go together with changes in subcategorisation. Consider the innocent looking example in (iv).

(iv) I have spoken to you on the phone.

The entailment tests instruct us to ask the following questions.

Q1 : Does I have just spoken to you on the phone entail I have spoken?

Q2a: Does I have just spoken on the phone entail I did something on the phone?

Q2b: Does I have just spoken to you entail I did something to you?

Note that a single negative outcome in either of the tests is sufficient for postulating a semantic interdependency between verb and PP. In these 18% of the cases, one annotator answered all questions in the positive, thus judging both PPs as independent, while the other annotator answered at least one question in the negative, thereby postulating interdependence

between the verb and (usually) one of the prepositional phrases. Applied to the example given in (iv), this means that both annotators agreed that Q1 I have just spoken to you on the phone clearly entails I have spoken. They furthermore agreed that Q2a I have just spoken on the phone entails I did something on the phone. However, there was disagreement concerning Q2b, i.e. the question of whether I have just spoken to you entails I did something to you. A third ‘yes’ would result in the judgment that no semantic dependency is to be posited, whereas a negative answer would lead to the postulation of a semantic dependency between the verb and that PP. Answering the question strikes us as non-trivial because the test sentence—while being the one appropriate for verbs with agentive subjects—seems to be somewhat infelicitous as doing something to somebody seems to evoke a sense in which the referent of the direct object is much more strongly and negatively affected by the doing of the agent than is the case in an act of speaking. The general problem seems to be that the use of generic verbs or pro-forms may sometimes lead to a change in the semantics of polysemous items. Having discussed all such contexts of disagreement, we found ourselves unable to settle on a single, fully satisfying solution. In consequence, rather than opting for either a more narrow, i.e. more restrictive interpretation or a more broad, i.e. less restrictive interpretation of the verbal semantics, we decided to entertain two coding strategies corresponding to these competing modes of interpretation, which translate into two different factors investigated in the statistical analysis. We will restrict our discussion of lexical domain minimization to the more conservative narrow operationalization, in which questions like Q2b are answered in the positive and fewer semantic dependencies are posited.⁷ Figure 2 shows the observed PCD and LDD minimizations.

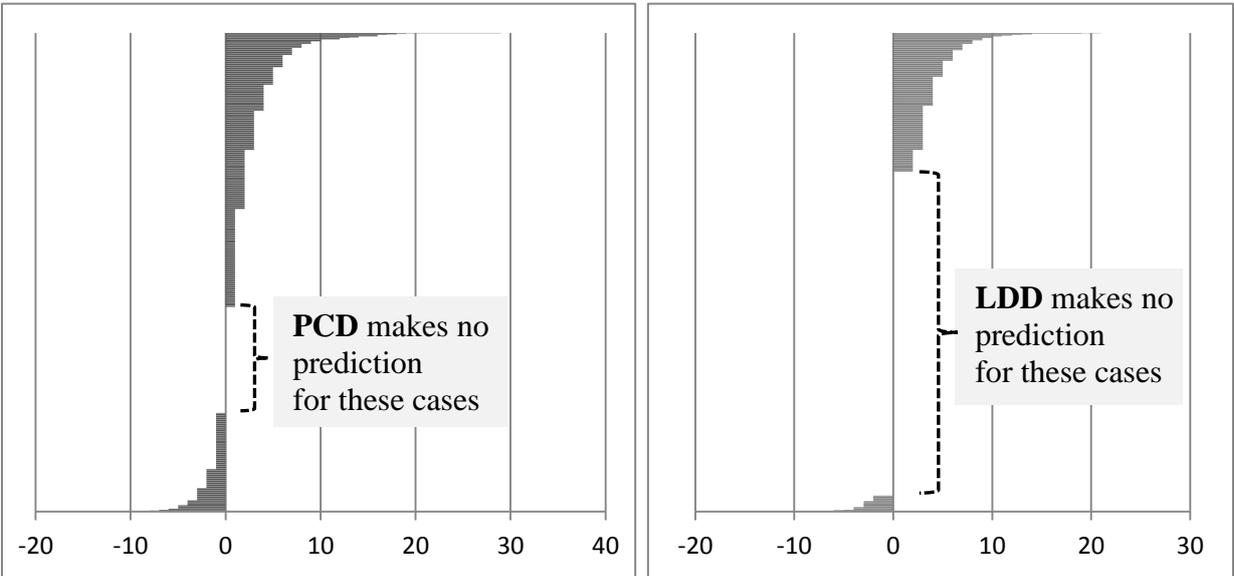


Figure 2: Observed PCD (left) and LDD minimizations (in words). Aligned on the vertical axis are the 1256 [V PP PP] sequences under analysis (sorted by magnitude). Values on the horizontal axis indicate the extent to which the observed ordering can be considered more efficient than its alternative.

Each of the 1256 double-PP constructions assumes a position on the vertical axes in Figure 2 and their arrangement is determined by the magnitude of its respective domain minimization, which can be read off the horizontal axes, which thus denote the degree to which a given instance can be said to be more efficient than its alternative. Thus, the higher the value, the greater is the magnitude of domain minimization for that example. Negative values indicate that the non-actualized, alternative ordering would in fact be more efficient in the respective domain. Cases in which the ordering choice has no impact on the efficiency of a pattern, i.e. if both PPs are of equal length (for PCD) or if there is no semantic dependency between the verb and either of the two prepositions (for LDD), receive the value 0 in the respective chart. Looking at these charts, we observe

1. that overall there are more cases with positive values than cases with negative ones. That is, if there is a difference in efficiency between the alternatives, the more efficient variant tends to be produced.
2. that PCD minimization makes ordering predictions for a larger share of the data as indicated by the fact that there are a lot more values $\neq 0$ in the PCD chart than in the LDD chart (PCD minimization makes an ordering prediction in $\sim 78\%$ of the cases, whereas LDD minimization does so in only about 30% of the cases).
3. that the mean value of the negative scores appears to be lower than that of the positive scores. That is, if the less efficient variant is in fact preferred over its more efficient alternative, the difference tends to be less pronounced.

Thus a brief look at these descriptive statistics already provides us with some interesting insights into our data. To arrive at a more nuanced understanding, we fit a statistical model to the data.

3. Statistical Modeling: Method & Results

In order to evaluate the distributions shown in Figure 2, and to assess the strength and importance of the two domain minimizations, we fit binomial logistic regression models without intercept to the data (cf. Benor & Levy 2006; Lohmann 2011; Levy in progress Ch. 6.8.4). Ordinary binomial logistic regression models, i.e. models with an intercept, are fairly

widely used in linguistic analyses to model linguistic choices with a binary outcome (cf., e.g., Baayen, 2008, for an introduction). However, such models cannot be applied in the present context for reasons that pertain to the nature of both the response and the predictor variables. Principally, ordinary logistic regression models have a “failure/success”-response variable, which requires that we have a clear definition in place of what exactly it means for an outcome to be a “success”. For example, when investigating genitive variation, we could just declare events that instantiate the analytic variant to be a “success” and treat the s-variant as a “failure”, or vice versa. There is no analogous way of defining our response variable when we wish to model the relative ordering of type-identical constituents: all we observe for a given instance is a particular relative ordering of the two PPs. Similarly, our covariates in the model are comparative properties (the observed order is k words more optimal than the non-observed order). To adapt the model form to this scenario, we effectively treated the response variable as a dummy variable whose value invariably was set to 1. Hence, the logit response variable is always a “success”. The predictor variables, LDD minimization and PCD minimization, were measured as already indicated: the value for PCD was determined by calculating the length differential (in words) between the two PPs. The sign of the values was aligned with the response, so positive values expressed high degrees of processing efficiency. For LDD, the value was equal to the number of words of the independent PP and its sign was negative if that PP was intervening.

3.1 Pitting LDD against PCD

The first model was set up to allow for a direct comparison of the two minimization principles and assumed the following form:

General model form: $\log\left(\frac{p}{1-p}\right) = \beta_1x_1 + \beta_2x_2, + \dots + \beta_kx_k$

where x_i represents a value of a given explanatory variable and β_i is a real-valued number corresponding to the weight of the i^{th} variable.⁸

LDD & PCD-model: $\log\left(\frac{p}{1-p}\right) = \text{weighted LDD minimization} + \text{weighted PCD minimization}$

The overall classification accuracy of the model is 74.6%, which clearly constitutes a statistically significant improvement over the performance of a null model which would achieve a 50% accuracy by simply guessing the order of PPs.⁹ However, it is also far from being fully predictive suggesting that an explanatorily complete account will have to include further predictors. We should also keep in mind that the phenomenon may very well resist any attempt at a complete explanation simply because one or even both PPs can be non-obligatory. If they are non-obligatory, it seems possible that they have been added to an originally planned, less complex message “on the fly” (rather than being part of the early planning stage). That is to say that the speaker might have updated his belief about his interlocutor’s state of knowledge and might have chosen to add some information which was deemed unnecessary at the time the original message was planned. We will return to such issues in our general discussion in Section 4. At this point, however, we should emphasize that what we are interested in here is not the overall predictive power of the model but rather the comparison of the two variables that pertain to domain minimization. The estimated regression coefficients and their statistics results of the model are presented in Table 2.

Table 2: Coefficient estimates, Standard error estimates, 95% confidence intervals and p-values for predictors in the model

	Estimate	Std. Error	95% Conf. Interval	Pr(> z)
LDD MINIMIZATION	0.65	0.06	0.53-0.77	< 0.001
PCD MINIMIZATION	0.39	0.03	0.33-0.45	< 0.001

As both predictors are measured on the same interval scale (processing advantage in number of words), we can directly compare the values of the coefficient estimates to assess their relative power. We observe that the LDD constraint is about $(0.65/0.39=)$ 1.7 times stronger than the PCD constraint. These results contrast with Hawkins’s findings as they suggest that the lexical-semantic dependency constitutes a stronger constraint on serialization than the weight-related syntactic one. More specifically, our results show that while syntactic minimization has much greater data coverage, the lexical-semantic factor has a much greater effect size, thus is much more seldomly violated. This difference in coverage is also reflected in the higher standard error and confidence interval for LDD, as compared to PCD.¹⁰ Figure 3 helps to illustrate that point.

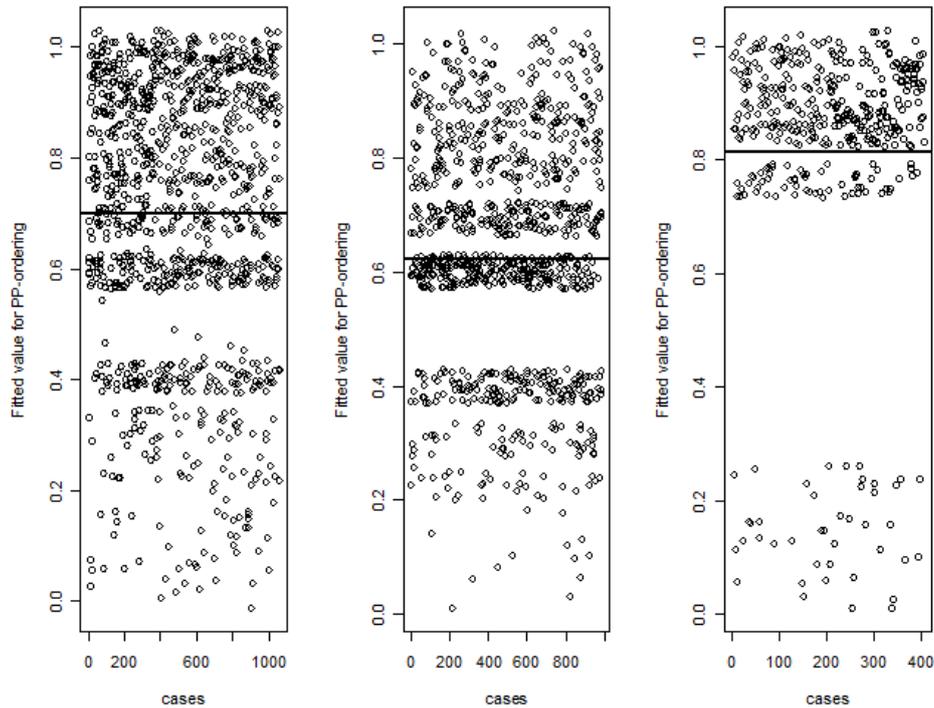


Figure 3: Predicted values of regression models (only cases are shown where predicted value is $\neq 0.5$: left: 'PCD & LDD'- model, middle: 'PCD-only'-model, right: 'LDD-only'- model. (For purposes of exposition, we added a small amount of noise to the value of each data point (using the jitter function in R), which is why some of the cases assume values smaller than 0 and larger than 1.)

Figure 3 contains three plots representing the performance of three different models. The plot on the left hand side represents the results of a model that includes both the syntactic (PCD) and the semantic (LDD) predictor. The other two plots can be viewed as a decomposition of that model: the plot in the middle represents a model that predicts the ordering choice only on the basis of PCD minimization, whereas the plot on the right predicts the ordering choice only on the basis of LDD minimization. Each dot in a given plot represents a fitted (viz. predicted) value of a given model, i.e. the output value (or answer data point) that is predicted by a regression equation. In other words, each dot's position on the vertical axis represents an estimation of the probability to produce the observed PP-order. The plots show the fitted values of only those instances for which the respective model makes an informed prediction, i.e. where it can use at least one predictor to move away from a purely chance-level prediction, which corresponds to a value of 0.5. The solid bars indicate the mean fitted values for each model: the closer the bars move towards the upper boundary, the greater is the model's overall degree of certainty. Comparing the mean fitted value across the three plots shows that whenever there is an informed prediction, it is most confident when it is based on semantic information, reflecting the greater effect size of LDD (see Table 2). The relatively

greater number of cases in the middle plot again shows the greater applicability of PCD in comparison to LDD. The rightmost plot in Figure 3 reveals that the LDD constraint, while being the strongest predictor, applies only in a subset of the data, as it affects only those data points where one of the two PPs is semantically dependent (see also Figure 2). The model we calculated above (see Table 2) is thus based on a sample in which LDD does not consistently apply. As a sanity check of our assessment of the difference in effect size between the two predictors we also fitted a model to only those data in which both PCD and LDD applied in every case, i.e. to only those cases where there is a dependent PP (N=406).¹¹ The results of this model confirm the results we obtained in the initial calculation: Coefficient values change only slightly and thus LDD is still the predictor with the greater effect size (LLD: Estimate 0.66; SE 0.06; $\Pr(>|z|) < 0.001$; PCD: Estimate 0.44; SE 0.03; $\Pr(>|z|) < 0.001$).¹²

3.2 Conflict cases

It is of particular interest to focus on the subset of cases in which PCD and LDD pull in different directions, i.e. scenarios where they make conflicting predictions. This is the case for 12% of the data (n=101). Consider (v) and (vi) for examples of such contexts:

- (v) [D]welling [for a few moments] [on what the police have done or have not done]

(ICE-GB:S2B-037 #59:1:A)

Δ PCD = +6, Δ LDD = -4

(being faithful to PCD leads to an overall improvement of 6 words and being faithful to LDD leads to an overall improvement of 4 words. The positive sign for PCD and the negative value for LDD, respectively, signify that in this example the former is being adhered to, while the latter has been violated.)

- (vi) But let 's just stick [with the nerve affecting the muscle] [for the moment]

(ICE-GB:S1B-009 #160:1:A)

Δ PCD = -3, Δ LDD = +3

In cases like (v) and (vi) LDD and PCD call for different orderings. These conflicts arise whenever the semantically dependent phrase is longer than the independent one. In these cases the LDD constraint calls for a long-before short ordering, which constitutes a violation of the PCD constraint. These contexts permit two outcomes: either PCD (as in (v)), or LDD wins the tug of war (as in (vi)). Note that either outcome results in a certain processing advantage on

one dimension and a disadvantage on the other. Recall that we calculated the size of this (dis-)advantage by counting the number of words intervening between the two dependent elements. Thus for (v) we obtain a processing advantage for PCD of six words, as the second phrase is six words longer than the first. At the same time LDD yields a processing disadvantage of four words, as the phrase intervening between the verb and the dependent preposition is four words long. In example (vi) both constraints yield the same value, viz. three words, but it is LDD which wins out in this example.

Within Hawkins’s framework it is assumed that in these cases of conflict the adherence to one or the other domain should be “in proportion to the extent of the minimization difference in each domain.” (Hawkins, 2004:111). This is to say that one would expect that factor to win which yields the greater processing advantage (in number of words), as is the case in (v). This does not have to be the case however, as it could be that one of the two factors is considerably stronger thereby overruling the other even in contexts where the processing advantage in number of words of its competitor is equal (as in (vi)) or even greater. In order to further explore this issue, we may sum the two distance values for LDD and PCD for all conflict cases, yielding for example $(+6 -4=) +2$ for (v) and $(-3 +3=) 0$ for (vi). We may then calculate the mean of this difference separately for (a) all cases in which LDD has won and (b) all cases in which PCD has won. The results of such a calculation are given in Figure 4.

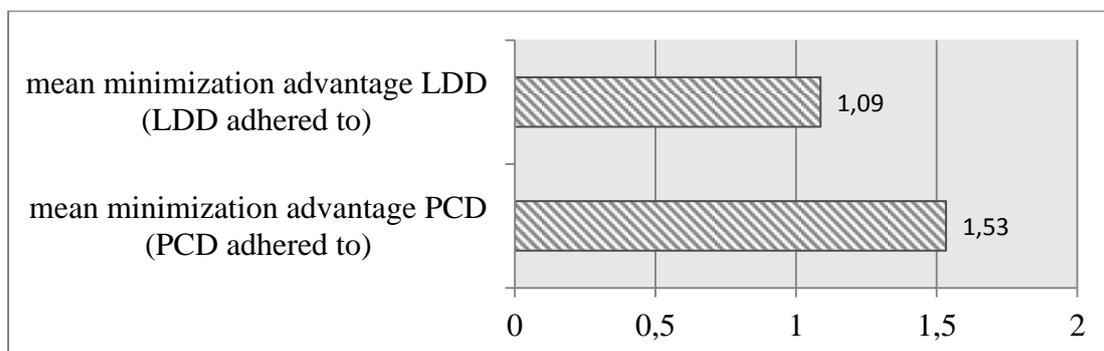


Figure 4: Mean minimization advantage in cases of conflict of LDD and PCD (in number of words)

The positive values in Figure 4 indicate that the constraint that succeeds typically exhibits a minimization advantage over its competitor. However, note that the minimization advantage is typically more pronounced when PCD wins over LDD: the value of 1.53 for PCD means that in cases of conflict that are won by PCD, the minimization advantage obtained through an adherence of PCD is 1.53 words larger than the minimization advantage obtained through an adherence to LDD. The value of 1.09 for LDD means that in cases of conflict that are won by

LDD, the minimization advantage obtained through an adherence of LDD is 1.09 words larger than the minimization advantage obtained through an adherence to PCD. In other words, while PCD on average needs a minimization advantage over LDD as pronounced as 1.53 words to succeed, LDD arises as the winner even if its advantage is only 1.09 words. This difference roughly corresponds to the variables' coefficients in the regression model, where the coefficient for LDD was found to be 1.7 times stronger (see Table 2).

3.3 Including MPT

As indicated earlier, we also investigated how the ordering choice is influenced by the MANNER > PLACE > TIME (MPT) generalization. The MPT explanation holds that adjacency of V and PP iconically mirrors the degree of semantic integration: manner phrases are typically more central to the action (or situation) being described than those of place and time, and place is usually more important than time. We followed the characterization of adverbial roles as described in Quirk, Greenbaum, Leech, and Svartvik (1985:649-650), where it is actually referred to as “PROCESS > PLACE > TIME”, and annotated our data with information regarding this semantic dimension. Quirk and colleagues introduce many fine-grained differentiations but since the MPT-generalization makes reference only to what are considered major superordinate categories in Quirk et al., the annotation process was rather straightforward and relatively unproblematic. The only noteworthy problem occurred with cases in which the “V PP”-sequence was motivated by conceptual metaphor (Lakoff and Johnson, 1980) and was noticeably non-literal as in to fall in love. Despite being headed by a spatial preposition, such cases were treated as specifying neither PLACE nor TIME (let alone MANNER) but were put into a fourth category, “other”. Of course, the usage of spatial prepositions like in is very often metaphorical in nature making it very difficult to motivate a principled cut-off point. For example, the preposition in in a phrase like in his reading of Joyce ultimately is also licensed by conceptual metaphor. However, in contrast to a phrase like in love, a phrase like in his reading of Joyce can be felicitously used to answer a where-question. Consider the examples in (vii) and (viii):

- (vii) Where did John adopt the method? – In his reading of Joyce.
- (viii) Where did John fall? – ?/*In love.

Using such wh-questions as tests, we would keep cases like (vii)—in this case assigning the role PLACE—but exclude from the MPT-candidate list cases like (viii). It is worth mentioning

that we included both static (in the room), as well as directional PPs (into the room) in the category PLACE, as both provide spatial information. Having assigned a role to each PP, we then derived a variable specifying (a) whether MPT applies and (b) whether the ordering constraint was adhered to or not. We assigned the value 0 whenever MPT made no prediction, i.e. when the PPs assumed the same role or if at least one role was specifying neither manner nor place nor time. When MPT was applicable and respected, we assigned the value 1; when it was applicable but violated, we assigned the value -1. The next step in our modeling was to add MPT into a model that also comprises the two domain minimization variables, LDD and PCD minimization and see if its inclusion results in a statistically significant improvement of the model. Since MPT, LDD and PCD do not run on the same scale (MPT may take on only three values while LDD and PCD exhibit a much wider range), the predictor variables were standardized, which allows for a straightforward comparison of the coefficients of the three predictors.¹³ Table 3 presents the standardized regression coefficients.

Table 3: Three predictor model – LDD, PCD, MPT (Standardized input variables)

	Estimate	Std. Error	95% Conf. Interval	Pr(> z)
LDD MINIMIZATION	3.19	0.29	2.62-3.76	< 0.001
PCD MINIMIZATION	2.43	0.21	2.02-2.84	< 0.001
MANNER > PLACE > TIME	1.15	0.14	0.88-1.42	< 0.001

The multifactorial model discloses a comparably weak effect of MPT as a predictor of PP-ordering, as both the coefficients of LDD and PCD are considerably higher. Moreover, including MPT does not affect the relation between LDD and PCD: LDD still remains the stronger effect.

Despite its relatively weak effect, including MPT does improve the predictive accuracy of the model: An analysis of deviance reveals that the 3-variable model performs significantly better than the model without MPT (Deviance = 1174.3-1245.4 = 71.1, $P(>|\text{Chi}|) < 0.001$). The 3-variable model also scores better in terms of Akaike’s Information Criterion, which punishes models with more parameters ($\text{AIC}_{2\text{-var model}} = 1196$; $\text{AIC}_{3\text{-var model}} = 1180$). Furthermore, including MPT raises the classification accuracy from 74.6% to 76.8%. Since the model includes two semantic variables which could be correlated, we tested for multicollinearity of the model. This however turned out not to be an issue as the model’s condition number is very low ($\kappa = 1.89$).

In summary, our results suggest that while being the weakest factor investigated so far, MPT is still relevant for an explanation of PP-ordering. This contrasts with Hawkins’s assessment which concludes that “MPT’s predictions [...] are not statistically significant [and exhibit] a success rate that is at chance level at best” (Hawkins, 2000:240).

3.4 Information Status

Another factor to potentially influence the linearization of syntactic constituents, which may therefore also be relevant to the order of prepositional phrases, concerns what we refer to here as the (pragmatic) *information status* of entities in a context of utterance (cf. Prince 1981, Gundel et al. 1993, Ariel 1990, *inter alia* for related proposals and overviews). The general idea underlying the inclusion of information status as a potential co-determinant of phrase order is this: information that the speaker of an utterance U can presume to be active in the hearer’s mind at the time U is produced will be expressed before information that presumably is not active in the hearer’s mind at the time U is produced.¹⁴ The degree of activation of a nominal concept can be derived from the linguistic form chosen to express the information unit in question:¹⁵ here I am basically, it is assumed that more active information can be communicated (=re-activated) with very little linguistic material, e.g. a pronoun, whereas the communication of new information requires more elaborate means of expression, e.g. a full lexical noun phrase. In our assessment of information status, we followed the operationalization used in Wolk, Bresnan, Rosenbach and Szmrecsanyi (2012) and assigned to each PP-internal NP one of four possible values for information status (ordered by degree of activation from highest to lowest): personal pronoun, proper name, definite NP, indefinite NP. Table 4 presents the results of a model that includes information status as a fourth predictor (again to maximize the comparability of the regression estimates all input data were standardized as described in the previous section).

Table 4: Four predictor model – LDD, PCD, MPT, INFORMATION STATUS (Standardized input variables)

	Estimate	Std. Error	95% Conf. Interval	Pr(> z)
LDD MINIMIZATION	3.33	0.30	2.74-3.92	< 0.001
PCD MINIMIZATION	2.21	0.21	1.80-2.62	< 0.001
MANNER > PLACE > TIME	1.16	0.15	0.87-1.45	< 0.001
INFORMATION STATUS	0.89	0.16	0.58-1.20	< 0.001

We observe that information status (INF) has a statistically significant effect on the ordering. This result contrasts with the findings reported in Hawkins (2000), who concludes that “[p]ragmatic information status [...] appears to add nothing to the predictions of EIC [here PCD] and lexical adjacency [here LDD]” (Hawkins 2000:257). In fact its inclusion raises the classification accuracy of the model from 76.8% to 78.7%. However, our results support the claim that the effect of information status is rather weak, in fact it is the weakest of the four variables tested. Multicollinearity is again not an issue with this model (Condition number $\kappa = 1.96$).

3.5 Comparing modalities

Our data comprise sufficient amounts of cases from both spoken and written language allowing us to compare the influence of the investigated variables across modalities ($N_{\text{spoken}} = 719$; phenomenon occurs ~113 times per 100k words; $N_{\text{written}} = 537$; phenomenon occurs ~127 times per 100k words). The statistics from the resulting models are given in Table 5.

Table 5: Four predictor models – LDD, PCD, MPT, INFORMATION STATUS (INF) – across modalities (Standardized input variables)

	Spoken sample (N= 719)				Written sample (N = 537)			
	Estimate	Std. Error	95% Conf.Interval	Pr(> z)	Estimate	Std. Error	95% Conf.Interval	Pr(> z)
LDD	3.12	0.36	2.41-3.83	< 0.001	3.87	0.56	2.77-4.97	< 0.001
PCD	2.58	0.30	1.99-3.17	< 0.001	1.83	0.30	1.24-2.42	< 0.001
MPT	1.39	0.21	0.98-1.80	< 0.001	1.00	0.22	0.57-1.43	< 0.001
INF	0.68	0.20	0.29-1.07	< 0.001	1.17	0.24	0.70-1.64	< 0.001

Comparing effects across modalities, we observe that the regression coefficients of LDD and INF are lower, whereas PCD and MPT are stronger in the model fitted to the spoken data. The predictive accuracy of the models is 80.7% for the spoken data but only 75.9% for the written data. With regards to the two domain minimization variables, PCD is a little more important in contexts of real-time pressure (spoken language) and the reverse is true for LDD. A look at the confidence intervals of the models’ coefficients reveals however that these strongly overlap across the two modalities for all variables, which means that the ‘true’ coefficient

values may in fact be the same. In other words, none of the differences between modalities is statistically significant.

4. General discussion

The most important finding of the present study is that semantic domain minimization constitutes the strongest of the tested constraints in the serialization of prepositional phrases. Globally speaking, i.e. in a model that comprised all data points and estimated the effects of both LDD and PCD minimization, the regression coefficient of LLD was about 1.7 times greater than that of PCD.¹⁶ The latter, weight-related syntactic factor needed to be about $(1.53 / 1.09 \Rightarrow) 1.42$ times more pronounced to override the semantic preference in cases where the two are in conflict (cf. Figure 4). MPT was shown to be a comparatively weak predictor, yet its inclusion still leads to a statistically significant improvement of the predictive power of the model. This suggests that semantic factors play a more important role than suggested by Hawkins (Hawkins 2000, 2004). Our findings regarding the relative strengths of the dependency domains, which are based on naturalistic usage data, are compatible with experimental results reported in Marblestone (2007). The relative importance of the three factors investigated turned out to be comparable across modalities. To see if and to what extent these findings are generalizable to other populations, they must, of course, be submitted to further testing against data from other phenomena (cf. Lohse, Hawkins & Wasow 2004 on particle placement) but the solid empirical basis on which the present results are based gives us some confidence in asserting that they are good approximations of the dynamics of the powers at hand. In the remainder of this section we shall discuss what we believe are the most interesting questions at this point.

4.1 Overall performance of the model

In light of the fact that even for the spoken domain, the predictive success of the models is limited (the prediction accuracy was 78.7%), it seems reasonable to assume that additional factors figure in an explanatorily fully adequate account of PP-ordering. One potentially relevant variable is rhetorical in nature and concerns the relation of the position of a phrase within the sentence and the degree to which the information encoded by that phrase is emphasized. That is to say that speakers may very well put a given phrase in sentence-final position to emphasize its contents (end-focus, cf. Givón, 1992; Lambrecht, 1994). Since information about intended focus is not available from the corpus data investigated here, we cannot measure its contribution to the co-determination of PP-order.

However, there is also reason to believe that PP-ordering represents a type of phenomenon that is categorically different from other structural alternations like, e.g. the dative alternation (*He gave Mary the book* vs. *He gave the book to Mary*; cf. Bresnan et al., 2007, who are able to predict up to 95% of the cases). The crucial difference between PP-ordering and such alternations appertains to the fact that in the vast majority of cases of ‘double PP constructions’ at least one of the two PPs is not obligatorily required by the semantics of the verbal head. Hence, planning the ordering is not strictly required in early phases of utterance planning. Speakers can simply elaborate their utterance and add additional information “on the fly”. This is not possible in the case of e.g. the dative alternation as this variation involves two VP-internal arguments both of which are semantically required to express a complete thought—in Frege’s sense (Frege, 1948) and, consequently, both are considered to belong to one and the same message in established models of language production (cf. e.g. Bock & Levelt 1994).¹⁷ From this it follows that at least one PP constituent may not have been part of the initial planning phase of the message to be communicated. It seems fair to say that typically, i.e. in the large majority of cases, only one of the PPs specifies information that is necessary to express the thought at hand while the other expresses additional information the planning of which may have followed the planning of the obligatory elements. This would happen if the speaker wanted to express some proposition that comprises one obligatorily required PP-element specifying, for example, spatial information and during speaking decides that it might be appropriate to add some more information to the message which was not part of the initial planning process. In such scenarios we can expect violations of PCD (as in I have been living [in this wonderful little town in the northern part of Germany] [since 1992]).

4.2 Why is LDD stronger than PCD?

The tendency to place semantically dependent PPs adjacent to the verb can be taken to also follow from the contingencies and the time-course of sentence planning. A greater likelihood of semantically dependent phrases to stand adjacent to the verb follows directly from the general architecture of current models of speech production. Basically, the rationale is that a semantically dependent PP has to be part of an early planning phase, while semantically independent ones may or may not be part of that early phase, which will then give rise to the observed distributional bias. Following established psycholinguistic theorizing in language production (cf. Levelt, 1989; Bock & Cutting, 1992; Bock & Levelt, 1994) we may assume (a) that clause level structures are the fundamental units of utterance planning and (b) that

within the preparation of clause level structures, conceptual preparation, i.e. choosing the conceptual building blocks, precedes both lexical access and, importantly, constituent assembly. For illustration, let us consider the sentence given in (ix):

(ix) John was counting PP_{dependent}[on his son] PP_{independent}[at the time]

We may assume that the speaker of (ix) knows that she wants to express roughly the idea that some individual, John, was relying on (the help of) his son in some situation of his life. What is minimally needed to verbalize this idea is the production of some linguistic form that is suited to express the thought ‘that John relies on his son’. The speaker may or may not choose to add additional information to that message but, crucially, he cannot say less than this. Now, our speaker apparently has selected the linguistic form ‘count on NP’ to express the semantic relation RELY.ON (x,y). As a result of this, we find ourselves in a situation where the preparation of the phrase [on his son] is necessarily part of some early planning phase, in which the verbalization of the most central proposition is done. In contrast, the second, semantically independent PP, [at the time], may or may not have been part of this early planning phase. It is certainly possible that our fictitious speaker has begun planning her utterance without it and she might have chosen to add this information to the message at some later point. So, we end up with only two possible situations: either [on his son] and [at the time] were planned at roughly the same time (during initial planning) or the preparation of [on his son] was chronologically prior. If we assume that constituents are produced as soon as possible, we would expect semantically dependent phrases to tend to occur more often in adjacency to the verb. So the argument assumes the following form: If (p) a semantically dependent PP must be ready at the time the processing system engages in the planning of constituent ordering (qua its being part of the verbalization of the minimal proposition to be expressed by the clause) and if (q) the planning of independent PPs can in principle begin at a later stage, then it follows—*ceteris paribus*—that semantically dependent phrases should tend to precede semantically independent phrases.¹⁸

4.3 LDD and PCD across modalities

Although we observed only non-significant differences across modalities for the tested constraints, we may nevertheless discuss the trends we found for the variables pertaining to domain minimization, as they relate to previous research in interesting ways. While the question of modality-specific differences has been disregarded in prior research on PP

ordering, there are studies on a related phenomenon, viz. particle placement, which are relevant in the present context. Like PP-ordering particle placement (He gave [prt up] the job vs. He gave the job [prt up]), too, exhibits both semantic and syntactic domains: the dependency between verb and particle constitutes a LDD whose minimization involves placing the particle adjacent to the verb. The PCD of the VP is minimized if long NPs are positioned after the particle.

Gries (2003: 87-88), as well as Lohse et al. (2004) find that the dependency relation between verb and particle (LDD) influences the choice of construction more strongly in written than in spoken language. These results are congruent with an earlier study by Kroch & Small (1978) who find that LDD plays out more strongly in more formal settings and interpret this effect as an influence of prescriptivism.¹⁹ Positioning the particle right after the verb reflects the semantic unity of these two elements, which corresponds to prescriptivist principles according to Kroch & Small (1978: 46-49). Although our results are not statistically significant and therefore not entirely conclusive, it is interesting to note that we observe the same trend in our data for PP ordering, which could thus be interpreted as a reflection of prescriptive norms.

In contrast, a possible interaction of modality and PCD is less clear. While both Gries (2003:84-85) and Lohse et al. (2004) report more pronounced effects of length for the written medium with particle placement, the size of a potential interaction effect appears to be rather weak. A recalculation of the data provided by Lohse et al. (2004) yields a non-significant result of the comparison across modalities.²⁰ Recall that we found PCD to yield a stronger result in speech, which thus contrasts with previous research. However, since the difference we found is not statistically significant, it cannot be conclusively interpreted.

Summarizing, there is some evidence for stronger semantic dependencies (LDD) in written language, while for PCD conflicting trends were found. As none of these results are entirely conclusive, a more detailed analysis is called for. Since there are pronounced differences between the registers that make up each modality in a general corpus, we believe that such an analysis should take into account these more fine-grained distinctions, beyond a mere spoken-written divide. For example, it seems likely that staged speech and drama (both written) are more similar to the language of direct communication (spoken) than telephone calls among friends are to legal cross examinations (both spoken).

5. Conclusion

This study set out to contribute to the further development of Hawkins's efficiency-based theory of constituent ordering. Specifically, our goal was to assess the relative importance of syntactic and semantic domain minimization constraints on the basis of an exhaustive multifactorial analysis of V PP PP-sequences in the ICE GB corpus. We have argued that prior attempts to determine the roles of these constraints are problematic because (1) they are based on less than optimal corpus data, and (2) they employ less than optimal methodologies to analyze these data. This has ultimately led to a descriptively inadequate picture of the state of affairs. Our results suggest that the minimization of lexical-semantic dependency domains (LDD) constitutes a stronger constraint on serialization (LDD exhibits a greater effect size) than the weight related syntactic one (PCD), even though the latter is characterized by a larger domain of application (PCD exhibits greater coverage). The importance of semantic factors is further corroborated by the finding that the MANNER > PLACE > TIME constraint emerged as a statistically significant predictor. Finally, we observed a statistically significant, but rather weak effect of information status (givenness, accessibility). Given this re-evaluation of the relative roles of the investigated factors, we would like to emphasize a general methodological point: we contend that the further development of Hawkins's theory can strongly benefit from a stronger commitment to empirical rigor and the application of multifactorial analysis like the one proposed here. Framing all statements in the unifying language of regression modeling avoids vague or ill-defined notions such as the 'strength' of a factor and provides instead a number of well-defined constructs (coverage, effect size, etc.) thereby allowing for a better comparison of empirical findings across different phenomena.

With regards to future research, it would be revealing to extend the study to other contexts to see whether the greater strength of semantic dependencies over syntactic ones we found is of general validity. Furthermore it would be of particular interest to investigate these effects in languages with verb-final clause structures, as Hawkins's theory makes different predictions for these. A language which lends itself particularly well to such an analysis is German, which is verb-second in main clauses, but verb-final in subordinate clauses. Therefore we predict that the influences of LDD and PCD interact with clause type in that language: While the ordering of two prepositional phrases in German main clauses should be influenced by LDD and PCD in the same way as in English main clauses, we predict reverse tendencies for subordinate clauses. We are currently working on a paper which empirically addresses this possible interaction of clause type and domain minimization in German.²¹

Endnotes

1 There was some discussion between the authors as to whether or not our treatment of the two tests does in fact faithfully reflect the way Hawkins conceived of them. It should be noted that Hawkins formally distinguishes different situations in which one but not the other test yields a negative outcome and he also distinguished those situation from situations where they both yield a negative outcome (e.g. Vd PPd PPI from Vi PPd PPI) suggesting that he has reason to believe that these different types describe different types of semantic dependencies. We considered these test to capture the same dependency domain. Compare Hawkins (2000:244): “The LDD for a dependent verb (Vd) or dependent preposition (Pd) consists of all terminal and non-terminal nodes dominated by VP on the path from Vd to the preposition on which it depends for semantic and/or syntactic property assignments, or on the path from a verb to Pd.”

2 The ICE GB is a tagged, parsed and checked corpus comprising one million words of spoken and written British English from the 1990s.

3 Contrary to Hawkins (2000:236-7), we did not exclude data points based on some preconceived notion of verb transitivity. Hawkins categorized all verbs in his sample as either ‘transitive’ or ‘intransitive’ and included only intransitive verbs and passivized transitives. In our opinion such a categorization is rather simplistic and problematic in light of the fact that many English verbs occur with various subcategorization frames. The acuteness of the problem is already apparent in (i)—Hawkins’s prime example—as it is not possible to straightforwardly assign a transitivity value to count in this example.

4 We approximated word units by counting strings of characters that are surrounded by white spaces. We agree with one anonymous reviewer that this is not the best possible operationalization of words. Given the fact that it is notoriously difficult to define the notion of ‘word’ in English, the present operationalization recommends itself due to its simplicity.

5 Hawkins (2000) proposes a third domain, the so called lexical matching domain (LMD), which also concerns a lexico-semantic dependency between V and PP. His testing for LMD involves looking up the word in question in the American Heritage Dictionary to see if “the PP in a [V (. . .) PP] sequence provides semantic content that gives more or equal semantic specification to the semantic content of V, as defined explicitly in any one lexical entry for V in the American Heritage Dictionary” (Hawkins, 2000:249). We are skeptical about LMD: It seems to aim at the very same semantic interrelationship assessed already by the entailment tests. Therefore we doubt that it captures a dependency domain distinct from LDD. Hence, following Occam’s razor, we decided to not consider it.

6 The magnitude of intercoder disagreement is similar to that reported in a related study by Lohse, Hawkins and Wasow (Lohse et al., 2004), who report agreement in over 80% of all cases with their initial coding.

7 We would like to add, however, that using the other operationalization does not yield a qualitatively different assessment of the relationship between PCD- and LDD-minimization.

8 Note that the intercept, typically denoted by β_0 , has been removed from the model (cf. Levy, in progress: 141-142 for details).

9 We arrived at the value for classification accuracy of 74.6% by letting the model distribute the number of pure guesses evenly across both correct and false predictions.

10 Standard errors and confidence intervals of coefficients inform us about the (un-)certainty of the coefficient estimates (Gelman & Hill 2007: 40). The 95%-confidence interval for LDD defines a range of 0.53-0.77, which means that there is a 95% probability that the true coefficient in the population falls in between this range. The higher standard error and greater range of the confidence interval for LDD as compared to PCD means that we can be less certain of its true value compared to the coefficient of PCD. This difference can be explained by the fact that LDD does not apply in all cases in the sample. The calculation of the LDD-coefficient is thus based on fewer data points, which renders it less certain. Note however that the 95%-confidence interval for the LDD coefficient estimate

ranges from 0.53-0.77, which crucially does not overlap with the 95%-confidence interval of PDD (0.33-0.45). We may thus still conclude that LDD's effect size is significantly greater.

11 We thank an anonymous reviewer for pointing this out to us

12 Note that a change can be observed in the standard errors and confidence intervals compared to the first model: While the standard error for LDD remains unchanged, the value for PCD has risen, along with an increased confidence interval. This can be explained by the fact that the calculation of the coefficients is based on a smaller sample, which reduces the certainty of the coefficients. The coefficient for LDD is unaffected, as in reducing the sample, we only omitted those cases in which LDD does not apply, which did not help in increasing the certainty of the LDD coefficient in the first model.

13 The standardization (or scaling) was done by dividing the original values of the predictors by two standard deviations, as suggested by Gelman & Hill (2007:57). We did not center the predictors by subtracting the mean, as this creates nonsensical results when entering them into a model without intercept.

14 But cf. Givon (1983) principle of task urgency, which predicts the reverse ordering preference.

15 We thank an anonymous reviewer for pointing out that the correlation between length/type of a given referential expression and the activation state of its associated mental representation is not perfect: speakers may, for example, choose to refer to a fully active representation non-pronominally.

16 The assessment of the relative effect size is taken from a model that includes only the two relevant factors (cf. Table 2).

17 One reviewer pointed out that such a "on the fly" elaboration is possible for other constructions as well: for example, He sent [a letter] could be elaborated on the fly with the phrase [to the president]. We would like to argue that the semantics of send necessitate the entertainment of three conceptual representations corresponding to the three required arguments of a sending-event (SENDER, SENT OBJECT, RECIPIENT). A sentence that does not include the recipient argument will still evoke the SEND-frame and, hence, by necessity evoke a constitutive RECIPIENT concept. In consequence, there is a categorical difference in the planning of dative constructions and double PP constructions.

18 Our results furthermore relate to a finding in psycholinguistic experimenting, namely the notion of verb disposition in phrasal ordering. Stallings, MacDonald and O'Seaghdha (1998) find that ordering phenomena are influenced by the "syntactic experiences" (Stallings et al., 1998:396) of the individual verb which denotes its frequency of occurrence in particular syntactic contexts e.g. in the verb particle construction. They do however not provide an answer to the question of what motivates these occurrences. Our results suggest that the frequent occurrence in the verb-particle construction may well be motivated by lexical dependencies between verb and preposition which may thus be the driving force underlying the verb dispositions Stallings et al. (1998) identified.

19 Neither Kroch & Small (1978), nor Gries (2003) employ the term 'Lexical Dependency Domain'. However both studies measure the semantic dependency of verb and particle which corresponds to our operationalization of LDD.

20 We tested this by fitting regression models using data that can be derived from Figures 6 & 7 in Lohse et al. (2004: 258). Crucially, the interaction between the PCD minimization and medium is not statistically significant ($p > 0.125$).

21 We are grateful to an anonymous reviewer who also mentioned this possible interaction in German which further encouraged us to empirically pursue this question.

References

Ariel, Mira. (1990). *Assessing Noun Phrase antecedents*. London: Routledge.

- Ariel, Mira. (2001). Accessibility theory: An overview. In T. Sanders et al., (eds.) *Text representation: Linguistic & psycholinguistic aspects*. Amsterdam: Benjamins. 29-87.
- Baayen, Rolf H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press.
- Behaghel, Otto. (1932). *Deutsche Syntax. Eine geschichtliche Darstellung. Vol. IV. Wortstellung, Periodenbau*. Heidelberg: Winter.
- Benor, Sarah & Levy, Roger. (2006). The chicken or the egg? A probabilistic analysis of English binomials. *Language* 82(2):233–278.
- Bock, Kathryn & Cutting, J.Cooper. (1992). Regulating mental energy: Performance units in language production. *Journal of Memory & Language* 31:99-127.
- Bock, Kathryn & Levelt, Willem J.M. (1994). Language production: grammatical encoding. In Morton Ann Gernsbacher (ed.), *Handbook of Psycholinguistics*. San Diego, CA: Academic. 945-984.
- Boisson, Claude. (1981). Hiérarchie universelle des spécifications de temps, de lieu, et de manière. *Confluents* 7:69–124.
- Bresnan, Joan, Cueni, Anna, Nikitina, Tatiana & Baayen, Harald. (2007). Predicting the Dative Alternation. In G. Boume, I. Kraemer, & J. Zwarts (eds.), *Cognitive Foundations of Interpretation*. Amsterdam: Royal Netherlands Academy of Science. 69-94.
- Diessel, Holger. (2005). Competing motivations for the ordering of main and adverbial clauses. *Linguistics* 43: 449-470.
- Frege, Gottlob. (1948). Sense & reference. *The Philosophical Review* 573:209–230.
- Gibson, Edward. (1998). Linguistic complexity: Locality & syntactic dependencies. *Cognition* 68:1-76.
- Gelman, Andrew & Hill, Jennifer. (2007). *Data analysis using regression & multilevel/hierarchical models*. Cambridge: Cambridge University Press.
- Givón, Talmy. (1992). The grammar of referential coherence as mental processing instructions. *Linguistics* 30:5–55.
- Gundel, Jeanette, Nancy Hedberg and Ron Zacharski. (1993). Cognitive Status and the Form of Referring Expressions in Discourse. *Language* 69:274-307.
- Hawkins, John. (1994). *A performance theory of order & constituency*. Cambridge: Cambridge University Press.
- Hawkins, John. (2000). The relative order of prepositional phrases in English: Going beyond manner-place-time. *Language Variation and Change* 11: 231-266.
- Hawkins, John. (2004). *Efficiency & Complexity in Grammars*. Oxford: Oxford University

- Press.
- Hawkins, John. (2009). Language universals and the performance-grammar correspondence hypothesis in Christiansen, Morten H., Collins, Chris & Edelman, Shimon. (eds.). *Language universals*. New York: Oxford University Press.
- Jaeger, T. Florian and Tily, Harry. (2011). Language Processing Complexity and Communicative Efficiency. *WIRE: Cognitive Science* 2(3), 323-335.
- Lakoff, George & Johnson, Mark. (1980). *Metaphors We Live By*. Chicago: University of Chicago Press.
- Lambrecht, Knut. (1994). *Information structure & sentence form: Topic, focus & the mental representation of discourse referents*. Cambridge: Cambridge University Press.
- Levelt, Willem J. M. (1989). *Speaking: from intention to articulation*. Cambridge, MA: MIT Press.
- Levy, Roger. (in progress). *Probabilistic Models in the Study of Language*. Cambridge, MA: MIT Press. (January 27, 2012 - <http://idiom.ucsd.edu/~rlevy/textbook/text.html>)
- Lohmann, Arne. (2011). *Constituent order in coordinate constructions - a processing perspective*. Doctoral dissertation, University of Hamburg.
- Lohse, Barbara, Hawkins, John A. & Wasow, Thomas. (2004). Processing Domains in English Verb-Particle Constructions. *Language* 80(2):238-261.
- Marblestone, Karen L. 2007. *Semantic and syntactic effects on double prepositional phrase ordering across the lifespan*. Doctoral Dissertation. University of Southern California.
- Nelson, Gerald, Wallis, Sean & Aarts, Bas. (2002). *Exploring Natural Language: Working with the British Component of the International Corpus of English*. Amsterdam/Philadelphia: Benjamins.
- Newmeyer, Frederick J. (1998). *Language Form & Language Function*. MIT Press.
- Prince, Ellen. (1981). Toward a taxonomy of given-new information. In Cole, Peter (ed.) *Radical Pragmatics*. NY: Academic Press. Pp. 223-56.
- Quirk, Randolph, Greenbaum, Sidney, Leech, Geoffrey & Svartvik, Jan. (1985). *A comprehensive grammar of the English language*. Harlow: Longman.
- R Development Core Team. (2011). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.
- Szmrecsanyi, Benedikt (2004). "On operationalizing syntactic complexity". In: Purnelle,

- Gérard, Cédric Fairon & Anne Dister (eds.), *Le poids des mots*. Proceedings of the 7th International Conference on Textual Data Statistical Analysis. Louvain-la-Neuve, March 10-12, 2004, Vol. 2. Louvain-la-Neuve, Presses universitaires de Louvain. 1032-1039.
- Stallings, Lynne, MacDonald, Maryellen, & O'Seaghdha, Padraig. (1998). Phrasal ordering constraints in sentence production: phrase length & verb disposition in heavy-NP shift. *Journal of Memory & Language*, 39:392-417.
- Wasow, Thomas. (2002). *Postverbal behavior*. Stanford: CSLI Publications.

List of Tables

Table 1: Comparison of data sets - Hawkins (2000) versus present study

	Hawkins's data set	Present data set
What?	500 page corpus of selected text from novels and some (popular) science	Exhaustive extraction from ICE-GB
Time	1949 - 1994	1990s
Variety	British and American	British
Medium	Written	Written and Spoken
Sample size	394	1256

Table 2: Coefficient estimates, Standard error estimates, 95% confidence intervals and p-values for predictors in the model

	Estimate	Std. Error	95% Conf. Interval	Pr(> z)
LDD MINIMIZATION	0.65	0.06	0.53-0.77	< 0.001
PCD MINIMIZATION	0.39	0.03	0.33-0.45	< 0.001

Table 3: Three predictor model – LDD, PCD, MPT (Standardized input variables)

	Estimate	Std. Error	95% Conf. Interval	Pr(> z)
LDD MINIMIZATION	3.19	0.29	2.62-3.76	< 0.001
PCD MINIMIZATION	2.43	0.21	2.02-2.84	< 0.001
MANNER > PLACE > TIME	1.15	0.14	0.88-1.42	< 0.001

Table 4: Four predictor model – LDD, PCD, MPT, INFORMATION STATUS (Standardized input variables)

	Estimate	Std. Error	95% Conf. Interval	Pr(> z)
LDD MINIMIZATION	3.33	0.30	2.74-3.92	< 0.001
PCD MINIMIZATION	2.21	0.21	1.80-2.62	< 0.001
MANNER > PLACE > TIME	1.16	0.15	0.87-1.45	< 0.001
DEFINITENESS	0.89	0.16	0.58-1.20	< 0.001

Table 5: Four predictor models – LDD, PCD, MPT, INFORMATION STATUS (INF) – across modalities (Standardized input variables)

	Spoken sample (N= 719)				Written sample (N = 537)			
	Estimate	Std. Error	95% Conf.Interval	Pr(> z)	Estimate	Std. Error	95% Conf.Interval	Pr(> z)
LDD	3.12	0.36	2.41-3.83	< 0.001	3.87	0.56	2.77-4.97	< 0.001
PCD	2.58	0.30	1.99-3.17	< 0.001	1.83	0.30	1.24-2.42	< 0.001
MPT	1.39	0.21	0.98-1.80	< 0.001	1.00	0.22	0.57-1.43	< 0.001
DEF	0.68	0.20	0.29-1.07	< 0.001	1.17	0.24	0.70-1.64	< 0.001

Figures: Descriptive legends/captions:

- Figure 1:** Differences in phrasal combination domain (PCD) length of alternative PP orders
- Figure 2:** Observed PCD (left) and LDD minimizations (in words). Aligned on the vertical axis are the 1256 [V PP PP] sequences under analysis (sorted by magnitude). Values on the horizontal axis indicate the extent to which the observed ordering can be considered more efficient than its alternative.
- Figure 3:** Predicted values of regression models (only cases are shown where predicted value is $\neq 0.5$): left: 'PCD & LDD'- model, middle: 'PCD-only'-model, right: 'LDD-only'- model. (For purposes of exposition, we added a small amount of noise to the value of each data point (using the jitter function in R), which is why some of the cases assume values smaller than 0 and larger than 1.)
- Figure 4:** Mean minimization advantage in cases of conflict of LDD and PCD (in number of words)

Artwork:

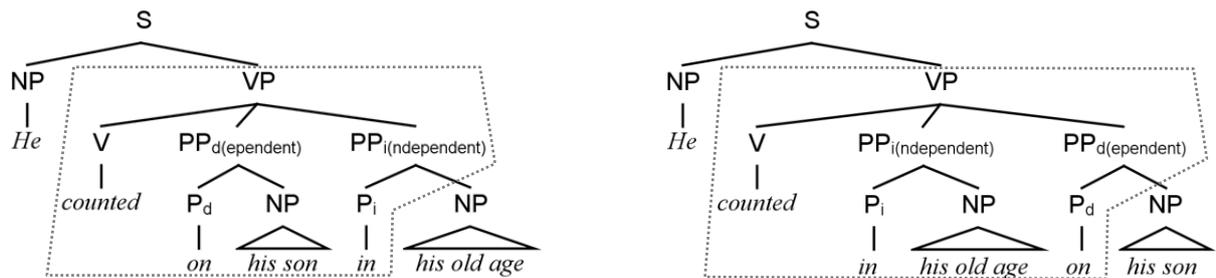


Figure 1:

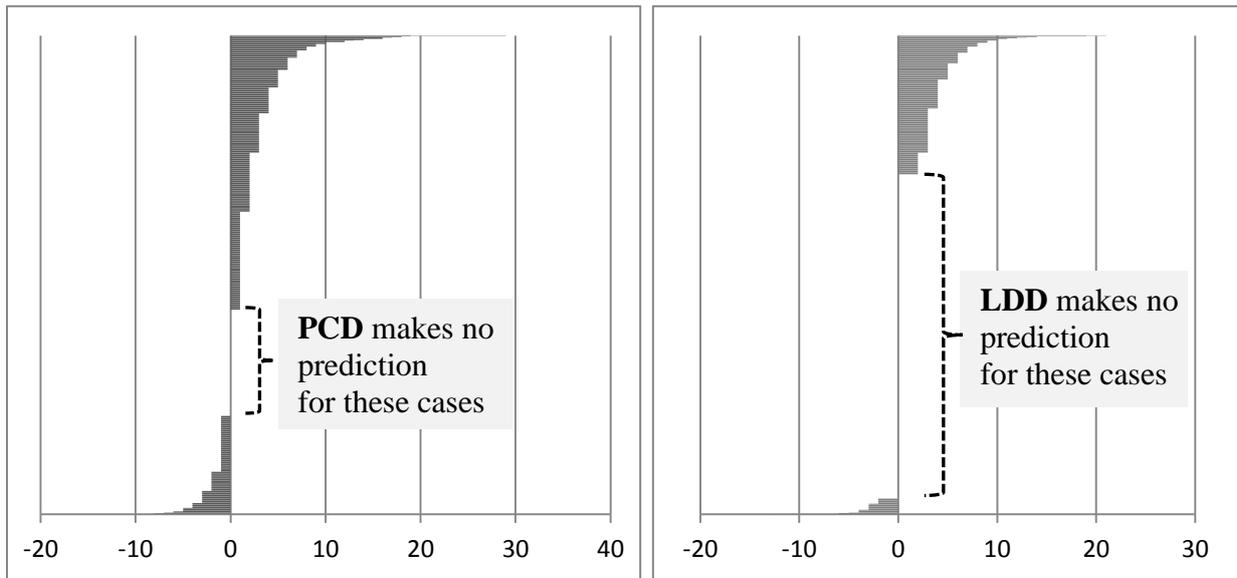


Figure 2:

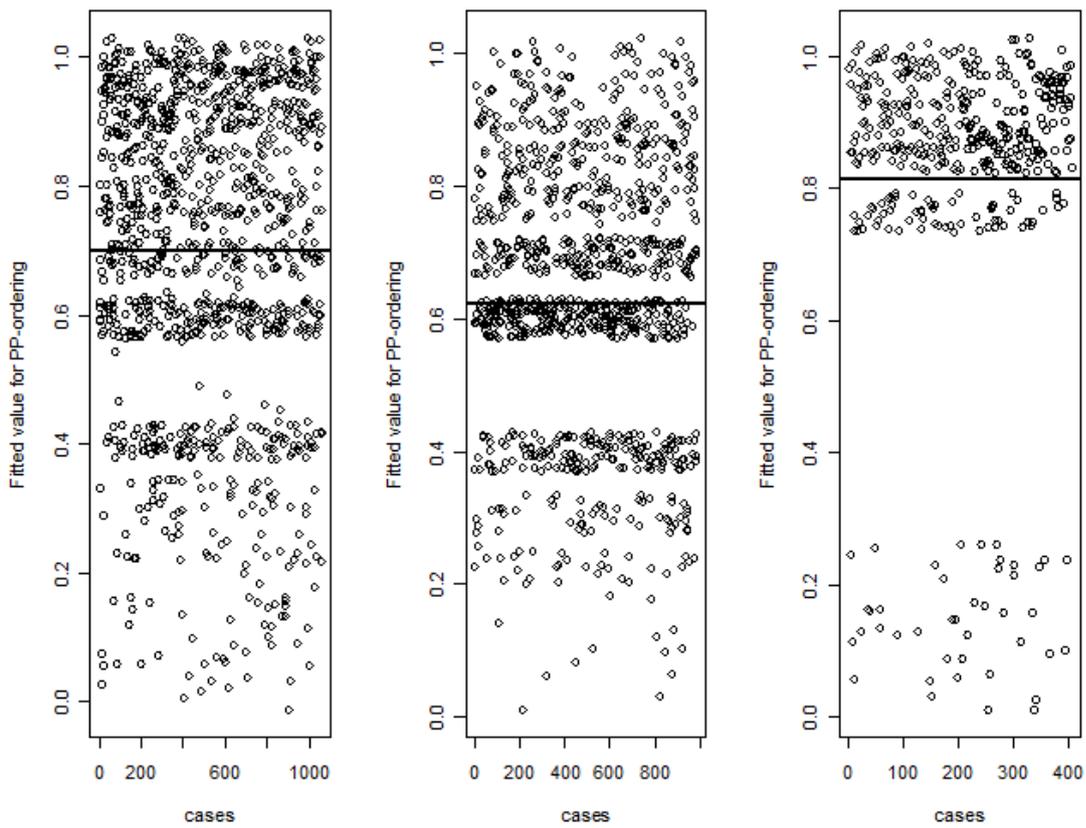


Figure 3:

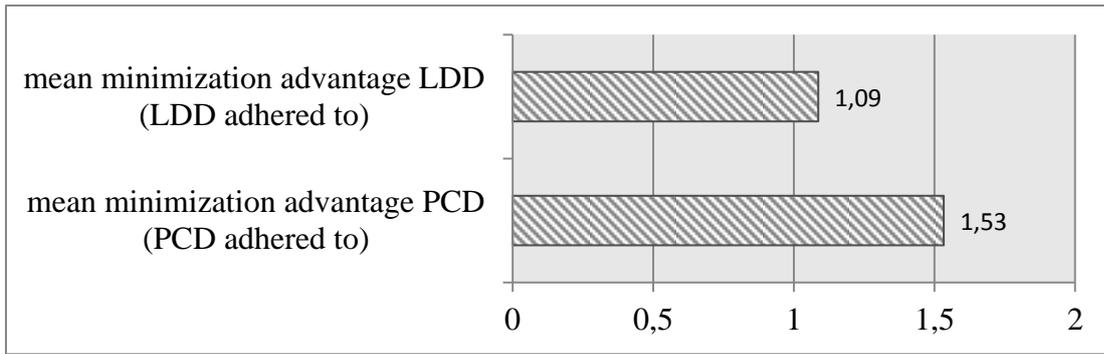


Figure 4: