

Article

The Radical Unacceptability Hypothesis: Accounting for Unacceptability without Universal Constraints

Peter W. Culicover^{1,2,*}, Giuseppe Varaschin³ and Susanne Winkler⁴¹ Department of Linguistics, The Ohio State University, Columbus, OH 43210, USA² Department of Linguistics, University of Washington, Seattle, WA 98195, USA³ Institut für deutsche Sprache und Linguistik, Humboldt University of Berlin, 10099 Berlin, Germany; giuseppe.varaschin@gmail.com⁴ Englishes Seminar, University of Tübingen, 72074 Tübingen, Germany; susanne.winkler@t-online.de

* Correspondence: culicover.1@osu.edu

Abstract: The Radical Unacceptability Hypothesis (RUH) has been proposed as a way of explaining the unacceptability of extraction from islands and frozen structures. This hypothesis explicitly assumes a distinction between unacceptability due to violations of local well-formedness conditions—conditions on constituency, constituent order, and morphological form—and unacceptability due to extra-grammatical factors. We explore the RUH with respect to classical islands, and extend it to a broader range of phenomena, including freezing, A' chain interactions, zero-relative clauses, topic islands, weak crossover, extraction from subjects and parasitic gaps, and sensitivity to information structure. The picture that emerges is consistent with the RUH, and suggests more generally that the unacceptability of extraction from otherwise well-formed configurations reflects non-syntactic factors, not principles of grammar.

Keywords: syntactic theory; island constraints; processing complexity; unacceptability and grammaticality; A' constructions; frequency; surprisal



Citation: Culicover, Peter W., Giuseppe Varaschin, and Susanne Winkler. 2022. The Radical Unacceptability Hypothesis: Accounting for Unacceptability without Universal Constraints. *Languages* 7: 96. <https://doi.org/10.3390/languages7020096>

Academic Editors: Anne Mette Nyvad, Ken Ramshøj Christensen, Juana M. Licerias and Raquel Fernández Fuertes

Received: 29 November 2021

Accepted: 6 April 2022

Published: 13 April 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Syntactic islands are syntactic configurations that in principle should permit extraction, but appear not to. A typical example is (1), which illustrates the unacceptability of extracting from a relative clause.

- (1) a. Sandy read [_{NP} a book [_S that deals with economic theory]].
b. *What subject_i did Sandy read [_{NP} a book [_S that deals with *t_i*]]

It is characteristic of islands that they *appear* to be well-formed, in that all LOCAL CONSTRAINTS ON FORM are satisfied. For example, in (1b) the *wh*-phrase *what subject* is in clause-initial position, where it should be in a *wh*-question. There is a gap in the complement position of the preposition that determines its function and allows the subcategorization requirements of the preposition to be met. All of the phrases are otherwise well-formed: e.g., the various categories are in the correct linear order and all conditions on subcategorization and morphological agreement are satisfied.

In the absence of a plausible alternative, linguists have hypothesized that the unacceptability of (1b) reflects a violation of a syntactic constraint on extraction from a relative clause configuration. Unlike the constraints that determine linear order, subcategorization and agreement, this constraint is NON-LOCAL in nature because the gap can be embedded at an arbitrary depth within the relative clause, as (2b) illustrates:

- (2) a. *What subject_i did Sandy read [_{NP} a book [_S that reveals [_S that Kim worked on *t_i*]]]?
b. ...

- b. *What subject_i did Sandy read [NP a book [_S that reveals [_S that Taylor knows ... [that Kim worked on *t_i*]]]]?

Any syntactic account of phenomena like (2) will typically require grammars of natural languages to include constraints whose domain of application goes well beyond local trees or phrases, encompassing pieces of structure that, though finite in principle, have no upper-bound (Kaplan and Zaenen 1995; Pullum 2019). A corollary of this is that the description language one uses to state syntactic constraints must be endowed with special devices that accomplish the feat of finitely characterizing the unbounded disjunction of paths that may separate a filler from its corresponding gap (devices like existential quantification over nodes or variables in the sense of early transformational grammar).

Ross (1967) showed that these constraints on extraction were general, and not features of particular rules or constructions. Given their abstract nature, a reasonable hypothesis is that such constraints are universal properties of the language faculty, and govern all constructions involving extraction. This hypothesis has driven much of syntactic theorizing since Ross (1967) and the option of attributing the unacceptability that results from violating constraints on extractions to general grammatical principles remains active in much contemporary theorizing (Bošković 2015; Chomsky 2001, 2008; Citko 2014; Nunes and Uriagereka 2000; Phillips 2013a, 2013b; Rizzi 1990; Sabel 2002; Villata et al. 2016, i.a.).¹

However, a plausible case can be made that these constraints are simply descriptive generalizations. On this view, certain syntactic configurations give rise to unacceptability without violating conditions on grammatical form (Boeckx 2008, p. 154). In fact, at this point there is a substantial literature that makes the case that many constraints on extraction do not reflect violation of grammatical principles, but non-syntactic factors such as processing complexity (Arnon et al. 2005; Chaves 2013, 2020; Chaves and Dery 2014, 2019; Chaves and Putnam 2020; Culicover 2013b, 2013c; Deane 1991; Goldberg 2006; Hofmeister et al. 2007, 2013a; Hofmeister and Sag 2010; Hofmeister et al. 2013b; Kluender 1991, 1992, 1998, 2004; Kluender and Kutas 1993b; Newmeyer 2016; Sag et al. 2006, 2007; Staum Casasanto et al. 2010, i.a.).

In this article we pursue this idea, extending the Radical Unacceptability Hypothesis of Culicover and Winkler (2018, p. 380):

Radical Unacceptability Hypothesis (RUH): If judgments of reduced acceptability in cases of otherwise well-formed (i.e., locally well-formed) extractions are due to processing complexity, not syntactic constraints.

The basic idea is that processing complexity is responsible for a broader class of judgments of unacceptability beyond islands per se. Processing complexity arises from such factors as parsing A' chains, referential processing and the management of information structure. We focus specifically on acceptability judgments which result from A' extractions (*wh*-movement, topicalization, etc.) from 'strong' islands and other configurations from which A' extractions are allegedly *never* allowed, such as relative clauses and subjects. The phenomena that we cite here are primarily those that we have addressed in our own prior work, in many cases complementing other research in the field.

This article is organized as follows. First we sketch out in Section 2 a picture of the relationship between acceptability judgments, on the one hand, and the various factors that determine these judgments. We take the position that unacceptability neither directly nor necessarily reflects ungrammaticality, in the sense of a violation of a grammatical condition. From this perspective, an understanding of the ways in which acceptability judgments may arise is essential in investigating the nature of grammar.

In Section 3 we discuss the theoretical basis for the distinction between grammaticality and acceptability. We also briefly review the classical island constraints of Ross (1967), pointing to the substantial literature that shows that these constraints are at best descriptive generalizations about phenomena that are better explained in terms of non-syntactic factors.

In Sections 4 and 5 we review patterns of unacceptability that do not all fall under the classical island constraints and argue that these, likewise, are not explained in terms of grammatical constraints, but non-syntactic factors. Among the phenomena that we consider are: freezing (Section 4.1), A' chain interactions (Section 4.2), topic islands (Section 4.3), zero relative clauses (Section 4.4), weak crossover (Section 5.1), parasitic gaps (Section 5.2), and sensitivity to information structure (Section 5.3).

Section 6 addresses phenomena for which accounts in terms of the RUH are *prima facie* incompatible with the RUH; we suggest ways in which they may ultimately be brought under the RUH.

Finally, on the basis of our review of the causes of unacceptability in cases of extraction, we conclude in Section 7 that there is strong evidence for the following extended version of the RUH.²

Extended Radical Unacceptability Hypothesis (ERUH): All judgments of reduced acceptability in cases of otherwise well-formed (i.e., locally well-formed) extractions are due to non-syntactic factors, not grammatical constraints.

2. Sources of Unacceptability

Let us consider the reasons for a judgment that a sentence is less than fully acceptable. Clearly, violation of a grammatical condition is one source of such a judgment. For example, in (3a) the verb and its complement are in the wrong order, in (3b) there is a subcategorization problem, while in (3c) there is a failure of subject-verb agreement.

- (3) a. *Sandy the beer drank;
- b. *Sandy relies about Kim;
- c. *Sandy are happy.

Such linear order, subcategorization, and morphological agreement constraints are what we call **LOCAL WELL-FORMEDNESS CONDITIONS (LWFC)**. A LWFC, as we understand it, is a constraint on a local piece of linguistic structure, such as adjacent sister nodes or mother-daughter configurations in a tree of depth-1. What defines a LWFC is the fact that it applies to structures of a pre-determined maximum finite size; within some frameworks, these may extend beyond local trees to include non-recursive clausal structures or sequences of phrasal projections, e.g., X' structures, understood as trees of depth-3 (Jackendoff 1977).

How does violation of an LWFC produce a judgment of unacceptability? The obvious answer is that the form of the example is incompatible with the form stipulated by the LWFC. It is useful to think of LWFCs in terms of experience and expectations. Speakers' prior exposure to their language contributes to the emergence of probabilistic expectations regarding what structures they are likely to hear next. Some of these expectations become consistent and stable enough so they can be described in terms of symbolic LWFCs (Bybee 2006, 2010; Bybee and Hopper 2001; Culicover 2005, 2015; Culicover and Nowak 2003). A LWFC is established on the basis of experience with examples that share certain characteristics, for example, that the order of a VP in English is V > NP, not NP > V. If a given example has these characteristics, then its form is expected on the basis of experience. But if it does not have these characteristics, then its form is surprising, and this leads to the judgment of unacceptability.

We assume, therefore, that there is a relationship between the degree of surprise triggered by a linguistic form, or **SURPRISAL**, and acceptability. Low surprisal corresponds to high levels of acceptability, higher levels of surprisal correspond to lower levels of acceptability (Hale 2001, 2003; Levy 2005, 2008, 2013; Levy and Jaeger 2007; Park et al. 2021). Surprisal is inversely related to frequency: the higher the frequency of a construction in a given context, the lower its surprisal; the lower the frequency of a construction in a context, the higher its surprisal.³

Clearly, the frequency of experience plays a role in determining the level of surprisal even when productive LWFCs are not at stake. There are special cases in English where the order NP > V is possible in VP, e.g., (4).

(4) One swallow does not a summer make.

This example contrasts sharply with (3a). For speakers who accept it, it is because they have encountered it in their experience; it is a special construction in their grammar (Culicover 2021). This experience leads to the probability of hearing the verb *make* follow the NP object *a summer* being much higher than it is for NP > V sequences in general. As a result, surprisal in the case of (4) is lower than it is in the case of the structurally identical (3a), and acceptability is higher.

So we have the relationship shown in Figure 1. Experience increases the frequency of particular constructions, and lack of experience corresponds to zero frequency. Frequency leads to expectations. Some of the expected patterns can be described as general LWFCs (i.e., principles of grammar), and some cannot, as we discuss below. Regardless of this, conformity to expectations leads to low surprisal, and low surprisal corresponds to acceptability.⁴

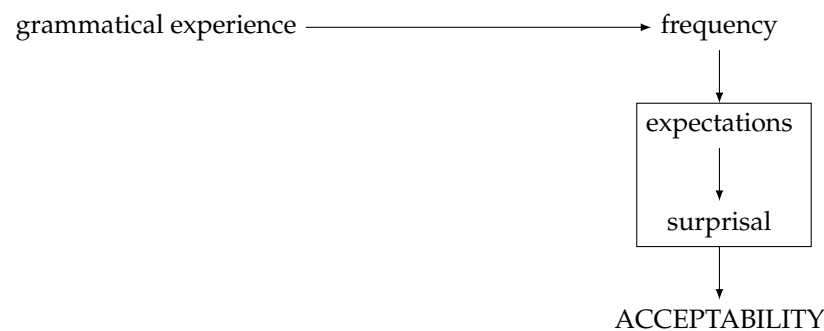


Figure 1. The logic of acceptability judgments for grammatical conditions.

Having established this relationship between grammatical experience and judgments of acceptability, we can now consider other sources of acceptability judgments. One source can be found in the early literature in generative grammar, which suggested that some instances of unacceptability may result from processing complexity and not grammar (e.g., Chomsky 1965; Jackendoff and Culicover 1972; Miller and Chomsky 1963). In particular Miller and Chomsky (1963) demonstrated clearly that unacceptability can arise due to processing complexity in a sentence that satisfies all LWFCs, arguably due to limitations of short-term memory.

It is plausible to assume that higher complexity leads to lower frequency, hence greater surprisal. Since LWFCs can themselves be understood as emergent byproducts of experience-driven expectations, we anticipate that high complexity should have a similar effect on judgments as violation of LWFCs. We therefore extend our picture to that in Figure 2.

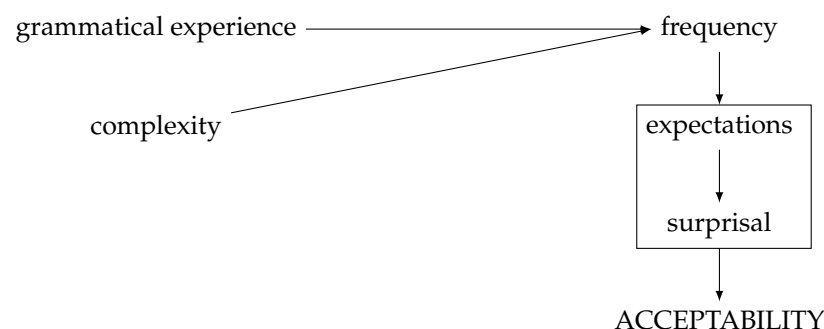


Figure 2. The logic of acceptability judgments for grammatical conditions, version 2.

As we proceed, we flesh out ‘complexity’ with a number of more specific factors.

Given this general framework, it is now possible to understand a wide range of cases of unacceptability judgments as responses to surprisal. Where the expectations come from that lead to such judgments is a complex question, and each case has to be evaluated on its own terms. In the discussion to follow we offer some suggestions, as well as pointers to relevant literature, recognizing that we are far from understanding all of the fine details. The property that is common to all sources of unacceptability is that lack of conformity to expectations leads to surprisal. In other words, surprisal acts like a CAUSAL BOTTLENECK between a wide range of independent factors that impinge on speakers’ expectations and a (behaviorally measurable) acceptability response (Levy 2008).

3. The Acceptability/Grammaticality Distinction and Standard Island Constraints

We suggested above that classical islands of the kind discovered by Ross (1967) may simply be useful generalizations about the kinds of extraction patterns that yield a high level of surprisal, giving rise to an unacceptability response from speakers. If in fact these island patterns are simply generalizations, the following question arises: what factors lead to such generalizations? One answer to this question is the RUH, which in the present framework amounts to the claim that the surprisal associated with island violations stems from the influence of non-syntactic factors in the frequency of particular structures. This hypothesis explicitly assumes a distinction between unacceptability due to violations of local well-formedness conditions (LWFCs)—conditions on constituency, linear order and morphological form—and unacceptability due to non-syntactic factors such as processing complexity as outlined in Section 2.

This distinction has a long lineage in the history of generative grammar (see, for example, Bever 1970; Chomsky 1965; Jackendoff and Culicover 1972; Miller and Chomsky 1963 for some early instances). As soon as language came to be viewed as a cognitive capacity integrated within the larger ecology of the mind, linguists were quick to speculate that grammatical constraints are not the only factors that contribute to the acceptability of sentences (e.g., Kluender 1991, 1998; Kluender and Kutas 1993b). Acceptability came to be viewed as a psychological effect that could be triggered by a host of disparate factors, grammaticality being just one among them (Chomsky 1965, pp. 11–12).

The first exploration of this idea was Miller and Chomsky’s (1963) account of the unacceptability of multiple center-embedding structures (e.g., *the man who the boy who the students recognized pointed out is a friend of mine* (Chomsky 1965) in terms of short-term memory limitations. The first attempt to apply this rationale to constraints on extraction was Jackendoff and Culicover’s (1972) proposal to explain the restrictions to movement out of ditransitive VPs in terms of perceptual strategies for identifying A’ dependency gaps. Their basic idea was that structures like (5a) are unacceptable because the verb-adjacent NP superficially satisfies the verb’s selectional requirement, and the parser expects a gap after the preposition *to* as in (5b)—this is arguably a type of ‘garden path’; see Pritchett (1988) for a range of examples. In terms of the model summarized in Figure 3, the absence of a preposition after the NP in (5a) contradicts the frequency-based expectations of the speaker, and, therefore, yields a surprisal effect that contributes to unacceptability.

- (5) a. * Who_i did Taylor give *t_i* a book?
- b. Who_i did Taylor give a book to *t_i*?

In order to explain these phenomena in purely grammatical terms, it would be necessary to enrich the language for stating syntactic constraints in non-trivial ways.⁵ Rather than appealing to ad hoc extensions, non-syntactic accounts along the lines of work cited above in Section 1 promise to allow us to keep syntactic theory reasonably simple and constrained. Given their potential to make syntax simpler, it is only natural that we consider the possibility that in some cases the unacceptability of extraction from classical islands reflects not grammar, but processing complexity that arises from particular syntactic configurations, as the RUH proposes.

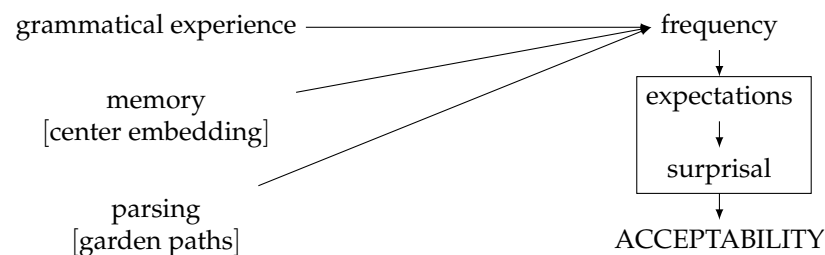


Figure 3. The logic of acceptability judgments for grammatical conditions, version 3.

The application of RUH to classical islands is inspired by two general observations. First, classical island constraints are, in general, TOO STRONG: they exclude sentences that are actually judged to be acceptable by speakers in many circumstances.⁶ As an illustration, consider the Complex NP Constraint discussed in connection to (1) above. The counterexamples to this constraint provided below come from [Erteschik-Shir and Lappin \(1979, p. 58\)](#), [Pollard and Sag \(1994, p. 206\)](#) and [Sag \(1997, p. 454\)](#).

- (6) a. This is the kind of weather_{*i*} that there are [NP many people [S who like *t_i*]].
 b. Which diamond ring_{*i*} did you say that there was [NP nobody in the world [S who could buy *t_i*]]?
 c. There were several old rock songs_{*i*} that she and I were [NP the only ones [S who knew *t_i*]]?

Second, classical island constraints are also TOO WEAK: they fail to exclude extraction patterns that speakers generally consider to be unacceptable. In Sections 4 and 5 we review several examples of A' extractions that do not fall under the classical accounts of islands but which, nonetheless, are unacceptable ([Chomsky 1973, 1977, 1986, 2008](#); [Ross 1967, i.a.](#)).

Furthermore, most, if not all, island constraints appear to function in a wide range of languages, and may be universal. If so, the question arises as to the source of such universals. Evolution is an unlikely explanation; island constraints are neither undecomposable features of language that could have arisen by a simple random mutation streamlined by economy constraints (like Merge Labeling and Agree are claimed to be ([Berwick and Chomsky 2016](#); [Chomsky et al. 2019](#))), nor the kinds of features that could have been selected for by adaptive pressures, leading to a gradual evolutionary process ([Corballis 2017](#); [Jackendoff 1999](#); [Pinker and Bloom 1990](#); [Progovac 2016](#)). It is, therefore, implausible that the human linguistic phenotype evolved specifically to exclude extraction from all of the specific configurations that have been proposed as islands in the literature. One alternative is that the causes of unacceptability in extractions are what biologists call SPANDRELS: phenotypic traits that are not directly selected, but emerge as byproducts of a complex interaction of independent functional adaptations ([Gould and Lewontin 1979](#)). In the case of islands, these may be general cognitive factors related to memory ([Kluender and Kutas 1993b](#)), attention ([Deane 1991](#)), and the management of information flow in discourse ([Erteschik-Shir 1977, 2007](#); [Erteschik-Shir and Lappin 1979](#)).⁷

[Chaves and Putnam \(2020\)](#) offer an extended discussion of classical islands. They review substantial evidence that virtually all of these allow acceptable violations. In addition, they document the factors that enter into judgments of unacceptability (see also [Newmeyer 2016](#)). The case they make supports the RUH as an alternative to the default syntactic approach to unacceptability of islands.⁸ To further support this view, in the next sections we review briefly a number of additional phenomena that fall outside of the traditional island constraints, or that are not traditionally categorized as islands, and argue that they too reflect non-syntactic factors. The conclusion that we draw is an extension of the RUH – if the sentence containing an extraction is locally well-formed and unacceptable, the unacceptability must be due to a non-syntactic factor.

4. Processing A' Chains

In this section, we will explore how several extra-grammatical factors related to the processing and parsing of A' chains increase processing complexity. This, in turn, contributes to reducing the frequency of the particular A' configurations in which these factors are manifested. According to the model outlined in Figure 3, lower frequency leads to higher surprisal and reduced acceptability.

4.1. Freezing

Classic freezing, noted first by Ross (1967, p. 305) is exemplified by the relative unacceptability of extracting from an extraposed prepositional phrase, as in (7b).

- (7) a. You saw [a picture t_j] yesterday [_{PP} of Thomas Jefferson] _{j} .
 b. *Who _{i} did you see [a picture t_j] yesterday [_{PP} of t_i] _{j} ?

Historically, explanations for freezing focus on identifying properties of the syntactic configurations from which extraction is not possible and a corresponding grammatical constraint that explicitly blocks such extraction (Corver 2017). For example, Ross (1967) formulated the Frozen Structure Constraint in (8).

- (8) a. The Frozen Structure Constraint: If a clause has been extraposed from a noun phrase whose head noun is lexical, this noun phrase may not be moved, nor may any element of the clause be moved out of that clause. (Ross 1967, p. 295)⁹
 b. If a prepositional phrase has been extraposed out of a noun phrase, neither that noun phrase nor any element of the extraposed prepositional phrase can be moved. (Ross 1967, p. 303)

Later, Wexler and Culicover (1980) proposed the Raising Principle and the Freezing Principle, based on considerations of language learnability. The Freezing Principle has the effect of blocking extraction from an extraposed PP, as in (7). The Raising Principle blocks extraction from a constituent raised from a lower clause, as in (9).

- (9) a. *Who _{i} did you say that [friends of t_i] _{j} , you dislike t_j ? (subextraction from embedded topicalization)
 b. *Who _{i} did you say that [friends of t_i] _{j} t_j dislike you? (subextraction from subject)

In (9a) a constituent is extracted from a topicalized constituent. Attribution of the unacceptability in (9b) to the Raising Principle of course depends on an analysis in which the subject is taken to be raised from its clause.¹⁰

The main point about constraints such as these is that they are categorical. In contrast, Hofmeister et al. (2015) and Culicover and Winkler (2018) argue on the basis of experimental evidence that the unacceptability of so-called 'freezing' configurations is gradient and reflects processing complexity, determined by such factors as DEPENDENCY LENGTH of filler-gap chains and the INTERACTION of overlapping A' chains.

Regarding the first, in the string *read the book*, there is a minimal dependency between *the* and *book*, and a slightly longer dependency between *read* and *book*. Work such as Gibson (1998, 2000) has suggested that longer dependency distance correlates with processing complexity. As far as we know, there is no consensus on how to measure dependency length; several measures of dependency length have been proposed in the literature, including as a function of number of intervening words (Gibson 1998; Lewis and Vasishth 2005; Liu 2008; Liu et al. 2017; Temperley 2007), of complexity of branching structure (Hawkins 1994, 2004, 2014), and of number of new discourse referents (Gibson 2000). Research has shown that in general languages tend to minimize the distance between dependent elements, measured in terms of hierarchical structure (Futrell et al. 2015; Hawkins 1994, 2004, 2014; Liu 2008; O'Grady et al. 2003; Yadav et al. 2021).¹¹

Dependency length is added in Figure 4.

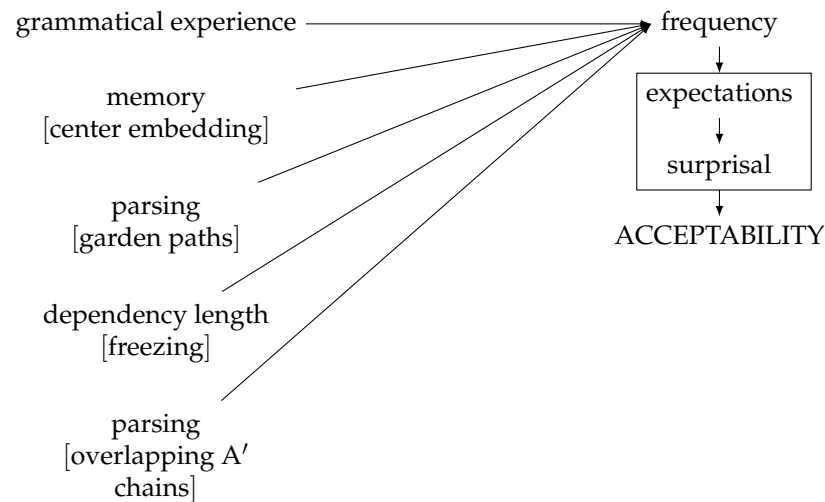
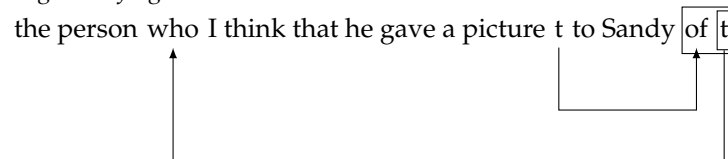


Figure 4. The logic of acceptability judgments for grammatical conditions, version 4.

4.2. Overlapping A' Chains

Regarding chain interaction, note that in the case of (7), for example, the configuration is that of ‘right surfing’ (10), where the tail of the extraposed constituent precedes the tail of the chain of the extracted *wh*-phrase in the linear order.¹²

(10) Right Surfing



Hofmeister et al. (2015) and Culicover and Winkler (2018) provide experimental evidence that the unacceptability of extraction from extraposed PP depends on the length of the A' chain and the extraposition chain. The acceptability of the A' chain alone is a linear function of the length of the dependency, as is the acceptability of PP extraposition alone. The acceptability of extraction from extraposition is determined by the sum of the two overlapping dependencies. Therefore there is no reason to believe that the most unacceptable cases are ungrammatical in a strict sense, to be ruled out by a syntactic constraint. Following the early insights of Miller and Chomsky (1963), the reasoning here presupposes that syntactic constraints as such are largely insensitive to quantitative properties of structures, such as the SIZE of a phrase, the NUMBER of embeddings or the LENGTH of a chain. If acceptability is sensitive to these factors, this is *prima facie* evidence that the source of the judgment is non-syntactic—plausibly related to working memory capacity.

Similar results were found for the freezing of Heavy NP Shift by Konietzko et al. (2018), as illustrated in (11). This is another case of RIGHT SURFING, where the trace of the constituent that appears in VP-final position contains the trace of the A' constituent.

- (11) a. You put [a picture of FDR]_j on the table.
 b. You put *t_j* on the table [a picture of FDR]_j.
 c. *Who_i did you put *t_j* on the table [a picture of *t_i*]_j?

The experimental results reported in Konietzko et al. (2018) suggest, again, that the unacceptability of extraction from the heavy NP is a function of the interaction of the overlapping chain dependencies, and not the configuration of the VP.

To the extent that multiple dependencies entail complexity, the model in Figure 3 leads us to expect that structures with multiple interacting chains will be progressively less frequent in a way that is inversely related to the total size of the interacting chains they contain. As a result, such structures are associated with high surprisal and, therefore, are expected to give rise to low acceptability. We summarize these results by adding the factor ‘parsing’ to Figure 4.¹³

Why multiple dependencies affect processing complexity is very much an active research question. The most explicit computational models that we are aware of that go beyond the formulation of constraints on parsers are those that appeal to interactions between activation and retrieval from memory, attentional focus, and activation decay (Lewis 1993, 1996; Lewis and Vasishth 2005; Lewis et al. 2006; van Dyke and Lewis 2003; Vasishth and Lewis 2006; Vasishth et al. 2019). No doubt a more fine-grained understanding of the processes involved in the computation of chain dependencies will shed considerably more light on the various phenomena that we have noted here.

4.3. Topic Islands

Topic island phenomena (Rochmont 1989) arguably reflect the interaction of chains in processing as well. Classical examples are given in (12).

- (12) a. * What_i does John think that Bill_j, Mary gave t_i to t_j ?
 b. * This is the man who_i that book_j, Mary gave t_j to t_i .
 c. * How_i did you say [that the car_j, Bill fixed t_j t_i]?
 d. * This book_i, I know that Tom_j, Mary gave t_i to t_j .
 (Rochmont 1989, p. 147)

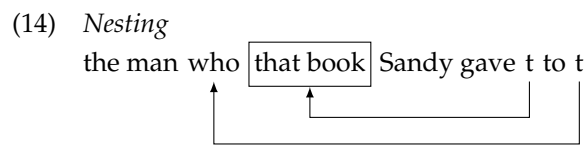
Rochmont’s account of the unacceptability of examples such as these relies on Chomsky’s (1973) Subadjacency condition, which blocks movement from a too deeply embedded position in the structure. Depth of embedding is determined by counting the number of barriers, where the notion ‘barrier’ is defined in terms of a variety of government called L-marking (Chomsky 1986).

An experimental study by Jäger (2018) confirms that extraction from embedded clauses in which topicalization has occurred is unacceptable. However, Jäger also demonstrates that topicalization alone is less acceptable than canonical SVO order in embedded clauses. Thus, it is plausible that the lower acceptability of embedded topicalization added to the processing cost of long A’ extraction is sufficient to account for the unacceptability of examples like (12).

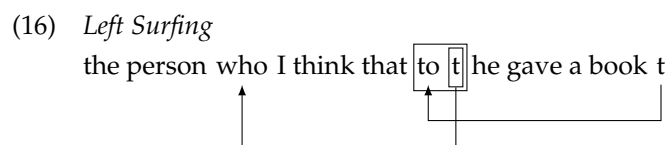
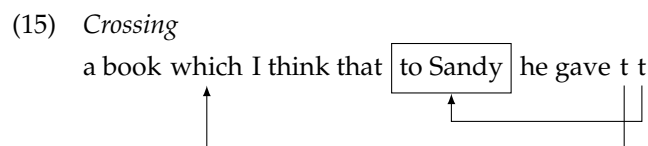
It is noteworthy that the examples in (12) involve overlapping chain interactions. What is topicalized in the embedded clause is an argument, and requires a trace in its canonical position. If we modify these examples as in (13) so that what appears in initial position in the embedded clause is a sentential adjunct (shown with underlining), acceptability increases. Crucially, a sentential adjunct can be interpreted as soon as it is encountered and does not have to form a chain.

- (13) a. ? What_i does John think [that at the concert, Mary proposed to sing t_i]?
 b. ? This is the man [who_i at the party, Mary insulted t_j].
 c. ? How_i did you say [that when he came home, Bill was feeling t_i]?
 d. ? This book_i, I know [that if the Times recommends it, Mary will buy t_i].

The chain interactions in (12) are different from those seen in the case of freezing. The latter are instances of right surfing, while the former are NESTING, illustrated in (14). In nesting, the fronted constituents are in reverse order to the traces that they form chains with, as shown in (14).



Like right surfing, nesting requires overlapping processing of two chains. Multiple chain processing is also required for CROSSING, illustrated in (15), and the LEFT SURFING configuration, illustrated in (16). In the more acceptable cases of crossing, the fronted constituents are in the same order as the traces that they form chains with, while in left surfing a constituent is extracted from a left extracted constituent.



Reasoning from the analogy of the freezing experiments, we expect that the processing of multiple overlapping chains to be more difficult than the processing of a single chain or of non-overlapping chains, and correspondingly more unacceptable. We expect the unacceptability to reflect the length of the overlapping chains. As suggested for nesting and crossing, the arrangement of the A' constituents with respect to their chains is also likely to play a role. Additional complications may arise when a preposition is stranded internally to another constituent, as in the case of left surfing illustrated in (16).¹⁴

To our knowledge, these factors have not been investigated systematically in the literature. Lewis (1993) proposed a computational model to account for the effects of multiple chains on processing complexity, but his model has not been further developed or brought to bear on the full range of chain interactions discussed here. While it is premature to rule out the possibility that there are grammatical constraints that account for the unacceptability of left surfing, crossing, and nesting, a processing explanation is promising and deserves a focused effort. For a review of recent proposals, see Chaves and Putnam (2020).

Another type of complexity associated with chain interactions is the extent to which the structure that the sentence processor assigns to a string faithfully reflects its semantic structure. This degree of CONGRUENCE determines how easily it is mapped to a semantic interpretation (Culicover and Nowak 2002). In part the ease of this mapping is determined by the extent to which constituents that are adjacent in the string correspond to semantic objects that form a larger semantic object. For example, an adjacent verb and NP in the string are more easily processed as a transitive predicate than a verb and a displaced NP. More complexity in processing would arise if parts of the NP were distributed to non-adjacent positions before and after the verb.¹⁵

Figure 5 reflects the contribution of congruence to complexity.

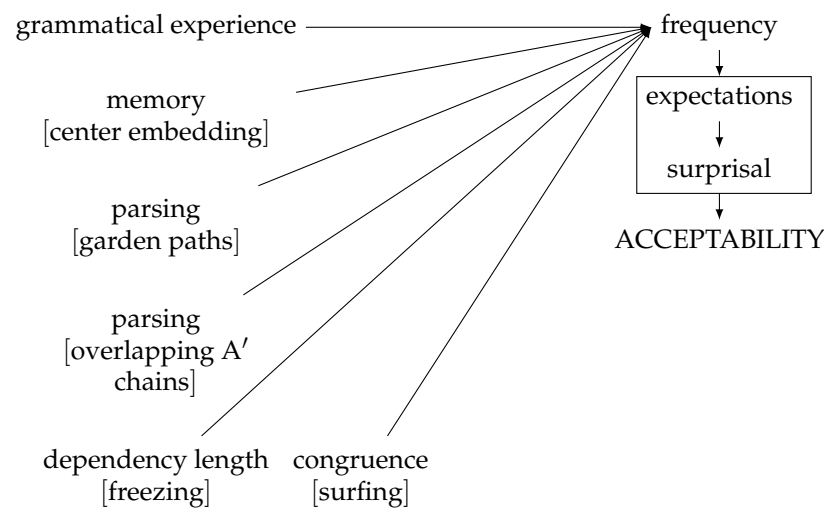


Figure 5. The logic of acceptability judgments for grammatical conditions, version 5.

4.4. Initial Non-Subjects in Zero-Relatives

Another way in which an A' chain might incur processing complexity is if speakers are unable to infer an appropriate structure on the basis of the cues provided by the overt string in which the chain is realized—i.e., if the string associated with the otherwise well-formed A' dependency lacks the kinds of overt signals that the processor relies on in order to parse correctly. A plausible instance of this is Jackendoff and Culicover's (1972) example in (5). This is also what happens in some instances of zero-relatives, as explored in Culicover (2013a). Consider the three relative clauses in (17).

- (17) *War and Peace* is
- a. a book which you should read.
 - b. a book that you should read.
 - c. a book \emptyset you should read.

These examples show that a relative may be introduced by a *wh*-form, *that* or zero (\emptyset). What we see in (18) is that an initial non-subject can occur in the first two, but not the zero-relative.

- (18) *War and Peace* is
- a. a book which if you have time you should read.
 - b. a book that if you have time you should read.
 - c. *a book \emptyset if you have time you should read.

Culicover (2013a) shows that the unacceptability seen in (18c) can arise in a number of other ways, as well. In (19a) there is an initial topicalized argument,¹⁶ in (19b) there is a initial negative constituent that triggers subject-aux-inversion, and in (19c) there is an initial predicate and stylistic inversion.

- (19) a. *He is a man liberty_j, we could never grant t_j to t_i . (Cf. ?He is a man that_i liberty_j, we could never grant t_j to t_i . (Baltin 1981))
- b. *He is a man under no circumstances would I give any money to t_i . (Cf. He is a man that_i under no circumstances would I give any money to t_i)
- c. *Detroit is a town in almost every garage can be found a car manufactured by GM. (Cf. Detroit is a town that in almost every garage can be found a car manufactured by GM.)

These, along with (18), illustrate four different constructions, with the initial constituent attached to a different position in the structure. The initial subordinate clause is very high up in the structure, and can be followed by a topicalized argument, as in (20).

- (20) If you have time to read a book, *War and Peace* you should definitely read.

The initial negative constituent may follow a topicalized argument, and would therefore appear to be attached lower.

- (21) To Sandy, not a single dollar would I give!

The initial predicate is arguably in Spec,IP, the conventional subject position (Culicover and Levine 2001).

Thus, there does not appear to be a single syntactic configuration that could be identified in a single syntactic constraint that accounts for the unacceptability of all of these cases. Given the diversity of syntactic configurations observed here, there would have to be a separate constraint for each case, which is clearly not an optimal account. There is a common factor, however: there is a non-subject or non-NP subject in the initial position in the zero-relative clause. As a consequence, in a zero-relative there is no reliable marker of the initial portion of the relative clause. As Culicover (2013a) argues, while zero-relatives with initial NP subjects are quite standard, non-NPs in initial position in relatives are rare. Thus, when the complementizer *that* is absent and there is a non-subject or non-NP in initial position, the processor has no way of reliably identifying and projecting the relative clause structure. We suggest that the unacceptability of topicalization in zero-relative clauses reflects processing complexity, not a set of grammatical constraints.

The factor at play in this case has to do with the prediction of syntactic structure in the course of processing. As suggested in the parsing literature (Hale 2001, 2003; Levy 2005, 2008, 2013; Levy and Jaeger 2007), the sentence processor makes predictions about the future trajectory of the parse based on frequency. In this case, surprisal reflects how expected a particular syntactic category is on the basis of the structure that has already been built. This notion of expectation covers cases such as certain garden paths, where on the basis of the currently parsed string—the PREFIX—the immediately processed constituent is strongly unexpected, leading to high surprisal. An example is (22), where the prefix *without her* creates the expectation that *her contributions* is the NP complement of the preposition.¹⁷

- (22) Without her contributions failed to come in.
(Pritchett 1988, p. 543)

There is strong evidence that the human sentence processor is continuously engaged in predicting words and structures (for a recent review, see Kuperberg and Jaeger 2016). A plausible model of such prediction is one in which at any point in the process, every possible well-formed continuation of the prefix is assigned a probability that reflects its likelihood (van Schijndel et al. 2013). In cases where the actual continuation deviates radically from what is most probable, a garden path occurs. However, when the flux of expectation is not dramatic, there is still variation in processing activity due to surprisal (Shain et al. 2020). It appears, therefore, that unacceptability judgments can be associated with levels of surprisal that exceed some threshold. (See Fodor's (1983, p. 190) discussion of "markedness" in GPSG parsing and Ross's (1987, p. 310) discussion of the accumulation of "losses in viability" for early proposals along these lines.)

In Figure 6 we add prediction of structure to the list of factors.

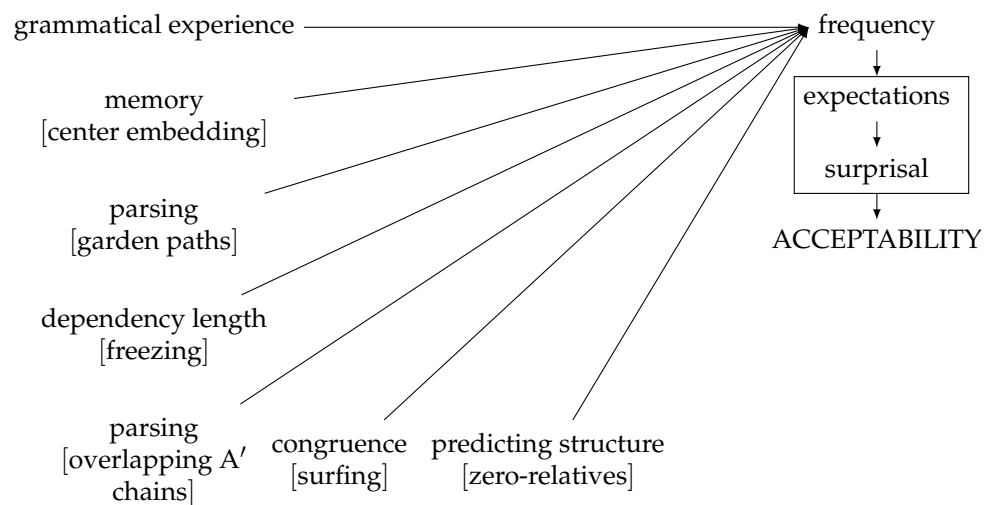


Figure 6. The logic of acceptability judgments for grammatical conditions, version 6.

5. Discourse and Information Structure

In order to communicate efficiently, speakers must provide hearers with just enough information to meet the particular goals of their conversational interaction (Grice 1975), which implies, among other things, identifying the referents the discourse is about (Ariel 1990, 2001, 2004; Roberts 2012). Excessive or irrelevant information leads to redundancy and puts the hearer through unnecessary effort, which increases processing complexity. Too little information leads to ambiguity, which also increases complexity. The management of several discourse referents at once can also lead to processing difficulties (Arnold and Griffin 2007; Gibson 2000; Kluender 2004; Warren and Gibson 2002). We argue below that these factors are plausible sources for the surprisal effect in several unacceptable A' extractions.

5.1. Weak Crossover

We start by looking at phenomena that have to do with the computation of reference in discourse. Notably, the relevance of discourse reference to phenomena covered by syntactic constraints was already argued for in detail by Kluender (1998).

The first phenomenon, weak crossover (WCO), is exemplified by the unacceptability of examples such as (23b), first observed by Postal (1971).

- (23) a. Who_i *t_i* loves his_i dog?
b. *Who_i does his_i dog love *t_i*?

Culicover (2013d), using data from Levine and Hukari (2006), argued that WCO violations such as (23b) do not reflect a principle of grammar. While the first such principle to be proposed was the Bijection Principle of Koopman and Sportiche (1983), the point is a general one: unacceptability of WCO shows the effect of referentiality and resolution of thematic assignment of chains in processing the linear string, not a syntactic constraint.¹⁸

Several extragrammatical factors appear to be at play. One factor is the discourse accessibility of the *wh*-phrase. The more specific the reference of the *wh*-phrase is, the more natural it is to refer to it with a pronoun, as seen in (24).

- (24) a. *Who_i did his_i dean publicly denounce *t_i*?
b. ?? Which professor_i did his_i dean publicly denounce *t_i*?
c. ? [Which distinguished molecular biologist that I used to work with]_i did his_i dean publicly denounce *t_i*?

Moreover, as has been often noted, the WCO configuration with a relative clause is reliably more acceptable than precisely the same configuration with a question. Compare (25) with (24b).

- (25) I plan to interview the professor who_i his_i dean publicly denounced t_i .

And an appositive relative is if anything even more acceptable (Lasnik and Stowell 1991); cf. (26).

- (26) I plan to interview Professor Smith_i, who_i his_i dean publicly denounced t_i .

This difference can be understood in terms of specificity as well, insofar as the head of the relative clause provides more specific information about the identity of the referent associated with the pronoun (Pesetsky 1987, 2000; Wasow 1979). The question, of course, is why this should be the case.

A second factor is whether the *wh*-phrase has a θ -role at the point in the processing of the sentence at which the pronoun is encountered. In (27a,c) the *wh*-phrase lacks a θ -role at the pronoun in the first conjunct, which contains the bound pronoun. However, in (27b,d) the *wh*-phrase gets a θ -role in the first conjunct and the pronoun is in the second conjunct.

- (27) a. ? Who_i does his_i mother love t_i and Sandy dislike t_i ?
 b. Who_i does Sandy dislike t_i and his_i mother love t_i ?
 c. ? a person who_i his_i mother loves t_i but Sandy dislikes t_i
 d. a person who_i Sandy dislikes t_i but his_i mother loves t_i

While the examples with the WCO violation in the first clause are somewhat marginal, those with WCO in the second clause are unobjectionable. Again, the question is why.

These factors are reducible to the degree of ACCESSIBILITY of the discourse representation corresponding to the *wh*-phrase at the point where the pronoun is encountered.¹⁹ Accessibility is understood as a property of non-linguistic mental representations that determines their ease of retrieval in real-time processing (Arnold 2010). In the case of discourse referents, accessibility is plausibly a consequence of *predictability*: i.e., a referent is more accessible to the extent that it is more likely to be mentioned in the context at hand (Arnold 2010; Arnold and Tanenhaus 2011; Givón 1983).

We noted above that economy in referential processing seems to favor an inverse correlation between the accessibility of a discourse referent and the amount of information conveyed by the expression used to refer to it. As a result, less informative NPs (e.g., pronouns) are optimal candidates for retrieving highly accessible referents and more informative NPs (e.g., names, definite descriptions) are optimal candidates for retrieving less accessible referents (Almor 1999, 2000; Almor and Nair 2007; Ariel 1988, 1990, 1991, 1994, 2001, 2004). Different types of NP function, thus, as specialized markers for different degrees of accessibility. Whenever speakers fail to match their choice of NPs to the degree of accessibility of the referent they intend to pick out, processing complexity ensues.

Personal pronouns, like the ones we see in the WCO examples, function as HIGH ACCESSIBILITY MARKERS – i.e., they must be paired with discourse referents that are highly accessible in the contexts where they appear. This occurs because pronouns are informationally impoverished; the only kind of information pronouns carry is their specification for features such as number, person, and gender (Almor 2000; Almor and Nair 2007; Ariel 2001; Bouchard 1984; Gundel et al. 1993; Levinson 1987, 1991).

As an illustration consider the example in (28):

- (28) Charlie and Frank finished watching a movie. Charlie was the one who picked it out. He didn't like it.

The personal pronoun *he* can successfully refer to Charlie in (28), because Charlie is a unique and highly accessible referent at the point where the pronominal is encountered. The discourse referent anchored to Frank is much less accessible, and, therefore, it would be odd for a speaker to use an uninformative form like *he* to refer to Frank in that context.²⁰

We suggest that this same factor contributes to the unacceptability of typical WCO structures like (23b): the referent of the *wh*-phrase is not accessible enough to be retrievable

by a high accessibility marker such as a pronoun at the point where the pronoun is encountered. The mismatch between the low degree of accessibility of the discourse referent and the high accessibility marking status of pronouns contributes to processing complexity (Almor 1999, 2000). This leads to lower frequency of the WCO configurations in speakers' prior experience, which, in turn, yields higher levels of surprisal.

Gernsbacher (1989) showed in a series of experiments how various linguistic factors may enhance the relative accessibility of discourse referents. One of her major findings is that more explicit expressions (i.e., low accessibility markers, in the sense of Ariel 1990) increase the accessibility of their mental representations more than less explicit expressions. In fact, as Ariel (2001, p. 68) points out, there is an inverse relationship between an NP's degree of accessibility marking and its potential to boost the future accessibility of its discourse referent: "the lower the accessibility marker used, the more enhanced the discourse entity coded by it will become".

What all of the amelioration effects in (24)–(27) share is that they increase the accessibility of the discourse representation corresponding to the *wh*-phrase in precisely this way. When the discourse representation of the *wh*-phrase becomes more accessible, subsequent retrieval by a high accessibility marker such as a pronoun becomes more acceptable. For example, in (27), we may think of the θ -role as contributing to more information about the referent of the *wh*-phrase, which, in turn, enhances the accessibility of the mental representation it corresponds to. Increasing specificity has a similar effect in (24b,c), (25), and (26). By providing a more adequate match between the accessibility status of the antecedent and the pronoun, these ameliorated WCO violations are less complex than the unacceptable cases. They are, therefore, expected to be more frequent and to be associated with lower degrees of surprisal, enhancing acceptability.

In Figure 7 we add discourse accessibility to the list of factors.

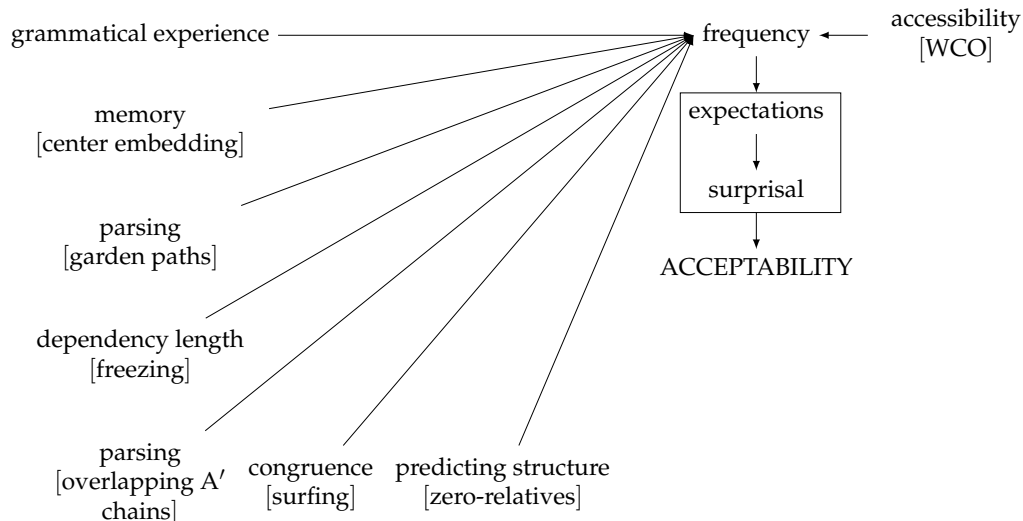


Figure 7. The logic of acceptability judgments for grammatical conditions, version 7.

5.2. The Uninvited Guest

We consider next the role that referential processing plays in the unacceptability of extraction from subjects, which conventionally falls under the Sentential Subject Constraint of Ross (1967), the Subject Condition of Chomsky (1973), and related formulations. The examples in (29) illustrate:

- (29) a. * a person who (not) shaking hands with *t* would really bother Sandy (gerund)
 b. * a person who us shaking hands with *t* would bother Sandy (gerund with pronominal subject)
 c. * a person who Terry shaking hands with *t* would bother Sandy (gerund with referential subject)

- d. * a person who Terry's shaking hands with *t* would bother Sandy (gerund with possessive)
- e. * a person who that Terry shakes hands with *t* would bother everyone (*that* clause)
- f. * a person who to shake hands with *t* would bother Sandy (infinitive)
- g. * a person who for $\left\{ \begin{array}{c} \text{us} \\ \text{Terry} \end{array} \right\}$ to shake hands with *t* would bother Sandy (*for-to* infinitive)
- h. * a person who offensive jokes about *t* would bother Sandy? (NP)
- i. * a person who the fact that Sandy shakes hand with *t* would bother Terry (sentential complement of N like *belief*, *claim*)

In spite of the unacceptability of examples such as these, there is a substantial literature that demonstrates that the extraction from subjects is grammatical and varies in acceptability according to a number of factors, including lexical choice and repeated exposure (Abeillé et al. 2020; Chaves 2013; Chaves and Dery 2019; Chaves and Putnam 2020; Kluender 2004; Polinsky et al. 2013). Culicover and Winkler (2022) argue that in some cases, the unacceptability of extraction from subjects reflects the complexity of such extraction combined with a novel referential expression in the predicate, which they call the UNINVITED GUEST. In (29a–c), for example, the Uninvited Guest is *Sandy*.

In terms of the general model in Figures 4–7, when the complexity of a subject extraction is coupled with the complexity afforded by having to process an additional referential argument, we get a more complex and, therefore, less frequent structure, which carries a high degree of surprisal. On the account proposed by Culicover and Winkler (2022), the amelioration effect we see in connection to parasitic gaps is a consequence of reducing complexity in referential processing by omitting an extra referential argument (the Uninvited Guest). This effect can be seen in (30a–c), compared with (29a–c). The notation *pg* indicates a parasitic gap.

- (30) a. a person who (not) shaking hands with *pg* would bother *t*
- b. ? a person who us shaking hands with *pg* would bother *t*
- c. ? a person who Terry shaking hands with *pg* would bother *t*²¹
- d. * a person who Terry's shaking hands with *pg* would bother *t*
- e. * a person who that Terry shakes hands with *pg* would bother *t*
- f. * a person who to shake hands with *pg* would bother *t*
- g. * a person who for $\left\{ \begin{array}{c} \text{us} \\ \text{Terry} \end{array} \right\}$ to shake hands with *pg* would bother *t*
- h. * a person who the fact that Sandy shakes hands with *pg* would never bother *t*

The fact that the parasitic gap configuration is not sufficient to render all of these extractions from subjects acceptable suggests that the unacceptability here is not a matter of grammaticality per se. It is simply not the case that the presence of an extra gap elsewhere in the sentence provides a syntactic means to make subject extractions automatically grammatical, as proposed in the syntactic theories of parasitic gaps (Chomsky 1986; Frampton 1990).

This observation is further supported by the fact that there are many acceptable extractions from subject in corpora in sentences that do not contain an extra gap that could syntactically license the gap within the subject. A few examples are given in (31).

- (31) a. ... with them—the people who love you and who you love, who you laugh with and **who spending time with** is enriching rather than exhausting.
- b. More than anything though, The Joker is a fascinating character **who spending time with** is a treat.

- c. There are some things **which fighting against** is not worth the effort. Concentrating on things which can create significant positive change is much more fruitful.
- d. That might be a good idea, the only way I could get her contact information would be through my SM though, **which asking for** would become a fiasco.

Chomsky (2008) attributes the difference in acceptability of extraction from subject to the underlying position of the subject. On Chomsky's account, if an NP is the underlying complement of a verb, extraction from subject is possible, but if it is an underlying subject, it is not. Passives would all be of the first type, as would unaccusatives, while unergatives would be of the second type. In this way, Chomsky preserves the view that subjects are islands in the grammatical sense.

However, Culicover and Winkler (2022) provide corpus evidence that extraction from subject may be acceptable even if the predicate is transitive, if the NP in the predicate does not denote a novel discourse referent, that is, if it is not an Uninvited Guest. This NP is an 'Invited Guest'—the discourse referent it invokes is highly accessible in the discourse context, implying that it carries less cost for referential processing. In every instance, the Invited Guest that has the discourse status 'given' or 'c-construable' (Rochemont and Culicover 1990), is by virtue of being part of the common ground.

A sample of Invited Guest examples is given in (32)–(36). When the object NP refers to an individual, that individual is always immediately available in the discourse, i.e., the speaker (32), the addressee or generic *you* (33), or a third party who is being discussed (34). Where the object NP does not refer to a person, it typically refers to a property of the general common ground such as *the day*, *my life* (35). The only apparent exceptions are *your playing*, *your patriotism* and *the postulated meaning* in (36), which bear on the topics of the discourse, and therefore have the discourse status 'given'.

(32) First person

- a. I've found people **who spending time with** isn't an exhausting experience and actually gives me a boost.
- b. However, there have been girls **who spending time with** and going places [sic] because we love them have made us happy.

(33) Second person

- a. In your head you're able to let the mind wander to all sorts of corners, day dreaming about the happy things you hope might happen one day, the good times you've enjoyed, and the people **who spending time with** makes you feel good.
- b. there are some people **who talking to** gives you a sort of high
- c. ... Deathstroke, and some other important characters, such as Alfred (**who talking to** gives you more ...), James Gordon, and Barbara Gordon.
- d. The purpose of a relationship (in my mind) is to find someone **who spending time with** makes you happier than you would be on your own, this guy's behaviour does not represent that in my opinion and it certainly doesn't sound like minor character traits that you may be able to change with time because it doesn't sound like he's at all willing to change.

(34) Third person

- a. But even if that were so, it would seem that he had at least one person in his life **who spending time with** and whose love made him feel pure bliss.
- b. ... But there was one part of Tim **which to describe as typical** rather undersells him, although it is an aspect of his being to which we would all aspire, because Tim's integrity—his sense of honour, his honesty, his deep sense of decency—was special and it was rare.

- c. Until Marinette, the shy classmate who tended to word-vomit in his vicinity and otherwise cease being able to function like a normal human for reasons he had yet to understand (and **which asking about** would get him sly looks from Alya and concerned looks from Nino), was there.
- (35) Common attribute
- a. Do you have vendors you work with that you truly enjoy? People who work hard for you, do a great job and **who spending time with** makes the day go by happily and productively?
 - b. Today, there was this person **who talking to** would make my life exponentially more complicated and fucked up.
- (36) Sentence topic
- a. Definitely the most important advice is to join an orchestra. You will not only meet likeminded individuals **who spending time with** will improve your playing, but friends and connections for life.
 - b. I desire that you accept of no offers of transportation from officials who deprived you of the very food, in some cases, which was necessary to supply your pressing wants, and who couple their offers of a free passage with conditions **which to accept** would cast a stain upon your patriotism as Irishmen and as free citizens, who are bound to sympathize with every struggling nationality.
 - c. For purposes of Proof the important distinction lies solely between assertions capable of denial with a meaning, and those **which to deny** would contradict the postulated meaning.

The data presented by Culicover and Winkler (2022) thus supports the position that there is no grammatical constraint that blocks extraction from subjects. Rather, the extraction varies in unacceptability due to a number of factors ultimately related to referential processing. When the extraction is marginally acceptable and the Uninvited Guest is absent, acceptability associated with parasitic gaps results. However, when the Uninvited Guest is present, it adds complexity to existing complexity, resulting in a judgment of unacceptability.

The Uninvited Guest analysis adds support to the claim that there does not appear to be strong evidence that non-local unacceptability in these cases is due to a grammatical constraint, although the question of why extraction from subject is complex remains open. One possible answer is that neither the *wh*-phrase nor the subject have a θ -role at the point at which the trace of the *wh*-phrase is encountered. We already saw in the case of WCO that interpreting an unresolved *wh*-chain appears to be relatively costly. Furthermore, Frazier and Clifton (1989); Kluender (2004); Kluender and Kutas (1993a) provided experimental evidence that initiating processing of an embedded sentence has a processing cost. Gibson (1998) showed that processing of referential expressions, including reference to specific times, has a cost when a *wh*-chain is not resolved. Thus it is not surprising that the most acceptable extraction from subject is from a gerund such as *shaking hands with NP*, less acceptable extraction is from a gerund with a subject such as *Terry shaking hands with NP*, and still less acceptable extraction is from a tensed S such as *that Terry shakes hands with NP*.

Figure 8 adds the processing of discourse referents to the list of factors.

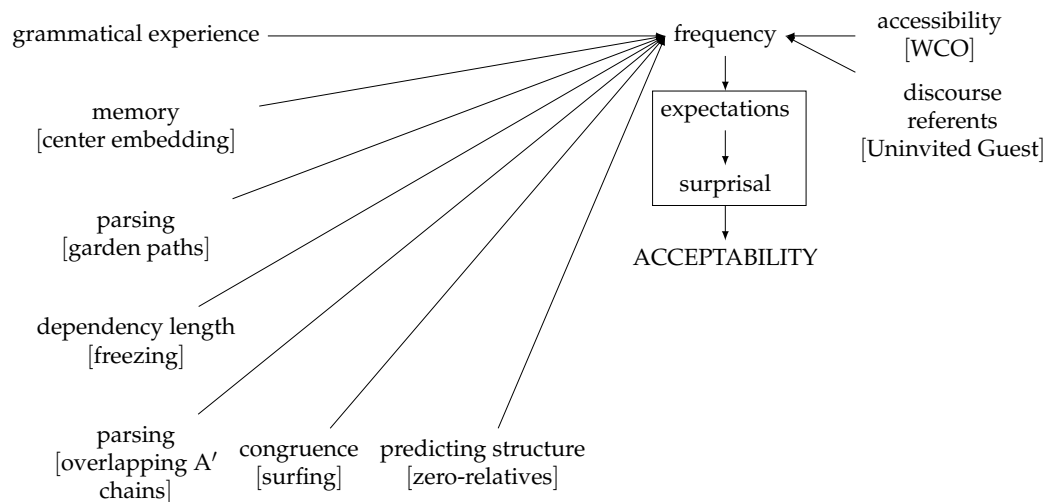


Figure 8. The logic of acceptability judgments for grammatical conditions, version 8.

5.3. Information Structure

One of the reasons to extend the RUH beyond sentence processing in a narrow sense is that there are cases in which it appears that an information structure mismatch contributes to judgments of unacceptability. The mis-management of information flow is, of course, also connected to processing complexity in a more holistic sense, having to do with the discourse as a whole.

Discrepancies between the at-issue content of utterances and the Question Under Discussion (QUD), as Roberts (2012) describes them, can cause processing difficulties (De Kuthy and Konietzko 2019; Konietzko et al. 2019). To take a simple example, the sentence (37b), while well-formed, is an inappropriate answer to the question preceding it, which functions as the QUD in that particular context. (Capitalization marks prosodic accent (focus).)

- (37) Who ate the pizza?
 a. SANDY ate the pizza.
 b. # Sandy ate the PASTA.

It is likely that such mismatches fall under the general category of surprisal, but whether they recruit the same resources as garden paths and other cases that involve structure as well as interpretation is an open question.

There is evidence that information structure mismatches of this sort also play a role in acceptability judgments in extraction constructions. We cite two studies that demonstrate this. First, Culicover and Winkler (2018), following Winkler et al. (2016), observe that the acceptability of extraction from the German *was-für* construction is higher if extraction is from a focus. Compare the examples in (38)/(39), due to Müller (2010, p. 61(36)).

- (38) *Was_i haben [DP *t_i* für Bücher] [DP den FRITZ] beeindruckt?
 what have [DP *t* for books.NOM] [DP the Fritz.ACC] impressed
 ‘What kind of books impressed Fritz?’
- (39) Was_i haben [DP den Fritz]_j [DP *t_j* für BÜCHer] *t_j* beeindruckt?
 what have [DP the Fritz.ACC] [DP *t* for books.NOM] *t* impressed
 ‘What kind of books impressed Fritz?’

On Müller’s account, *was für Bücher* in (38) is frozen, because it is last-merged in the specifier-position of vP, and hence blocks extraction. However, it is not frozen in (39), because the movement of *den Fritz* over it by scrambling removes the offending configuration that froze it—this is what Müller calls ‘melting’.

However, Winkler et al. (2016) note that in German, the immediate preverbal position is a focus position (Haider and Rosengren 2003; Höhle 1982; Reis 1993; Selkirk 2011; Truckenbrodt 1995, among others). Extraction from focus in the German Mittlefeld has been independently shown to be more acceptable than extraction from non-focus (Bayer 2004). Thus, (38) is unacceptable because *Bücher* is not a focus, while (39) is more acceptable. They show that judgments of extraction from immediate preverbal and scrambled position can be manipulated by changing the context to change the focus, which rules out an explanation in structural terms.

Second, Konietzko (forthcoming) explores in detail PP extraction from subjects in German. He shows that such extraction is also sensitive to information structure and context—extraction from a focus is more acceptable than extraction from a non-focus. Konietzko shows as well that PP extraction from NP in German is sensitive to the argument type of the NP. Extraction from unaccusative subjects is best, followed by unergative subjects, transitive objects, and transitive subjects. A summary of Konietzko’s results for extraction of *von wem* ‘by whom’ appears in Figure 9.

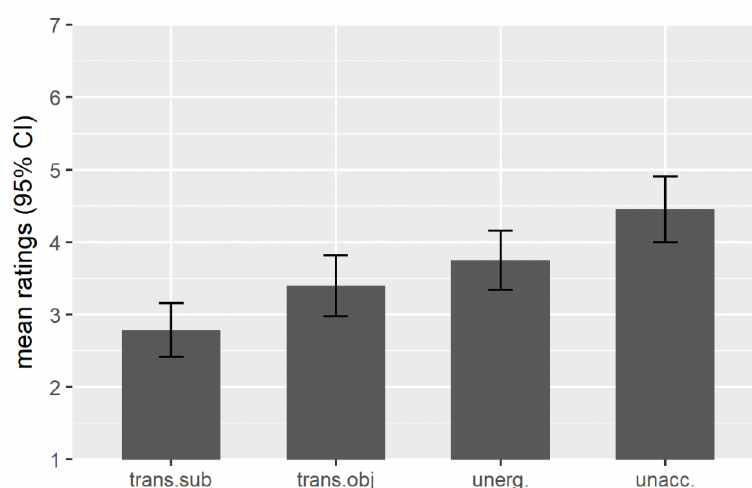


Figure 9. Acceptability of extraction of *von wem* ‘by whom’ from NP in German (Konietzko forthcoming).

Extraction of *über wen* ‘about whom’ from an NP shows sensitivity to the argument type as well (Figure 10). Most acceptable is extraction from the subject of a passive, followed by subject of unaccusative, transitive, and psych-verb. The differences between these types of subjects have been dealt with in mainstream generative grammar in derivational terms. Konietzko concludes that there is a basis for attributing the unacceptability of at least some cases of extraction from subject to structural configuration.

Note that *wh*-constituents are canonically associated with the status of discourse *foci* (Culicover and Rochemont 1983). What happens in (38)–(39) as well as in the cases of PP extraction from subjects examined by Konietzko (forthcoming), is that full acceptability only occurs if the focus implied by the *wh*-construction is coherent with the focus associated with the structural position from which extraction takes place (the immediate pre-verbal position in (39)).

What we see in (38) is a non-optimal alignment between the information structure status of the *wh*-phrase and *den Fritz*, both of which are assumed to be foci by default. The suggestion of multiple conflicting foci arguably makes the example harder to process than (39). As a result, structures like (38) are expected to be less frequent, to give rise to higher surprisal and, correspondingly, lower acceptability.

Based on the observations in this section and Section 4, we complete our picture of the sources of unacceptability in Figure 11.

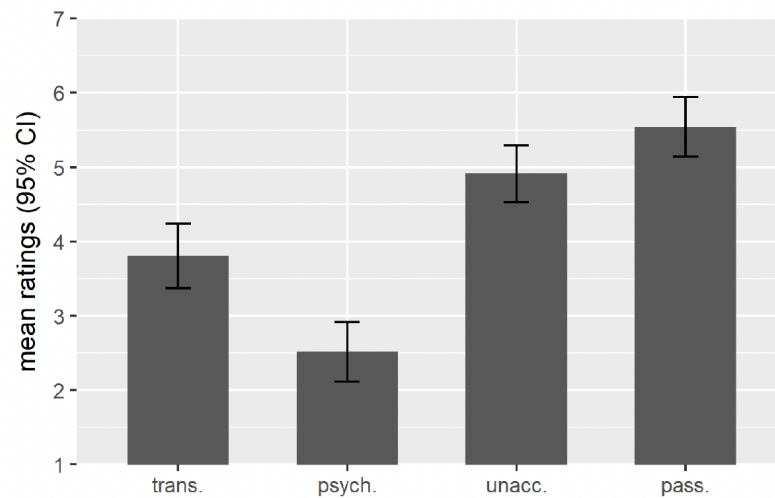


Figure 10. Acceptability of extraction of *über wen* ‘about whom’ from NP in German (Konietzko forthcoming).

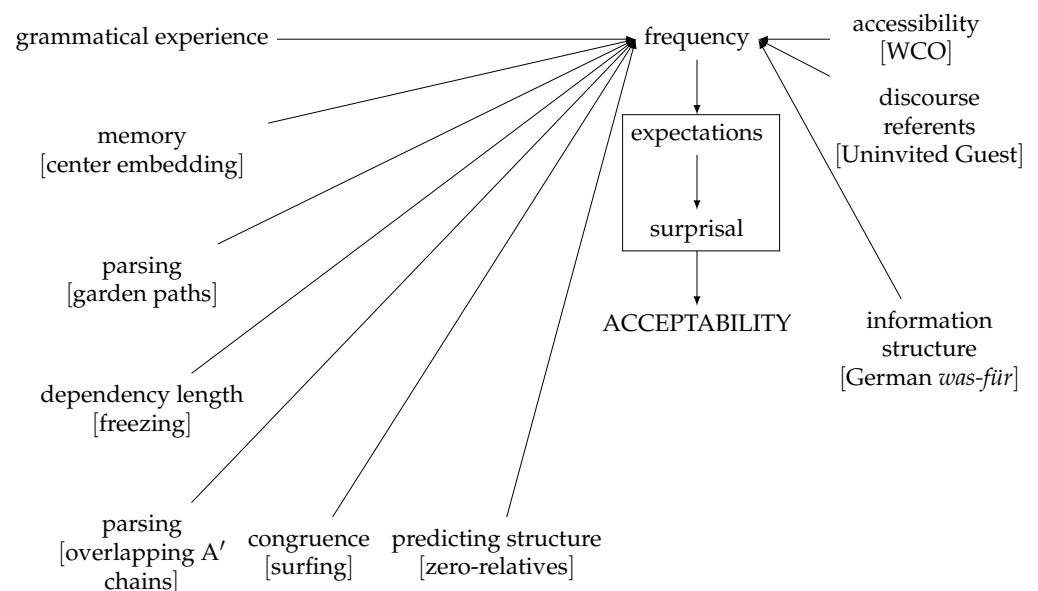


Figure 11. The logic of acceptability judgments for grammatical conditions, final version.

6. Processing Factors and Problematic Cases

As mentioned above, there are some classical island constraints that do not seem to be so readily amenable to a non-syntactic treatment. In this section, we examine specifically the Coordinate Structure Constraint and the Left Branch Condition. The phenomena covered by these constraints are *prima facie* counterexamples to the strongest interpretation of our hypothesis. We argue that, while there are still many open questions, there is suggestive evidence that these principles are still compatible with the ERUH.

We start by noting that it is possible that the grammar itself is a source of low frequency in a way that does not imply the existence of non-local constraints. A plausible case for this can be made for the Coordinate Structure Constraint, stated in (40) in a form that incorporates the familiar across-the-board (ATB) exceptions:

(40) *Coordinate Structure Constraint* (Ross 1967, p. 89)

In a coordinate structure, (a) no conjunct may be moved, (b) nor may any element contained in a conjunct be moved out of that conjunct unless the same element is moved out of both conjuncts.

Following previous work (Grosu 1973; Oda 2017; Pollard and Sag 1994), we distinguish between the CONJUNCT CONSTRAINT (clause (a) in (40)) and the ELEMENT CONSTRAINT (clause (b) in (40)). The former is illustrated in (41a) and the latter in (41b):

- (41) a. *Who_i did you see [Joanne and *t_i*] yesterday.
 b. *Which book_i did you say that [Amy wrote *t_i* and Harry bought the magazine]?

Though there are numerous counter-examples to (40b) which suggest that it might be reduced to a discourse-level principle (Goldsmith 1985; Kehler 1996; Kubota and Lee 2015; Lakoff 1986), (40a) seems to be a solid generalization about how coordination works in various languages (Chaves and Putnam 2020).²²

There are, however, several alternative explanations for the robust effect illustrated in (41a) that do not involve a non-local grammatical constraint on extraction. As many authors point out, this effect follows automatically from two independently motivated proposals: the traditional analysis of coordinating conjunctions as non-heads (Bloomfield 1933; Borsley 2005; Chaves 2007; Gazdar 1980; Gazdar et al. 1985; Pesetsky 1982; Ross 1967) and the traceless account of filler-gap dependencies that is the hallmark of HPSG since the mid-1990s (Bouma et al. 2001; Chaves 2020; Chaves and Putnam 2020; Ginzburg and Sag 2000; Pickering and Barry 1991; Pollard and Sag 1994; Sag 1997; Sag and Fodor 1994).

The former idea is motivated by the basic observation that the distribution of a coordinate phrase is mainly determined by that of its conjuncts (a conjunction of NPs functions like an NP, a conjunction of VPs functions like a VP, etc.). The latter idea, in turn, is based on the hypothesis that unbounded dependency gaps are introduced by heads, rather than by phonologically null constituents (i.e., traces). This proposal requires a lexical rule which allows a head to omit one of its arguments from surface realization while at the same time introducing a corresponding gap in its argument structure. The general point is the following: If A' gaps are not syntactic constituents, but are licensed as syntactically unrealized arguments of a head *via* a lexical rule, then coordinating conjunctions, *qua* non-heads, will not be able to co-occur with gaps.

An alternative account of the Conjoint Constraint that does not presuppose a traceless theory of extraction is suggested by Levine (2017, pp. 317–18) and Kubota and Levine (2020, pp. 302–3). They argue that the effects of (40a) can be derived from a prosodic restriction on coordinate structures requiring that each coordinated phrase contain at least one stressed syllable (see also Zwicky 1986). This is motivated by the observation that phonologically reduced cliticized pronouns cannot occur in coordinations like (42):

- (42) I don't know what happened to Taylor, but it's been years since I heard from
 Sandy { or him }
 { *or'm }.

Since extraction gaps are never phonologically realized, they cannot bear stress on their own. Therefore, in the context of NP coordinations, they cannot avoid violating this prosodic constraint.

Regardless of which theory is ultimately correct, both traceless and prosodic accounts derive the empirically robust part of the Coordinate Structure Constraint without appealing to a non-local grammatical constraint. These accounts explain the effects of (40) by means of what amount to LWFCs, thereby preserving the ERUH. The traceless theory appeals to the nature of the rule that establishes extraction gaps and the prosodic account appeals to a constraint on the prosody of the local sisters of coordinators.

Another of the classical island constraints that has resisted analysis as a consequence of non-syntactic factors is Ross's (1967) Left Branch Condition, stated in (43) and illustrated in (44):

- (43) *Left Branch Condition (LBC)* (Ross 1967, p. 207)
 No NP which is the leftmost constituent of a larger NP can be reordered out of this NP by a transformational rule.

- (44) a. *Whose_i did you read [NP *t_i* book]?
 b. *His_i, I don't think you liked [NP *t_i* food].
 c. *How much_i did she earn [NP *t_i* money].

In Ross's formulation, the LBC blocks the extraction of the left branch of an NP, and requires that the phrase be pied-piped. Ross also noted that the LBC appears to be more general, and extends to examples such as (45). On the basis of such cases, Gazdar (1981) formulated a GENERALIZED LEFT BRANCH CONDITION, whose purpose is to block extraction of *any* element to the left of a lexical head (see also Emonds 1985).

- (45) a. *How_i is Sandy [AP *t_i* tall]? (Cf. How tall is Sandy?)
 b. *[How big]_i did you buy [AP *t_i* a house]? (Cf. How big a house did you buy?)

Chaves and Putnam (2020, pp. 196–200) point out that their traceless account of movement also derives these effects. If gaps do not originate as traces, but on the argument structure of heads, elements that cannot be construed as arguments of a head (determiners and other pre-nominal specifiers), will not be able to appear as gaps—i.e., they are predicted to be unextractable.

This strategy of using the rule that introduces gaps to derive the LBC faces challenges. Chief among these is the fact that, as Ross (1967 pp. 236–38) himself recognized, there are counterexamples to even the more restrictive statement of the LBC in (43) in languages like Russian and Latin.

- (46) Čuju_i ty čitaješ [NP *t_i* knigu]?
 whose you read book
 'Whose book are you reading?'
 (47) Cuius_i legis [NP *t_i* librum]?
 whose read.2SG book
 'Whose book are you reading?'

The fact that the LBC can be systematically violated in some languages suggests that it should be handled with a different strategy from the Conjunct Constraint, which is basically exceptionless. In particular, we certainly do not want to derive it from the very mechanism that builds A' chains like Chaves and Putnam (2020) do, as this would either make wrong predictions about (46) and (47) or force us to adopt otherwise unmotivated structures for these languages.²³

Thus, in spite of the robustness of the LBC, there are reasons to think, with Ross, that it is not a universal constraint on extraction. There is additional evidence to support this hypothesis. First, as has been recognized for some time, extraction of a subject (widely thought of as a left branch position) is acceptable in English (48).

- (48) Who_i do you believe [_S *t_i* will win]?

As Grosu (1974, p. 309) observes, extraction of a possessive NP is impossible even when it is not on a left branch, as in (49) (compare with (44a)).

- (49) a. *Whose_i did you read [NP some books of *t_i*]? (Cf. You read some books of Susan's.)
 b. *Your wife's_i, I met [NP an uncle of *t_i*]. (Cf. I met an uncle of your wife's.)

These last examples suggest that the problem is not with left branch extraction per se. It is reasonable to conclude, then, that there is no grammatical constraint along the lines of Ross's LBC or its generalized variant.

The explanation for the ungrammatical examples in (44)–(49) remains unclear, of course. That said, the ungrammaticality of (49a) and (49b) suggests that the problem is that the A' constituent is by default processed as a phrasal argument with an elided

nominal head, e.g., [_{NP} *whose* [_N \emptyset]]. Such an analysis renders cases such as (44a) and (49a) unparseable, since there is no suitable gap for the A' chain and no suitable parse of the NPs [*t book*] and [*some books of t*]. Something similar plausibly applies to the other cases: e.g., there is a tendency to parse the displaced constituent at the left edge of the NP in (44b) as [_{NP} *his* [_N \emptyset]] (as in *I liked most of the food they brought to the party, but his_i I did not like t_i*), and, in (44c), as [_{NP} *how much* [_N \emptyset]] (as in *How much_i did she earn t_i?*).

The general principle at work here seems to be a preference for parsing strings in A' positions as full phrasal projections. This gives rise to a garden-path effect when the speaker encounters an NP missing a left branch. Whether this idea is on the right track, and whether it can be extended to all other cases handled by the LBC is a question that we leave open here.

7. Summary

Let us summarize. For almost every constraint on extraction that has been noted in the literature, including classical strong islands, we have suggested that it is possible to identify a plausible non-syntactic cause or causes. For the single case where a non-syntactic cause seems implausible (the Conjunct Constraint), a purely local well-formedness condition seems to be sufficient. The picture that emerges is consistent with the ERUH.

Extended Radical Unacceptability Hypothesis: All judgments of reduced acceptability in cases of otherwise well-formed (i.e., locally well-formed) extractions are due to non-syntactic factors, not syntactic constraints.

Thus, it appears that there is limited support for grammatical constraints as accounts of the unacceptability of extraction from islands. It is in fact reasonable to hypothesize that in virtually every case of unacceptability, if the local well-formedness conditions of the grammar are satisfied, the reason for the unacceptability is non-syntactic. Processing complexity appears to be the most prominent candidate, which is sensitive to syntactic configuration, discourse accessibility, pragmatic plausibility including relevance, contextual factors such as information structure, and frequency.

That said, there are several major open questions that have to be dealt with. One is to see whether our ERUH-compliant explanations for the Coordinate Structure Constraint and Left Branch Condition hold up under closer empirical scrutiny. There are also cases of apparent freezing that involve chain interactions different from the sort discussed in the English freezing cases discussed in Section 4.1. In these cases we must seek alternate sources of low frequency, which would be sufficient to account for low acceptability in the model sketched in Figure 11.

A second question concerns cross-linguistic variation: if island and similar effects are the consequence of non-syntactic factors, why do different languages reflect differences in the extent to which they show sensitivity to island constraints? Still more problematic is evidence for inter-individual variation in judgments for particular island violations (Kush et al. 2017). One would assume that non-syntactic factors would be constant across languages and individuals. In order to account for the variation, we would suggest pursuing an explanation in terms of language-specific differences in frequency in the specific constructions that show differences in acceptability judgments. Again, the key idea is that acceptability correlates with frequency.

Finally, ERUH is a very strong hypothesis—it says that there are no purely syntactic constraints that are not LWFCs. This strong localist outlook is characteristically associated with GPSG, SBCG, and variants of Categorical Grammar (Gazdar et al. 1985; Kubota and Levine 2020; Sag 2012)—theories which confine syntactic constraints to local chunks of representation of the kind that could be encoded in a single phrase-structure rule. In those cases where there is putative evidence that syntax per se is responsible for acceptability judgments in non-local dependencies (e.g., Kush et al. 2017; Phillips 2006, 2013a), we

would always want to see if it is possible to rule out all plausible aspects of processing, pragmatics and semantics as potential explanations.

As we saw throughout our discussion here, many of the phenomena that were once plausibly analyzed as requiring syntactic constraints on non-local configurations are actually better explained in terms of non-syntactic factors. We believe that this kind of approach is plausible not only for the empirical reasons we mentioned in this paper, but also for conceptual and heuristic ones. Conceptually, a theory of grammar that subscribes to the ERUH excludes a *prima facie* source of complexity that would impose a heavy burden on evolutionary accounts of the syntactic component of the language faculty (Berwick and Chomsky 2016; Hauser et al. 2002; Jackendoff 2002). In addition, heuristically, the questions that the ERUH raises open a fruitful avenue of cross-disciplinary dialogue between theories of linguistic representation and theories of processing and general cognitive capacities.

Author Contributions: Conceptualization, P.W.C., G.V. and S.W.; methodology, P.W.C., G.V. and S.W.; formal analysis, P.W.C., G.V. and S.W.; writing—original draft preparation, P.W.C., G.V. and S.W.; writing—review and editing, P.W.C., G.V. and S.W. All authors have read and agreed to the published version of the manuscript.

Funding: The collaborative research reported here was partially funded by the Alexander von Humboldt Stiftung and the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—SFB 833—A7—Project ID 75650358.

Acknowledgments: We wish to express our profound thanks to the reviewers for the time, effort and care that they devoted to reading our paper and pointing out errors, omissions, relevant literature, and passages in need of clarification. We are very much in debt to them. Any errors that remain are solely our responsibility.

Conflicts of Interest: The authors declare no conflict of interest.

Notes

- ¹ Subsequent syntactic theories of islands have, of course, evolved well beyond Ross's early efforts. The main thrust of the literature on islands after Ross, as far as we can see, is to derive the central results of Ross's classical syntactic account in a more principled way, often with the goal of unifying various locality conditions (see Boeckx (2012) for an overview). However, this tradition inherits from Ross and other early work like Chomsky (1973) the idea that the patterns underlying island effects are syntactic in nature (Bošković 2015; Chomsky 2008; Phillips 2013a, 2013b; Sprouse 2007, 2012a, 2012b). Our discussion here—as well as the more lengthy arguments in Kluender (1991), Goldberg (2006), Hofmeister and Sag (2010), Chaves and Putnam (2020) and Kubota and Levine (2020)—targets this basic assumption, rather than the details of specific syntactic accounts.
- ² Due to space considerations we are unable to survey every phenomenon that bears on this hypothesis. For research on a broad array of phenomena that appear to be consistent with the ERUH, see Francis (2022). Additionally, it appears plausible that the ERUH applies to other kinds of putative non-local constraints, such as Condition C and the binding of long-distance anaphors (Reinhart and Reuland 1991; Varaschin 2021; Varaschin et al. 2022). We also do not deal with weak islands such as *wh*-islands and negative islands, for which a range of both syntactic and semantic accounts have been proposed. For a review, see Szabolcsi and Lohndal (2017), who conclude that “it seems true beyond reasonable doubt that a substantial portion of this large [weak island—PWC, GV and SW] phenomenon is genuinely semantic in nature”, and Abrusán (2014). This work suggests that weak islands are consistent with the ERUH. See also Kroch (1998) for a pragmatic account of weak islands, and Gieselman et al. (2013) for experimental evidence that the unacceptability of extraction from negative islands arises from the interaction of various processing demands.
- ³ More formally, Levy (2008) defines the surprisal associated with a given linguistic expression e_n as its negative log probability conditional on all the previous expressions in the discourse and the relevant features of the extra-sentential context (written as CONTEXT):

$$(i) \quad \text{surprisal}(e_n) = -\log P(e_n \mid e_1, \dots, e_{n-1}, \text{CONTEXT})$$

Our use of surprisal is different in several respects from Levy's. First, Levy (2008) defines surprisal relative to *words*. We are generalizing the notion to linguistic expressions in general, including words and phrases. Second, Levy documents the correlation between surprisal and performance measures such as reaction times, while we are focusing on the underlying processing and acceptability responses. In this respect we are following a line of research pursued by Park et al. (2021), who use surprisal to measure a deep learning language model's knowledge of syntax. They explore the extent to which a language model's surprisal score for pairs of sentences matches with standard acceptability contrasts found in textbooks. They found that “the accuracy of

BERT's acceptability judgments [i.e., the correspondence between the surprisal value assigned by the language model, BERT, and the acceptability reported in textbooks] is fairly high" (Park et al. 2021, p. 420).

- 4 The frequency that determines expectations is not that of sequences of strings, but, rather, of linguistic expressions, minimally construed as correspondences of phonological, syntactic, and semantic information (Goldberg 1995, 2006; Jackendoff 2002; Michaelis 2012; Sag 2012). This caveat is necessary in order to avoid the objection Chomsky (1957, pp. 15–17) raised to statistical approaches. In the context *I saw a fragile* _i, the strings *bassoon* and *of* may share an equal frequency in the past linguistic experience of an English speaker (≈ 0). However, since the speaker independently knows that *bassoon* is a noun and *of* is a preposition and the sequence *fragile* NP is much more frequent than *fragile* P, the expectation (and, therefore, the acceptability) for the former is much higher than for the latter.
- 5 For instance, in order to state a syntactic restriction against multiple center-embeddings, we would need some way of counting the number of embedded clauses; in order to account for (5), we would need the syntactic constraint on A' movement to be sensitive to the position of the gap in the linear order of the string (which contradicts the widespread assumption that transformations are structure-dependent). The very idea of syntactic constraints on unbounded dependencies also entails a non-trivial extension of the vocabulary of syntactic theory insofar as it requires ways of referring to chunks of syntactic representations of an indeterminate size, as discussed in connection to (2) above.
- 6 A reviewer correctly points out that in principle failure of a particular example to observe a proposed syntactic constraint could be a 'grammatical illusion' (Christensen 2016; de Dios-Flores 2019; Engelmann and Vasishth 2009; Phillips et al. 2011; Trotzke et al. 2013). Clearly, such a possibility always exists where there are differences in judgments of acceptability. However, in order to appeal to a grammatical illusion to account for the acceptability of an island violation it is important to show that doing so results in a simplification of the theory of grammar; otherwise, one can always appeal to a grammatical illusion in order to get around any counterexample to a proposed syntactic constraint. Quite the opposite appears to be the case for islands. As Phillips (2013a, p. 54) puts it, "[n]atural language grammars would probably be simpler if there were no island constraints". The reasons relate to the point we made above about how syntactic accounts require extending the descriptive vocabulary of grammatical theory.
- 7 We note that evolutionary considerations are not incompatible per se with a syntactic approach to islands. On such a view, it would be necessary to show that island effects follow from an interaction of general architectural features of the syntactic part of language that could independently be justified on evolutionary grounds. We are not aware of such a demonstration. Hauser et al. (2002) suggest an alternative view, where island constraints arise automatically from solutions to the problem of optimizing the syntactic outputs constructed by the "narrow" faculty of language to the constraints imposed by the "broad" faculty of language – i.e., the cognitive systems that the syntax interacts with. If the latter are understood to include processing systems, Hauser et al.'s (2002) hypothesis can be seen as an instance of the RUH.
- 8 In fact, the experiments they report demonstrate that manipulation of frequency has an effect on acceptability judgments for island extractions.
- 9 Ross's formulation of the constraint reflects the fact that it is not possible to extract from an extraposed relative clause, even though it is not in a configuration that would fall under the Complex NP Constraint. Thus we see right at the start the treatment of freezing as a special type of island phenomenon.
- 10 For other proposals that take chain interactions to result in ungrammaticality, see Chomsky's (1977) discussion of the interaction of wh-movement and tough-movement and also Fodor (1978), and Pesetsky (1982). In contrast, Collins (2005) proposes an account of the English passive that requires movement of a sub-constituent from a larger, moved constituent.
- 11 It should also be noted that there are phenomena where greater distance between dependent elements appears to improve acceptability (see, for example, Vasishth and Lewis 2006). Such 'anti-locality' effects suggest that there are yet other factors at play, such as predictability related to selection (Levy and Keller 2013; Rajkumar et al. 2016). Moreover, research on the processing of relative clauses in languages such as Japanese and Korean suggests that there may be a preference of extraction of subjects over objects even though the gaps corresponding to the subjects are arguably further from the head (see, for example, Nakamura and Miyamoto 2013; Ueno and Garnsey 2008). These data favor the view that dependency length should be measured in terms of complexity of branching structure, given that in head-final languages the position of subject gaps is linearly farther but hierarchically closer to the position of the filler noun.
- 12 The term 'surfing' is due to Sauerland (1999).
- 13 For completeness we note that there is a range of cases of purported freezing that do not immediately lend themselves to explanations in terms of non-syntactic factors. Among these are phenomena in German (Bayer 2018; Müller 2018), and Dutch (Corver 2018). These phenomena await a more extensive analysis than we can provide here.
- 14 Crossing is also seen in another type of example that fell under the freezing account of Wexler and Culicover (1980):

(i) Which table_i did you put t_j on t_i [a picture of FDR_j] ?

- 15 The dependency length literature suggests that minimization of dependency length alone is not sufficient to account for structural preferences reflecting degree of congruence (Kuhlmann and Nivre 2006). Also relevant are the degree of adjacency of dependent constituents, measured by GAP DEGREE, which measures the number of discontinuities within a subtree, EDGE DEGREE, which

- measures the number of intervening constituents spanned by a single edge, and the disjointness of constituents, measured by WELL-NESTEDNESS (Kuhlmann and Nivre 2006, p. 511).
- 16 For analyses of the relationship between topicalization and the complementizer in terms of Optimality Theory, see Pesetsky (1998) and Grimshaw (1997).
- 17 For a review of a range of types of garden paths, see Pritchett (1992, 1988).
- 18 For a comprehensive review of WCO effects and of proposals to account for WCO, see Safir (2017). Safir notes a number of cases that are more complex than (23b) that the current proposal does not address.
- 19 For a computational account of crossover effects in terms of linear order processing, see Shan and Barker (2006).
- 20 The lower accessibility of Frank would justify repeating the name *Frank* or using some other referential phrase carrying a higher degree of informativity. Repetition of *Charlie* in (28), in turn, would have been redundant and would, as a result, contribute to increase processing complexity (Gordon and Hendrick 1998).
- 21 We show below that the relative unacceptability of (30c–e) vs. (30a) is related to the Uninvited Guest in virtue of the presence of additional referring expressions as subjects as well as finite tense (cf. Kluender 1998).
- 22 Throughout most of the history of transformational grammar, the Coordinate Structure Constraint has resisted an integration into general syntactic theories of islands like the ones proposed by Chomsky (1973, 1977, 1986, 2008). However, it did play an important role in non-transformational theories like GPSG and HPSG (Gazdar 1981; Pollard and Sag 1994). More recently, minimalist accounts of both parts of (40) have been proposed which make critical use of NON-LOCAL GRAMMATICAL CONSTRAINTS, such as Chomsky's (2000) Phase Impenetrability Condition and Rizzi's (1990) Relativized Minimality (Bošković 2020; Oda 2021). Relativized minimality counts as a non-local constraint in our sense because, even in the absence of interveners, the distance between a target position and a movement trace can still be arbitrarily large. A similar observation applies to the size of the domain of a phase (i.e., the spell-out domain), from which extraction is ruled out by the Phase Impenetrability Condition (Chomsky 2000).
- 23 This is ultimately the strategy advocated by Chaves and Putnam (2020, pp. 102–3).

References

- Abeillé, Anne, Barbara Hemforth, Elodie Winckel, and Edward Gibson. 2020. Extraction from subjects: Differences in acceptability depend on the discourse function of the construction. *Cognition* 204: 104293. [CrossRef] [PubMed]
- Abrusán, Márta. 2014. *Weak Island Semantics*. Oxford: Oxford University Press.
- Almor, Amit. 1999. Noun-phrase anaphora and focus: The informational load hypothesis. *Psychological Review* 106: 748–65. [CrossRef] [PubMed]
- Almor, Amit. 2000. Constraints and mechanisms in theories of anaphor processing. In *Architectures and Mechanisms for Language Processing*. Edited by Matthew W. Crocker, Martin J. Pickering and Charles Clifton. Cambridge, MA: MIT Press, pp. 341–54.
- Almor, Amit, and Veena A. Nair. 2007. The form of referential expressions in discourse. *Language and Linguistics Compass* 1: 84–99. [CrossRef]
- Ariel, Mira. 1988. Referring and accessibility. *Journal of linguistics* 24: 65–87. [CrossRef]
- Ariel, Mira. 1990. *Accessing Noun-Phrase Antecedents*. London: Routledge.
- Ariel, Mira. 1991. The function of accessibility in a theory of grammar. *Journal of Pragmatics* 16: 443–63. [CrossRef]
- Ariel, Mira. 1994. Interpreting anaphoric expressions: A cognitive versus a pragmatic approach. *Journal of linguistics* 30: 3–42. [CrossRef]
- Ariel, Mira. 2001. Accessibility theory: An overview. In *Text Representation: Linguistic and Psycholinguistic Aspects*. Edited by Ted. J. M. Sanders, Joost Schilperoord and Wilbert Spooren. Amsterdam: John Benjamins Publishing Company, pp. 29–87.
- Ariel, Mira. 2004. Accessibility marking: Discourse functions, discourse profiles, and processing cues. *Discourse Processes* 37: 91–116. [CrossRef]
- Arnold, Jennifer E., and Zenzi M. Griffin. 2007. The effect of additional characters on choice of referring expression: Everyone counts. *Journal of Memory and Language* 56: 521–36. [CrossRef]
- Arnold, Jennifer E. 2010. How speakers refer: The role of accessibility. *Language and Linguistics Compass* 4: 187–203. [CrossRef]
- Arnold, Jennifer E., and Michael K. Tanenhaus. 2011. Disfluency effects in comprehension: How new information can become accessible. In *The Processin and Acquisition of Reference*. Edited by Edward A. Gibson and Neal J. Pearlmutter. Cambridge, MA: MIT Press, pp. 197–217.
- Arnon, Inbal, Philip Hofmeister, T. Florian Jaeger, IvanA. Sag, and Neal Snider. 2005. Rethinking superiority effects: A processing model. Paper presented at the 18th Annual CUNY Conference on Human Sentence Processing, University of Arizona, Tucson, AZ, March 2.
- Baltin, Mark. 1981. Strict bounding. In *The Logical Problem of Language Acquisition*. Edited by C. Leroy Baker and John McCarthy. Cambridge, MA: MIT Press.
- Bayer, Josef. 2004. Was beschränkt die Extraktion? Subjekt-Objekt vs. Topic-Fokus. In *Deutsche Syntax: Empirie und Theorie*. Edited by Franz-Josef D'Avis. Volume 46 of *Acta Universitatis Gothoburgensis*. Göteborg: Göteborger Germanistische Forschungen, pp. 233–57.

- Bayer, Josef. 2018. Critical freezing in the syntax of particles. In *Freezing: Theoretical Approaches and Empirical Domains*. Edited by Jutta Hartmann, Marion Jäger, Andreas Konietzko and Susanne Winkler. Berlin: De Gruyter, pp. 224–63.
- Berwick, Robert C., and Noam Chomsky. 2016. *Why Only Us: Language and Evolution*. Cambridge, MA: MIT Press.
- Bever, Thomas. 1970. The cognitive basis for linguistic structures. In *Cognition and the Development of Language*. Edited by John R. Hayes. New York: John Wiley & Sons, pp. 279–362.
- Bloomfield, Leonard. 1933. *Language*. New York and London: Henry Holt and Co. and Allen and Unwin Ltd.
- Boeckx, Cedric. 2008. Islands. *Language and Linguistics Compass* 2: 151–67. [\[CrossRef\]](#)
- Boeckx, Cedric. 2012. *Syntactic Islands*. Cambridge: Cambridge University Press.
- Borsley, Robert D. 2005. Against Conj. *Lingua* 115: 461–82. [\[CrossRef\]](#)
- Bošković, Željko. 2015. From the complex NP constraint to everything: On deep extractions across categories. *The Linguistic Review* 32: 603–69. [\[CrossRef\]](#)
- Bošković, Željko. 2020. On the Coordinate Structure Constraint, across-the-board-movement, phases, and labeling. In *Recent Developments in Phase Theory*. Edited by Jeroen van Craenenbroeck, Cora Pots and Tanja Temmerman. Berlin: De Gruyter Mouton, pp. 133–82.
- Bouchard, Denis. 1984. *On the Content of Empty Categories*. Dordrecht: Foris.
- Bouma, Gosse, Robert Malouf, and Ivan Sag. 2001. Satisfying constraints on extraction and adjunction. *Natural Language and Linguistic Theory* 19: 1–65. [\[CrossRef\]](#)
- Bybee, Joan. 2006. From usage to grammar: The mind's response to repetition. *Language* 82: 711–33. [\[CrossRef\]](#)
- Bybee, Joan. 2010. *Language, Usage and Cognition*. Cambridge: Cambridge University Press.
- Bybee, Joan L., and Paul J. Hopper. 2001. *Frequency and the Emergence of Linguistic Structure*. Amsterdam and Philadelphia: John Benjamins Publishing Company.
- Chaves, Rui P. 2007. Coordinate Structures: Constraint-Based Syntax-Semantics Processing. Ph.D. thesis, Universidade de Lisboa, Lisbon, Portugal.
- Chaves, Rui P. 2013. An expectation-based account of subject islands and parasitism. *Journal of Linguistics* 49: 285–327. [\[CrossRef\]](#)
- Chaves, Rui P. 2020. Island phenomena and related matters. In *Head-Driven Phrase Structure Grammar: The Handbook*. Edited by Stefan Müller, Anne Abeillé, Robert D. Borsley and Jean-Pierre Koenig. Berlin: Language Science Press.
- Chaves, Rui P., and Jeruen E. Dery. 2014. Which subject islands will the acceptability of improve with repeated exposure? In *Proceedings of the Thirty-First Meeting of the West Coast Conference on Formal Linguistics, Arizona State University, February 7–9, 2013*. Edited by Robert E. Santana-LaBarge. Somerville, MA: Cascadia Project, pp. 96–106.
- Chaves, Rui P., and Jeruen E. Dery. 2019. Frequency effects in subject islands. *Journal of linguistics* 55: 475–521. [\[CrossRef\]](#)
- Chaves, Rui P., and Michael T. Putnam. 2020. *Unbounded Dependency Constructions: Theoretical and Experimental Perspectives*. Oxford: Oxford University Press.
- Chomsky, Noam. 1957. *Syntactic Structures*. The Hague: Mouton.
- Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.
- Chomsky, Noam. 1973. Conditions on transformations. In *A Festschrift for Morris Halle*. Edited by Stephen Anderson and Paul Kiparsky. New York: Holt, Reinhart & Winston, pp. 232–86.
- Chomsky, Noam. 1977. On *wh*-movement. In *Formal Syntax*. Edited by Peter W. Culicover, Thomas Wasow and Adrian Akmajian. New York: Academic Press, pp. 71–132.
- Chomsky, Noam. 1986. *Barriers*. Cambridge, MA: MIT Press.
- Chomsky, Noam. 2000. Minimalist inquiries: The framework. In *Step by Step: Essays on Minimalist Syntax in Honor of Howard Lasnik*. Edited by Roger Martin, David Michaels and Juan Uriagereka. Cambridge, MA: MIT Press, pp. 89–156.
- Chomsky, Noam. 2001. Derivation by phase. In *Ken Hale: A Life in Linguistics*. Edited by Michael Kenstowicz. Cambridge, MA: MIT Press, pp. 1–52.
- Chomsky, Noam. 2008. On phases. In *Foundational Issues in Linguistic Theory: Essays in Honor of Jean-Roger Vergnaud*. Edited by Robert Freidin, Carlos P. Otero and Maria Luisa Zubizarreta. Cambridge, MA: MIT Press, pp. 133–66.
- Chomsky, Noam, Ángel J. Gallego, and Dennis Ott. 2019. Generative grammar and the faculty of language: Insights, questions, and challenges. *Catalan Journal of Linguistics* 18: 229–61. [\[CrossRef\]](#)
- Christensen, Ken Ramshøj. 2016. The dead ends of language: The (mis)interpretation of a grammatical illusion. In *Let Us Have Articles Betwixt Us: Papers in Historical and Comparative Linguistics in Honour of Johanna L. Wood*. Edited by Sten Vikner, Henrik Jørgensen and Elly van Gelderen. Aarhus: Aarhus University, pp. 129–59.
- Citko, Barbara. 2014. *Phase Theory: An Introduction*. Cambridge: Cambridge University Press.
- Collins, Chris. 2005. A smuggling approach to the passive in English. *Syntax* 8: 81–120. [\[CrossRef\]](#)
- Corballis, Michael C. 2017. The evolution of language. In *APA Handbook of Comparative Psychology: Basic Concepts, Methods, Neural Substrate, and Behavior*. Edited by Josep Call, Gordon M. Burghardt, Irene M. Pepperberg, Charles T. Snowdon and Thomas R. Zentall. Washington, DC: American Psychological Association, pp. 273–97.
- Corver, Norbert. 2017. Freezing effects. In *The Blackwell Companion to Syntax*, 2nd ed. Edited by Martin Everaert and Henk van Riemsdijk. Malden: Blackwell Publishing.
- Corver, Norbert. 2018. The freezing points of the (Dutch) adjectival system. In *Freezing: Theoretical Approaches and Empirical Domains*. Edited by Jutta Hartmann, Marion Jäger, Andreas Konietzko and Susanne Winkler. Berlin: De Gruyter, pp. 143–94.

- Culicover, Peter W. 2005. Squinting at Dali's Lincoln: On how to think about language. In *Proceedings of the Annual Meeting of the Chicago Linguistic Society*. Chicago: Chicago Linguistic Society, vol. 41, pp. 109–28.
- Culicover, Peter W. 2013a. English (zero-)relatives and the competence-performance distinction. *International Review of Pragmatics* 5: 253–70. [\[CrossRef\]](#)
- Culicover, Peter W. 2013b. *Explaining Syntax*. Oxford: Oxford University Press.
- Culicover, Peter W. 2013c. *Grammar and Complexity: Language at the Intersection of Competence and Performance*. Oxford: Oxford University Press.
- Culicover, Peter W. 2013d. The role of linear order in the computation of referential dependencies. *Lingua* 136: 125–44. [\[CrossRef\]](#)
- Culicover, Peter W. 2015. Simpler Syntax and the mind. In *Structures in the Mind: Essays on Language, Music, and Cognition in Honor of Ray Jackendoff*. Edited by Ida Toivonen, Piroska Csurí and Emile Van Der Zee. Cambridge: MIT Press, pp. 3–20.
- Culicover, Peter W. 2021. *Language Change, Variation and Universals—A Constructional Approach*. Oxford: Oxford University Press.
- Culicover, Peter W., and Robert D. Levine. 2001. Stylistic inversion in English: A reconsideration. *Natural Language & Linguistic Theory* 19: 283–310.
- Culicover, Peter W., and Andrzej Nowak. 2002. Markedness, antisymmetry and complexity of constructions. *Linguistic Variation Yearbook* 2: 5–30. [\[CrossRef\]](#)
- Culicover, Peter W., and Andrzej Nowak. 2003. *Dynamical Grammar: Minimalism, Acquisition and Change*. Oxford: Oxford University Press.
- Culicover, Peter W., and Michael Rochemont. 1983. Stress and focus in English. *Language* 59: 123–65. [\[CrossRef\]](#)
- Culicover, Peter W., and Susanne Winkler. 2018. Freezing, between grammar and processing. In *Freezing: Theoretical Approaches and Empirical Domains*. Edited by Jutta Hartmann, Marion Jäger, Andreas Konietzko and Susanne Winkler. Berlin: De Gruyter, pp. 353–86.
- Culicover, Peter W., and Susanne Winkler. 2022. Parasitic gaps aren't parasitic or, the Case of the Uninvited Guest. *The Linguistic Review* 39:1–35. [\[CrossRef\]](#)
- de Dios-Flores, Iria. 2019. Processing sentences with multiple negations: Grammatical structures that are perceived as unacceptable. *Frontiers in Psychology* 10: 2346. [\[CrossRef\]](#) [\[PubMed\]](#)
- De Kuthy, Kordula, and Andreas Konietzko. 2019. Information structural constraints on PP topicalization from NPs. In *Architecture of Topic*. Edited by Valéria Molnár, Verner Egerland and Susanne Winkler. Berlin and New York: De Gruyter, pp. 203–22.
- Deane, Paul. 1991. Limits to attention: A cognitive theory of island phenomena. *Cognitive Linguistics* 2: 1–63. [\[CrossRef\]](#)
- Emonds, Joseph E. 1985. *A Unified Theory of Syntactic Categories*. Dordrecht: Foris.
- Engelmann, Felix, and Shravan Vasishth. 2009. Processing grammatical and ungrammatical center embeddings in english and german: A computational model. Paper presented at Ninth International Conference on Cognitive Modeling, Manchester, UK, July 24–26. pp. 240–45.
- Erteschik-Shir, Nomi. 1977. On the Nature of Island Constraints. Ph.D. thesis, MIT, Cambridge, MA, USA.
- Erteschik-Shir, Nomi. 2007. *Information Structure: The Syntax-Discourse Interface*. Oxford: Oxford University Press.
- Erteschik-Shir, Nomi, and Shalom Lappin. 1979. Dominance and the functional explanation of island constraints. *Theoretical Linguistics* 6: 43–84. [\[CrossRef\]](#)
- Fodor, Janet Dean. 1978. Parsing strategies and constraints on transformations. *Linguistic Inquiry* 9: 427–73.
- Fodor, Janet Dean. 1983. Phrase structure parsing and the island constraints. *Linguistics and Philosophy* 6: 163–223. [\[CrossRef\]](#)
- Frampton, John. 1990. Parasitic gaps and the theory of wh-chains. *Linguistic Inquiry* 21: 49–78.
- Francis, Elaine J. 2022. *Gradient Acceptability and Linguistic Theory*. Oxford: Oxford University Press.
- Frazier, Lyn, and Charles Clifton. 1989. Successive cyclicity in the grammar and the parser. *Language and Cognitive Processes* 4: 93–126. [\[CrossRef\]](#)
- Futrell, Richard, Kyle Mahowald, and Edward Gibson. 2015. Large-scale evidence of dependency length minimization in 37 languages. *Proceedings of the National Academy of Sciences* 112: 10336–41. [\[CrossRef\]](#) [\[PubMed\]](#)
- Gazdar, Gerald. 1980. A cross-categorical semantics for coordination. *Linguistics* 3: 407–9. [\[CrossRef\]](#)
- Gazdar, Gerald. 1981. Unbounded dependencies and coordinate structure. *Linguistic Inquiry* 12: 155–84.
- Gazdar, Gerald, Ewan Klein, Geoffrey Pullum, and Ivan A. Sag. 1985. *Generalized Phrase Structure Grammar*. Oxford and Cambridge: Blackwell Publishing and Harvard University Press.
- Gernsbacher, Morton. 1989. Mechanisms that improve referential access. *Cognition* 32: 99–156. [\[CrossRef\]](#)
- Gibson, Edward. 1998. Linguistic complexity: Locality of syntactic dependencies. *Cognition* 68: 1–76. [\[CrossRef\]](#)
- Gibson, Edward. 2000. The dependency locality theory: A distance-based theory of linguistic complexity. In *Image, Language, Brain*. Edited by Alec P. Marantz, Yasushi Miyashita and Wayne O'Neil. Cambridge, MA: MIT Press, pp. 95–126.
- Gieselmann, Simone, Robert Kluender, Ivano Caponigro, Yelena Fainleib, Nicholas LaCara, and Yangsook Park. 2013. Isolating processing factors in negative island contexts. *Proceedings of NELS* 41: 233–46.
- Ginzburg, Jonathan, and Ivan A. Sag. 2000. *Interrogative Investigations*. Stanford: CSLI publications.
- Givón, Talmy. 1983. *Topic Continuity in Discourse: A Quantitative Cross-Language Study*. Amsterdam: John Benjamins Publishing Company.
- Goldberg, Adele E. 1995. *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago: University of Chicago Press.
- Goldberg, Adele E. 2006. *Constructions at Work: The Nature of Generalization in Language*. Oxford: Oxford University Press.

- Goldsmith, John. 1985. A principled exception to the Coordinate Structure Constraint. In *Proceedings of the Chicago Linguistic Society* 21. Edited by William H. Eilfort, Paul D. Kroeber and Karen L. Peterson. pp. 133–43.
- Gordon, Peter C., and Randall Hendrick. 1998. The representation and processing of coreference in discourse. *Cognitive Science* 22: 389–424. [\[CrossRef\]](#)
- Gould, Stephen Jay, and Richard C. Lewontin. 1979. The spandrels of San Marco and the Panglossian paradigm: A critique of the adaptationist programme. *Proceedings of the Royal Society of London. Series B. Biological Sciences* 205: 581–98.
- Grice, H. Paul. 1975. Logic and conversation. In *Speech Acts*. Edited by Peter Cole and Jerry L. Morgan. Volume 3 of *Syntax and Semantics*. New York: Academic Press, pp. 43–58.
- Grimshaw, Jane. 1997. Projection, heads, and optimality. *Linguistic Inquiry* 28: 373–422.
- Grosu, Alexander. 1973. On the nonunitary nature of the Coordinate Structure Constraint. *Linguistic Inquiry* 4: 88–92.
- Grosu, Alexander. 1974. On the nature of the Left Branch Condition. *Linguistic Inquiry* 5: 308–19.
- Gundel, Jeanette, Nancy Hedberg, and Ron Zacharski. 1993. Cognitive Status and the Form of Referring Expressions in Discourse. *Language* 69: 274–307. [\[CrossRef\]](#)
- Haider, Hubert, and Inger Rosengren. 2003. Scrambling: Nontriggered chain formation in OV languages. *Journal of Germanic Linguistics* 15: 203–67. [\[CrossRef\]](#)
- Hale, John. 2001. A probabilistic Earley parser as a psycholinguistic model. Paper presented at Second Meeting of the North American Chapter of the Association for Computational Linguistics, Morristown, NJ, USA, June 1–7. Stroudsburg, PA: Association for Computational Linguistics, pp. 1–8.
- Hale, John. 2003. The information conveyed by words in sentences. *Journal of Psycholinguistic Research* 32: 101–23. [\[CrossRef\]](#)
- Hauser, Marc D., Noam Chomsky, and W. Tecumseh Fitch. 2002. The faculty of language: What is it, who has it, and how did it evolve? *Science* 298: 1569–79. [\[CrossRef\]](#)
- Hawkins, John A. 1994. *A Performance Theory of Order and Constituency*. Cambridge: Cambridge University Press.
- Hawkins, John A. 2004. *Efficiency and Complexity in Grammars*. Oxford: Oxford University Press.
- Hawkins, John A. 2014. *Cross-Linguistic Variation and Efficiency*. Oxford: Oxford University Press.
- Hofmeister, Philip, Peter W. Culicover, and Susanne Winkler. 2015. Effects of processing on the acceptability of frozen extraposed constituents. *Syntax* 18: 464–83. [\[CrossRef\]](#)
- Hofmeister, Philip, T. Florian Jaeger, Inbal Arnon, Ivan Sag, and Neal Snider. 2013. The source ambiguity problem: Distinguishing the effects of grammar and processing on acceptability judgments. *Language and Cognitive Processes* 28: 48–87.
- Hofmeister, Philip, T. Florian Jaeger, Ivan A. Sag, Inbal Arnon, and Neal Snider. 2007. Locality and accessibility in *wh*-questions. In *Roots: Linguistics in Search of Its Evidential Base*. Edited by Sam Featherston and Wolfgang Sternefeld. Berlin: Mouton de Gruyter, pp. 185–206.
- Hofmeister, Philip, and Ivan A. Sag. 2010. Cognitive constraints and island effects. *Language* 86: 366–415. [\[CrossRef\]](#)
- Hofmeister, Philip, Laura Staum Casasanto, and Ivan A. Sag. 2013. Islands in the grammar? Standards of evidence. In *Experimental Syntax and the Islands Debate*. Edited by Jon Sprouse and Norbert Hornstein. Cambridge: Cambridge University Press, pp. 42–63.
- Höhle, Tilman. 1982. Explikation für normale Betonung und normale Wortstellung. In *Satzglieder im Deutschen*. Edited by Werner Abraham. Tübingen: Gunter Narr Verlag, pp. 75–153.
- Jackendoff, Ray. 1977. *X' Syntax*. Cambridge, MA: MIT Press.
- Jackendoff, Ray. 1999. Possible stages in the evolution of the language capacity. *Trends in Cognitive Sciences* 3: 272–79. [\[CrossRef\]](#)
- Jackendoff, Ray. 2002. *Foundations of Language: Brain, Meaning, Grammar, Evolution*. Oxford: Oxford University Press.
- Jackendoff, Ray, and Peter Culicover. 1972. A reconsideration of dative movement. *Foundations of Language* 7: 397–412.
- Jäger, Marion. 2018. An experimental study on freezing and topicalization in English. In *Freezing: Theoretical Approaches and Empirical Domains*. Edited by Jutta Hartmann, Marion Knecht, Andreas Konietzko and Susanne Winkler. Berlin and New York: Mouton de Gruyter, pp. 430–50.
- Kaplan, Ronald M., and Annie Zaenen. 1995. Long-distance dependencies, constituent structure, and functional uncertainty. In *Formal Issues in Lexical-Functional Grammar*. Edited by Mary Dalrymple, Ronald M. Kaplan, John T. Maxwell III and Annie Zaenen. Stanford, CA: CSLI Publications. pp. 137–65.
- Kehler, Andrew. 1996. Coherence and the coordinate structure constraint. *Annual Meeting of the Berkeley Linguistics Society* 22: 220–31. [\[CrossRef\]](#)
- Kluender, Robert. 1991. Cognitive Constraints on Variables in Syntax. Ph.D. thesis, University of California, San Diego, La Jolla, CA, USA.
- Kluender, Robert. 1992. Deriving islands constraints from principles of predication. In *Island Constraints: Theory, Acquisition and Processing*. Edited by Helen Goodluck and Michael Rochemont. Dordrecht: Kluwer, pp. 223–58.
- Kluender, Robert. 1998. On the distinction between strong and weak islands: A processing perspective. In *Syntax and Semantics 29: The Limits of Syntax*. Edited by Peter W. Culicover and Louise McNally. San Diego: Academic Press, pp. 241–79.
- Kluender, Robert. 2004. Are subject islands subject to a processing account? Paper presented at 23rd West Coast Conference on Formal Linguistics, University of California, Davis, CA, USA, April 23–25. Edited by Benjamin Schmeiser, Veneeta Chand, Ann Kelleher and Angelo J. Rodriguez. Somerville: Cascadilla Press. pp. 475–99.
- Kluender, Robert, and Marta Kutas. 1993a. Bridging the gap: Evidence from ERPs on the processing of unbounded dependencies. *Journal of Cognitive Neuroscience* 5: 196–214. [\[CrossRef\]](#)

- Kluender, Robert, and Marta Kutas. 1993b. Subjacency as a processing phenomenon. *Language and Cognitive Processes* 8: 573–633. [\[CrossRef\]](#)
- Konietzko, Andreas. forthcoming. PP extraction from subject islands in German. *Glossa* 7.
- Konietzko, Andreas, Janina Radó, and Susanne Winkler. 2019. Focus constraints on relative clause antecedents in sluicing. In *Information Structure and Semantic Processing*. Edited by Sam Featherston, Robin Hörnig, Sophie von Wietersheim and Susanne Winkler. Berlin and New York: Mouton de Gruyter, pp. 105–27.
- Konietzko, Andreas, Susanne Winkler, and Peter W. Culicover. 2018. Heavy NP shift does not cause freezing. *Canadian Journal of Linguistics/Revue Canadienne de Linguistique* 63: 454–64. [\[CrossRef\]](#)
- Koopman, Hilda, and Dominique Sportiche. 1983. Variables and the bijection principle. *The Linguistic Review* 2: 139–60.
- Kroch, Anthony. 1998. Amount quantification, referentiality, and long *wh*-movement. In *Penn Working Papers in Linguistics*, Vol. 5.2. Edited by Alexis Dimitriadis, Hikyoung Lee, Christine Moisset and Alexander Williams. Philadelphia: University of Pennsylvania.
- Kubota, Yusuke, and Jungmee Lee. 2015. The coordinate structure constraint as a discourse-oriented principle: Further evidence from Japanese and Korean. *Language* 91: 642–75. [\[CrossRef\]](#)
- Kubota, Yusuke, and Robert D. Levine. 2020. *Type-Logical Syntax*. Cambridge: MIT Press.
- Kuhlmann, Marco, and Joakim Nivre. 2006. Mildly non-projective dependency structures. Paper presented at COLING/ACL 2006, Main Conference Poster Sessions, Sydney, Australia, July. pp. 507–14.
- Kuperberg, Gina R., and T. Florian Jaeger. 2016. What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience* 31: 32–59. [\[CrossRef\]](#)
- Kush, Dave, Terje Lohndal, and Jon Sprouse. 2017. Investigating variation in island effects. *Natural Language and Linguistic Theory* 36: 1–37. [\[CrossRef\]](#) [\[PubMed\]](#)
- Lakoff, George. 1986. Frame semantic control of the Coordinate Structure Constraint. In *Proceedings of the Chicago Linguistic Society* 22. Edited by Anne M. Farley, Peter T. Farley and Karl-Erik McCullough. Chicago: Chicago Linguistic Society. pp. 152–67.
- Lasnik, Howard, and Tim Stowell. 1991. Weakest crossover. *Linguistic Inquiry* 22: 687–720.
- Levine, Robert D. 2017. *Syntactic Analysis: An HPSG-Based Approach*. Cambridge: Cambridge University Press.
- Levine, Robert D., and Thomas Hukari. 2006. *The Unity of Unbounded Dependency Constructions*. Stanford: CSLI Publications.
- Levinson, Stephen C. 1987. Pragmatics and the grammar of anaphora: A partial pragmatic reduction of binding and control phenomena. *Journal of Linguistics* 23: 379–434. [\[CrossRef\]](#)
- Levinson, Stephen C. 1991. Pragmatic reduction of the binding conditions revisited. *Journal of Linguistics* 27: 107–61. [\[CrossRef\]](#)
- Levy, Roger. 2005. Probabilistic Models of Word Order and Syntactic Discontinuity. Ph.D. thesis, Stanford University, Stanford, CA, USA.
- Levy, Roger. 2008. Expectation-based syntactic comprehension. *Cognition* 106: 1126–77. [\[CrossRef\]](#)
- Levy, Roger. 2013. Memory and surprisal in human sentence comprehension. In *Sentence Processing*. London: Psychology Press, pp. 78–114.
- Levy, Roger, and T. Florian Jaeger. 2007. Speakers optimize information density through syntactic reduction. *Advances in Neural Information Processing Systems* 19: 849.
- Levy, Roger P., and Frank Keller. 2013. Expectation and locality effects in German verb-final structures. *Journal of Memory and Language* 68: 199–222. [\[CrossRef\]](#)
- Lewis, Richard. 1993. An architecturally-based theory of human sentence comprehension. Paper presented at 15th Annual Conference of the Cognitive Science Society, Boulder, CO, USA, June 18–21. Mahwah: Erlbaum, pp. 108–13.
- Lewis, Richard. 1996. Interference in short-term memory: The magical number two (or three) in sentence processing. *Journal of Psycholinguistic Research* 25: 93–115. [\[CrossRef\]](#)
- Lewis, Richard, and Shravan Vasishth. 2005. An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science* 29: 1–45. [\[CrossRef\]](#)
- Lewis, Richard, Shravan Vasishth, and Julie Van Dyke. 2006. Computational principles of working memory in sentence comprehension. *Trends in Cognitive Science* 10: 447–54. [\[CrossRef\]](#)
- Liu, Haitao. 2008. Dependency distance as a metric of language comprehension difficulty. *Journal of Cognitive Science* 9: 159–91.
- Liu, Haitao, Chunshan Xu, and Junying Liang. 2017. Dependency distance: A new perspective on syntactic patterns in natural languages. *Physics of Life Reviews* 21: 171–93. [\[CrossRef\]](#) [\[PubMed\]](#)
- Michaelis, Laura A. 2012. Making the case for construction grammar. In *Sign-Based Construction Grammar*. Edited by Hans C. Boas and Ivan A. Sag. Stanford: CSLI Publications, pp. 31–68.
- Miller, George, and Noam Chomsky. 1963. Finitary models of language users. In *Handbook of Mathematical Psychology*. Edited by R. Duncan Luce, Robert R. Bush and Eugene Galanter. New York: Wiley, vol. 2, pp. 419–92.
- Müller, Gereon. 2010. On deriving CED effects from the PIC. *Linguistic Inquiry* 41: 35–82. [\[CrossRef\]](#)
- Müller, Geron. 2018. Freezing in complex pre-fields. In *Freezing: Theoretical Approaches and Empirical Domains*. Edited by Jutta Hartmann, Marion Jäger, Andreas Konietzko and Susanne Winkler. Berlin: De Gruyter, pp. 105–39.
- Nakamura, Michiko, and Edson T. Miyamoto. 2013. The object before subject bias and the processing of double-gap relative clauses in Japanese. *Language and Cognitive Processes* 28: 303–34. [\[CrossRef\]](#)
- Newmeyer, Frederick J. 2016. Nonsyntactic explanations of island constraints. *Annual Review of Linguistics* 2: 187–210. [\[CrossRef\]](#)

- Nunes, Jairo, and Juan Uriagereka. 2000. Cyclicity and extraction domains. *Syntax* 3: 20–43. [\[CrossRef\]](#)
- Oda, Hiromune. 2017. Two types of the Coordinate Structure Constraint and rescue by PF deletion. *Proceedings of the North East Linguistic Society* 47: 343–56.
- Oda, Hiromune. 2021. Decomposing and deducing the coordinate structure constraint. *The Linguistic Review* 38: 605–44. [\[CrossRef\]](#)
- O’Grady, William, Miseon Lee, and Miho Choo. 2003. A subject-object asymmetry in the acquisition of relative clauses in Korean as a second language. *Studies in Second Language Acquisition* 25: 433–48. [\[CrossRef\]](#)
- Park, Kwonsik, Myung-Kwan Park, and Sanghoun Song. 2021. Deep learning can contrast the minimal pairs of syntactic data. *Linguistic Research* 38: 395–424.
- Pesetsky, David. 1982. Paths and Categories. Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, MA, USA.
- Pesetsky, David. 1987. Wh-in-situ: Movement and unselective binding. In *The Representation of (In)definiteness*. Edited by Eric J. Reuland and Alice G. B. ter Meulen. Cambridge, MA: MIT Press, pp. 98–129.
- Pesetsky, David. 1998. Some optimality principles of sentence pronunciation. In *Is the Best Good Enough?* Edited by Pilar Barbosa, Danny Fox, Paul Hagstrom, Martha McGinnis and David Pesetsky. Cambridge, MA: MIT Press, pp. 337–84.
- Pesetsky, David. 2000. *Phrasal Movement and Its Kin*. Cambridge, MA: MIT Press.
- Phillips, Colin. 2006. The real-time status of island phenomena. *Language* 82: 795–823. [\[CrossRef\]](#)
- Phillips, Colin. 2013a. On the nature of island constraints i: Language processing and reductionist accounts. In *Experimental Syntax and Island Effects*. Edited by Jon Sprouse and Norbert Hornstein. Cambridge: University Press Cambridge, pp. 64–108.
- Phillips, Colin. 2013b. Some arguments and non-arguments for reductionist accounts of syntactic phenomena. *Language and Cognitive Processes* 28: 156–87. [\[CrossRef\]](#)
- Phillips, Colin, Matthew W. Wagers, and Ellen F. Lau. 2011. Grammatical illusions and selective fallibility in real-time language comprehension. In *Experiments at the Interfaces*. Bingley: Emerald, vol. 37, pp. 147–80.
- Pickering, Martin, and Guy Barry. 1991. Sentence processing without empty categories. *Language and Cognitive Processes* 6: 229–59. [\[CrossRef\]](#)
- Pinker, Steven, and Paul Bloom. 1990. Natural language and natural selection. *Behavioral and Brain Sciences* 13: 707–27. [\[CrossRef\]](#)
- Polinsky, Maria, Carlos G. Gallo, Peter Graff, Ekaterina Kravtchenko, Adam Milton Morgan, and Anne Sturgeon. 2013. Subject islands are different. In *Experimental Syntax and Island Effects*. Edited by Jon Sprouse and Norbert Hornstein. Cambridge: Cambridge University Press, pp. 286–309.
- Pollard, Carl, and Ivan A. Sag. 1994. *Head-Driven Phrase Structure Grammar*. Chicago: University of Chicago Press and CSLI Publications.
- Postal, Paul. 1971. *Crossover Phenomena*. New York: Holt, Rinehart and Winston.
- Pritchett, Bradley L. 1992. *Grammatical Competence and Parsing Performance*. Chicago: University of Chicago Press.
- Pritchett, Bradley L. 1988. Garden path phenomena and the grammatical basis of language processing. *Language* 64: 539–76. [\[CrossRef\]](#)
- Progovac, Ljiljana. 2016. A gradualist scenario for language evolution: Precise linguistic reconstruction of early human (and neandertal) grammars. *Frontiers in Psychology* 7: 1714. [\[CrossRef\]](#) [\[PubMed\]](#)
- Pullum, Geoffrey K. 2019. What grammars are, or ought to be. Paper presented at 26th International Conference on Head-Driven Phrase Structure Grammar, București, Romania, July 25–26. Edited by Stefan Müller and Petya Osenova. Stanford: CSLI Publications, pp. 58–79.
- Rajkumar, Rajakrishnan, Marten van Schijndel, Michael White, and William Schuler. 2016. Investigating locality effects and surprisal in written English syntactic choice phenomena. *Cognition* 155: 204–32. [\[CrossRef\]](#) [\[PubMed\]](#)
- Reinhart, Tanya, and Eric Reuland. 1991. Anaphors and logophors: An argument structure perspective. In *Long-Distance Anaphora*. Edited by Jan Koster and Eric Reuland. Cambridge: Cambridge University Press, pp. 283–322.
- Reis, Marga. 1993. *Wortstellung und Informationsstruktur*. Berlin: Walter de Gruyter.
- Rizzi, Luigi. 1990. *Relativized Minimality*. Cambridge, MA: MIT Press.
- Roberts, Craige. 2012. Information structure: Towards an integrated formal theory of pragmatics. *Semantics and Pragmatics* 5: 6–1. [\[CrossRef\]](#)
- Rochemont, Michael. 1989. Topic islands and the subadjacency parameter. *Canadian Journal of Linguistics—Revue Canadienne de Linguistique* 34: 145–70. [\[CrossRef\]](#)
- Rochemont, Michael, and Peter W. Culicover. 1990. *English Focus Constructions and the Theory of Grammar*. Cambridge: Cambridge University Press.
- Ross, John R. 1967. Constraints on Variables in Syntax. Ph.D. thesis, MIT, Cambridge, MA, USA.
- Ross, John R. 1987. Islands and syntactic prototypes. *Chicago Linguistic Society Papers* 23: 309–20.
- Sabel, Joachim. 2002. A minimalist analysis of syntactic islands. *Linguistic Review* 19: 271–315. [\[CrossRef\]](#)
- Safir, Ken. 2017. Weak crossover. In *The Wiley Blackwell Companion to Syntax*. Edited by Martin Everaert and Henk van Riemsdijk. Hoboken, NJ: Wiley Online Library, pp. 1–40.
- Sag, Ivan A. 1997. English relative clause constructions. *Journal of Linguistics* 33: 431–84. [\[CrossRef\]](#)
- Sag, Ivan A. 2012. Sign-based construction grammar—A synopsis. In *Sign-Based Construction Grammar*. Edited by Hans C. Boas and Ivan A. Sag. Stanford: CSLI Publications, pp. 61–197.
- Sag, Ivan A., Inbal Arnon, Bruno Estigarribia, Philip Hofmeister, T. Florian Jaeger, Jeanette Pettibone, and Neal Snider. 2006. *Processing Accounts for Superiority Effects*. Stanford, CA: Stanford University, Unpublished ms.

- Sag, Ivan A., and Janet D. Fodor. 1994. Extraction without traces. In *Proceedings of the Thirteenth West Coast Conference on Formal Linguistics, Stanford University*. Edited by Raul Aranovich, William Byrne, Susanne Preuss and Martha Senturia. Stanford University: CSLI Publications. pp. 365–84.
- Sag, Ivan A., Philip Hofmeister, and Neal Snider. 2007. Processing complexity in Subjacency violations: The complex noun phrase constraint. Paper presented at 43rd Annual Meeting of the Chicago Linguistic Society, Chicago, IL, May 3–5. Chicago: University of Chicago.
- Sauerland, Uli. 1999. Erasability and interpretation. *Syntax* 2: 161–88. [\[CrossRef\]](#)
- Selkirk, Elisabeth. 2011. The syntax-phonology interface. In *The Handbook of Phonological Theory*. Edited by John A. Goldsmith, Jason Riggle and Alan C. L. Yu. Oxford: Wiley-Blackwell, vol. 2, pp. 435–83.
- Shain, Cory, Idan Asher Blank, Marten van Schijndel, William Schuler, and Evelina Fedorenko. 2020. fmri reveals language-specific predictive coding during naturalistic sentence comprehension. *Neuropsychologia* 138: 107307. [\[CrossRef\]](#) [\[PubMed\]](#)
- Shan, Chung-Chieh, and Chris Barker. 2006. Explaining crossover and superiority as left-to-right evaluation. *Linguistics and Philosophy* 29: 91–134. [\[CrossRef\]](#)
- Sprouse, Jon. 2007. Continuous acceptability, categorical grammaticality, and experimental syntax. *Biolinguistics* 1: 118–29. [\[CrossRef\]](#)
- Sprouse, Jon, Matt Wagers, and Colin Phillips. 2012a. A test of the relation between working-memory capacity and syntactic island effects. *Language* 88: 82–123. [\[CrossRef\]](#)
- Sprouse, Jon, Matt Wagers, and Colin Phillips. 2012b. Working-memory capacity and island effects: A reminder of the issues and the facts. *Language* 88: 401–7. [\[CrossRef\]](#)
- Staum Casasanto, Laura, Philip Hofmeister, and Ivan A. Sag. 2010. Understanding acceptability judgments: Distinguishing the effects of grammar and processing on acceptability judgments. Paper presented at 32nd Annual Conference of the Cognitive Science Society, Portland, OR, USA, August 11–14. Edited by Stellan Ohlsson and Richard Catrambone. Austin: Cognitive Science Society, pp. 224–29.
- Szabolcsi, Anna, and Terje Lohndal. 2017. Strong vs. weak islands. In *The Wiley Blackwell Companion to Syntax*, 2nd ed. Edited by Martin Everaert and Henk van Riemsdijk. New York: John Wiley.
- Temperley, David. 2007. Minimization of dependency length in written english. *Cognition* 105: 300–33. [\[CrossRef\]](#)
- Trotzke, Andreas, Markus Bader, and Lyn Frazier. 2013. Third factors and the performance interface in language design. *Biolinguistics* 7: 1–34. [\[CrossRef\]](#)
- Truckenbrodt, Hubert. 1995. Phonological Phrases: Their Relation to Syntax, Focus, and Prominence. Ph.D. thesis, MIT, Cambridge, MA, USA.
- Ueno, Mieko, and Susan M. Garnsey. 2008. An ERP study of the processing of subject and object relative clauses in japanese. *Language and Cognitive Processes* 23: 646–88. [\[CrossRef\]](#)
- van Dyke, Julie, and Richard Lewis. 2003. Distinguishing effects of structure and decay on attachment and repair: A cue-based parsing account of recovery from misanalyzed ambiguities. *Journal of Memory and Language* 49: 285–316. [\[CrossRef\]](#)
- van Schijndel, Marten, Andy Exley, and William Schuler. 2013. A model of language processing as hierarchic sequential prediction. *Topics in Cognitive Science* 5: 522–40. [\[CrossRef\]](#) [\[PubMed\]](#)
- Varaschin, Giuseppe. 2021. A Simpler Syntax of Anaphora. Ph.D. thesis, Universidade Federal de Santa Catarina, Florianopolis, Brazil.
- Varaschin, Giuseppe, Peter W. Culicover, and Susanne Winkler. 2022. In pursuit of condition C. In *Information Structure and Discourse in Generative Grammar: Mechanisms and Processes*. Edited by Andreas Konietzko and Susanne Winkler. Berlin: Walter de Gruyter, to appear.
- Vasishth, Shravan, and Richard L. Lewis. 2006. Argument-head distance and processing complexity: Explaining both locality and antilocality effects. *Language* 82: 767–94. [\[CrossRef\]](#)
- Vasishth, Shravan, Bruno Nicenboim, Felix Engelmann, and Frank Burchert. 2019. Computational models of retrieval processes in sentence processing. *Trends in Cognitive Sciences* 23: 968–82. [\[CrossRef\]](#) [\[PubMed\]](#)
- Villata, Sandra, Luigi Rizzi, and Julie Franck. 2016. Intervention effects and Relativized Minimality: New experimental evidence from graded judgments. *Lingua* 179: 76–96. [\[CrossRef\]](#)
- Warren, Tessa, and Edward Gibson. 2002. The influence of referential processing on sentence complexity. *Cognition* 85: 79–112. [\[CrossRef\]](#)
- Wasow, Thomas. 1979. *Anaphora in Generative Grammar*. Amsterdam: John Benjamins Publishing Company.
- Wexler, Kenneth, and Peter W. Culicover. 1980. *Formal Principles of Language Acquisition*. Cambridge, MA: MIT Press.
- Winkler, Susanne, Janina Radó, and Marian Gutscher. 2016. What determines ‘freezing’ effects in *was-für* split constructions? In *Firm Foundations: Quantitative Approaches to Grammar and Grammatical Change*. Edited by Sam Featherston and Yannick Versley. Boston and Berlin: Walter de Gruyter, pp. 207–31.
- Yadav, Himanshu, Samar Husain, and Richard Futrell. 2021. Do dependency lengths explain constraints on crossing dependencies? *Linguistics Vanguard* 7: 1–7. [\[CrossRef\]](#)
- Zwicky, Arnold M. 1986. The unaccented pronoun constraint in English. *Ohio State University Working Papers in Linguistics* 32: 100–13.