Third Factors and the Performance Interface in Language Design

Andreas Trotzke, Markus Bader & Lyn Frazier

This paper shows that systematic properties of performance systems can play an important role within the biolinguistic perspective on language by providing third-factor explanations for crucial design features of human language. In particular, it is demonstrated that the performance interface in language design contributes to the biolinguistic research program in three ways: (i) it can provide additional support for current views on UG, as shown in the context of complex center-embedding; (ii) it can revise current conceptions of UG by relegating widely assumed grammatical constraints to properties of the performance systems, as pointed out in the context of linear ordering; (iii) it can contribute to explaining heretofore unexplained data that are disallowed by the grammar, but can be explained by systematic properties of the performance systems.

Keywords: center-embedding; ellipsis; linguistic performance; third factor; word order

1. Introduction

Asking why natural languages are built ('designed') the way they are by considering how systematic properties of the performance systems, the acquisition, production and comprehension systems, connect to the representation of grammars is anything but new. One prominent attempt in the pre-minimalist era to approach this issue is Berwick & Weinberg's (1984) influential book *The Grammatical Basis of Linguistic Performance*. Since they try to provide an "explanation for the form of certain axioms of the grammatical system in terms of extragrammatical principles" (Berwick & Weinberg 1984: 143), they deal with an explanatory relationship between human sentence parsing and linguistic con-

Andreas Trotzke thanks Josef Bayer, Anna Maria Di Sciullo, and John A. Hawkins for useful comments and discussion. He gratefully acknowledges financial support from the German Academic Exchange Service (DAAD) and from the German Excellence Initiative (Young Scholar Fund, University of Konstanz, project no. 65411). Markus Bader thanks Jana Häussler for helpful discussions. His work was supported by a Heisenberg scholarship from the German Research Foundation (DFG). Lyn Frazier is very grateful to Greg Carlson, Chuck Clifton, Pat Keating, Jason Merchant, Chris Potts, and Ivan Sag for discussion of the ideas presented here. Her work was supported by R01HD18708 to the University of Massachusetts, Amherst. Finally, we would like to thank two anonymous reviewers for their helpful comments.

ISSN 1450–3417

straints imposed by the grammar that this paper is concerned with, too. However, we want to depart from their approach in central respects, following Fodor (1985). Berwick & Weinberg's work is part of the long attempt to directly translate formal models of linguistic competence, using modern computer-science algorithms, into a performance model (cf. Miller & Chomsky 1963). Recently, this tradition has been continued by Di Sciullo (2000), Phillips (2004), Berwick (2011), and many others. Although both the theory of grammar and the idea of what counts as an explanation of language design have shifted significantly since Berwick & Weinberg's work, we want to point out in this section that Fodor's (1985) objections to Berwick & Weinberg's theory are in the spirit of current methodology in biolinguistics. To see why, let us briefly recall their approach.

Based on a set of parsing problems created by the dependencies between fillers and gaps, Berwick & Weinberg claim that Universal Grammar (UG) should include a locality constraint whose functional source is the parsing mechanism. In particular, they argue in favor of intermediate traces (i.e. Subjacency conditions) in terms of keeping the left-context of a structural configuration active at every derivational cycle. They claim that precisely such parsing problems gave rise to the evolution of this linguistic constraint, which is part of UG and hence must have evolved, according to them, by selection pressures. Note that their parser is intended to be a natural implementation of the rules and representations of the grammar. However, the grammar, at this time, was characterized by Government/Binding (GB)-theory. Due to its highly modular structure, GB-theory contains a rich UG, with several constraints and principles operating on different levels of representation. Thus, Berwick & Weinberg's move to assign a locality constraint to UG by demonstrating a corresponding constraint in the parser poses no problem for the general character of the assumed theory of grammar.

Fodor (1985) raised several objections to Berwick & Weinberg (1984). Her main points were: (i) Berwick & Weinberg assume a constraint in the parser to provide a functional explanation for the constraint in the grammar and hence do not take into account that the constraint might be motivated independently. (ii) Their constraint cannot account for all possible constructions. On evolutionary grounds, then, Fodor argues that their functional account of this aspect of language design is not convincing because "the fit that can be established between the linguistic phenomenon to be explained and the functional problem that is identified as its source […] is not close" (Fodor 1985: 20). Based on these criticisms, she asks what could count as a performance explanation of this aspect of language design and provides some useful distinctions that may help in approaching this issue.

First, she states that the weakest claim would be that a parser P can incorporate some constraint C, that is, P can obey C (Type 1). Second, according to Fodor, a stronger claim would be that P benefits from C, that is, P can not only incorporate C but it also operates more efficiently if it does incorporate C than if it does not. If other kinds of parsers could be shown not to benefit from C, then P could offer a stronger functional motivation for C than these other parsers do (Type 2 = Berwick & Weinberg). The third explanatory option is that P must incorporate C, that is, C is entailed by the defining properties of P (Type 3). Accord-

ing to Fodor, an adequate statement about language design only follows when there are reasons to believe that human sentence parsing has exactly these properties. That is, the statement must be of Type 3. In other words, "we would recognize explanatoriness to the extent that the relevant properties of P are independently motivated in some fashion" (Fodor 1985: 5). By postulating that an adequate ('deep') explanation should take into account independent motivations and by assuming that a constraint assigned to UG should not only serve as a solution that solves part of an evolutionary problem but instead serves as an 'optimal solution,' Fodor, in many respects, anticipates the biolinguistic perspective on language design.

According to Chomsky (2005: 6), three factors have to be explored when one aims at an explanation of language design:

(i) The genetic endowment (= UG)
(ii) Linguistic experience
(iii) Principles not specific to the language faculty

According to this distinction, factor (i) contains the components of the faculty of language that are both language- and species-specific; thereby it roughly corresponds to what Hauser *et al*. (2002) call the 'faculty of language — narrow sense' (FLN). Factor (ii) refers to the linguistic input, which is the source of variation within this parcellation of language design. Factor (iii) contains principles of biological and computational systems not specific to the faculty of language. According to Chomsky (2005: 6), these are "(a) principles of data analysis that might be used in language acquisition and other domains; (b) principles of structural architecture and developmental constraints […] including principles of efficient computation," and, as Chomsky (2007: 3, fn. 4) adds, "properties of the human brain that determine what cognitive systems can exist."

Given this factor distinction, Berwick & Weinberg's (1984) claim that UG includes the locality constraint is unwanted. Within the biolinguistic perspective on language design, a third-factor explanation offers a benchmark for what counts as a genuine explanation and thereby corresponds to Fodor's (1985: 30) "full-blooded Type 3 explanation." In other words, as pointed out also by Richards (2008: 134), biolinguistics is characterized by a trend away from factor (i), that is, UG must be small and simple, on evolutionary grounds. The faculty of language, according to Chomsky, arose too recently for there to have been enough time (in evolutionary terms) for the development of a rich UG containing several language-specific principles, constraints, etc. Accordingly, as Chomsky (2007) argues, the more we can ascribe to third factors and the less to UG, the more tractable the issue of language evolution becomes. Given this shift in perspective, it is reasonable to assume that UG only contains properties such as recursive Merge, binary branching structure, and the valued-unvalued feature distinction. All other universal properties might follow from the interaction between UG and principles of extralinguistic components that belong to factor (iii). These principles, by definition, do not depend on UG and are independently motivated.

Note that these principles, unlike the principles of UG in the GB-theory/Principles and Parameters (P&P)-theory, are presumably invariant be-

cause, according to Boeckx (2011: 210), "[t]here is simply no way for principles of efficient computation to be parametrized […], it strikes me as implausible to entertain the possibility that a principle like 'Shortest Move' could be active in some languages, but not in others. Put differently, […] there can be no parameters within the statements of the general principles that shape natural language syntax." Of course, that does not mean that third factors cannot contribute to explaining parameters. Consider, for instance, the head-complement parameter (cf. Holmberg 2010, Richards 2008): UG allows that X can merge with an XP, but UG does not fix their linear order, that is, X can either precede or follow the complement XP. When Merge is maximally unspecified — when it is symmetrical (cf. Chomsky's 2000a set-Merge) — it cannot specify any particular ordering. However, the physics of speech, that is, the nature of the articulatory and perceptual apparatus require one of the two logical orders, since pronouncing or perceiving the head and the complement simultaneously is impossible. Thus, the head-complement parameter, according to this approach, is a third-factor effect.

As the above reasoning concerning the head-complement parameter has shown, third-factor explanations sometimes refer to rather abstract design features. In this paper, we will show how concrete properties of the performance systems can contribute to explaining the design features of language. In particular, we will be concerned with non-trivial systematic processing phenomena and argue that they are part of an implicit knowledge of human language performance systems, and thus, they lend themselves to third-factor explanations of the design features of human language.

In section 2, we will show how performance data support current conceptions of UG. In particular, we will investigate recent claims that the grammar includes special constraints on center-embedding and ask whether their properties follow from independently established constraints on sentence processing. In section 3, we will discuss how properties of the performance systems can revise current approaches to UG. Specifically, we will discuss a third-factor explanation of a constraint on linear ordering that is widely assumed to be part of UG. After having shown that the assumption of independently motivated performance systems is methodologically beneficial, given the biolinguistic framework to reduce UG, in section 4 we will present empirical evidence that such systems are needed anyway in order to account for data in the context of acceptable ungrammaticality. Section 5 summarizes the main results and concludes.

2. The Performance Interface and Complex Center-Embedding

Although Miller & Chomsky (1963), as mentioned above, argued in favor of a general transparency between (theories of) grammar and (theories of) linguistic performance, they also provide arguments for keeping grammar and performance strictly separate. In other words, they claimed that some design features of human language, like recursive embedding, cannot be explained by any of the three types of explanations pointed out by Fodor (1985) because these design features do not show any relationship between a grammatical constraint and properties of the performance systems.

One prominent argument, initially developed by Chomsky & Miller (1963), in favor of drawing a sharp distinction between processes on the level of performance and formal mechanisms on the level of grammar rests on the property of recursive self-embedding and the observation that multiple center-embedding leads to structures that can no longer be produced or comprehended under normal on-line conditions, as illustrated by (1):

(1) The rat the cat the dog chased killed ate the malt.
(Chomsky & Miller 1963: 286)

The fact that such sentences are quite incomprehensible has no bearing on the possibility of generating them on the level of grammar because, as Chomsky (1963: 327) points out by means of an analogy, "the inability of a person to multiply 18,674 times 26,521 in his head is no indication that he has failed to grasp the rules of multiplication." The overall conclusion, then, is that such structures are excluded by performance factors that limit the realization of our grammatical competence. In particular, Miller & Chomsky (1963) showed that, given certain reasonable assumptions about language processing, this construction, unlike other relative clause configurations, creates a major strain on working memory. They therefore concluded that it is a performance violation, not a competence violation.

This conclusion was disputed by Reich (1969), who claimed that a sentence such as (2) is not just unacceptable — that is, beyond the processing capabilities of the human sentence processor — but downright ungrammatical, where the term 'ungrammatical' is understood in the classical way of meaning 'not within the set of sentences derivable by the mental competence grammar.'

(2) The rat that the cat that the dog worried killed ate the malt.
(Reich 1969: 831)

The dispute about sentences as in (1) and (2) points to a deeper problem. Even if we know that a certain sentence is ungrammatical, we cannot know *a priori* what to blame for the unacceptability: The performance mechanisms, which do not have the capacity required for processing the sentence, or the competence grammar, which does not generate the sentence? Chomsky & Miller (1963) opted for the first alternative and attributed the unacceptability of sentences with double center-embedding to limitations on working memory. Reich (1969) took the opposite way. He proposed a finite-state grammar capable of generating sentences with degree-1 center-embedding but not center-embeddings of degree 2 or higher (for related ideas in a connectionist setting, cf. Christiansen & Chater 1999).

Data from language processing — either data from psycholinguistic experiments or corpus data — are no different in this regard. They cannot show whether a sentence is unacceptable due to performance limitations or because it is outside the scope of the grammar. Such data can nevertheless be quite helpful in cases where the source of unacceptability is under dispute. In particular, performance data can provide evidence on whether the limited use made of certain syntactic structures can plausibly be attributed to performance factors, or

whether grammatical constraints are necessary for this purpose.

With regard to multiple center-embedding, Roeck *et al*. (1982) argued that corpus data provide clear evidence against Reich's (1969) claim that the competence grammar cannot generate more than one level of center-embedding. They presented several corpus examples of doubly center-embedded clauses and thus showed that such sentences are produced from time to time in actual language use. Recently, such empirical approaches to multiple center-embedding have regained attention in the context of Hauser *et al*.'s (2002) claim that recursive syntactic embedding is the only human- and language-specific component of the human language faculty. In what follows, we will show, based on our own empirical data, that such approaches do not provide evidence against recent biolinguistic claims that infinite recursive nesting is a central part of UG (cf. Sauerland & Trotzke 2011 for a recent collection of papers).

In a recent volume on recursion, Karlsson (2010: 55) claims that "[m]ultiple nesting cannot […] reasonably be considered a central design feature of language, as claimed by Hauser *et al*. (2002)." His claim is based on a corpus study of multiple center-embedded clauses, where he analyzed 132 doubly center-embedded clauses from seven European languages (cf. Karlsson 2007). Given these data, he proposed specific grammatical constraints on multiple center-embedding and claimed that they reveal that "more aspects of competence (i.e. grammar) are involved in multiple center-embedding than Chomsky and his followers have been assuming" (Karlsson 2007: 385). Thus, by formulating grammatical constraints, Karlsson objects to the view that any constraint on center-embedding must solely follow from the performance systems. Like Berwick & Weinberg (1984), he assumes that properties of the performance systems provide a functional explanation for the constraints in the grammar, since he claims that "the constraints are epiphenomenal consequences of more basic cognitive properties, especially short-term memory limitations" (Karlsson 2007: 385). Thus, according to Fodor's (1985) typology, he offers a 'type 2 explanation' and does not take into account that the constraints might be motivated independently and do not exist in the grammar ('type 3 explanation').

In this section, we will argue that the properties of these grammatical constraints follow from independently motivated constraints on sentence processing and that they are therefore superfluous. To show that we are dealing with systematic properties of the performance systems, we will present data from both production and comprehension, our hypothesis being that it is precisely the collusion of speakers and hearers that yields such systematic properties. In particular, we will discuss corpus data and results from associated acceptability experiments that have investigated doubly center-embedded relative clauses in German. The major question addressed by the corpus data is whether doubly center-embedded relative clauses have special properties that call for specific grammatical constraints on multiple center-embedding (e.g., Karlsson 2007), or whether their properties follow from independently established constraints on sentence processing (e.g., Gibson 2000).

In order to address these questions, Bader (2012) analyzed the deWaC corpus (cf. Baroni *et al*. 2009) for the occurrence of multiply center-embedded relative clauses (RCs) in German. This study goes beyond Karlsson (2007) not only

by looking at a larger number of examples but also by taking into account structural variants involving extraposition. This makes it possible for the first time to determine empirically whether multiply center-embedded RCs have unique properties requiring specific grammatical constraints.

Four sentence structures were investigated. Sentence (3) is an original corpus example with a doubly center-embedded RC (RC-low within RC-high, intraposed relative clauses).

(3) *German*
RC-low within RC-high, intraposed relative clauses
Internationale Studien belegen, dass Medizinstudenten, denen
International studies prove that medical-students to-whom
identische Krankenakten, die nur in Bezug auf Alter und Geschlecht
identical patient's files that only in relation to age and gender
variieren, vorgelegt werden, unterschiedlich entscheiden.
vary presented are unequally decide
'International studies show that medical students decide unequally if they are confronted with patient's files that only differ with respect to age and gender.'

A search of the deWaC corpus with its 1,278,177,539 tokens of text revealed 351 instances of doubly center-embedded RCs as in (3). In accordance with Karlsson (2007), sentences with more deeply embedded RCs were practically absent. Thus, doubly center-embedded RCs do occur, but they are rare.

However, doubly center-embedded RCs are not only special by involving two degrees of clausal center-embedding, they are also special on several other measures. For example, a doubly center-embedded RC disrupts the dependency between its head noun (*Medizinstudenten* 'students of medicine' in (3)) and the corresponding clause-final verb (*entscheiden* 'decide') much more severely than a simple RC not containing a second RC. Since the disruption of dependencies is a major source of sentence complexity — as captured by the notion of structural integration cost in the *Dependency Locality Theory* (DLT) of Gibson (2000) — the rareness of doubly center-embedded RCs cannot be attributed to the degree of center-embedding as such without further justification. In order to determine whether doubly center-embedded RCs have special properties due to their high degree of center-embedding, it is crucial to compare them to other RC structures that are matched as closely as possible but at the same time involve no center-embedding or only a single degree of center-embedding.

Such a comparison was made possible in Bader (2012) by analyzing three further types of complex RCs which differ from doubly center-embedded RCs only with regard to the position of the RCs. This was achieved by applying extraposition to RC-high, RC-low, or both. Schematic tree structures for the four sentence types that were thus investigated in the corpus study are given in Figure 1. Original corpus examples are shown in (3) above and (4)–(6) below.

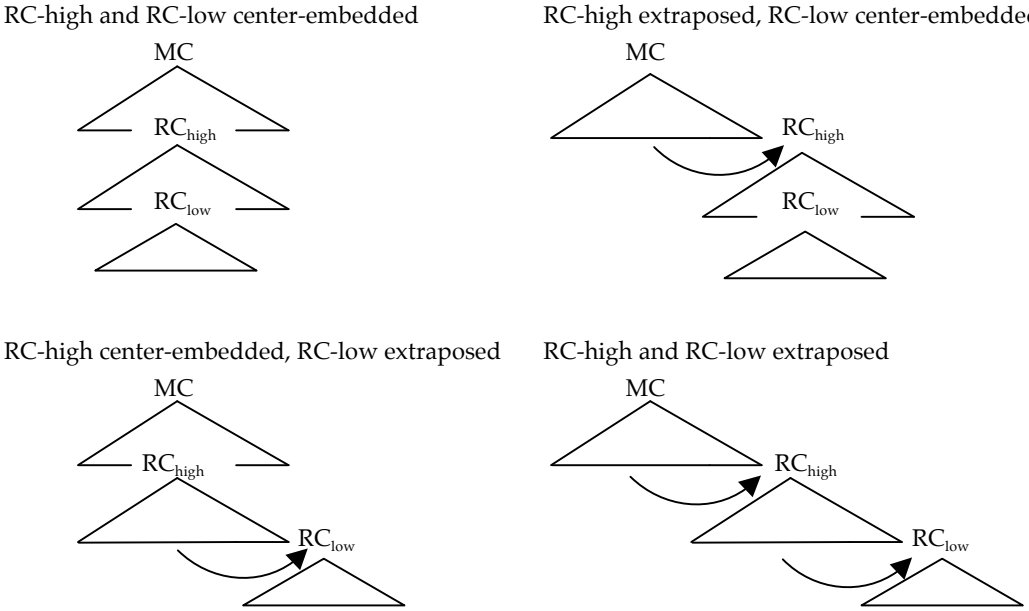

Figure 1: Schematic tree structures for the four sentence types investigated in the corpus study (MC = matrix clause; RC = relative clause)

(4) *German*

RC-low behind RC-high, intraposed relative clauses

Ihr werdet bemerkt haben, dass Völker, die in Ländern leben, in
You will realized have that peoples who in countries live in
denen ein besseres Verständnis von Leben und Tod herrscht, den
which a better understanding of life and death governs the
Weggang eines geliebten Menschen oftmals zelebrieren.
departure of-one loved person often celebrate.

'You will have realized that peoples who live in countries where there exists a better understanding of life and death often celebrate the passing away of a beloved person.'

(5) *German*

RC-low within RC-high, extraposed relative clauses

Hector Sanchez ist davon überzeugt, daß der Geist von Tom Donovan
Hector Sanchez is by-that convinced that the ghost of Tom Donovan
zurückgekehrt ist, der vor zehn Jahren während einer Explosion, die
returned is who before ten years during an explosion which
Annie, Dan und er versehentlich ausgelöst hatten, ums Leben kam.
Annie Dan and he accidentally caused had over life came

'Hector Sanchez is convinced that the ghost of Tom Donovan has returned, who was killed in an explosion that was accidentally caused by Annie, Dan and himself, has returned.'

(6) *German*

RC-low behind RC-high, extraposed relative clauses

Grundlegend ist hier die Annahme, dass es keine allgemeingültige
Essentially is here the assumption that it no universal
Definition gibt, die die Lieder beschreibt, die sich für den Einsatz
definition gives that the songs describes that itself for the use
im Unterricht eignen.
in class suit.

'Basically here is the assumption that there is no generally accepted definition describing the songs that may suit for the usage in class.'

In (4), RC-low has been extraposed behind RC-high, but the two relative clauses are still center-embedded within the matrix clause (RC-low behind RC-high, intraposed relative clauses). In (5), RC-low is again center-embedded within RC-high, but the relative clauses have been extraposed behind the matrix clause (RC-low within RC-high, extraposed relative clauses). In (6), RC-low has been extraposed behind RC-high and the relative clauses as a whole have been extraposed (RC-low behind RC-high, extraposed relative clauses).

The existence of doubly center-embedded RCs raises two major questions. First, why do doubly center-embedded RCs occur so rarely, or, put more generally, what factors affect the frequency with which they are produced? Second, why do doubly center-embedded RCs occur at all, that is, why are they not avoided completely by means of extraposition? If it is not the degree of center-embedding as such, but the processing cost induced by clausal embedding in general, then these two questions should find answers that are not specifically tailored to the case of double center-embedding. Instead, the answers should be general enough to also cover the RC structures in (4)–(6).

We begin with the first question: Why do doubly center-embedded RCs occur so rarely? If performance constraints are responsible for this, and not grammatical constraints on multiple center-embedding, then sentences with intraposed complex relative clauses should be rare in general because they introduce a lengthy dependency between the antecedent NP of RC-high and the clause-final verb (e.g., Gibson 2000). This should be true whether RC-low occurs within RC-high (degree of center-embedding = 2) or behind RC-high (degree of center-embedding = 1). In accordance with this prediction, the corpus study revealed that doubly center-embedded relative clauses as well as intraposed relative clauses with RC-low behind RC-high ((3) and (4)) are rare in comparison to similar sentences with the relative clauses extraposed ((5) and (6)).

As far as the particular constraints proposed in Karlsson (2007) were found to hold, it turned out that they reflect more general properties of complex RCs, properties that are not specific to doubly center-embedded RCs. As a case at hand, consider the NO-MULTIPLE-OBJECT-RELATIVIZATION constraint which is given in (7) (from Karlsson 2007: 383):

(7) **O–O constraint*

Direct objects must not be multiply relativized in C^2s.

Among the doubly center-embedded RCs analyzed in Bader (2012), there were only approximately 4% in which both the relative pronoun of the higher RC and the relative pronoun of the lower RC were objects. The *O–O constraint thus seems to hold, not as an absolute constraint but as a very strong preference.

However, a closer analysis revealed that the *O–O constraint is just a descriptive generalization that applies not only to doubly center-embedded RCs but to the other types of complex RCs as well. For sentences as in (4)–(6), the rate of O–O RCs was also about 4% or even less. Furthermore, the rareness of complex RCs in which both relative pronouns are objects could be shown to follow from the rareness of object relativization in general. This is shown in Table 1 for the case of doubly center-embedded RCs.

	p(rel-pro/high) *p(rel-pro/low)	Predicted proportion	Observed proportion	Predicted frequency	Observed frequency
Subject-Subject	0.835*0.785	0.66	0.66	155.4	157
Subject-Object	0.835*0.215	0.18	0.17	42.6	41
Object-Subject	0.165*0.785	0.13	0.12	30.6	29
Object-Object	0.165*0.215	0.04	0.04	8.4	10

Table 1

The row labeled "Subject-Subject" shows the relevant data for RCs in which both the relative pronoun of the higher RC and the relative pronoun of the lower RC are subjects. Overall, the probability (proportion) of subjects in higher RCs was 0.835 and the probability of subjects in lower RCs was 0.785. If these probabilities were independent of each other, then the joint probability of both relative pronouns being a subject is the product of the two individual probabilities. This joint probability is shown in the column labeled "predicted proportion." As a comparison with the observed proportions in the next columns shows, the predicted proportions and the observed proportions are quite close together, as also shown by the predicted and observed frequencies in the last two columns. For the remaining three combinations of subject and object relative pronoun, the same considerations apply.

What these considerations show is that there is no need to invoke a constraint like the *O–O constraint in order to account for the low proportion of object-object RCs. Instead, this low proportion follows from the low overall proportions of object RCs. Furthermore, since a low proportion of object-object RCs was not only observed for doubly center-embedded RCs, and calculations similar to those shown in Table 1 lead to the same results for the other RC types, we can conclude that we are dealing with a more general phenomenon here which is not related to the degree of center-embedding.

We turn now to the second question: Why are doubly center-embedded RCs not avoided completely by means of extraposition? The null hypothesis is that the decision of whether or not to extrapose a doubly center-embedded relative clause should be governed by the same factors that are also at work for relative clauses with a single degree of center-embedding. This hypothesis was also

confirmed. In accordance with prior findings for simple RCs (e.g., Hawkins 1994), the main determinant of extraposition is the amount of material to be crossed by extraposition. If extraposition is only across the clause-final verb, it is almost obligatory. If extraposition is across some non-verbal material in addition to the verb, it becomes highly unlikely. Importantly, this is true both when RC-low occurs within RC-high and when RC-low occurs behind RC-high, as confirmed by a logistic regression model.

Overall, the present corpus results argue that constraints on multiple center-embedding follow from processing limitations and that accordingly grammatical constraints on multiple center-embedding are not needed. In order to corroborate this claim, Bader (2012) presents an experiment that required participants to judge the grammaticality of all four major structures investigated in the current corpus study (see (3)–(6)).[1] The experiment used speeded grammaticality judgments, a method which has been used before both within psycholinguistics (e.g., Warner & Glass 1987) and within experimental syntax (Bader & Schmid 2009). In experiments using this procedure, participants have to quickly judge sentences as either grammatical or ungrammatical. Sentences are presented word-by-word on a computer screen with a presentation rate that leaves enough time for fully comprehending sentences but no time for deliberate reasoning. In the present context, this method is particularly appropriate because, as amply documented in Karlsson (2007: 379–380), the linguistic literature abounds with claims to the effect that sentences containing multiply center-embedded relative clauses are ungrammatical. As already pointed out by Karlsson, the finding of such sentences in authentic texts provides strong reasons to reject these claims. However, given their absence from spoken language and their rareness in written language, it cannot be excluded that such sentences are grammatically degraded, even if they are not outright ungrammatical.

The results of Bader's experiment can be summarized as follows: (i) Sentences with extraposed relative clauses ((5) and (6)) were judged as grammatical most of the time, with no significant differences depending on whether RC-low was center-embedded within RC-high or extraposed behind RC-high. (ii) In comparison to sentences with extraposed relative clauses ((5) and (6)), sentences with center-embedded relative clauses ((3) and (4)) received lower percentages of grammatical judgments, whether RC-low occurred within or behind RC-high. This effect was highly significant, but its magnitude was quite moderate, amounting to a difference of about 9%. Thus, even sentences with doubly center-embedded relative clauses were judged as grammatical in almost three quarters of all cases.

In this section, we have shown how the performance interface in language design provides support for Chomsky & Miller's (1963) claim that there are no specific grammatical constraints on multiple center-embedding. In particular, by presenting data from both language production and comprehension, we have demonstrated that there are systematic properties of the performance systems that constrain multiple center-embedding. Accordingly, following Chomsky's

[1] In addition to complete sentences, the experiment also included sentences in which the verb cluster of RC-high was missing. This issue of missing-VPs will be discussed in section 4.1.

(2005) three-factor parcellation, constraints on multiple center-embedding can be accounted for by a third-factor explanation and do not require us to complicate our theory of grammar (UG). Thus, in contrast to Karlsson's (2007, 2010) 'type 2 explanation,' we provided a 'type 3 explanation,' which refers to the explanatory power of the independently motivated systematicity of the performance systems.[2]

After having shown that performance studies can serve to support common views of UG, let us now look at one linguistic constraint that is widely assumed to be part of UG, but that can possibly be relegated to third-factor principles by taking into account the performance interface.

3. The Performance Interface and Linear Ordering: FOFC

In this section, we will look at a linguistic constraint that is widely assumed to be part of UG. Recently, however, some scholars have attempted to relegate this constraint to third-factor principles by referring to the performance systems, that is, to efficient processing.

Based on the fact that the word order V–O–Aux does not exist either synchronically or diachronically in Germanic, while all other orders are attested, Holmberg (2000: 124) formulated the following generalization that predicts that head-final phrases can occur embedded in head-initial phrases, but head-initial phrases cannot occur embedded in head-final phrases:

(8) *The Final-Over-Final Constraint (FOFC)*

If α is a head-initial phrase and β is a phrase immediately dominating α, then β must be head-initial. If α is a head-final phrase, and β is a phrase immediately dominating α, then β can be head-initial or head-final.

Thus, the FOFC states that head-finality must be lower in the structure than head-initiality. The generalization can be formally stated as follows, where αP is the complement of β and ΥP is the complement of α:

(9) a. $[_{\beta P}\ \beta\ [_{\alpha P}\ \alpha\ \Upsilon P]\ ]$ harmonic order
b. $[_{\beta P}\ [_{\alpha P}\ \Upsilon P\ \alpha]\ \beta\ ]$ harmonic order
c. $[_{\beta P}\ \beta\ [_{\alpha P}\ \Upsilon P\ \alpha]\ ]$ disharmonic order
d. $*[_{\beta P}\ [_{\alpha P}\ \alpha\ \Upsilon P\ ]\ \beta\ ]$ disharmonic order & violating FOFC

We will abstract away from exceptions that are discussed in the literature and that have yielded refinements of FOFC (cf. Biberauer *et al*. 2007 *et seq.*). Because

[2] In addition to constraints on sentence processing pointed out in this section, other factors, such as alignment constraints between syntax and prosody (cf. Fodor & Nickels 2011), might also play a crucial role in explaining the limitations of multiple center-embedding. However, it is beyond the scope of the present paper to decide if such constraints could be motivated on performance theoretic grounds ('type 3 explanation'), as suggested by an anonymous reviewer, or if these alignment constraints are an integral part of the grammar that entails advantages for parsing the structures ('type 2 explanation').

this generalization is widely assumed to hold (though see Hawkins to appear), we assume that there is something universal to it. Moreover, it has been claimed that there are cases where FOFC holds without exception. The most prominent case is the fact about VO-languages that they do not permit sentence-final complementizers (cf., e.g., Hawkins 1990). Referring to our formal statements above, this fact can be formulated as follows:

(10)	a.	V–O & Comp-TP	harmonic order	(= 9a)
	b.	O–V & TP-Comp	harmonic order	(= 9b)
	c.	O–V & Comp-TP	disharmonic order	(= 9c)
	d.	*V–O & TP-Comp	disharmonic order & violating FOFC	(= 9d)

Having introduced a constraint that is assumed to be part of UG, let us now turn to an alternative explanation in terms of properties of the performance systems.

Recently, Walkden (2009), Biberauer *et al*. (2010), and Sheehan (2010, to appear) have pointed out that Hawkins' (1994, 2004) Performance-Grammar Correspondence Hypothesis provides a potential processing account of FOFC. Hawkins' theory of early immediate constituents provides an alternative to the formal accounts that claim that FOFC is part of UG (for a more detailed description of what follows, cf. Sheehan to appear: 13–19).[3] In particular, the following performance-based efficiency principle correctly predicts that both (9a) and (9b) are strongly preferred (cf. Hawkins 1994: 58–59; 77):

(11) a. *Early Immediate Constituents (EIC)*

The human parser prefers linear orders that maximize the IC-to-non-IC ratios of constituent recognition domains.

b. *Constituent Recognition Domain (CRD)*

The CRD for a phrasal mother node M consists of the set of terminal and non-terminal nodes that must be parsed in order to recognize M and all ICs of M, proceeding from the terminal node in the parse string that constructs the first IC on the left, to the terminal node that constructs the last IC on the right, and including all intervening terminal nodes and the non-terminal nodes that they construct.

The EIC is a local complexity metric that predicts linear orderings. Given (11a) and (11b), the IC-to-word [= non-IC] ratio can be calculated as in (12).

[3] Of course, in performance-oriented linguistics, Hawkins' locality theory is controversial and it faces the same fundamental issues as, for instance, Gibson's (2000) theory sketched in section 2 (cf. Konieczny 2000 for a prominent critique). However, even alternative theories operating with constrained activation rather than with locality-driven complexity metrics point out, with reference to locality theories like Hawkins', that "it is clear that locality plays a critical role in sentence comprehension" (Vasishth & Lewis 2006: 788). Given the programmatic nature of our paper, we will not be concerned with (nor will we subscribe to) all the details of Hawkins' theory. However, we assume that Hawkins' approach can be regarded as an influential theory of the performance systems that should be taken into account when dealing with third-factor explanations.

(12) IC-to-word ratio = $\frac{\text{Number of ICs in domain}}{\text{Number of words in domain}}$

Applying this metric to our cases in (9), harmonic ordering such as (9a) and (9b) is preferred (cf. Hawkins 1994: 96–97, Sheehan to appear: 16):

(13) a. $[_{VP}$ V $[_{PP}$ P NP]]IC-to-word ratio = 1/1, 2/2, => average: 100% (9a)
b. $[_{VP}$ $[_{PP}$ NP P] V]IC-to-word ratio = 1/1, 2/2, => average: 100% (9b)

As pointed out by Hawkins (1994: 96), both (13a) and (13b) have optimal IC-to-word-ratio ratios of 1/1 and 2/2 (average: 100%). More specifically, in (13a), V constructs the first IC (V), resulting in a ratio of 1/1. Since P occurs immediately to the right of V and constructs the second IC (PP), the adjacent constituents V and P both construct two ICs (V and PP), thus resulting in the second IC-to-word-ratio of 2/2. In sum, the number of words is equal to the number of ICs built at each structural level. The same holds for (13b), except that in these cases of head-final languages, a bottom-up parse takes place (for elaboration on this, cf. also Sheehan to appear: 16). Let us now look at the disharmonic constructions given in (9c) and (9d).

Hawkins (1994) discusses the following disharmonic structures, where NP complements of P are within the CRD of VP. Concerning these 'non-optimal CRDs,' Hawkins (1994: 82) calculates the IC-to-word ratios from left to right in order to make the appropriate discriminations among these configurations. He gives the following (Left-to-Right) IC-to-word ratios (cf. Hawkins 1994: 255, Sheehan to appear: 17):

(14) a. $[_{VP}$ V $[_{PP}$ $[_{NP}$ Det N] P]] IC-to-word ratio = 1/1, 2/4, => average: 75% (9c)
b. $[_{VP}$ $[_{PP}$ P $[_{NP}$ Det N] V]] IC-to-word ratio = 1/3, 2/4, => average: 42% (9d)

In (14a), as in (13a), the first word V constructs the first IC (V), resulting in an IC-to-word ratio of 1/1. The IC-to-word ratio of the second IC (PP), however, is 2/4, since PP is constructed by the fourth word (i.e. P) in the CRD. Taken together, (14a) has an average ratio of 75%. In contrast to (14a), which corresponds to the configuration (9c), (14b) corresponds to the FOFC-violating ordering (9d). Since the three words (P, Det, N) dominated by PP (first IC) fall within the CRD, they are counted both in construction of PP (ratio = 1/3) and, together with the word V, in construction of the two ICs VP and PP (ratio = 2/4). Taken together, (14b) has an average ratio of 42% (for more elaboration, cf., again, Sheehan to appear: 17).

Accordingly, the EIC correctly predicts that the FOFC-violating order (14b = 9d) is more difficult to process and thus dispreferred. More recently Hawkins (2004, to appear) reformulates this left-to-right calculation procedure in terms of a separate principle of Maximize On-line Processing (which penalizes (14b) for the processing delay in the long first IC, the PP, compared to the short V in (14a)) and so defines a similar discrimination between (14a) and (14b). EIC is mean-

while converted to a more general principle stating that all structural domains that need to be accessed in the processing of grammatical relations of combination and dependency are preferably minimal, i.e. Minimize Domains. These two principles predict that the harmonic orders, (9a) and (9b) above, will be much preferred relative to the two disharmonic ones, (9c) and (9d), with (9c), e.g., (14a), having some on-line processing advantages compared with (9d), e.g., (14b). Hawkins argues that the relative quantities of language-particular grammars that exemplify the different ordering possibilities for e.g. the head-complement parameter can be predicted from these processing efficiency calculations. In sum, Hawkins' processing theory can account for FOFC and explain the facts without referring to UG, and instead his theory predicts the distribution of language-particular grammars, including FOFC, from systematic properties of the performance systems. Note that EIC and the more general Minimize Domains "is a comprehension-oriented principle of production, with numerous […] correspondences between ease of comprehension and ease of production" (Hawkins 1994: 427). Thus, it can be viewed as a systematic property of the performance systems that provides, in Fodor's terms, a 'full-blooded Type 3 explanation,' which does not resort to UG and provides independent, 'third-factor' motivations for FOFC.

However, while this explanatory power of efficient processing is acknowledged, Walkden (2009: 69–71) goes on to recast the metrics of Hawkins (1994, 2004) in order to fall back to an "UG-based FOFC." Accordingly, by reformulating the metrics of Hawkins in order to make no reference to processing, Walkden (2009) proposes a 'type 2 explanation,' insofar as he assumes that the parser benefits from a UG-constraint, that is, the parser operates more efficiently if it incorporates the constraint than if it does not.

In the following, we want to depart from Walkden's (2009) (and other's) reasoning by taking issue with their objections to an explanatory account based on processing. We want to argue in favor of a third-factor explanation that refers to systematic properties of the performance systems that are supported by experimental processing data and by corpus studies. It appears that the reason why Walkden reformulated Hawkins' account is in order to avoid any reference to processing that is not well grounded.

The first problem pointed out by Walkden (2009: 68, cf. also references cited there) is that cases like (15) exist, where O-V and D-NP are no less common than V-O and NP-D, and so, there is no evidence that FOFC holds for DP complements of V.

(15) *German*

Johann	hat	[$_{VP}$ [$_{DP}$ den	Mann]	gesehen]
John	*has*	*the*	*man*	*seen*

'John has seen the man.'

According to the classical formulation of FOFC (cf. (8)) and according to the EIC, Walkden argues that the ordering O-V and D-NP are not predicted. The non-existence of FOFC effects between DP and V is regarded a problematic case for an account based on Hawkins' processing principles, since formal UG-approaches can now deal with these exceptions (cf. Biberauer *et al*. 2007 *et seq.*), while the

processing theory of Hawkins cannot. However, as Hawkins (to appear) points out, there are also exceptions in typological samples such as Dryer (1992) to the current UG approaches to FOFC:

(16) V–O & VP–T disharmonic order & violating FOFC (= 9d)

Accordingly, further refining constraints have to be added to UG anyway, which is methodologically undesirable, given that UG should be reduced to a minimum in biolinguistics. Moreover, why should exceptions such as (15) pose a serious problem for a processing account at all? To our mind, it is precisely the strength of theories referring to processing preferences that they define a preference scale and a frequency ranking and predict, in contrast to UG-accounts, that violations like (15) can occur, since they only state that they are much less frequent and certainly less frequent than the harmonic orders (9a) and (9b) and less than the inverse-FOFC order (9c).

The second problem Walkden (2009) mentions is that the more absolute cases such as *V-O & S-TP (see (10d) above) seem to point in the direction of a UG-explanation, because Hawkins' principles cannot make any claim about absolute non-occurrence. He argues that "[f]or such cases a prohibition within UG […] is more satisfactory" (Walkden 2009: 69). Again, we don't see the plausibility of this argument. First, even in the 'absolute' cases, there seem to exist exceptions that force scholars to qualify their statements. For instance, Zwart (2009) argues that in the 214 languages he has taken into account, he finds no 'true' final coordinating conjunctions in head-initial languages. Of course, he has to introduce a definition and then (a restriction) of what counts as 'true'. Furthermore, as Biberauer *et al*. (2007) point out themselves, there do seem to be some — if only very few — exceptions. Accordingly, the FOFC seems to reflect a tendency anyway and does not lend itself to being an 'absolute' statement.

Based on the above objections, our suggestion is that it is more in the biolinguistic spirit to assume that FOFC, as a distinct constraint, is simply not located in the grammar anyway. Instead of assigning additional refinements to the grammar, we concur with Hawkins (to appear: 17) that "stipulations of formal models can become less stipulative by shifting their ultimate motivation away from an innate UG towards (ultimately innate and neurally predetermined) processing mechanisms."

To sum up, after having shown, in section 2, that systematic properties of the performance systems can provide additional evidence for common views of UG, we have discussed the possibility that a linguistic constraint that is widely assumed to be part of UG — FOFC — can possibly be relegated to independently motivated principles of efficient processing. In contrast to the 'type 2 explanation' proposed by Walkden, Biberauer, and colleagues, implying that the parser benefits from a UG-constraint, we have argued in favor of a 'type 3 explanation,' which relegates some language universals to the independently motivated systematicity of the performance systems. To our mind, this is in the spirit of Chomsky's three-factor parcellation, which aims at reducing UG to a minimum.

Up to this point, we have been arguing that there are constraints and strategies that are not part of UG, but show systematic properties and determine,

in interaction with UG, both how we understand and how we produce sentences. Our arguments were mainly based on methodological grounds, however. In particular, we argued that, according to both Fodor (1985) and Chomsky (2005), it is reasonable to reduce UG to a minimum when aiming at a 'deep' explanation of language design. In the next section, we will present empirical evidence that implicit knowledge of the human language performance systems is systematic and is needed anyway in the context of acceptable ungrammaticality.

4. The Performance Interface and Acceptable Ungrammaticality

The claim that an adequate theory of language design needs to take into account a systematic level of performance principles that is not transparent to the grammar goes back to Fodor *et al*. (1974), who presented "a body of phenomena which are systematic but not explicable within the constructs manipulated by formal linguistics" (Fodor *et al*. 1974: 369).

Recently, however, there has been a tendency in generative linguistics to return to the axioms of the derivational theory of complexity. To our mind, the clearest statement in this direction is formulated by Phillips (2004), who tries to show that the crucial arguments against the derivational theory of complexity are not as compelling as one might think (for similar discussion of what follows, cf. Marantz 2005, Boeckx 2009: 133-146). Let us briefly illustrate this reasoning and then argue that it cannot account for the findings to be presented in this section.

One prominent argument, discussed by Phillips (2004: 23–26), for the separation of grammar and parser has been that the systems for comprehension and production are operating in time and are thus prone to errors, while the grammar is defined to be precise. The famous garden path sentences are prominent cases posing a comprehension breakdown (cf. Bever 1970):

(17) The horse raced past the barn fell.

As is well known, the sentence yields an improper parse because of the tendency to interpret *The horse raced past the barn* as a complete clause, not as an NP containing a modifying clause. However, Phillips (2004) argues that, in these cases, hearers do not construct hypotheses that violate grammatical rules or principles, since the grammar clearly allows building structures such as *The horse raced past the barn*. In other words:

> Garden path sentences arise in circumstances of structural ambiguity, where two or more possible grammatical analyses are available. If the parser makes the wrong choice and subsequently breaks down when it becomes clear that the choice was the wrong one, this reflects lack of telepathy, not lack of grammatical precision. (Phillips 2004: 263)

According to Phillips (2004: 264-265), a more serious issue for the claim that hearers do not construct hypotheses that go against the grammar are sentences investigated by Christianson *et al*. (2001) and Ferreira *et al*. (2001):

(18) While the man hunted the deer ran into the woods.

Speakers go down the garden path here, since they misinterpret *the deer* as the object of *hunted*. What is crucial here, however, is that even after realizing this wrong interpretation, Ferreira and colleagues report that participants continue to believe that the man hunted the deer. Accordingly, they seem to interpret *the deer* as both the object NP of the embedded clause and the subject NP of the main clause. Since this is, of course, a grammatically impossible representation, Phillips (2004: 264) points out that "[i]f true, these findings present a serious challenge to the widespread assumption that the parser constructs only grammatically sanctioned representations."

In the following, we will present empirical evidence showing that such discrepancies between performance systems and grammar are more widespread than often assumed. We will argue that this evidence supports our general claim that we have to assume systematic performance systems that are independent from the grammar and that could, therefore, lend themselves to third-factor explanations, in the sense of Chomsky (2005).

4.1. Missing-VP Effect

Multiple center-embedding normally leads to processing breakdown when the degree of center-embedding exceeds a rather small limit, and sentences containing multiply center-embedded clauses therefore tend to be judged as ungrammatical despite being derivable by the mental grammar. Surprisingly, however, multiple center-embedding can also have the reverse effect. As first discussed in Frazier (1985) (based on an observation attributed to Janet Fodor), a sentence as in (19) seems to be grammatical at first sight.

(19) The patient the nurse the clinic had hired ____ met Jack.

In fact, however, sentence (19) is ungrammatical because it does not contain a VP for the NP *the nurse*. As also pointed out by Frazier (1985), this grammatical illusion only arises if the middle VP (the VP of the higher relative clause in sentences with a doubly center-embedded relative clause) is missing. If either the VP of the superordinate clause or the VP of the lower relative clause is omitted, the ungrammaticality is detected easily.

The missing-VP effect was later confirmed experimentally. In the first experimental investigation of this effect, Gibson & Thomas (1999) had participants rate the complexity of sentences like those in (20) on a scale ranging from 1 ("easy to understand") to 5 ("hard to understand").

(20) a. *All three VPs present* (mean rating = 2.90)

The ancient manuscript that the graduate student who the new card catalog had confused a great deal was studying in the library was missing a page.

b. *VP of the higher RC missing* (mean rating = 2.97)

The ancient manuscript that the graduate student who the new card catalog had confused a great deal ____ was missing a page.

Further experimental confirmations of the missing-VP effect were provided by Christiansen & MacDonald (2009) and Vasishth *et al*. (2010) for English and by Gimenes *et al*. (2009) for French. For the case of VO languages, the missing-VP effect is thus well established.

The only OV-language for which experimental evidence on the missing-VP effect is available, as far as we know, is German. In addition to English sentences, for which they adduced evidence for the missing-VP effect, Vasishth *et al*. (2010) also investigated German sentences as in (21) with the strikethrough verb either included or omitted.

(21) *German*

Der Anwalt, den der Zeuge, den der Spion betrachtete, ~~schnitt~~,
the laywer who the witness who the spy watched avoided
überzeugte den Richter.
convinced the judge
'The lawyer that the witness that the spy watched avoided convinced the judge.'

With both self-paced reading and eye-tracking, Vasishth *et al*. (2010) found increased reading times in the region following the higher relative clause for sentences with a missing VP in comparison to complete sentences, indicating that readers detected the ungrammaticality caused by the missing verb. Vasishth *et al*. (2010) hypothesize that the reason for this purported difference between English and German is that because of the head-final nature of German, readers of German have a stronger expectation of a VP and are therefore less prone to overlook the fact that a verb is missing.

The results of Vasishth *et al*. (2010) contrast with experimental results of Bader *et al*. (2003) and Bader (2012) as well as findings from the corpus study of Bader (2012). These experiments made use of the procedure of speeded grammaticality judgments, which we already introduced in section 2. One of the reasons for using this method for investigating the missing-VP effect is that this effect is one of a number of grammatical illusions, that is, ungrammatical sentences which are nevertheless perceived as grammatical under certain conditions. By using a method that explicitly asks for judgments of grammaticality, it is possible to obtain quantitative evidence on how often a grammatical illusion is experienced by native speakers.

All sentences investigated in Bader *et al*. (2003) had the head-noun of the complex relative clause located within the so-called German midfield, that is, the part of the sentences between C^0 and the clause-final verb(s) (the complex relative clause always consisted of a higher relative containing a lower relative clause in a center-embedded position). Two sample sentences illustrating this for the case of main clauses are shown in (22).

(22) a. *German*

Extraposed: Complete

Heute ist das Programm abgestürzt, das den Programmierer
today is the program crashed which the programmer
geärgert hat, der die Dokumentation ohne irgendeine Hilfe
annoyed has who the documentation without any help

erstellen musste.
write had-to
'Today the program crashed which had annoyed the programmer who had to write the documentation without any help.'

b. **Center-embedded: Complete or missing-VP**

Heute ist das Programm, das den Programmierer, der die
today is the program which the programmer who the
Dokumentation ohne irgendeine Hilfe erstellen musste,
documentation without any help write had-to
~~geärgert hat~~, abgestürzt.
annoyed has crashed
'Today the program crashed which had annoyed the programmer who had to write the documentation without any help.'

In (22a), the relative clauses have been extraposed. In (22b), the relative clauses occur center-embedded. Sentences as in (22b) were presented to participants either completely or with the struckthrough verbal complex omitted.[4] Complete center-embedded sentences, (22b), were judged as grammatical less often than extraposition sentences, (22a), although they still received a majority of grammatical responses. When the VP of the higher relative clause in center-embedded position was missing, sentences were judged as grammatical about half of the time. Other experiments showed acceptance rates of similar size. This indicates that comprehenders of German often, although not always, perceive missing-VP sentences as grammatical.

In the experiment reported in Bader (2012), a further comparison concerned the position of the complex relative clause that was missing the higher VP. Here, a striking difference between sentences with extraposed and sentences with center-embedded relative clauses showed up. When the relative clauses were extraposed, participants rarely overlooked the fact that a VP was missing. In contrast, when the relative clauses were center-embedded, participants often did not notice that the sentences were incomplete and therefore ungrammatical. More than half of the time (58%), participants judged sentences of this type as grammatical.

In sum, for sentences in which the head noun of the higher relative clause was located in the middle field, the experimental evidence shows that the missing-VP effect also occurs in the head-final language German. Although most of this evidence comes from experiments using grammaticality judgments, a recent experiment using self-paced reading (cf. Bader & Häussler 2012) supported the same conclusion. The different conclusions arrived at in the experiments of Vasishth *et al*. (2010) and in our experiments thus do not seem to be caused by different experimental procedures. The differences are probably due to the different syntactic positions of the relative clauses. A relative clause in SpecCP seems to be easier to process than a relative clause in the middle field (cf. Bader & Häussler 2010 for corpus evidence). This seems to make it easier to notice that a VP is miss-

[4] Other experimental conditions cannot be discussed here for reasons of space. See Bader *et al*. (2003) for further information.

ing in sentences like those investigated by Vasishth *et al*. (2010).

Further evidence for the reality of the missing-VP effect in German comes from the corpus study of Bader (2012), which was already discussed in section 2. In 15% of all corpus instances with a doubly center-embedded relative clause, the VP of the higher relative clause was missing, as in the original corpus example in (23).

(23) *German*

Missing-VP example

Dieser	Typ	entsteht,	wenn	lin-3	oder	ein	Gen,	das	für	die
this	*type*	*emerges*	*when*	*lin-3*	*or*	*a*	*gene*	*that*	*for*	*the*
Induktion,	die	von	der	Ankerzelle	ausgeht,	_____		mutiert	ist.	
induction	*that*	*from*	*the*	*anchor-cell*	*originates*			*mutated*	*is*	

'This type emerges when lin-3 or a gene that _____ for the induction that originates from the anchor cell has mutated.'

(Dewac-1/95201, http://www.zum.de)

In the other three sentences types investigated in Bader (2012), VPs were also sometimes missing, but with a substantially lower rate ranging from 0-2%. Of the three VPs involved, the VP of the lower relative clause was almost never omitted. The VP of the superordinate clause was missing in a small number of cases, but only in sentences with center-embedded relative clauses and never in sentences with extraposed relative clauses. The VP of the higher relative clause was missing in a substantial number of cases in doubly center-embedded relative clauses and also sometimes when the higher relative clause was extraposed but the lower relative clause still center-embedded within the higher relative clause. When the lower relative clause was extraposed behind the higher relative clause, the VP of the higher relative clause was never missing. The generalization that emerges is that VPs are missing only under circumstances of high processing load. Processing load is highest in sentences containing doubly center-embedded relative clauses, and the rate of missing VPs is accordingly highest in these sentences. Processing load is lowest in sentences in which the higher relative clause is extraposed behind the superordinate clause and the lower relative clause behind the higher relative clause, and there was not a single missing VP in these sentences.

In light of the overall pattern of missing VPs, the claim that the high rate of VPs missing from the higher relative clause in doubly center-embedded relative clauses is just a side effect of such sentences being particularly prone to grammatical errors in general can be rejected. Thus, we conclude that the missing-VP effect, which had previously only been reported for language comprehension, also occurs during language production.

As discussed above, Vasishth *et al*. (2010) have proposed that due to experience with the head final order of German, German comprehenders may maintain a prediction of an upcoming verb in a more highly activated state permitting the prediction to persist longer than in a head initial language. As shown by the corpus data reviewed above, language producers of German regularly forget the prediction of a VP and thus produce incomplete sentences. Thus, the prediction

of a VP in German is clearly not strong enough to prevent the omission of a grammatically required VP. Since comparable data are not available for English, we do not know whether producers of English forget to produce all VPs even more frequently, as would be predicted by the hypothesis of Vasishth *et al*. (2010).

In sum, the fact that the missing VP2 phenomena appears in both English and German shows that the regularities are deep and not a reaction to the particular configuration created by the word order of one language. If it were otherwise, it might lead to rather dramatically different processing systems in different languages, making the biolinguistic program somewhat less plausible. But, as in other domains, the processing system looks largely the same across languages, modulo differences in the grammar itself.

4.2. Mismatch Ellipsis

After having demonstrated that the missing-VP effect is not due to particular ordering in specific languages but, instead, points toward deep regularities that belong to the biologically grounded performance systems, we now turn to another case of acceptable ungrammaticality: mismatch ellipsis. As in the case of the missing-VP effect, we will present evidence from both production and comprehension, thereby supporting our view that the properties that can be attested for this case of acceptable ungrammaticality are part of an abstract knowledge of the performance systems that constrains both production and comprehension. Let us first introduce the phenomenon we are dealing with here.

Focusing on Verb Phrase Ellipsis, it is well known that the grammar requires the elided constituent and its antecedent to match syntactically, apart from certain morphological features (Sag 1976, Williams 1977). Counter-examples to this claim include the prominent example in (24):

(24) This information could have been released by Gorbachov, but he chose not to.

(Daniel Shorr, NPR, 10/17/92, reported by D. Hardt)

Examples without a matching antecedent raise two problems for the approach advocated here. One problem is to explain why listeners and readers tend to accept certain mismatch ellipsis examples like (24) if indeed they are ungrammatical, and one problem is to explain why speakers equipped with a grammar prohibiting 'mismatch ellipsis' sentences like (24) would produce them anyway. In what follows, we will argue that the solutions to these two problems are related: speakers utter mismatch ellipsis examples as speech errors and listeners repair such errors, finding them relatively acceptable under particular conditions where they are easy to repair, they sound like a form the human language production system would produce, and the repaired meaning is plausible. Let us first turn to evidence for repairs and show that the acceptability of mismatch ellipsis depends on the number of repairs and on the amount of evidence for each repair.

When an elided VP has a syntactically mismatching antecedent, the processor attempts to repair the antecedent. If this can be done easily (with only a small

number of operations for which there is plentiful evidence in the input), then the ellipsis will be repaired. This predicts gradient acceptability depending on the number of repairs. Arregui *et al*. (2006) provided experimental evidence for this prediction, showing in a written acceptability judgment study that acceptability drops as one moves from VP ellipsis examples containing a matching VP antecedent in predicate position (25a), to a VP in subject position (25b), to examples requiring a trace to be replaced by its ultimate binder (25c) to very low acceptability for examples where the required VP antecedent could only be built by deconstructing a word to create the verb needed to head the VP antecedent (25d).

(25) a. None of the astronomers saw the comet, /but John did.
(Available verb phrase)

b. Seeing the comet was nearly impossible, /but John did.
(Embedded verb phrase)

c. The comet was nearly impossible to see, /but John did.
(Verb phrase with trace)

d. The comet was nearly unseeable, /but John did.
(Negative adjective)

Percentage Acceptable Responses, Experiment 1	
Condition	% Acceptable
a. Available verb phrase	82.8
b. Embedded verb phrase	66.1
c. Verb phrase with trace	43.9
d. Negative adjective	17.1

Table 2

In subsequent studies, Arregui *et al*. showed that the results could not be due to the antecedent alone, but implicated repair of the antecedent, and they presented further findings, e.g., showing that VP ellipsis examples with verbal gerundive antecedents were more acceptable than ones with nominal gerundive antecedents.

Fanselow & Frisch (2006) show that processing difficulty may decrease the acceptability of a sentence in uncontroversially grammatical sentences such as object-initial German sentences. So it is not terribly surprising that repair difficulty influences rated acceptability if comprehension of the sentence involves identifying and making the hypothesized repair(s). Though the complexity of the repair operation (the number of repairs and the amount of evidence for them) is clearly related to acceptability, it is insufficient to give a full picture of the acceptability of mismatch ellipsis. For this we need to consider what is known about acceptability judgments in other cases of ungrammatical sentences (see Otero 1972 for an early example involving confusion among different *se*s in Spanish).

In general, acceptability ratings are higher if even one example of a structure has been encountered before rating a novel example. In five experiments, Luka & Barsalou (2005) had participants read grammatical sentences first and then rate novel sentences. Mere exposure to a sentence, or to a sentence structure with different words, resulted in higher ratings. For example, *Egor lugged Dr.*

Frankenstein the corpse was rated higher after reading unrelated sentences with double object structures, and *What the pharmacist recommended is to read the directions* was rated higher after reading sentences with a pseudo-cleft structure.

Frazier (2008a) considered the question of whether a speech error, essentially a blending of two forms, would be rated as more acceptable if it sounded like a natural or motivated speech error than if it did not. In a small experiment with Chuck Clifton, participants were asked to rate the acceptability of two items like (28), where (28a) is an actual speech error, heard on National Public Radio. The actual error (28a) involves a switch from impersonal *you* to *we* that would avoid the unwanted implication that the addressee was deluding himself. The unmotivated error in (28b) involves the same switch of subjects, now from *we* to *you*, but suggests that the speaker has gone out of his way to insult the addressee.

(28) a. If you think this is going to solve the terrible problems in Najaf, we're deluding ourselves.
b. If we think this is going to solve the terrible problems in Najaf, you're deluding yourself.

Acceptability judgments (1–5, where 5 means perfectly acceptable) were indeed influenced by how natural the error was: Natural errors like (28a) received a mean rating of 4.05, which was significantly higher than the mean of the unnatural errors (3.37). The implication of these studies is that familiarity of novel forms, and even how natural a particular form is as an output of the human language system, influences acceptability. The studies thus reinforce the conclusion that acceptability ratings reflect the judged goodness of utterances based on both grammatical and performance factors. By contrast, the classification of an utterance as grammatical or ungrammatical is a theoretical matter. The status of an utterance as being grammatical or ungrammatical follows from the best most explanatory overall theory of language, that is, if the judged badness of an utterance follows from independently needed grammatical constraints, the utterance is ungrammatical; if its badness follows from independently known performance factors, the utterance is unacceptable.

In the case of mismatch ellipsis, it is clear why speakers might utter an antecedent clause and an elided clause that don't match. Memory may lead them to misremember. Since it is known that passives are misremembered more often as actives than the other way around (Mehler 1963), this predicts an asymmetry: Passive-active mismatches, as motivated errors, should be more acceptable than active-passive mismatches. This prediction was confirmed (cf. Arregui *et al*. 2006). With coreferential subjects, conjoined clause antecedents might be misremembered as conjoined VP antecedents. This predicts that listeners and readers might choose conjoined VP antecedents for an elided VP even when one didn't actually occur. This too has been confirmed (cf. Frazier & Clifton 2011a).[5]

[5] Whether argument structure alternations other than passive-active ones give rise to acceptable mismatches has not been investigated thoroughly. For example, 'X and Y collided' might compete with 'X collided with Y' resulting in a sentence like *Sue and Mary will collide, I think, and John will with George*, which sounds relatively acceptable to us. However, on the present account, only alternations that are alternative expressions of the SAME proposition

Moving beyond ellipsis, we will now see that the notion of repair or speech error reversal advocated here is not intrinsically tied to ellipsis and that repairing syntactic blends can be established in other domains as well: when alternative linguistic forms compete in production, the listener repairs the input based in part on the speaker's likely intent.

Frazier & Clifton (2011b) investigated doubled quantifiers. In a language like English, speakers must choose between using a determiner quantifier (*everyone, nobody*) or an adverbial quantifier (*always, never*). Sometimes both show up in the same (blended) utterance, as in (34).

(34) *Doubled quantifiers in attested blends*

a. Many people often thought that you use whipped cream pie.
(National Public Radio, discussion of clowns and pie throwing)

b. Typically when I meet people I often ask people what they would talk about if this wasn't a job talk.
(Introduction to a University of Massachusetts colloquium, 3-22-10)

c. …and it might not require scientific research to infer that the majority of sarcasm one encounters is usually spoken.
(Undergraduate paper, University of Massachusetts, Spring, 2010)

In a written interpretation study, participants were asked to choose the interpretation they gave to sentences like (35) with both a determiner and an adverbial quantifier. Four types of examples were tested: *many–often*, as in (35), *every–always*, negation examples of various types, and *few–seldom*. The data are presented in Table 3.

(35) Many students often turn in their assignments late.
What did that mean?

a. The number of students who turn in their assignments late is large. (Undoubled)

b. The number of students who frequently turn in their assignments late is large. (Doubled)

Percentage Choices of Undoubled Paraphrases, by Item Set (with Standard Errors in Parentheses) Item Set			
many (many-often)	universal (every-always)	negation (no-never)	few (few-seldom)
77	64	36	73

Table 3

Apart from negation, the majority of interpretations undoubled the quantifier, as would be expected if participants reversed the speech error before interpreting the sentence. Notice that without the speech error reversal, only the interpreta-

should be acceptable. Hence, *John drove the car*/*The car drove well* or *Ina melted the ice*/*The ice melted* would not be expected to give rise to blended utterances and thus not be expected to give rise to acceptable ungrammaticality.

tion like (35b) is grammatical according to the compositional semantics (cf. Frazier & Clifton 2011 for problems with an attempt to interpret the doubled forms as emphatic).[6]

Another example of speech error reversal, pointed out to us by Greg Carlson, involves sentences like (36). People fail to notice the grammatical interpretation of the sentence, where the mother kills the child by preventing her from almost drowning.

(36) Mother saves child from nearly drowning.

(37) *Prevent X from happening/X almost happened*
  a. Mother saved the child from nearly drowning.
  b. Mother saved the child from drowning.

In a written study (Clifton & Frazier, in progress) where participants indicate whether sentences were "o.k., acceptable," overwhelmingly participants accepted sentences like (37a) as well as sentences like (37b). This result, like the quantifier undoubling result, suggests that comprehenders reverse common or natural errors, assigning interpretations that would be unacceptable if the reversal did not take place.

To sum up, various sources of evidence suggest predictable speech errors involving a blend of two competing forms are repaired by listeners. The interpretation that goes with the repaired utterance is only accepted as a possible interpretation when it is a plausible interpretation. This stands in stark contrast with unrepaired utterances that are paired with their meanings by the compositional semantic interpretation of the actual utterance. In other words, the performance based pairing of form and meaning is token-based, and it relies on the performance systems (competing morphological or syntactic forms in production, comprehension repair mechanisms based in part on knowledge of the speaker's probable intent) together with the compositional semantics to pair form and meaning. In the case of mismatch ellipsis, the repair involves licensing of a later form based on the unselected form of the antecedent clause, as illustrated in (40). Given a particular message, two forms are available for expressing the message (active-passive; conjoined VP-conjoined clause). The speaker chooses one form but a later (ellipsis) clause is licensed only by the unchosen form.

(40) Meaning — Form 1–Speaker chooses Form 1
  — Form 2–Licenses later form (e.g., ellipsis)

By hypothesis, it is implicit knowledge of human language performance systems that allows the comprehender to repair the form, as if 'reading through' the error.[7] Adopting this explanatory strategy, certain attested utterances are explained

[6] We suspect that the double negative examples were treated differently because American students are taught in school to avoid doubling negation. Perhaps if we were to test less educated subjects, the rates of undoubling for negation would be on a par with those for the other quantifiers.

[7] Error reversal repair is NOT a form of sloppy or 'good enough' processing. Ferreira & Pat-

outside the grammar proper. This permits a much simpler grammar than would otherwise be possible, as in the case of mismatch ellipsis. It also offers a more nuanced account of the data, explaining why particular tokens of a structure may be acceptable even though other tokens are not. In the case of doubled quantification and prevent-near-culmination sentences (37), it is an ungrammatical interpretation that is acceptable.

To sum up, the findings presented in this section demonstrate that alongside the type based grammatical system for pairing forms and meaning, we must recognize the existence of a token based system that allows certain utterances to be patched up before they are interpreted. Let us now broaden the perspective again and ask how cases of acceptable ungrammaticality like the missing-VP effect and mismatch ellipsis can contribute to a biolinguistic explanation of the design features of human language.

4.3. The Performance Interface and Acceptable Ungrammaticality

In the sections above, we presented empirical evidence that implicit knowledge of the human language performance systems is systematic and is needed in the context of acceptable ungrammaticality. Together with the methodological benefit of relegating certain grammatical constraints to the performance systems, which we have demonstrated in the context of center-embedding (section 2) and linear ordering (section 3), this is a strong argument to factor performance systems into the biolinguistic approach to language, since the overall goal of this approach is to reduce UG and to focus on third-factor explanations instead.

Both the missing-VP effect and the case of mismatch ellipsis are grammatical illusions, a class of phenomena that have not been systematically studied until recently. For many scholars, grammatical illusions are a good reason to distinguish between structures built on-line and structures generated by the grammar. In other words, the phenomena subsumed under the term 'grammatical illusions' seem to contradict the view that grammatical constraints are transparently implemented in real-time language processes. Accordingly, the recent attempts to return to the axioms of the derivational theory of complexity mentioned at the outset of section 4 seem to be challenged by grammatical illusions.

Phillips *et al*. (2011), however, take issue with this class of phenomena and argue in favor of a "systematic account of selective fallibility that can predict the on-line effects of an individual [grammatical] constraint based on its structural description" (Phillips *et al*. 2011: 168). In other words, committed to the general view that 'the parser is the grammar' (Phillips 1996), they assume that grammatical illusions can be ultimately traced back to properties of the grammar. As the research overview by Phillips *et al*. (2011) shows, grammatical illusions is a matter of intense investigation, with a lot of specific proposals, and we cannot do justice to them here. We do not deny the view that both parsing and production make heavy use of the same syntactic mechanisms in the grammar, as Phillips

---

son (2007) among others have argued that the processor assigns only as much structure or interpretation to a sentence as is required for a particular task (cf. Frazier 2008b for a critique).

and colleagues claim. However, both tasks are subject to several external constraints and we advance the claim that (many of) these constraints can be accounted for in terms of systematic properties of the performance systems rather than by resorting to a single grammatical module with a strong (and rich) predictive component that is embedded in a noisy cognitive architecture (cf. Phillips to appear). At this point, we would like to argue that this reasoning is unwanted, given the biolinguistic framework to reduce UG. As we already have made clear at the outset of the paper, approaches that try to translate formal models of grammar into performance models by assuming constraints in the grammar that could also be accounted for by a 'type 3 explanation,' that is, by no reference to the grammar at all, are not in line with the general impetus of biolinguistics to ascribe as many properties of language design as possible to third factors. More concretely, since Phillips & Lewis (to appear: 15) claim that "'generate-and-filter' mechanisms are familiar from many grammatical theories […], and hence are plausible components of a real-time grammar," they try to provide an account of grammatical illusions that unnecessarily complicates the grammar, thereby deviating from the methodological standards set by both Fodor (1985) and Chomsky (2005).

In contrast to Phillips and colleagues, we argue that what makes illusory cases acceptable is not located in the grammar at all but is due to systematic properties of the performance systems. In this context, let us recall that Chomsky (1965: 4) pointed out that the actual behavior of a speaker-hearer is "the interaction of a variety of factors, of which the underlying competence of the speaker-hearer is only one." That is, the competence grammar interacts with other cognitive components when language production and comprehension take place. One of these components is the implicit knowledge of human language performance systems. Taking this perspective of interacting systems seriously has an interesting consequence for the subject of acceptable ungrammaticality and for one of the main data sources of theoretical linguistics — acceptability judgments — in general: Since judging a sentence is also an interaction effect, it may well be that the fact that speakers have remarkably stable judgments about a large amount of sentences is not per se an indication of the nature of the grammar. As we saw in this section, other cognitive components such as the performance systems may well boost the acceptability of sentences. In other words, judgments always involve both the grammar and the processor. It's just in many simple examples it's harmless to ignore the contribution of the processor. But once one gets into more complicated examples (longer, more complex, less well understood), it becomes apparent that grammar and processor are always implicated in judgments.

5. Conclusion

In this paper, we have shown that systematic properties of performance systems can play an important role within the biolinguistic perspective on language by providing third-factor explanations for crucial design features of human language. In particular, we have demonstrated that the performance interface in language design contributes to the biolinguistic research program in three ways:

(i) it can provide additional support for current views on UG, as shown in section 2 in the context of complex center-embedding; (ii) it can revise current conceptions of UG by relegating widely assumed grammatical constraints to properties of the performance systems, as shown in section 3 in the context of linear ordering; (iii) it can contribute to explaining heretofore unexplained data that are disallowed by the grammar, but can be explained by systematic properties of the performance systems.

At the outset of our paper, we referred to Berwick & Weinberg (1984) as a prominent case of attempting to directly translate formal models of linguistic competence into a performance model. Recently, many scholars point out that a revitalized version of the derivational theory of complexity may be the best way for the Minimalist Program to move forward (cf. Marantz 2005). It strikes us as particularly interesting that the recent minimalist literature appeals to notions of 'efficiency' and 'computational economy' and refers to derivations as 'actual computations'. The most telling case in this regard might be the notion of 'phase', basically (re)introducing the concept that syntactic derivations proceed in incremental chunks. In particular, Chomsky (2000a: 106) claims that "at each stage of the derivation a subset […] is extracted, placed in active memory (the 'workspace')." As we have exemplified throughout the paper, there are, according to Fodor (1985), roughly two explanatory strategies concerning such cases where the nature of linguistic constraints obviously suggests a connection between the grammar and the properties of the performance systems: (i) assuming that properties of the performance systems provide a functional explanation for the constraints in the grammar (type 2 explanation), or (ii) taking into account that the constraints might be motivated independently and do not exist in the grammar (type 3 explanation). In this paper, we provided a new perspective on 'type 3 explanations'. Specifically, we introduced a notion of performance systems that dovetails well with the biolinguistic methodology of reducing UG by referring to third-factor explanations. We hope our paper thereby encourages addressing other features of current syntactic theory from the perspective of the performance interface in language design.

Importantly, our notion of performance properties does not contradict basic axioms of linguistic theory. As we highlighted in section 4, linguistic behavior like acceptability judgments, are, according to Chomsky (1965), interaction effects. Crucially, however, according to the biolinguistic perspective, which is characterized by focusing on how the language faculty is biologically grounded in the brain, "[t]here is good evidence that the language faculty has at least two different components: a 'cognitive system' that stores information in some manner, and performance systems that make use of this information for articulation [and] perception" (Chomsky 2000b: 117). Consequently, within biolinguistics, the theory of grammar and the theory of performance characterize two objects at the same level of description, since interaction in terms of information flow is postulated between the grammar and the performance systems. In this paper, we have strengthened this view that the performance systems, beside the grammar, constitute a distinct cognitive component of biolinguistic inquiry by showing that these systems are not random but characteristic. Performance in our sense involves systematic properties of the language processing system, not just the

study of phenomena like errors made when drunk or after stubbing a toe. In other words, non-trivial systematic processing phenomena will be part and parcel of understanding the grammar (its boundaries, i.e., what it must account for and what not). To uncover these systematic properties, the examples discussed in the paper involve both comprehension and production (the role of errors in acceptability judgments and repair, corpus and comprehension experiments for center-embedding and the missing-VP effect).

In sum, this paper contributes to the biolinguistic explanation of language design by shifting away from a rich innate UG towards ultimately innate and neurally determined processing mechanisms that belong to the domain of third-factor effects — a domain that offers a promising perspective for future collaboration and cross-fertilization of linguistic theory and psycholinguistics.

References

Arregui, Ana, Charles Clifton, Lyn Frazier & Keir Moulton. 2006. Processing elided verb phrases with flawed antecedents: The recycling hypothesis. *Journal of Memory and Language* 55, 232–246.

Bader, Markus. 2012. Complex center-embedded relative clauses in German. Ms., Goethe-Universität Frankfurt.

Bader, Markus, Josef Bayer & Jana Häussler. 2003. Explorations of center-embedding and missing VPs. Poster presented at the *16th CUNY Conference on Sentence Processing*, MIT, Cambridge, MA. [27–29 March 2003]

Bader, Markus & Jana Häussler. 2010. Word order in German: A corpus study. *Lingua* 120, 717–762.

Bader, Markus & Jana Häussler. 2012. Missing-VP effects: headedness does not matter. Poster presented at *25th CUNY Conference on Sentence Processing*, CUNY, New York. [14–16 March 2012]

Bader, Markus & Tanja Schmid. 2009. Verb clusters in Colloquial German. *The Journal of Comparative Germanic Linguistics* 12, 175–228.

Baroni, Marco, Silvia Bernardini, Adriano Ferraresi & Eros Zanchetta. 2009. TheWaCkyWide Web: A collection of very large linguistically processed web-crawled corpora. *Language Resources and Evaluation Journal* 23, 209–226.

Berwick, Robert C. 2011. Syntax facit saltum redux: Biolinguistics and the leap to syntax. In Anna Maria Di Sciullo & Cedric Boeckx (eds.), *The Biolinguistic Enterprise: New Perspectives on the Evolution and Nature of the Human Language Faculty*, 65–99. Oxford: Oxford University Press.

Berwick, Robert C. & Amy S. Weinberg. 1984. *The Grammatical Basis of Linguistic Performance: Language Use and Acquisition*. Cambridge, MA: MIT Press.

Bever, Thomas G. 1970. The cognitive basis for linguistic structures. In John R. Hayes (ed.), *Cognition and the Development of Language*, 279–360. New York: Wiley.

Biberauer, Theresa, Anders Holmberg & Ian Roberts. 2007. Disharmonic word-order systems and the Final-over-Final Constraint (FOFC). In Antonietta

Bisetto & Francesco E. Barbieri (eds.), *Proceedings of the XXXIII Incontro di Grammatica Generativa*, 86–105. Bologna.

Biberauer, Theresa, Anders Holmberg & Ian Roberts. 2010. A syntactic universal and its consequences. Ms., University of Cambridge & Newcastle University.

Boeckx, Cedric. 2009. *Language in Cognition: Uncovering Mental Structures and the Rules Behind Them*. Malden, MA: Wiley-Blackwell.

Boeckx, Cedric. 2011. Approaching parameters from below. In Anna Maria Di Sciullo & Cedric Boeckx (eds.), *The Biolinguistic Enterprise: New Perspectives on the Evolution and Nature of the Human Language Faculty,* 205–221. Oxford: Oxford University Press.

Chomsky, Noam. 1963. Formal properties of grammars. In R. Duncan Luce, Robert R. Bush & Eugene Galanter (eds.), *Handbook of Mathematical Psychology: Vol. 2*, 323–418. New York: Wiley.

Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

Chomsky, Noam. 2000a. Minimalist inquiries: The framework. In Roger Martin, David Michaels & Juan Uriagereka (eds.), *Step by Step: Essays on Minimalist Syntax in Honor of Howard Lasnik,* 89–155. Cambridge, MA: MIT Press.

Chomsky, Noam. 2000b. *New Horizons in the Study of Language and Mind*. Cambridge: Cambridge University Press.

Chomsky, Noam. 2005. Three factors in language design. *Linguistic Inquiry* 36, 1–22.

Chomsky, Noam. 2007. Approaching UG from below. In Uli Sauerland & Hans-Martin Gärtner (eds.), *Interfaces + Recursion = Language? Chomsky's Minimalism and the View from Syntax–Semantics*, 1–29. Berlin: Mouton de Gruyter.

Chomsky, Noam & George A. Miller. 1963. Introduction to the formal analysis of natural languages. In R. Duncan Luce, Robert R. Bush & Eugene Galanter (eds.), *Handbook of Mathematical Psychology: Vol. 2,* 269–321. New York: Wiley.

Christiansen, Morton H. & Nick Chater. 1999. Toward a connectionist model of recursion in human linguistic performance. *Cognitive Science 23*, 157–205.

Christiansen, Morten H. & Maryellen C. MacDonald. 2009. A usage-based approach to recursion in sentence processing. *Language Learning* 59, 126–161.

Christianson, Kiel, Andrew Hollingworth, John F. Halliwell & Fernanda Ferreira. 2001. Thematic roles assigned along the garden path linger. *Cognitive Psychology* 42, 368–407.

Clifton, Charles & Lyn Frazier. In progress. Registering acceptability judgments.

De Roeck, Anne, Roderick Johnson, Margaret King, Michael Rosner, Geoffrey Sampson & Nino Varile. 1982. A myth about centre-embedding. *Lingua* 58, 327–340.

Di Sciullo, Anna Maria. 2000. Parsing asymmetries. In Dimitris N. Christodoulakis (ed.), *Proceedings of the Second International Conference on Natural Language Processing,* 1–15. London: Springer.

Dryer, Matthew S. 1992. The Greenbergian word order correlations. *Language* 68, 81–138.

Fanselow, Gisbert & Stefan Frisch. 2006. Effects of processing difficulty on

judgements of acceptability. In Gisbert Fanselow, Caroline Féry, Ralf Vogel & Matthias Schlesewsky (eds.), *Gradience in Grammar: Generative Perspectives*, 291–316. Oxford: Oxford University Press.

Ferreira, Fernanda, Kiel Christianson & Andrew Hollingworth. 2001. Misinterpretations of garden-path sentences: Implications for models of reanalysis. *Journal of Psycholinguistic Research* 30, 3–20.

Ferreira, Fernanda & Nikole D. Patson. 2007. The 'Good Enough' approach to language comprehension. *Language and Linguistics Compass* 1, 71–83.

Fodor, Janet D. 1985. Deterministic parsing and subjacency. *Language and Cognitive Processes* 1, 3–42.

Fodor, Janet D. & Stefanie Nickels. 2011. Prosodic phrasing as a source of center-embedding difficulty. Poster presented at *Architectures and Mechanisms of Language Processing (AMLaP) 2011*, Paris. [Paris Descartes University, 1-3 September 2011.]

Fodor, Jerry A., Thomas G. Bever & Merrill F. Garrett. 1974. *The Psychology of Language: An Introduction to Psycholinguistics and Generative Grammar*. New York: McGraw-Hill.

Frazier, Lyn. 1985. Syntactic complexity. In David Dowty, Lauri Karttunen & Arnold Zwicky (eds.), *Natural Language Processing: Psychological, Computational, and Theoretical Perspectives*, 129–189. Cambridge: Cambridge University Press.

Frazier, Lyn. 2008a. Processing ellipsis: A processing solution to the undergeneration problem. In Charles Chang & Hannah Haynie (eds.), *Proceedings of WCCFL 26*, 21–32. Somerville, MA: Cascadilla.

Frazier, Lyn. 2008b. Is 'Good Enough' processing good enough? In Luciano Arcuri, Pietro Boscolo & Francesca Peressotti (eds.), *Festschrift in Honor of Ino Flores d'Arcais*. Padova: CLEUP.

Frazier, Lyn & Charles Clifton. 2011a. Dynamic interpretation: Finding an antecedent for VPE. In Jesse A. Harris & Margaret Grant (eds.), *University of Massachusetts Occasional Papers in Linguistics 38*, 23–36.

Frazier, Lyn & Charles Clifton. 2011b. Quantifiers undone: Reversing predictable speech errors in comprehension. *Language* 87, 158–171.

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. Papers from the First Mind Articulation Project Symposium*, 95–126. Cambridge, MA: MIT Press.

Gibson, Edward & James Thomas. 1999. Memory limitations and structural forgetting: the perception of complex ungrammatical sentences as grammatical. *Language and Cognitive Processes* 14, 225–248.

Gimenes, Manuel, François Rigalleau & Daniel Gaonac'h. 2009. When a missing verb makes a French sentence more acceptable. *Language and Cognitive Processes* 24, 440–449.

Hauser, Marc D., Noam Chomsky & W. Tecumseh Fitch. 2002. The faculty of language: What is it, who has it, and how did it evolve? *Science* 298, 1569–1579.

Hawkins, John A. 1990. A parsing theory of word order universals. *Linguistic Inquiry* 21, 223–261.

Hawkins, John A. 1994. *A Performance Theory of Order and Constituency*. Cam-

bridge: Cambridge University Press.
Hawkins, John A. 2004. *Efficiency and Complexity in Grammars*. Oxford: Oxford University Press.
Hawkins, John A. To appear. Disharmonic word orders from a processing efficiency perspective. In Theresa Biberauer & Michelle Sheehan (eds.), *Theoretical Approaches to Disharmonic Word Order*. Oxford: Oxford University Press.
Holmberg, Anders. 2000. Deriving OV order in Finnish. In Peter Svenonius (ed.), *The Derivation of VO and OV*, 123–152. Oxford: Oxford University Press.
Holmberg, Anders. 2010. Parameters in minimalist theory: The case of Scandinavian. *Theoretical Linguistics* 36, 1–48.
Karlsson, Fred. 2007. Constraints on multiple center-embedding of clauses. *Journal of Linguistics* 43, 365–392.
Karlsson, Fred. 2010. Syntactic recursion and iteration. In Harry van der Hulst (ed.), *Recursion and Human Language*, 43–67. Berlin: Mouton de Gruyter.
Konieczny, Lars. 2000. Locality and parsing complexity. *Journal of Psycholinguistic Research* 29, 627–645.
Luka, Barbara J. & Lawrence W. Barsalou. 2005. Structural facilitation: Mere exposure effects for grammatical acceptability as evidence for syntactic priming in comprehension. *Journal of Memory and Language* 52, 436–459.
Marantz, Alec. 2005. Generative linguistics within the cognitive neuroscience of language. *The Linguistic Review* 22, 429–445.
Mehler, Jacques. 1963. Some effects of grammatical transformations on the recall of English sentences. *Journal of Verbal Learning and Verbal Behavior* 2, 346–351.
Miller, George A. & Noam Chomsky. 1963. Finitary models of language users. In R. Duncan Luce, Robert R. Bush & Eugene Galanter (eds.), *Handbook of Mathematical Psychology: Vol. 2*, 419–491. New York: Wiley.
Otero, Carlos P. 1972. Acceptable ungrammatical sentences in Spanish. *Linguistic Inquiry* 3, 233–242.
Phillips, Colin. 1996. Order and structure. Cambridge, MA: MIT dissertation.
Phillips, Colin. 2004. Linguistics and linking problems. In Mabel L. Rice & Steven F. Warren (eds.), *Developmental Language Disorders: From Phenotypes to Etiologies*, 241–287. Mahwah, NJ: Lawrence Erlbaum.
Phillips, Colin. To appear. Parser-grammar relations: We don't understand everything twice. In Montserrat Sanz, Itziar Laka & Michael Tanenhaus (eds.), *Language Down the Garden Path: The Cognitive and Biological Basis for Linguistic Structure*. Oxford: Oxford University Press.
Phillips, Colin, Matthew W. Wagers & Ellen F. Lau. 2011. Grammatical illusions and selective fallibility in real-time language comprehension. In Jeffrey T. Runner (ed.), *Experiments at the Interfaces*, 147–180. Bingley, UK: Emerald.
Phillips, Colin & Shevaun Lewis. To appear. Derivational order in syntax: Evidence and architectural consequences. In Cristiano Chesi (ed.), *Directions in Derivations*. Amsterdam: John Benjamins.
Reich, Peter A. 1969. The finiteness of natural language. *Language* 45, 831–843.
Richards, Marc. 2008. Two kinds of variation in a minimalist system. *Linguistische Arbeitsberichte* 87, 133–162.

Sag, Ivan. 1976. Deletion and Logical Form. Cambridge, MA: MIT dissertation.
Sauerland, Uli & Andreas Trotzke. 2011. Biolinguistic perspectives on recursion: Introduction to the special issue. *Biolinguistics* 5, 1–9.
Sheehan, Michelle. 2010. Formal and functional approaches to disharmonic word orders. *Newcastle Working Papers in Linguistics* 16, 146–166.
Sheehan, Michelle. To appear. Explaining the Final-Over-Final Constraint: Formal and functional approaches. In Theresa Biberauer & Michelle Sheehan (eds.), *Theoretical Approaches to Disharmonic Word Order*. Oxford: Oxford University Press.
Vasishth, Shravan & Richard L. Lewis. 2006. Argument-head distance and processing complexity: Explaining both locality and antilocality effects. *Language* 82, 767–794.
Vasishth, Shravan, Katja Suckow, Richard L. Lewis & Sabine Kern. 2010. Short-term forgetting in sentence comprehension: Crosslinguistic evidence from verb-final structures. *Language and Cognitive Processes* 25, 533–567.
Walkden, George. 2009. Deriving the final-over-final constraint from third factor considerations. *Cambridge Occasional Papers in Linguistics* 5, 67–72.
Warner, John & Arnold L. Glass. 1987. Context and distance-to-disambiguation effects in ambiguity resolution: Evidence from grammaticality judgements of garden path sentences. *Journal of Memory and Language* 26, 714–738.
Williams, Edwin. 1977. Discourse and logical form. *Linguistic Inquiry* 9, 101–139.
Zwart, Jan-Wouter. 2009. Relevance of typology to minimalist inquiry. *Lingua* 119, 1589–1606.

Andreas Trotzke
Universität Konstanz
Fachbereich Sprachwissenschaft
Universitätsstraße 10
78457 Konstanz
Germany
andreas.trotzke@uni-konstanz.de

Markus Bader
Universität Frankfurt
Institut für Linguistik
Grüneburgplatz 1
60629 Frankfurt a. M.
Germany
bader@em.uni-frankfurt.de

Lyn Frazier
University of Massachusetts
Department of Linguistics
226 South College
Amherst, MA 01003
USA
lyn@linguist.umass.edu