Effects of processing on the acceptability of ‘frozen’ extraposed constituents

Abstract In syntactic theory, ‘freezing’ refers to the idea that a constituent extraposed to a non-canonical position is resistant to extraction of any its subconstituents (What did Terry see a movie yesterday about ____ ?). The unacceptability of such examples, compared to minimally different sentences without extraposition, has been claimed to be a result of a grammatical constraint on dependency formation, e.g. Ross’ (1967) Frozen Structure Constraint. Here, we argue that the gradient pattern of acceptability associated with such examples is better explained in terms of processing complexity. Experimental evidence from controlled acceptability tasks indicate that extraction and extraposition each independently lower judgments (Experiments 1 & 2), as does increasing the distance between a noun phrase and an in-situ (non-displaced) preposition that subcategorizes for it (Experiment 3). Indeed, Experiment 4 shows that the total penalty for extraction from extraposed constituents is predictable from the summed acceptability penalties independently linked to extraction and extraposition. As prior research links both of these displacement phenomena to increased processing costs, we conclude that a grammar of unbounded dependencies in English does not require a ‘freezing’ constraint and that an analysis rooted in processing costs better accounts for variance in judgment data.

1. Introduction

Dependencies in English, as in other languages, are theoretically unbounded in length. This means that there is no theoretical limit to the number of words that can separate a wh-word like what and the point at which it is thematically interpreted. The example in (1) could, in theory, continue on and on with more embedded complement clauses ad nauseum and not violate any rules of English grammar:

(1) What did the police officer say that it was commonly known that it was illegal to claim that . . .

In light of this unboundedness of filler-gap dependencies, there are a number of otherwise puzzling instances of ill-formedness that involve long-distance wh-dependencies. Some syntactic accounts, such as those of Chomsky (1964) and Ross (1967), propose that grammatical constraints account for the unacceptable status of items like (2b), derived from the minimally different (2a):
Ross (1967), for instance, proposes that a grammatical constraint called the Complex NP Constraint rules out extraction from a relative clause. The conclusion that there is such a grammatical constraint follows from the observation that questions in English are formed by moving the wh-phrase (in this case what) to the front of the sentence. Since such a movement occurs in (2b) and ungrammaticality results, a straightforward explanation is that a syntactic constraint renders extractions from such contexts ungrammatical. The research program that explores the nature of such constraints has been enormously productive, and has led to fundamental developments in syntactic theory (represented by such works as Chomsky 1973, 1977, 1981, 1986, 1995).

An alternative perspective on such phenomena recognizes that not all cases of unacceptability are necessarily ungrammatical and must be ruled out by grammatical principles. This perspective was articulated at the very beginning of contemporary syntactic theory by Miller & Chomsky (1963), who observed that sentences with multiple center-embedding were formally well-formed, that is, ‘grammatical’, but difficult or impossible to process, and hence unacceptable.

Subsequently, there have been a number of attempts to account for apparent ungrammaticality in terms of processing difficulty. For example, Jackendoff & Culicover (1972) offer a processing explanation for the intuition that A movements of indirect objects in English are mildly unacceptable, as in (3) & (4):

(3)  a. Terry bought Robin an ice cream cone.
    b. ??Who did Terry buy _ an ice cream cone?

(4)  a. Terry will send Robin a letter.
    b. ??Who will Terry send _ a letter?

Kluender (1992, 1998, 2004) subsequently proposed that unacceptability attributed to certain apparent grammatical constraints were actually the consequence of processing difficulty. On the basis of reading time and acceptability evidence, a case has been made for processing accounts of a wider range of phenomena by Sag, Hofmeister and colleagues (Hofmeister 2007; Hofmeister & Sag 2010; Hofmeister, Jaeger, Sag, Arnon & Snider 2007; Hofmeister, Staum Casasanto & Sag in press; Hofmeister, Jaeger, Arnon, Sag & Snider in press).

Our focus here is on explanations that appeal to ‘freezing’ as a grammatical constraint. In syntactic theory, freezing refers to the idea that a constituent displaced to a non-canonical position is resistant to extraction of any its
subconstituents (Wexler & Culicover 1980). A typical example of a frozen structure is a PP that has been extraposed. As (5b) shows, extraction from such a PP causes a reduction in acceptability (the use of symbols like $t$ is merely to indicate the position where the wh-word is thematically interpreted, but it is not meant to endorse the syntactic or psychological validity of movement traces).

(5)  
   a. You saw [a picture ___] yesterday [PP of Thomas Jefferson].
   b. ?Who$_t$ did you see [a picture ___] yesterday [PP of t$_t$]?

The notion that some syntactic configurations produce freezing first appears in Ross (1967:305), and has been invoked subsequently to account for a substantial range of phenomena (see Corver 2006 for a review). Ross’ (1967) version of the freezing constraint in (6) deals specifically with examples like those above:

(6) The Frozen Structure Constraint (FSC): If a constituent C, where C is a clause or a prepositional phrase, has been extraposed from a noun phrase whose head is lexical, this noun phrase may not be moved, nor many any element of C be moved out of C (pp. 160, 165).

From the perspective of linguistic theory, such a grammatical principle rules out extraction from extraposed structures. Nonetheless, such structures are also characterized by properties that complicate sentence processing. In (5b), for example, the processor is holding who in memory when it encounters picture. It has no information at that point that the gap is in a PP, and because of the extraposition, there is no information at yesterday that a modifying PP is upcoming that will modify the NP a picture. When of is encountered the processor must simultaneously link who to the gap, link this reconstructed PP to the extraposition site, and interpret the PP as a complement of picture. The question, therefore, is: Is the lack of acceptability in freezing configurations a matter of grammar per se, or does it have a processing explanation?

To address this, we present evidence from an array of acceptability studies showing that key features of freezing cause acceptability reductions even when these examples do not violate any putative grammatical principle. Specifically, extraposition results in lower judgments, even in the absence of extraction, and extraction too lowers judgments independently of extraposition. Furthermore, we demonstrate that the reduced acceptability of freezing cases is equivalent to the summed acceptability penalties for extraction and extraposition. The net effect of these penalties thus leaves little variation or unacceptability to be explained by a grammatical constraint.

Our hypothesis is that freezing effects emerge from the combination of independent sources of processing costs: extraction and extraposition. As Hawkins (1999) notes, filler-gap dependencies, i.e. extractions, introduce numerous
costs, including the search for the correct point of retrieval and integration — the “gap” — while processing intervening material. In this sense, filler-gap dependency processing generally gives rise to a multi-tasking problem, and a variety of evidence suggests that filler-gap processing makes sentence processing more difficult.

On top of this, freezing violations involve extraposition or the rightward dislocation of a constituent from its canonical position. Here, too, the available evidence suggests that extraposition is a source of processing difficulty. Like extraction, extraposition separates two syntactically and semantically related constituents. Levy, Fedorenko, Breen & Gibson (2012) demonstrate through self-paced reading experiments that extraposition of relative clauses across VPs and PPs creates processing difficulty, compared to in-situ relative clauses. As discussed by Levy, Fedorenko, Breen & Gibson (2012), there are various possible reasons why extraposition creates comprehension difficulty, including derivational complexity, activation decay/retrieval interference, and structural expectations. Our goal here is not to decide which of these theories best capture effects of extraposition, particularly since our only measures are acceptability judgments, which do not always capture differences in online sentence processing. Rather, we note that all of these views on sentence processing complexity predict that processing extraposed constituents is harder than in-situ constituents. However, at least locality-based theories and structural expectation theories predict the difficulty ascribed to extraposition should increase as extraposition distance increases. This is because greater distance not only allows for more activation decay and/or interfering constituents, but the expectation for a modifying extraposed PP decreases with distance.

Yet another way of interpreting the difficulty attached to extraposition is in terms of STRUCTURAL REANALYSIS. That is, in an example like (5b), the processing of yesterday may lead the parser to conclude that the NP constituent a picture is completed. Subsequent processing of the PP of contradicts this initial analysis, leading to a reanalysis process. In this vein, Ferreira & Henderson (1991) illustrate that recovery from garden path misanalysis is more difficult as more words are added to the ambiguous region. Thus, like locality- and expectation-based theories, a structural reanalysis interpretation of extraposition effects also predicts these effects to be gradient in nature. A variety of approaches to complexity in sentence processing, therefore, converge on the prediction that extraposition of PPs leads to difficulty, compared to in-situ PPs. In the concluding section, we return to the issue of how to best analyze the difficulty associated with PP extraposition.

The present work builds on Culicover & Winkler (submitted) who argue that the unacceptability of freezing constructions arises from an interaction between the grammatical structure and the mechanism by which the structure is interpreted in real time. Culicover and Winkler observe that not all such extractions are equally unacceptable; in
fact, some are quite acceptable. For example, consider the following contextualized example from Huck & Na (1990) with contrastive emphasis on the sentence-final preposition:

(7) I know Alger found letters in the file TO Chambers, certainly, but I’m not sure I can remember whom he found letters in the files FROM ti. [Huck and Na 1990:66 (39c)]

Building on Huck and Na’s observations, Culicover and Winkler claim on the basis of intuitions that examples with the same syntactic structure with respect to freezing conditions vary in acceptability as a function of non-structural factors such as the distance between the extracted element and its corresponding gap. Our investigations here expand upon Culicover & Winkler’s work by testing whether freezing effects are predictable from the independent penalties due to extraction and extraposition.

Freezing effects and the related grammatical proposals may strike some as being off the beaten path and narrowly concerned with peripheral data, particularly those dealing in non-transformational approaches to syntax. To the contrary, we believe the topic sheds light on an issue of quite general importance: how should judgments of unacceptability be analyzed? At the moment, the field lacks clear guidelines for deciding between an analysis of unacceptability in terms of grammar versus processing costs (but see Staum Casasanto, Hofmeister & Sag (2010) & Hofmeister, Staum Casasanto, & Sag (submitted) for some suggestions). Why a particular structure sounds unnatural could thus theoretically be because of either grammatical principles or the processing costs the structure gives rise to. When we further consider that many structural generalizations appear to have some functional purpose (e.g. case-, number-, and person-marking all serve to explicitly mark the relationship between constituents that can be arbitrarily far from each other), grammatical analyses often compete with functional or processing-oriented explanations of the same phenomena (see, for instance, Sprouse & Hornstein (2012) and papers therein).

How can we adjudicate between such competing explanations of acceptability patterns? The approach we take here is to find independent evidence of general processing costs tied to a structure, see how much variation in acceptability judgments these costs can explain, and then ask what’s left for a hypothetical grammatical constraint to explain. We arrive at the conclusion that if one accepts that independently motivated processing costs can lower judgments, then little remains for grammar to explain in the case of freezing violations. As discussed in the general discussion section, our perspective is that an analysis based on independently motivated processing principles is ultimately superior to one based on grammatical principles that are without analogy in any other cognitive domain, because the former is cost-free in the sense that it does not introduce any new mechanisms or processes to account for human behavior and cognition.
The preference for an analysis in terms of general cognitive constraints, however, is not only based on pre-suppositions about explanatory elegance. As we argue in the concluding section, if one concludes that grammatical principles ultimately determine the unacceptability of freezing violations, then one needs to assume that processing costs are suspended or somehow irrelevant in violations of the FSC to make room for grammar—an assumption we find dubious at best. Secondly, an analysis which says that cognitive constraints determine the acceptability of these items attractively combines with other analyses that identify similar factors at play in other syntactic constructions. That is, a significant component of the variation in acceptability judgments for a range of items can be uniformly attributed to a basic set of processing costs. The challenge for such a cost-driven account, as we shall see, is to clearly identify what these processing costs are, what the mechanisms are that produce the costs, and how they aggregate within a sentence.

In the sections that follow, we test how the key features of freezing — extraction and extraposition — independently affect judgments. The first two experiments focus on gradient effects of extraposition: in Experiment 1, extraposition distance is varied but without extraction; in Experiment 2, we manipulate extraposition distance again, but this time in the context of extraction. Experiment 3 explores how distance between an NP and a syntactically and semantically related in-situ PP created by interposing material affects judgments of acceptability. Thus, Experiment 3 is designed to identify whether there is a general acceptability decrement tied to separating related NPs and PPs. Experiment 4 essentially combines the manipulations of Experiments 1 and 2: we manipulate both extraction and extraposition with the same set of participant and items to determine whether the freezing effect is tantamount to the summed independent decrements caused by extraction and extraposition.

## 2. Experiment 1: Extraposition Distance

Experiment 1 tests the hypothesis that extraposition by itself (without extraction) lowers judgments, and that as extraposition distance increases, acceptability judgments decrease. This hypothesis is motivated by a substantial body of evidence from self-paced reading and eye-tracking tasks showing that increasing the distance between syntactically-related linguistic units slows processing times at the point of integration (Gibson 1998; Gibson 2000; Grodner & Gibson 2005; Bartek, Lewis, Vasishth & Smith 2011). Moreover, if extraposition increases the likelihood of syntactic misanalysis, then prior evidence suggests that the necessary reanalysis process should be harder as extraposition distance increases (Ferreira & Henderson 1991).
We consequently reason that extraposition may give rise to processing difficulty associated with long-distance dependency processing and reanalysis. If extraposition gives rise to lower judgments because of processing difficulty, it is expected that this effect should be gradient: longer extrapositions should lower judgments more than extraposition over shorter distances. In contrast, grammatical accounts of freezing effects do not typically assume that extraposition in and of itself results in ungrammaticality. It is only when material is extracted from the extraposed constituent that ungrammaticality results. A grammatical account of freezing is thus ill-equipped to account for differences due to extraposition distance.

2.1 Materials

Distance here is measured in terms of the number of phrases intervening between an NP and an extraposed PP.\(^1\) Participants read examples where the head noun of an NP was separated from a syntactically and semantically related PP by zero, one, or two phrases. For instance, in the sample item in (8c), the NP a story is separated from the PP about that actor by two phrases, anxiously and while having breakfast:

(8) a. My friend read a story about that actor anxiously while having breakfast. [= SHORT]
   
b. My friend read a story anxiously about that actor while having breakfast. [= MEDIUM]
   
c. My friend read a story anxiously while having breakfast about that actor. [= LONG]

Each participant saw only one condition of each of the 24 items and an equal number of each condition (=8) throughout the entire experiment. 72 distracter items accompanied the critical items. These distractors spanned the spectrum of grammaticality from highly natural items (Republicans and Democrats in the State Senate appeared close to a power-sharing deal) to highly unnatural items (Police have released no information what on they may have found at scene the). 28 of the fillers belonged to an unrelated experiment.

\(^1\)Note that this distance metric differs from the way that Gibson and colleagues measure distance. They measure distance in terms of the new discourse referents (individuals or events). Hence, a word like anxiously would not increase the integration difficulty of a story and about that actor as it does not introduce a new discourse referent. The fact that we nevertheless find significant, gradient effects of extraposition may thus not be easily reconcilable with locality-based theories without revising how distance is calculated. Our use of ‘distance’ here thus bears a greater resemblance to its use in Hawkins (1994, 1999) where distance is calculated in terms of node quantity. As in our case, Hawkins remains non-committal about what processing mechanisms make longer filler-gap dependencies more complex, observing that the number of syntactic nodes is ‘just one index of the relative complexity of an FGD, but it is a fundamental one with numerous correlating properties that involve additional processing operations.’ See concluding section for further discussion.
2.2 Procedure

In all the experiments described here, we collected acceptability judgments via Amazon.com’s Mechanical Turk marketplace. Gibson, Piantadosi & Fedorenko (2011), Sprouse (2011), and others have demonstrated that this means of collecting judgments produces results similar to traditional laboratory methods, while simultaneously providing the advantages of large participant numbers and relatively short data gathering times.

At the beginning of each experiment, participants identified their physical location and language background. In order to remove any incentive to lie about language background, participants were not excluded from the experiment on the basis of language history. However, the data from non-English-speaking participants were excluded prior to analysis. This removal process generally affected only a handful of participants per study. The instructions also asked participants not to complete the Mechanical Turk assignment if they had completed a related experiment; however, some individuals disregarded these instructions, and we discarded the data for these participants as well.

The instructions preceding each experiment asked participants to judge the subsequent sentences in terms of naturalness on either a 5 or 7 point scale. Following each sentence, we included a comprehension question to ensure that participants read the items carefully. The data from participants with accuracy rates below 75% were eliminated entirely before analysis. Participants received between $1.50 and $3 for their participation.

We analyzed all acceptability data by using linear mixed effects models with participants and items modeled as random variables (Pinheiro & Bates 2000; Baayen 2007; Baayen, Davidson & Bates 2008). Prior to analysis, all predictors were centered and higher order variables were based on these centered predictors. In addition, we utilized the maximum random effect structure that converged for each experiment. Thus, for a design with two factors, F1 and F2, the random effect structure included random intercepts for participants and items, as well as by-participant and by-item random slopes for each factor and their interaction (F1 × F2).

In Experiment 1, we coded distance as a polynomial variable, because distance effects may not be strictly linear, i.e. the acceptability contrast between the short and medium conditions may not be equivalent to the acceptability contrast between the medium and long conditions. The quadratic component of this variable thus captures whether there are non-linear effects of the predictor variable.

2.3 Participants

60 participants took part in the experiment. All participants identified their location as the US and indicated that they were native English speakers.
2.4 Results

The acceptability results show a significant linear effect of distance, i.e. judgments decrease as the distance between the NP and PP increases (see Figure 1). Additionally, the significance of a quadratic term in the model reflects the fact that the difference between the short and medium conditions is larger than the difference between the medium and long conditions. In other words, the effect of interposing a second phrase was less than the effect of interposing the first one.

2.5 Discussion

Even without extraction, acceptability judgments drop significantly in the presence of extraposition. This suggests that a fundamental property of freezing constructions – extraposition – by itself introduces an acceptability penalty and that this penalty intensifies with distance. As the standard explanation for distance effects appeals to psycholinguistic rather than grammatical principles, the current findings suggest that processing difficulty — in particular, retrieval and integration difficulty — accounts for a substantial part of the unacceptability of freezing constructions.
Nonetheless, these results are not incompatible with an account of freezing effects based on grammatical principles. After all, processing effects on acceptability judgments can co-exist with grammatical effects. Thus, the net unacceptability of freezing violations, such as (5b), may include some penalty from the processing cost due to extraposition; however, this may be insufficient by itself to explain the degree of unacceptability. Room is left over, therefore, for grammatical constraints to work in tandem with processing constraints in shaping judgments of linguistic acceptability.

The gradient effects of extraposition distance, however, are not straightforwardly predicted by an grammatical account. In contrast, such gradient effects are expected in light of several different, non-exclusive psycholinguistic principles that predict processing costs for extraposition. Locality-based theories of dependency processing predict that retrieval and integration difficulty should increase as the distance between the NP & PP increases. Similarly, expectation-based accounts of online processing difficulty predict that an extraposed constituent is generally less and less likely as the extraposition distance grows. A structural reanalysis interpretation also predicts extraposition effects to be gradient. If the parser is incorrectly led to believe that the NP is finished or syntactically complete, then this belief grows stronger as more and more input argues in this direction. In this sense, what we observe here may be comparable to the ‘digging in’ effects of Tabor & Hutchins (2004): the more committed the parser becomes to a particular analysis due to features in the input, the harder it is to back out of the analysis. Our results do not distinguish between these alternative treatments of extraposition difficulty; we return to the issue of how to best treat extraposition effects in the general discussion.

3. Experiment 2: Extraction Distance

If there are general processing costs ascribed to extraposition, then extraposition should affect not only judgments for sentences without extraction but also for sentences with extraction. This is particularly relevant for examples of freezing-violations, where two long-distance dependencies exist: one between the extracted or fronted element and the extraposed PP, and a second between the frozen NP and the extraposed PP. As in the previous experiment, a cost-based analysis of freezing effects predicts distance effects to be gradient, whereas a grammatical account merely predicts that extraposition plus extraction is ungrammatical without comment on the relative degrees of extraposition.
Processing complexity and freezing effects

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Standard Error</th>
<th>t-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>WH-COMPLEXITY</td>
<td>0.141</td>
<td>0.077</td>
</tr>
<tr>
<td>DISTANCE</td>
<td>−0.520</td>
<td>0.067</td>
</tr>
<tr>
<td>WH-COMPLEXITY × DISTANCE</td>
<td>−0.002</td>
<td>0.093</td>
</tr>
</tbody>
</table>

Table 2: LME summary from Experiment 2

3.1 Materials

Experiment 2 thus looks at how distance affects judgments for sentences with extraposition and extraction. Sentences with the same meaning but greater distance between an extracted *wh*-phrase and the preposition (due to extraposition) should yield lower judgments of acceptability. In addition, we also manipulated the complexity of the *wh*-filler, based on prior work showing that more complex *wh*-phrases (a) facilitate processing at retrieval sites and (b) yield higher judgments than minimally different sentences with bare *wh*-words (Hofmeister 2007; Hofmeister & Sag 2010; Hofmeister 2011).

(9)  
   a. Kenneth finally revealed (which President/who) he overheard a nasty remark about earlier while on the subway.
   b. Kenneth finally revealed (which President/who) he overheard a nasty remark earlier about while on the subway.
   c. Kenneth finally revealed (which President/who) he overheard a nasty remark earlier while on the subway about.

Participants saw 24 such items in the experiment, but each participant rated only 1 condition from each item. Each condition was seen by each participant an equal number of times throughout the experiment. 72 distracter items appeared along with the critical items.

3.2 Procedure

Procedure was the same as in Experiment 1, except that the two fixed effect factors were treatment coded (a.k.a. dummy coding).
3.3 Participants

94 participants in total completed the experiment and received payment. The data from 6 participants were eliminated because they were not native English speakers or indicated they were outside the US. Another 2 failed to indicate their language history, 2 completed the experiment more than once, and a further 2 had already completed Experiment 1. We eliminated the data from all these individuals, leaving 82 participant datasets.

3.4 Results

As in the previous experiment, increasing dependency length via extraction systematically lowers acceptability judgments: two intervening phrases between the NP and PP yields the lowest judgments, followed by sentences with one intervening phrases, while those with no extraposition give rise to the highest judgments as seen in Figure 2. Wh-complexity, though, has only a marginal effect on judgments: judgments are higher for items with which-N phrases as opposed to bare wh-words.

3.5 Discussion

Experiment 2 adds to the results of Experiment 1 by showing that the same general decrements associated with extraposition without extraction occur in contexts with extraction. Again, this pattern of results follows naturally from the aforementioned principles of language processing, but has no apparent explanation in terms of grammatical constraints. The gradient effects of distance, moreover, are not predicted by an analysis which says that extraposition plus extraction equals ungrammaticality.

Although which-NPs yielded higher judgments, in accord with previous findings, this trend failed to reach significance. Since the trend was in the predicted direction, and the effect was statistically marginal, we tentatively treat these results as being consistent with the view that it is the processing costs of freezing violations that accounts for their unacceptability. However, as this factor is not central to our understanding of freezing effects, we leave aside any further discussion of wh-complexity.

As in Experiment 1, the finding of gradient effects of extraposition does not by itself remove the possibility that grammatical constraints account for freezing effects. The data here remain compatible with the hypothesis that a

---

Note that if structural reanalysis or structural expectation costs, rather than locality-based retrieval difficulty, accounts for the unacceptability of extraposition, the marginal status of these effects is unsurprising. Hofmeister (2011) interprets effects due to the complexity of fillers, such as wh-phrases, in terms of retrieval difficulty. If retrieval difficulty plays a relatively minor role in extraposition effects, then it is reasonable to expect that wh-phrase complexity should have a correspondingly weak effect on the acceptability of sentences with extraposed constituents.
grammatical ban on extraction from extraposed constituents exists, so long as the hypothesis allows extraposition to independently lower judgments due to processing costs. That is, the gradient effects of extraposition can, in principle, simply add to whatever the penalty for extraction from extraposition is.\(^3\)

### 4. Experiment 3: In-situ PP distance

Assuming that dependency length effects reflect general cognitive constraints on language processing, we expect them to occur whether or not extraposition occurs. That is, other means of separating a preposition from content that it subcategorizes for ought to produce similar results as extraposition. The purpose of Experiment 3 is thus to test how dependency length differences affect the acceptability of sentences with in-situ PPs.

\(^3\)It is theoretically possible that processing difficulty effects on acceptability judgments can gang up with grammar-based effects. Staum Casasanto, Hofmeister & Sag (2010) looked for such ganging up effects in an acceptability experiment that varied grammaticality and processing difficulty. They found that the processing difficulty difference between (1a) & (1b) had no effect on judgments when the sentence contained an ungrammatical verb form (defending), but did have a significant effect when the sentences contained the grammatical verb form (defended):

1. They couldn’t remember which lawyer that the reporter interviewed had defending/defended the elderly man at the courthouse.
2. They couldn’t remember which lawyer had defending/defended the elderly man that the reporter interviewed at the courthouse.

It thus remains to be seen under what conditions, if any, processing costs can gang up with grammatical constraints to lower acceptability judgments.
4.1 Materials

Here, we manipulated distance by altering the complexity of the material between a subject NP and its in-situ subcategorizing preposition. Thus, the distance between the critical NP and the preposition was either short (10c), medium length (10b), or long (10a). As the examples illustrate, relative clauses were used to modify dependency length. The long condition thus contained one more subject relative clause than the medium condition, which in turn contained one more subject relative clause than the short condition.

(10)  

a. The pit that the homeowner pushed the garbage that smelled of meat that was rotting into was seeping ooze into the groundwater.

b. The pit that the homeowner pushed the garbage that smelled of meat into was seeping ooze into the groundwater.

c. The pit that the homeowner pushed the garbage into was seeping ooze into the groundwater.

d. The pit into which the homeowner pushed the garbage that smelled of meat that was rotting was seeping ooze into the groundwater.

e. The pit into which the homeowner pushed the garbage that smelled of meat was seeping ooze into the groundwater.

f. The pit into which the homeowner pushed the garbage was seeping ooze into the groundwater.

Processing sentences with more relative clauses (or simply longer sentences), however, may lead to drops in acceptability independently of the dependency length between the subject NP and in-situ PP. Consequently, as a means for controlling for effects of overall sentence complexity, we included control versions without a stranded preposition (10d)-(10f). If dependency length has an effect above and beyond sentence complexity, we predict an interaction between sentence complexity and preposition stranding: the effect of sentence complexity should be greater given a stranded preposition.

24 critical items appeared in the experiment, but each participant rated only 1 condition from each item. 72 distracter items appeared along with the critical items.

4.2 Procedure

Procedure was identical to that used in Experiment 2.
Processing complexity and freezing effects

Table 3: LME summary from Experiment 3

<table>
<thead>
<tr>
<th></th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>t-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>IN-SITU</td>
<td>-0.417</td>
<td>0.115</td>
<td>-3.63</td>
</tr>
<tr>
<td>COMPLEXITY</td>
<td>-1.050</td>
<td>0.082</td>
<td>-12.74</td>
</tr>
<tr>
<td>IN-SITU × COMPLEXITY</td>
<td>-0.435</td>
<td>0.096</td>
<td>-4.55</td>
</tr>
</tbody>
</table>

Figure 3: Mean acceptability judgments from Experiment 3. Error bars show ± 1 standard error.

4.3 Participants

72 participants completed the experiment. 3 subjects’ data were eliminated due to identifying as non-native English speakers. The data from 5 additional subjects were removed because they had participated in one of the previous studies.

4.4 Results

Greater sentence complexity/length elicited lower ratings, i.e. (10c) & (10f) were better than (10b) & (10e), which were, in turn, better than (10a) & (10d). More importantly, preposition stranding caused a significant decrease in judgments (see Figure 3), and preposition stranding also interacted with sentence complexity: the effect of complexity was larger in the context of preposition stranding. In other words, the penalty for increased distance in the preposition stranding case cannot be ascribed entirely to sentence complexity – it also reflects a significant penalty for separating a NP from its subcategorizing prepositional head.
4.5 Discussion

In accord with the results from the first two experiments, these data confirm that increasing the distance between an NP and a syntactically and semantically related PP systematically lowers acceptability judgments. In this sense, distance-based effects on acceptability judgments are not specific to extraposition. In contrast to the previous experiments where we manipulated distance via extraposition, we observe distance effects here in the context of a filler-gap dependency: additional intervening relative clauses between the filler and gap lower acceptability judgments. These results, therefore, strongly resemble findings from the psycholinguistics literature on effects of dependency locality (Gibson 1998; Gibson 2000; Grodner & Gibson 2005).

To be clear, it is possible that the distance-based effects in Experiment 3 are underlyingly distinct from those in Experiments 1 & 2. In the former, distance is created via interposing different amounts of material between a filler and its gap. In the latter, we manipulated distance via extraposition. Consequently, the distance-based effects in Experiment 3 may strongly reflect locality-based effects due to activation decay/retrieval interference, whereas the distance-based effects in Experiments 1 & 2 may be more representative of structural expectations and reanalysis. Whatever the proper psycholinguistic analysis of these effects, our primary intention here is to show that separating syntactically and semantically related constituents systematically lowers judgments.

5. Experiment 4: Extraction and Extraposition

Both extraposition and extraction have independent and gradient effects on acceptability judgments, according to the previous results. As noted in the discussion section for each, though, these experiments do not remove the possibility that grammar says something special about the combination of extraction and extraposition. In Experiment 4, we crossed manipulations of extraction and extraposition to assess whether the unacceptability of freezing violations is predictable from the acceptability decrement due to extraction plus that due to extraposition.

5.1 Materials

For these materials, we systematically varied whether or not extraction happens, as well as whether the critical PP is extrapoosed or not. Crossing these two factors leads to example items like the one below:

(11) a. Tell me which actor your friend read a story about twice while having breakfast.
    b. Tell me which actor your friend read a story twice about while having breakfast.
Table 4: LME summary from Experiment 4

<table>
<thead>
<tr>
<th></th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>t-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>EXTRACTION</td>
<td>−1.44</td>
<td>0.167</td>
<td>−8.64</td>
</tr>
<tr>
<td>EXTRAPPOSE</td>
<td>−0.45</td>
<td>0.111</td>
<td>−4.01</td>
</tr>
<tr>
<td>EXTRACT x EXTRAPPOSE</td>
<td>0.09</td>
<td>0.200</td>
<td>−0.46</td>
</tr>
</tbody>
</table>

c. You told me your friend read a story about an actor twice while having breakfast.
d. You told me your friend read a story twice about an actor while having breakfast.

This design allows us to determine how much extraposition and extraction independently lower judgments, and whether combining the two lowers judgments beyond what is expected on the basis of each independent source of unacceptability.

Participants read 24 such items, but only saw 1 version of each item. The experimental materials were presented along with 54 unrelated filler items.

5.2 Procedure

Procedure was identical to that used in Experiment 2 & 3.

5.3 Participants

71 participants completed the experiment. 1 subject identified as a non-native speaker, 1 reported being outside the US, and 1 completed the experiment twice. An additional 12 participants completed the experiment who had already finished one of the previous experiments. In total, the analysis contained the data from 56 participants.

5.4 Results

As shown in Table 4, extraposition lowers judgments significantly. Examples like (11b) & (11d) are judged to be worse than those like (11a) & (11c). Extraction also lowers judgments: (11a) & (11b) are worse than (11c) & (11d). Crucially, there is no interaction: extraposition is no worse in contexts with extraction, (11b), than in contexts without, (11d). Indeed, the freezing violations (11b) have an average rating that is predictable on the basis of the independent average penalties for extraposition and extraction (see Figure 4).
5.5 Discussion

The data suggest that the low ratings for freezing violations are attributable to the combined penalties resulting from extraction and extraposition. Each of these syntactic phenomena independently lower judgments, and putting them together does not result in any “extra badness”. To explain the perceived unacceptability of these freezing violations, therefore, one does not need any special mechanism. So while the previous results from Experiments 1-3 leave a certain amount of room for grammatical constraints, what we find here leaves very little, if anything, for grammatical constraints to explain. Had the effects of extraction and extraposition combined “superadditively”, producing something worse than expected on the basis of the individual penalties, an argument could be made that the unexplained drop in acceptability results from the violation of a grammatical constraint (but it would also be consistent with processing costs becoming greater in the context of other processing costs, due to a processing bottleneck or resource limitations). As there are no such effects, we are left with the conclusion that freezing effects arise due to the independent effects of extraction and extraposition.

6. General Discussion

Showing that processing costs rather than grammatical principles explain the unacceptability of some construction is complicated by the fact that there are no direct tests of what makes a sentence unacceptable. In support of a processing-based analysis of freezing effects, our approach here has been to suggest that fundamental characteristics of freezing violations – extraposition and extraction – are associated with processing difficulty and that the costs
associated with each of these play a major role in how examples of freezing violations are judged. Our main finding, in this regard, is that the overall unacceptability of freezing violations is predictable from the independent, summed penalties attributed to each of these key features of freezing violations, as shown in Experiment 4.

To summarize, the results of our experiments are as follows:

- Experiment 1 shows that extraposition by itself (without extraction) lowers judgments, and that as extraposition distance increases, acceptability judgments decrease.

- Experiment 2 shows that extraposition in the context of a \(wh\)-dependency has a gradient effect on judgments: the longer the extraposition, the more acceptability judgments decrease.

- Experiment 3 shows that the longer an extraction from PP is, the lower the judgment, even when the PP is not extraposed. Thus extraposed PPs share a property with unextraposed PPs that has nothing to do with freezing.

- Experiment 4 shows that extraction and extraposition independently lower judgments and that the combined penalty for extraction from extraposed PPs is straightforwardly predictable from these summed independent penalties.

We initiated this research to determine whether there is a reason to attribute the unacceptability of extraction from extraposed PPs to a grammatical ‘freezing’ constraint. What would support such an account would be a component of unacceptability that cannot be attributed to independent factors, i.e., extraction, extraposition, and stranding of the preposition. The evidence that we have collected suggests that these factors are sufficient to account for all of the unacceptability, without appealing to a grammatical constraint, on the straightforward assumption that these factors contribute to the online processing difficulty and thereby lower acceptability judgments. The acceptability of freezing violations, according to our results, is thus tied up with factors that are not explicable on standard grammatical accounts.

Purely grammatical accounts of freezing, on the other hand, lead us to overlook the gradient decrements attributed to the general processing costs of forming long-distance dependencies. Such accounts, which typically make categorical assertions about grammaticality, certainly do not predict gradient effects. Unless we make the radical assumption that the independently verified acceptability penalties for extraction and extraposition do not occur when they appear in tandem, then we must conclude that the net unacceptability of freezing violations owes much, if not everything, to the combined costs of extraction plus extraposition. Thus, while it is never possible to completely
rule out a grammatical principle as a possible factor, our results suggest that such a grammatical principle would have little, if anything, to explain.

If freezing is a special case of a more general situation, a grammatical principle like the FSC become otiose and the grammar of long-distance dependencies can be simplified. The grammar need only specify that filler-gap dependencies are licensed without comment on what structural configurations such dependencies can enter or what boundaries they can cross.

What remains open to further investigation is precisely why these factors play a role in determining acceptability and to what extent other grammatical principles can be supplanted by appeals to processing complexity. The processing costs of filler-gap dependencies have been well-documented in the literature (Wanner & Maratsos 1978; Kluender & Kutas 1993; King & Kutas 1995). It thus seems uncontroversial to attribute the acceptability-lowering effects of extraction to processing costs linked to locating the appropriate integration site and retrieval difficulty.

Here, we have suggested that extraposition also lowers judgments due to the gradient processing costs of linking an extraposed constituent with its syntactic host. But unlike extraction costs, which are strongly tied to retrieval difficulty, the precise reason why extraposition increases processing difficulty is less well-understood. Creating discontinuous constituency may, in principle, give rise to processing difficulty because of retrieval difficulty, a high degree of surprisal, or the need to revisit a prior syntactic analysis. As noted previously, our experiments here were not designed to tease apart such theories. Nonetheless, we offer below some tentative views on the source of difficulty in sentences with extraposed constituents.

While distance clearly has a role to play in dependency processing due to memory retrieval costs (Gibson 1998; Gibson 2000; Bartek, Lewis, Vasishth & Smith 2011), it is less obvious why the relatively short distance between the NP *a story* and the PP *about ___*, as in (12) below, should create serious retrieval difficulty:

(12) My friend read a story twice about that actor while having breakfast.

The distance between the NP *a story* and *about that actor* here is minimal by just about any reckoning. The standard metric of distance in Gibson’s Distance Locality Theory (DLT), for instance, would not predict that integrating the NP and PP would be any harder in the presence of an intervening modifier like *twice* than in its absence. This is because DLT does not ascribe any cost to the intervening modifier because it does not introduce a discourse entity, i.e. it does not introduce a discourse-new individual or event referent. Moreover, it is unlikely that the modifier has the potential to interfere substantially with the retrieval of the NP, given their general dissimilarity.

Given these observations, we suggest that it is more likely that structural reanalysis costs or structural expectation
costs can better account for the difficulty of extraposition. On a structural reanalysis account, extraposition places demands on processing resources because (a) the adverb following the NP suggests that the NP representation is complete, i.e. the parser commits to an analysis, and (b) the presence of the PP forces the parser to revisit the original analysis. In cases of freezing, the memory costs of *wh*-dependency processing compound the reanalysis process as the point of retrieval and integration coincides with the reanalysis point. On a structural expectation account, extraposed constituents are difficult to process because they are unexpected. Levy, Fedorenko, Breen & Gibson (2012), in fact, found that the online difficulty associated with extraposed relative clauses varied systematically with structural expectations: strong expectations for a modifying relative clause facilitates comprehension of an extraposed relative clause. Given that our materials were not designed to decide between these hypotheses, we leave it for future research to determine the best interpretation of extraposition costs.

The account sketched here fits together with much recent work which has reinterpreted classical syntactic phenomena in terms of general cognitive constraints (Kluender 1991; Kluender 1992; Kluender 1998; Hofmeister & Sag 2010; Hofmeister, Jaeger, Arnon, Sag & Snider in press; Gieselman, Kluender & Caponigro 2011). Gieselman, Kluender & Caponigro (2011), for instance, demonstrate that the unacceptability of so-called negative islands is predictable from the independent acceptability drops due to negation, extraction, and the use of non-referential *wh*-phrases:

\[
\text{(13) How many projects didn’t the interns complete during the internship?}
\]

By obviating the need for a grammatical constraint to explain the unacceptability of such items, we ultimately arrive at a more streamlined grammar that overgenerates the set of possible sentences with long-distance dependencies, leaving it to processing considerations to cull the possible set of long-distance dependencies. This has the advantage that grammars ultimately become simpler and integrated with modern conceptions of cognitive architectures (Culicover & Jackendoff 2005). Perhaps the most notable advantage of such an account is that it deals directly with not only generalizations about syntactic phenomena, but also exceptions to those generalizations. That is, a proposal like the Frozen Structure Constraint captures the general unacceptability of extraction from extraposition. What it does not do, however, is capture apparent counterexamples — cases where extraction from extraposition sounds relatively acceptable. Often times, syntactic accounts must treat such troublesome data as “peripheral” or due to secondary syntactic principles. In contrast, the relative acceptability of some sentences with extraction from extraposition can be dealt with straightforwardly on an account such as ours: we make the strong prediction that (judgments for)
these sentences arise when various factors conspire to make the structure easier to process, e.g. supportive context, complex *wh*-phrases, highly predictable extraposed constituents, etc. A substantial advantage of a processing-based approach to freezing effects, therefore, is that we can view syntactic generalizations and variance surrounding those generalizations through the same mechanistic lens.

References


