# Frequency Effects in Subject Islands*

**Rui P. Chaves**
*Department of Linguistics*
*University at Buffalo, SUNY*

**Jeruen E. Dery**
*Zentrum für Allgemeine*
*Sprachwissenschaft, Berlin*

**Abstract**

This work provides evidence that Subject Island violation effects vanish if subject-embedded gaps are made as frequent and pragmatically felicitous as non-island counterpart controls. We argue that Subject Island effects are caused by the fact that subject-embedded gaps are pragmatically unusual – as the informational focus does not usually correspond to a dependent of the subject phrase – and therefore are highly contrary to comprehenders' expectations about the distribution of filler-gap dependencies (Hofmeister et al., 2013; Chaves, 2013). This not only explains why sentences with subject-embedded gaps often become more acceptable 'parasitically', in the presence of a second gap outside the island, but also explains why some Subject Island violations fail to exhibit any amelioration with repetition (Sprouse, 2009; Goodall, 2011; Crawford, 2011), some ameliorate marginally (Snyder, 2000, 2017) or moderately (Hiramatsu, 2000; Clausen, 2011; Chaves and Dery, 2014), and others become fully acceptable, as in our case. This conclusion extends to self-paced reading Subject Island studies (Stowe, 1986; Pickering et al., 1994; Kurtzman and Crawford, 1991; Phillips, 2006), which sometimes find evidence of gap-filling, sometimes do not.

**Keywords**: Filler-Gap Dependencies, Parasitism, Experimental Syntax

## 1 Introduction

Subject Islands are one of the strongest constraints on extraction. Whereas it is often possible to extract an NP from an NP complement, as in (1), extracting an NP from an NP subject as in (2) is generally regarded as impossible (e.g. Chomsky (1977, 106), Kayne (1981, 114), Huang (1982, 497), Lasnik and Saito (1992, 42), Nunes and Uriagereka (2000, 21), Jackendoff (2002, 42)).

*We thank audiences at the Laboratoire de Linguistique Formelle, Université Paris Diderot, the Department of Linguistics at the University of Rochester, and the *21st International Conference on Head-driven Phrase Structure Grammar* for feedback about some of the preliminary results reported in this work. We would also like to acknowledge useful discussions with Colin Phillips, Gail Mauner, Gregory Ward, Jean-Pierre Koenig, Maryellen MacDonald, Monica Do, Philip Hofmeister, Robert Levine, and Thomas Wasow. The usual exculpations apply. Finally, we wish to thank all reviewers for helping us to significantly improve this paper. Jeruen E. Dery was supported by German Federal Ministry of Education and Research Grant No. 01UG0711.

(1) a. Which celebrity$_x$ did you hire [the sister of _$_x$]?

b. It was ADELE$_x$ who you hired [the sister of _$_x$].

(2) a.*Which celebrity$_x$ did [the sister of _$_x$] hire you?

b.*It was ADELE$_x$ who [the sister of _$_x$] hired you.

Indeed, whereas NP extractions from NP complements are well-attested in corpora as the sample in (3) taken from the Corpus of Contemporary American English (Davies, 2008) illustrates, there are to our knowledge no English attestations of NP extraction from NP subjects whatsoever.[1]

(3) a. (...) this was something James didn't seem to have [a problem with _].

b. Others, we're going to have to find [some housing for _].

c. There was one last question my editor was dying to know [the answer to _].

A number of experimental studies have confirmed that Subject Island violations like those in (2) yield significantly lower acceptability ratings than uncontroversially grammatical controls (Sprouse, 2009; Goodall, 2011; Crawford, 2011; Clausen, 2011; Sprouse et al., 2015), and although comprehenders are known to postulate gaps as soon as possible (Crain and Fodor, 1985; Frazier, 1987; Stowe, 1986; Stowe et al., 1991) – including inside NP objects like (1) (Tollan and Heller, 2016) – there is no clear evidence that comprehenders actively postulate gaps inside NP subjects like those in (2) (Pickering et al., 1994; Stowe, 1986).

Matters are made more complex by the fact that Subject Island effects can often be alleviated if the subject-embedded gap is 'parasitic' (Engdahl, 1983) on a second gap outside the island environment, as in (4) and (5).[2]

(4) a.*Which rebel leader did [the rivals of _ ] shoot Castro?

b. Which rebel leader did [the rivals of _ ] shoot _ ?

(5) a.*What did [the attempt to repair _ ] ultimately damage the car?

b. What did [the attempt to repair _ ] ultimately damage _?

Not all Subject Islands can be avoided via parasitism, however. Subject-embedded gaps in tensed relative clauses like (6) are widely believed since Engdahl (1983) to be impossible to rescue via a secondary gap. The contrasts in (5) and (6) were experimentally validated by Kurtzman and Crawford (1991) and Phillips (2006).

[1] This contrasts with PP extractions from NP subjects, which are known to be acceptable since at least Ross (1967), and are attested in natural text (Huddleston et al., 2002; Santorini, 2007). See §2.2 for more discussion. We focus on NP extractions from NP subjects because they are unattested and widely regarded to be grammatically impossible since Chomsky (1973).

[2] See Levine and Hukari (2006, ch.4) for empirical arguments against accounts that propose that parasitic gaps are null resumptive pronouns rather than true filler-gap dependencies.

(6) a.*What did [the mechanic [who repaired _]] ultimately damage the car?

b.*What did [the mechanic [who repaired _]] ultimately damage _?

Phillips (2006) additionally provides evidence that comprehenders actively postulate gaps in subject-embedded infinitival environments like those in (5), but not in subject-embedded finite environments like those in (6). From this, Phillips (2006, 808) concludes that the language parser does not postulate gaps in positions where no gap is licit, parasitic or otherwise, which in turn lends support to the view that Subject Island effects are due to grammatically illicit filler-gap dependencies, rather than to cognitive constraints or performance limitations.

It is sometimes unclear what can be concluded from a null effect, however, because an experiment can fail to detect an extant effect either because the methodology was not sensitive enough, or because of a design flaw. In the present paper we provide evidence that Phillips (2006), Sprouse (2009), Goodall (2011), Crawford (2011), and others have overstated the significance of their null effect findings. We show that subject-internal gaps can in fact become as acceptable as uncontroversially acceptable controls (Experiment 1) and are actively postulated during on-line sentence comprehension (Experiment 2), by simply increasing their frequency, provided that the propositions they express are highly felicitous to begin with. We then examine gaps that are claimed to be illicit even parasitically, such as tense-embedded gaps like those in (6b), and again find evidence suggesting that increased exposure can not only induce comprehenders to view such gaps as acceptable as uncontroversially acceptable controls (Experiment 3), but also to actively postulate such gaps during on-line sentence comprehension (Experiment 4). We argue that the simplest interpretation of the facts is that extraction from subjects is grammatically licit, but usually deemed unacceptable for two functional reasons: (i) they are strongly contrary to comprehenders' expectations about where gaps usually reside, a factor that can be mitigated with increased exposure; (ii) subject-internal gaps often – though not always – pragmatically presuppose very unusual discourse circumstances.

More specifically, we claim that if a proposition is highly felicitous then the mention of the referents therein are necessarily relevant as well. Consequently, such a proposition more naturally lends itself to subject internal gaps, simply because the mention of such subject-internal referents is relevant for the proposition. In such cases, repetition can more easily induce the amelioration of the island effect, or even alleviate it completely. Our work thus highlights the importance of making sure the items employed in experimental island research express equally highly felicitous propositions to begin with, independent of extraction, so that the effect caused by extraction can be better isolated from orthogonal stylistic, semantic and pragmatic differences. We conclude by arguing that Subject Island effects are epiphenomenal, and that linguistic theory need not be made more complex with constraints that block such filler-gap dependencies.

# 2 Two broad views of Subject Island phenomena

In this section we discuss two major opposing views of what Subject Island effects are and what they mean for linguistic theory, and discuss what we consider to be their strengths and weaknesses.

## 2.1 The architectural view

Mainstream generative grammar assumes that islands are due to the very architecture of the human language faculty. Thus, extraction from subjects has been argued to be impossible because of basic constraints on movement (Chomsky, 1995; Takahashi, 1994; Nunes and Uriagereka, 2000; Sabel, 2002; Boeckx, 2003; Nunes, 2004; Rackowski and Richards, 2005; Rizzi, 2007; Stepanov, 2007; Chomsky, 2008; Müller, 2011; Haegeman et al., 2014; Chesi and Bianchi, 2014). Although no consensus yet exists about the exact nature of Subject Island effects, the architectural view has been claimed to be on the right track because of supporting psycholinguistic evidence: Phillips (2006), Sprouse (2007), Sprouse (2009), Wagers and Phillips (2009), Phillips et al. (2011), Phillips (2013), Sprouse et al. (2015) and others argue that Subject Island violations cannot be constructed by the language processor during language processing, or primed or ameliorated, as there are no mental representations to prime or ameliorate in the first place. As Sprouse (2007, 123) puts it: 'extra-grammatical factors that affect the acceptability – and are predicated on the existence of a representation – such as syntactic priming, should not affect the acceptability of ungrammatical sentences'.

The architectural view faces two kinds of potential challenges. First, Do and Kaiser (2017) shows that some Island violations can in fact be primed when the prime and the target were separated by only one sentence.[3] Second, whereas Sprouse (2009), and Crawford (2011) found no evidence that the acceptability of Subject Island violations increases by making such filler-gap dependencies more frequent, others have found clear amelioration effects (Hiramatsu, 2000; Clausen, 2011; Chaves and Dery, 2014).[4] The fact that different experiments have arrived at different results suggests that the amelioration of Subject Island violations is highly sensitive to the particular experimental items and methodologies (Snyder, 2017). On the other hand, no experiment has until now shown that Subject Island violations can become as acceptable as uncontroversially grammatical controls.

## 2.2 The functional view

Hofmeister et al. (2013) and Chaves (2013) more recently propose an alternative view in which Subject Island effects (and certain other phenomena) are instead due to probabilistic knowledge about the distribution of gaps: if the correct location of a gap is syntactically, semantically, or pragmatically highly unlikely in that particular utterance, then it is less likely for the sentence to be acceptable. For instance, if an object gap is strongly predicted in a given sentence context (given

[3]There is also evidence that comprehenders can adapt to some ungrammatical input. For example, Kaschak (2006) shows that speakers very rapidly adapt to previously ungrammatical structure such as the *needs VERBed* construction, used only in Northern Midlands dialects of American English, and similarly, Ivanova et al. (2011) provides evidence suggesting that sentences with illicit verbs of various kinds can can be primed.

[4]This amelioration phenomenon is referred to as 'satiation', drawing a supposed analogy to semantic satiation (see http://www.csi.uottawa.ca/tanka/files/judg_fatigue). But whereas semantic satiation is basically a reactive inhibition phenomenon, the increase in acceptability during sentence processing seems to be a facilitatory phenomenon. In fact, at least for some islands, the amelioration persists at least up to four weeks Snyder (2017). We therefore avoid the term 'satiation' to describe increases in acceptability, or decreases in reading/reaction time.

local cues and past experience with such constructions), a gap within a preceding complex subject would have low probability, making signs of a dependency inside the complex subject unlikely. This scenario leaves open the possibility that gaps within complex subjects can be predicted and posited, given sufficient cues that change the parser's expectations (Hofmeister et al., 2013, 49).

Indeed, there is good experimental evidence that speakers attend to probabilistic information about the syntactic distribution of filler-gap dependencies (van Schijndel et al., 2014), and that gap predictability is crucial for on-line processing of islands (Michel, 2014). More broadly, there is ample evidence that speakers attend and adapt to probabilistic information when processing a variety of linguistic input, including upcoming words (Altmann and Kamide, 1999; Arai and Keller, 2013; Creel et al., 2008; DeLong et al., 2005; Kutas and Hillyard, 1984), lexical categories (Gibson, 2006; Levy and Keller, 2013; Tabor et al., 1997), syntactic structures (Fine et al., 2013, 2010; Farmer et al., 2014; Fine and Jaeger, 2013; Levy et al., 2012; Lau et al., 2006; Levy, 2008; Staub and Clifton, 2006), semantics (Altmann and Kamide, 1999; Federmeier and Kutas, 1999; Kamide et al., 2003), and pragmatics (Ni et al., 1996; Mak et al., 2008; Roland et al., 2012).[5]

In order to understand what can make a gap more plausible than another, we must briefly turn our attention to the function of filler-gap dependencies in general. Erteschik-Shir (1981), Van Valin (1986), Kuno (1987), Takami (1992), Deane (1992), and Goldberg (2006, ch. 7) have argued that extraction is in general restricted to the informational focus of the proposition, i.e. to referents that are new or important in the sense that the speaker assumes the hearer cannot predict that they would be key participants in the given state of affairs. Perhaps the clearest example of the role of information focus in islands is the Coordinate Structure Constraint (CSC) and its (non-)ATB exceptions, illustrated in (7).

(7) a.*This is the whiskey that Sam bought the beer and Robin spilled _.

b. This is the whiskey that Sam bought _ and Robin spilled _.

c. This is the whiskey that Sam went to the store and bought _.

As Kehler (2002, ch.5) demonstrates, the CSC and its ATB exceptions follow from whether there is a parallel discourse coherence relationship between the conjuncts (7a,b), and non-ATB exceptions follow from whether the information focus is established only with reference to one of the conjuncts, as in (7c). For further empirical evidence that the CSC is pragmatic see Kubota and Lee (2015). The Complex NP Constraint (CNPC) is another type of island that is arguably constrained in very much the same way, since the only known acceptable exceptions to the CNPC

[5]The role of pragmatic expectations in language processing is demonstrated by Mak et al. (2008) and Roland et al. (2012), which show that the well-known processing advantage of subject relatives over object relatives is partly due to the differing degrees to which subject and object relative clauses violate more discourse expectations when encountered in isolated contexts, typical of experimental stimuli. Corpora evidence shows that the embedded noun phrase in object relative clauses nearly always has a discourse-old referent, even if the noun phrase is not pronominal, while the referent of the embedded noun phrase in subject relative clauses is typically discourse-new. Roland et al. (2012) found that noun phrase object relative clauses are not more difficult to process than subject relatives if preceded by appropriate discourse contexts.

involve relative clauses that express new information rather than backgrounded information; see Erteschik-Shir and Lappin (1979), Kuno (1987), Deane (1992), and Goldberg (2013). This is illustrated by the relative clauses in (8), which express assertions and therefore easily contain the information focus.

(8) a. There were several old rock songs that [she and I were [the only [two [who knew _ ]]]]. (Chung and McCloskey, 1983)

b. Which diamond ring did you say there was [nobody in the world [who could buy _ ]]? (Pollard and Sag, 1994, 206)

Now, given that subject phrases are typically reserved for topic continuity rather than for introducing new referents (Chafe, 1994; Kuno, 1972; Lambrecht, 1994; Bayer and Salzmann, 2013), it follows that subjects are more likely to be pronominal or elliptical than objects, and that extraction from English subject phrases is difficult because the information focus is unlikely to reside in the subject.[6] Crucially, whether a subject-internal referent can be the informational focus or not, is a matter of degree, and more likely in some propositions than in others. For example, classic Subject Island violations like **Which rebel leader did the rivals of shoot Castro?* are arguably unacceptable because out-of-the-blue it is unclear why the rebel leader is worthy of mention in the proposition *The rivals of a rebel leader shot Castro*. Without a suitable context, the rebel leader has no obvious bearing on the assertion that people shot Castro. As Kluender (2004, 495) insightfully noted: 'Subject Island effects seem to be weaker when the *wh*-phrase maintains a pragmatic association not only with the gap, but also with the main clause predicate, such that the filler-gap dependency into the subject position is construed as of some relevance to the main assertion of the sentence'.[7] Indeed, recent functional magnetic resonance imaging evidence from Matchin et al. (2018) suggests that Subject island effects activate networks involved in conceptual-semantic processing, as expected in the present account.

In our view, the items employed by previous Subject Island experiments tend to fail this pragmatic constraint. Consider for example *What does John know that a bottle of fell on the floor?* from Snyder (2000). Out of the blue, there is no clear *a priori* reason for why the content of a bottle is important for the fact that John knows that a bottle fell on the floor. This kind of problem can be avoided using suitable contextualizations or experimental items normed to be as plausible as non-island controls, i.e. items that have highly felicitous declarative counterparts. No prior research on islands has normed items in this way, to our knowledge.

Although it is not easy to construct sentences where a dependent of the subject can be easily deemed as the informational focus, it is by no means impossible. The key to achieving this is to make sure that in the declarative form of the utterance all the referents in the sentence (especially those embedded in the subject phrase) are highly relevant for the overall proposition. For instance,

[6] Michaelis and Francis (2007) found a bias in the distribution of lexical versus pronominal subjects in the Switchboard corpus (Marcus et al., 1993): of approximately 31,000 subjects of declarative sentences, only 9% are lexical NPs, while 91% are pronouns. In contrast, of approximately 7500 objects of transitive verbs, 34% are pronominal.

[7] See Shimojo (2002) for a very similar observation concerning Japanese Subject Island effects.

the extraction in (9a) is particularly acceptable because whether or not an impeachment causes outrage crucially depends on who is impeached (cf. with *Would the impeachment of Donald Trump cause outrage?*). Similarly, in (9b) whether or not an attempt failed or succeeded crucially depends on what was attempted (cf. with *The attempt to find the culprit ended in failure*).

(9) a. Which President would [the impeachment of _ ] cause outrage?
(Chaves, 2012)

b. What did [the attempt to find _ ] end in failure?
(Hofmeister and Sag, 2010, 370)

We know of no empirical reason for assuming that such extractions are syntactically different from standard Subject Island violations. See Chomsky (2008, 160,ft.39), Jiménez–Fernández (2009, 111), Chaves (2013), and Haegeman et al. (2014) for more examples of acceptable English extractions from a variety of subject phrases, including from transitive constructions.

The functional account also explains why a subject-embedded gap usually becomes more acceptable in the presence of a second non-island gap: since the two gaps are co-indexed, then the fronted referent is trivially relevant for the main assertion, as it is predicated by the main verb. For example, the low acceptability of (10a) is arguably caused by the lack of plausibility of the described proposition: without further contextual information, it is unclear how the attempt to repair an unspecified thing *x* is connected to the attempt causing damage to a car.

(10) a.*What did [the attempt to repair _ ] ultimately damage the car?

b. What did [the attempt to repair _ ] ultimately damage _?
(Phillips, 2006)

Our informants report that (10a) becomes more acceptable if it is contextually established that *x* is a component of the car. In contrast, (10b) is felicitous even out-of-the-blue because it conveys a proposition that is readily recognized as being plausible according to world knowledge: attempting to fix *x* can cause damage to *x*. If Subject Island effects are indeed contingent on how relevant the extracted subject-embedded referent is for the assertion expressed by the proposition, then a wide range of acceptable patterns is to be expected, parasitic or otherwise. This includes cases like (11), where both gaps are in Subject Island environments. As Levine and Sag (2003), Levine and Hukari (2006, 256), and Culicover (2013, 161) note, cases like (11) should be completely unacceptable, contrary to fact. The conclusion that Subject Island effects are contingent on the particular proposition expressed by the utterance thus seems unavoidable.

(11) This is a man who [friends of _ ] think that [enemies of _ ] are everywhere.

Further corroborating evidence for the role of pragmatics in extractions from subject phrases comes from the fact that all Subject Island violations attested in corpora that we are aware of involve relative clauses that describe new information rather than background information, underlined in (12). For example, (12a) asserts that 'five of their eight children are still living at home',

rather than merely asserting that 'they have eight children'. For more about relative clauses that express assertions rather than background information see Menn (1974), McCawley (1981), Lambrecht (1988), and Lambrecht (2000).

(12) a. They have eight children [of whom] [[five _ ] are still living at home].
(Huddleston et al., 2002, 1093)

b. (...) a letter [of which] [[every line _ ] was an insult].
(Santorini, 2007)

c. (...) ran a documentary featuring a young Auckland family [of which] [[the father _ ] earned $70,000 a year (...)].
(NOW Corpus)

d. It is believed the suspects left the scene with three bags containing new cellphones [of which] [[the value _ ] was unknown].
(NOW Corpus)

e. '(...) nearby Nemaha, a town that [[to describe _ as tiny] would be to overstate its size].
(Huddleston et al., 2002, 1093,1094)

f. In his bedroom, [which] [to describe _ as small] would be a gross understatement], he has an audio studio setup.
(Chaves, 2013)

In addition to being attested in corpora, Abeillé et al. (2018) provide experimental evidence that extractions like (12) are in fact acceptable.[8] Finally, note that Kluender and Kutas (1993) and Van Valin (1995) point out that PP extractions in general have an additional advantage over NP extractions in that the grammatical function of the extracted phrase is clearer from the onset: the presence of the preposition leads to fewer potential gap sites that are consistent with the extracted constituent, especially when the filler-gap dependency is short as in (12).

All of the above evidence is consistent with the proposal in Hofmeister et al. (2013) and Chaves (2013) that the more syntactically, semantically, and pragmatically plausible the subject-internal gap, the more acceptable such extractions should be, especially if comprehenders have the chance to adjust to the unusual syntactic location of the gap by being exposed to multiple exemplars of

[8] Also attested in relatives that do not express background information are *in situ* subject-embedded *wh*-phrases, as illustrated by the suppletive relative in (i) and (ii) from COCA and Google Books. Such constructions are formal in style, but not very rare. Abeillé et al. (2018) also found such constructions to be acceptable.

i. Maletskos was one of several scientists involved in the Fernald experiment, [[the results of which] proved that certain cereals can block the absorption of calcium (...)]

ii. Upon the corner lot he had built a block of buildings, [[some portions of which] extended over upon the second lot].

such sentences. We argue that such amelioration should be facilitated when the very mention of the subject-embedded referent is a highly felicitous discourse move to begin with. For the mention of any given referent to be felicitous and in accordance to the Gricean Maxims of *Quantity* ('Be brief') and *Manner* ('Avoid Prolixity'), the context must be such that the referent is not irrelevant for the proposition. In addition, for a referent to be extractable it must not only be relevant for the proposition, it must in addition play a pivotal role in it, as the informational focus. Hence, the more pragmatically felicitous the mention of a given referent is (subject-embedded or not) the more easily it can be deemed the informational focus, and the more felicitous the extraction. Because subjects tend to be topics, it follows that subject-embedded referents must be rarer than object-embedded referents. Using Tregex (Levy and Andrew, 2006), we have confirmed that NPs consisting of a nominal head followed by a PP occur about twice as often immediately after the verb (as objects) than immediately before the verb (as subjects) in the Wall Street Journal corpus (16159 occurrences vs. 8426), Brown corpus (4667 occurrences vs. 2222), and Switchboard corpus (1595 occurrences vs. 890). Of course, other languages may deploy different pragmatic biases, and regard subject-embedded referents as more likely informational foci than English does, potentially explaining at least some of the cross-linguistic variation in Subject Island effects surveyed in Stepanov (2007).[9]

The experiments we describe below collectively suggest that ensuring that a proposition with subject-embedded referents is highly felicitous is sufficient to allow the extractability of such subject-embedded referents when their frequency is increased. More specifically, Experiment 1 shows that sentences with subject-embedded gaps and sentences with object-embedded gaps that express essentially the same (highly felicitous) propositions rapidly become equally acceptable, within as much as 8 exposures. Experiment 2 offers supporting online evidence for the very same effect of adaptation to sentences with subject-embedded gaps in a self-paced reading task. In Experiments 3 and 4 we show that the similar effects arise in more extreme Subject Islands that are embedded in tensed relatives, which are allegedly impossible to ameliorate parasitically or otherwise. Our results are consistent with the functional view, but not with the architectural view. As a consequence, we argue that pragmatics and probabilistic knowledge must be taken into consideration in experimental research about filler-gap dependencies, if a explanatory and comprehensive account of island effects is to be reached.

# 3 Experiment 1: Extraction from simple subjects

The paradigm in (4), repeated below in (13), is traditionally assumed since Engdahl (1983) to show that subject-internal gaps are licit only if they are parasitic on a non-island gap. We reject this conclusion for two reasons.

(13) a.*Which rebel leader did [the rivals of _ ] shoot Castro?

b. Which rebel leader did [the rivals of _ ] shoot _ ?

[9]For experimental evidence about cross-linguistic variation in Subject Island effects see Jurka et al. (2011), Goodall (2011), Sprouse et al. (2015) and Fukuda et al. (2018).

First, many authors such as Kuno and Takami (1993), Pollard and Sag (1994, 195,ft.32), Culicover (1999, 230), Levine et al. (2001, 204), Levine and Sag (2003, 252, ft.6) and others have noted that some Subject Island violations are rather acceptable, like those in (9) above or (14) below.

(14) a. Which problem would [the resolution of _ ] surprise you the most?

b. Which disease did [the vaccine for _ ] suddenly stop working?
(Chaves and Dery, 2014)

Second, and more importantly for the present work, (13a) and (13b) express very different propositions and have different pragmatic requirements, and therefore there is no *a priori* reason to expect that (13a) and (13b) should receive identical acceptability judgements. The contexts in which (13a) is felicitous are stricter than those of (13b) because the pragmatic status of subjects is different from that of objects, as already discussed in §2.2. Thus, it is unclear what conclusions can legitimately be drawn from comparing sentences like (13a,b). Rather, we argue that it is far more appropriate to compare the acceptability of near-truth-conditionally equivalent sentence pairs such as (15), where the order of the subject and the object can be switched without a radical difference in truth-conditions because the verb is a symmetric predicate.

(15) a. Which country does the King of Spain resemble [the President of _ ]?

b. Which country does [the President of _ ] resemble the King of Spain?

In Experiment 1 below we show that although the acceptability (15b) starts out as much lower than (15a), after about 8 exposures the former has become as acceptable as the latter. Crucially, no such radical and sharp amelioration occurs with uncontroversially ungrammatical controls.

## 3.1 Method

### 3.1.1 Participants

We analyzed data provided by 74 participants with IP addresses originating from the United States that were recruited through Amazon.com's Mechanical Turk (AMT) crowdsourcing marketplace. For evidence that sentence acceptability data obtained via AMT parallel data obtained in the laboratory see Munro et al. (2010), Gibson et al. (2011), Melnick et al. (2011), and Sprouse (2011). All self-reported as native speakers of English, and had accuracy levels of at least 85% in comprehension questions, with a mean accuracy level of 90%.

### 3.1.2 Design and materials

We constructed 22 experimental items, each of which had two versions, as shown in (16). Every item involved a symmetric predicate. The location of the gap is made explicit below, though no

such indication was made in the actual stimuli. See Appendix A for a complete list.[10]

(16) a. Which committee does the report of _ supposedly contradict the recommendations of the experts? (*subject condition*)

b. Which committee does the report of the experts supposedly contradict the recommendations of _? (*object condition*)

To ensure that the subject and object conditions of our 22 experimental items expressed equally plausible propositions, we conducted a norming task over the declarative counterparts of the items. For example, the sentence pair in (16) was converted into *The report of this committee supposedly contradicts the recommendations of the experts* and *The report of the experts supposedly contradicts the recommendations of this committee*, respectively, and participants were asked to rate the plausibility of the sentences. This plausibility norming stage ensures that the propositions described by the items are equally highly felicitous to begin with. Hence, any difference in acceptability must come from extraction itself, not from other sources independent from extraction. A group of 100 participants recruited through AMT were asked to rate the plausibility of these declarative counterparts of our items using a 1 – 7 Likert scale. An overall t-test revealed that the plausibility of subject and object conditions did not differ, $t = -0.9, p = 0.366$. We also conducted pairwise t-tests on the responses for the subject and object conditions of each experimental item pair. These tests revealed that only two pairs exhibited a statistically significant difference in plausibility, but the remaining 20 pairs did not (all $p$s $> 0.05$). We nevertheless used all 22 pairs in the experiment, and included their plausibility ratings as an additional predictor in our analysis. In doing so, any difference attributable to the location of the gap will not be confounded by differences in plausibility which vary individually across experimental items.

Items were counterbalanced across two lists using a Latin Square design so that each participant only responded to one version of each experimental item. Experimental items were interspersed among 44 distractor items, pseudorandomized so that different participants saw items in different orders. A sample of the distractors is seen in (17) and (18), which were either object gaps or subject gaps, some of which were clause-embedded. Half of the distractors were ungrammatical, as in (18).

(17) a. Which cabinet does the stack of papers belong to according to the secretary?

b. Which type of music does the flyer say that the band of traveling musicians play?

c. Which jury members does the judge reportedly consider to be problematic for the trial?

[10]We used *which* phrases because such Subject Island violations are slightly more acceptable than those with *who*/*what* (Clausen, 2011; Chaves and Dery, 2014), perhaps because they aid the resolution of the filler-gap dependency: the latter *wh*-phrases are less informative (i.e. more indefinite) and therefore plausible with more candidate gap sites, as well as less likely to resist memory decay while processing the rest of the sentence, and thus are harder to retrieve downstream (Kluender and Kutas, 1993; Van Valin, 1995; Sag et al., 2007; Hofmeister et al., 2013).

(18) a.*Which complaint does the tenant of the condo rarely hear at each of the caretakers?

b.*Which artifact does the Museum of Fine Arts wish purchases about the British Museum?

c.*Which boat does the report unexpectedly reveal that the soldiers were thinking?

The ungrammatical distractors were relatively homogeneous, and therefore can be taken to function as additional controls: they all consisted of *which* interrogatives with an adverb and a clause-embedded gap. The source of oddness always involved an incorrect word in the embedded VP.

### 3.1.3 Procedure

Participants were asked to judge how natural each sentence was, by giving it a rating from 1 (very unnatural) to 7 (very natural). In order to ensure that comprehenders were attending to the structure and meaning of the experimental items, half of the grammatical distractors were immediately followed by a (nontrivial) True/False question; e.g. (17c) was followed by (19).

(19) The judge presiding over the trial was claimed to have no concerns about the jury. [True/False]

## 3.2 Results

The mean response for the Subject condition was 3.99 ($SD = 1.69$), and for the Object condition 5.08 ($SD = 1.51$), confirming the existence of an island effect. The mean response for the grammatical distractors was 5.7 ($SD = 1.46$), and 2.67 ($SD = 1.59$) for the ungrammatical distractors.[11]

An LMER model with gap location, presentation order, plausibility, and all possible interactions between the three factors as fixed factors revealed a main effect of gap location in which subject gaps were rated lower than object gaps ($\beta = -1.065$, $t = -7.69, p < 0.0001$), and a significant interaction between gap location and presentation order, suggesting that the the acceptability of subject and object items changed differently during the experiment ($\beta = 0.061$, $t = 5.23, p < 0.0001$). No significant effect of plausibility was found, nor of interactions. Since no two participants saw the items in the same order, the changes in acceptability cannot be due to any particular item order.

Simple effects analyses were conducted separately on each gap location condition by conducting LMER models with presentation order as fixed predictors. Presentation order was not significant for the object gap condition ($\beta = 0.004$, $t = 0.58, p = 0.56$), and highly significant for the subject gap condition ($\beta = 0.065$, $t = 7.84, p < 0.0001$), even in models that include plausibility

[11] Linear mixed-effect regression (LMER) models for this experiment and subsequent ones were implemented using the `lme4` package version 1.1–12 (Bates et al., 2014) in `R` version 3.3. Unless otherwise noted, the intercept was allowed to be adjusted by items, subjects, and lists in all models reported below, in order to account for random effects. All p-values were calculated by Satterthwaite approximation, using the `lmerTest` package version 2.0-30 (Kuznetsova et al., 2015). Throughout, we follow the American Psychological Association guidelines (Section 2.07 in the 6th edition) in using raw values, as they are easier to interpret relative to the original scale. Regardless, we have verified that using z scores instead does not change the overall results.

as a (non-significant) predictor. Separate models revealed that while the acceptability of the ungrammatical distractors did not increase during the experiment ($\beta = -0.003$, $t = 1.41$, $p = 0.24$), that of the grammatical distractors did ($\beta = 0.005$, $t = 1.98$, $p = 0.04$). Figure 1 shows regression lines for all conditions and distractors, with 95% confidence intervals. Each point corresponds to the average acceptability rating for items in the given order of representation. Since no two participants saw the experimental items in the same order, each point corresponds to the average rating that different participants gave to different items in the given order as the experiment progressed.

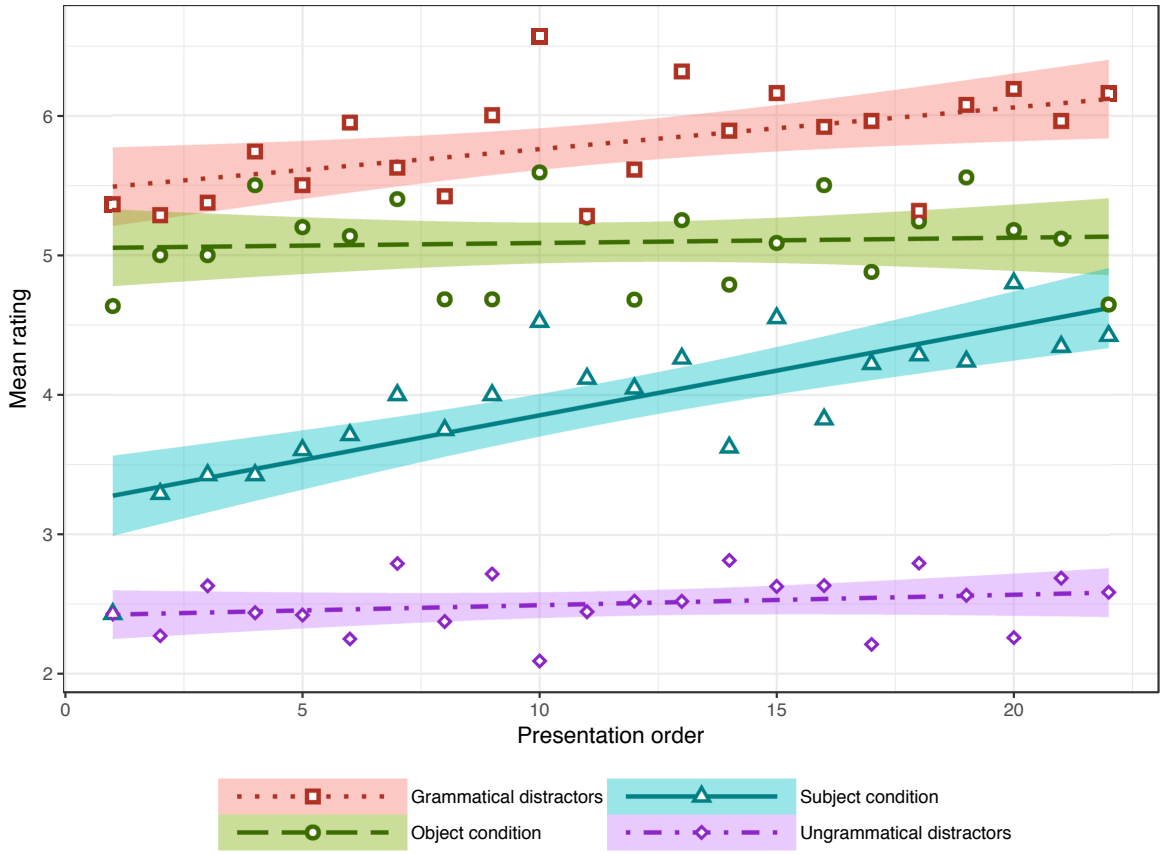


Figure 1: LMER of each item type with presentation order as a fixed predictor (Experiment 1)

Recall that in this experiment each participant read 11 subject island violations, and 11 object counterparts. We can therefore use the interaction estimate between the subject gap condition and presentation order to predict when the acceptability difference between subject gap items and object difference will disappear: $\approx$ 17 presentations (i.e. -1.06/0.061 = subject gap coefficient / interaction coefficient). In other words, the model suggests that if the experiment had been 5 items longer then the acceptability of the Object and Subject items would completely overlap, assuming the acceptability increase is linear. In fact, a *post-hoc* analysis already suggests that the acceptability of the subject items became as high as the acceptability of the object items in a subset of the data consisting of the last four presentation orders (144 responses). There, the mean

acceptability was 4.48 ($SD = 1.62$) for subject condition items, and 5.03 ($SD = 1.55$) for object condition items. An LMER model was run on this subset of data, with gap location as a fixed predictor, and found no significant difference between the two conditions ($\beta = -0.55$, $t = -1.68$, $p = 0.1$). The plots in Figure 2 illustrate the overlap between Subject item ratings at the end of the experiment with Object item ratings at the end of the experiment.

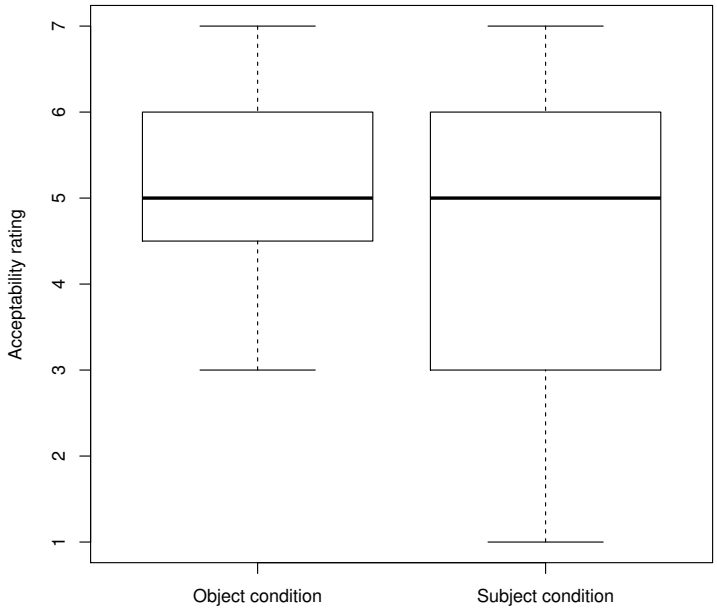


Figure 2: Ratings at the end of Experiment 1

Focusing on a smaller window consisting of the last two presentation orders, the mean acceptability was 4.41 ($SD = 1.57$) for subject condition items, and 4.92 ($SD = 1.71$) for object condition items (96 responses). An LMER model ran on this subset of data with gap location as a fixed predictor found it to not be significant ($\beta = -0.39$, $t = -1.26$, $p = 0.21$). By selecting only these late presentations we have reduced the sample size considerably, and therefore raised the odds of not finding a significant difference even though one may exist. This problem can be mitigated by using robust statistics, such as bootstrapping (Wilcox, 2005; Larson-Hall and Herrington, 2010). Accordingly, we sampled with replacement from the 96 total data points (48 per condition) to create 1056 data points (528 per condition). This process was repeated a total of 100 times, and for each upsampled dataset we ran an LMER model with gap location as a fixed predictor, allowing the random effects to have different slopes for the main factor. We again found no significant difference between subject and object conditions (mean $t = -0.95$, SD $= 0.01$).

Finally, we compared the acceptability of subject condition items seen towards the end of the experiment with the acceptability of object condition items at the beginning of the experiment. If Subject Island effects are as strong as standardly assumed, and due to the architecture of the

language faculty, then repeated exposure should not cause Subject Island violations to be as acceptable as the 'first-impression' acceptability ratings of grammatical controls. After all, linguists routinely collect 'first-impression' judgements, and deem those reliable. Thus, a different subset of data was formed containing 78 responses for subject items with presentation order of 20 or higher (mean response 4.41; $SD = 1.66$), and 78 responses for object items with presentation order of 3 or lower (mean response 4.75; $SD = 1.68$). An LMER model with gap location as a fixed predictor revealed that gap location was again not significant ($\beta = -0.29$, $t = -1.35$, $p = 0.17$).

Our findings indicate that the acceptability of subject items became indistinguishable from the acceptability of object items. A replication of this experiment without comprehension questions yielded the same overall results.

## 3.3 Discussion

Our findings suggest that the acceptability contrast between extracting from NP complements and extracting from NP subjects vanishes if the underlying propositions are equally plausible, and comprehenders are exposed to the two types of extraction in equal and sufficient amounts (at least 8 exposures). This result shows how extreme the amelioration caused by repeated exposure can be.

The fact that the acceptability of the grammatical distractors also increased is unremarkable, since there is no reason to assume that acceptability judgements of grammatical sentences should be constant. If such structures are sufficiently complex, stylistically marked, or unexpected in some way it is likely that comprehenders initially experience difficulty parsing such sentences but gradually become more proficient at it with repeated exposure. As Fine et al. (2013, 2010); Farmer et al. (2014); Fine and Jaeger (2013) show for garden-path sentences, comprehenders can quickly adapt to overcome unusual grammatical structures. Moreover, the fact that the grammatical controls were rated higher than the Object condition items is also unremarkable, as there is no reason for structurally different grammatical sentences expressing different propositions to have identical acceptability ratings; see for example Keller (2003), Luka and Barsalou (2005), and Lau et al. (2015) for independent evidence that sentence acceptability is correlated with sentence probability.

There are three possible interpretations for our findings. One is that Subject Island violations are ungrammatical but somehow can be deemed as acceptable as uncontroversially ungrammatical sentences, as 'grammatical illusions'.[12] We find such a view implausible given that the naive native speakers we have consulted over the years have no problem providing an appropriate context for Subject Island violations like those in (20), which they deem as completely acceptable. There is no reason to believe there is anything illicit about such sentences, nor that they are structurally special.

(20) a. Which president would [the impeachment of _ ] cause outrage?

[12]To be sure, Phillips et al. (2011) assert that island violations are not prone to such grammatical illusions: 'most evidence on the real-time status of island constraints indicates that the parser successfully avoids constructing illicit filler-gap dependencies.'

b. Which problem will [the solution to _ ] never be found?

c. Which airline is [the crew of _ ] currently on strike?
(Chaves and Dery, 2014)

Moreover, what would make comprehenders start hallucinating only after a number of exposures? It is unclear what the underlying mechanism would be. If data like (20) and our experimental findings are nothing but 'grammatical illusions' then there is no guarantee, for example, that parasitic gaps embedded in infinitival subjects which Phillips (2006, 808) found to be acceptable are in fact grammatical. More generally, in the absence of a clear criterion to distinguish true from illusionary grammaticality, it becomes impossible to objectively determine whether any (un)acceptable sentence is in fact (un)grammatical or not.

An alternative interpretation is one where Subject Islands start out as ungrammatical, but somehow become grammatical (e.g. perhaps due to the change of a microparameter for subject subextraction out of certain verbal structures). If so, our results would speak against the architectural approach because they are inconsistent with constraints on extraction being intrinsic properties of the language faculty. However, such an account flies in the face of the fact that some Subject Island violations like (20) are quite acceptable, even without repetition. Of course, any model can always be made more complex in order to account for additional facts. The question is whether such an extension is independently motivated or not. We are unaware of any configurational account that is consistent with the full range of empirical and behavioral facts.

The simplest interpretation, in our view, is one where Subject Island violations are simply not architecturally illicit. Their oddness, whenever any exists, is due to the subject-internal gap being unlikely, and contrary to comprehenders' prior expectations about the distribution of filler-gap dependencies. In this view, it is unnecessary to complicate theories of grammar with syntactic constraints that prevent extraction from subjects, as there are no such constraints. As already discussed in §2.2, subject phrases are typically reserved for topic continuity rather than for introducing new referents, and therefore are much less likely to involve the informational focus. Hence, it is pragmatically unlikely for sentences to contain subject-embedded gaps, simply because extraction targets the informational focus, not the topic.

Another, perhaps related factor that may further preempt speakers from considering the possibility of subject-embedded gaps is that sentences with complex subjects are harder to process than sentences with complex complements, even in the absence of extraction (Kynette and Kemper, 1986; Clark and Wasow, 1998; Ferreira, 1991; Tsiamtsiouris and Cairns, 2009; Amy and Noziet, 1978; Eady and Fodor, 1981), and that more cognitive effort is required in order to process unexpected input (Boston et al., 2008; Demberg and Keller, 2008; Roark et al., 2009; Smith and Levy, 2008), especially at the beginning of the utterance (Petten and Kutas, 1990). Matters are made worse by the fact that processing long-distance dependencies is cognitively costly (Chen et al., 2005; Frazier, 1987; Pickering and Barry, 1991; Kluender, 1998; Wanner and Maratsos, 1978; Phillips et al., 2005; Sussman and Sedivy, 2003; Warren and Gibson, 2002, 2005). The net effect of all these pragmatic and cognitive pressures is that speakers may simply prefer to avoid positing subject-embedded gaps, causing them to be extremely rare, though not grammatically impossible.

But by increasing the frequency of such constructions comprehenders can adapt and revise their expectations about the plausibility of subject internal gaps, all else being equal. For a maximum entropy model consistent with the linear increase of acceptability in the presence of increased exposure see Chaves (2018).

# 4 Experiment 2: Gap-filling inside simple subjects

The previous off-line experiment suggests that Subject Island effects are influenced by the probability of the input. In the next experiment we provide on-line sentence processing evidence that Subject Island effects are sensitive to the probabilities of the input. Together, the two experiments suggest that the functional view is superior to the architectural view.

It is well known that comprehenders postulate gaps and attempt to fill them as soon as possible (Crain and Fodor, 1985; Frazier, 1987; Stowe, 1986; Stowe et al., 1991), but such an effect has not been systematically observed inside subject phrases. For example, Stowe (1986) and Experiment 2 of Pickering et al. (1994) probed sentences like (21), and found no slowdown in reading time at the regions in bold, suggesting that comprehenders do not postulate subject-internal gaps.

(21) a. The teacher asked what [the silly story about **Greg's** older brother] was supposed to mean.

b. I know what [a book about **the local election**] discussed the most.

Although these null effects would seem to support the view that subject-internal gaps are not postulated during sentence comprehension because they are not grammatical, such a view is inconsistent with the fact that subject-internal gaps *are* licit, at least parasitically, as in (4b) and (5b) above.

Crucially, the experiments in Stowe (1986) and Pickering et al. (1994) did not contain any items with subject-embedded gaps, but given that probabilistic knowledge influences active gap-filling (van Schijndel et al., 2014; Michel, 2014), then including such gaps in the experimental items might make their postulation more likely. Experiment 2 below confirms this.

## 4.1 Method

### 4.1.1 Participants

We analyzed data provided by 41 participants who were recruited in AMT, using recruitment protocols identical to Experiment 1. All had accuracy levels of at least 80% in comprehension questions, with a mean accuracy level of 87%. There were 15 additional participants whose accuracy scores were lower than the 80% threshold, as well as 20 additional participants who participated in similar experiments in the past. Data collected from these participants were discarded.

### 4.1.2 Design and materials

A between-subject block design was employed to investigate whether gap-filling inside subject phrases is contingent on the probability of such gaps. Participants were randomly assigned to one

of two groups, which we will call the Subject group and the Object group. In both groups, participants were exposed to two blocks of sentences, reading a total of 45 sentences in the first block, and 30 sentences in the second block. Participants in the Subject group read 15 Subject Island violation sentences and 30 distractor sentences in the first block, while participants in the Object group read 45 distractors in the first block. The second block was identical for both groups, and contained 10 Subject Island 'violation' sentences and 20 distractors. (22) illustrates the Subject items that participants saw. The symbol '|' indicates the regions that were shown during the experiment. See Appendix B for a complete list.

(22) a. Which animal $_1$| does $_2$| the song of $_3$| reportedly $_4$| mimic $_5$| the Gray Catbird's sounds? $_6$|

b. Which athlete $_1$| does $_2$| the manager of $_3$| clearly $_4$| resemble $_5$| Tiger Woods' agent? $_6$|

c. Which company $_1$| do $_2$| the employees of $_3$| allegedly $_4$| reject $_5$| salary increases? $_6$|

The adverb (region 4) is consistent with a gap-less parse because adverbs can modify an upcoming NP (e.g. in attested examples like *the infants of reportedly non-smoking mothers*, or in *the execution of allegedly innocent people*). It is the main verb region (region 5) that is incompatible with a gap-less parse.

A sample of distractors is in (23). There were three types in each of the three blocks: object extractions like (23a,b), where the gap is in a complement phrase instead of the subject, adverbial extractions like (23c), where the gap is an adverb that modifies the main verb, and matrix subject interrogatives like (23d), where the *wh*-phrase immediately precedes the verb that selects it. Some distractors contain the preposition *of* in the subject, others in the object, others none at all.

(23) a. Which condition | does | the treatment of | eczema | typically | mimic the effects of? |

b. Which magazine | does | the media | openly | regard | as the best | on the market? |

c. When | does | the State of | Illinois | plan | to accuse | someone | of the crime? |

d. Which gymnast | does | a floor routine | at the sound of | a rock song | during practice? |

### 4.1.3 Procedure

Participants read sentences on a self-paced moving window display (Just et al., 1982). Programming and presentation of the experimental stimuli was done using Ibex 0.3.9 (Drummond, 2013). See Futrell (2012, 25–35) for validation studies showing that Ibex self-paced reading experiments with AMT participants can replicate classic self-paced reading experiments run in the laboratory. See also Enochson and Culbertson (2015) for an AMT self-paced reading validation study of the classic filler-gap experiment by Wanner and Maratsos (1978).

Each trial began by presenting a sequence of dashes representing the non-space characters in the sentences. Pressing the space bar caused the dashes corresponding to the first region to

be replaced by words. Subsequent presses revealed subsequent regions, while the previous region reverted to dashes. Reading times between each pair of button presses were recorded. The comprehension question that was presented after half of the distractor items had only two possible answers, *yes* or *no*, which appeared in random order. The responses to the comprehension questions were recorded, and participants were informed of any incorrect answers. Participants completed four practice trials at the beginning of the experiment to familiarize themselves with the task. The experiment immediately followed the practice trials. Presentation of sentences was quasi-randomized within blocks, the structure of which was entirely implicit (i.e. from the participants' perspective, 75 sentences were shown one at a time, without breaks or any explicit indication of the block structure of the experiment).

## 4.2 Results

For each region of interest, we excluded reading times that were less than 100 ms and more than 1,500 ms long, and replaced data points that were greater or less than 2.5 standard deviations from each participant's mean with these boundary values. We then computed residual reading times by subtracting from the actual reading time for a region the reading time predicted by a regression equation relating region length to reading time (Trueswell and Tanenhaus, 1994). This regression equation was computed separately for each participant, using all regions in the experimental and distractor items. Residual reading times for all sentence regions were analyzed using an LMER model to test whether reading times of sentences with Subject Island violations in Block 2 was affected by the sentences in Block 1. Figure 3 shows mean residual reading times, where bars represent 95% confidence intervals.

Region 5 exhibited a significant effect ($\beta = 86.80$, $t = 2.07$, $p = 0.04$), suggesting that participants in the Subject group (i.e. who saw 15 Subject Island violation sentences in Block 1) read the region of interest of the Subject Island violations in Block 2 faster than participants in the Object group (i.e. who saw zero Subject Island violation sentences in Block 1). No region before region 5 exhibited a significant effect.

## 4.3 Discussion

Since the only difference between the two groups of participants consisted of whether there were Subject Island violations in the Block 1 of items, the observed difference in reading times can only be attributed to that difference. The reading time differences appear after the critical region (the gap site), suggesting that comprehenders that were not trained to process Subject Island violations had more difficulty processing such sentence regions. The fact that the participants in the Subject group processed the critical region faster than those in the Object group is consistent with the view that the former comprehenders adjusted their expectations about subject-embedded gaps during their exposure to Block 1, and therefore were less surprised by such gaps in Block 2 than the participants in the Object group. In sum, our results are consistent with the findings of Experiment 1, in that Subject Island effects are sensitive to the probability of the input. Together, the two experiments

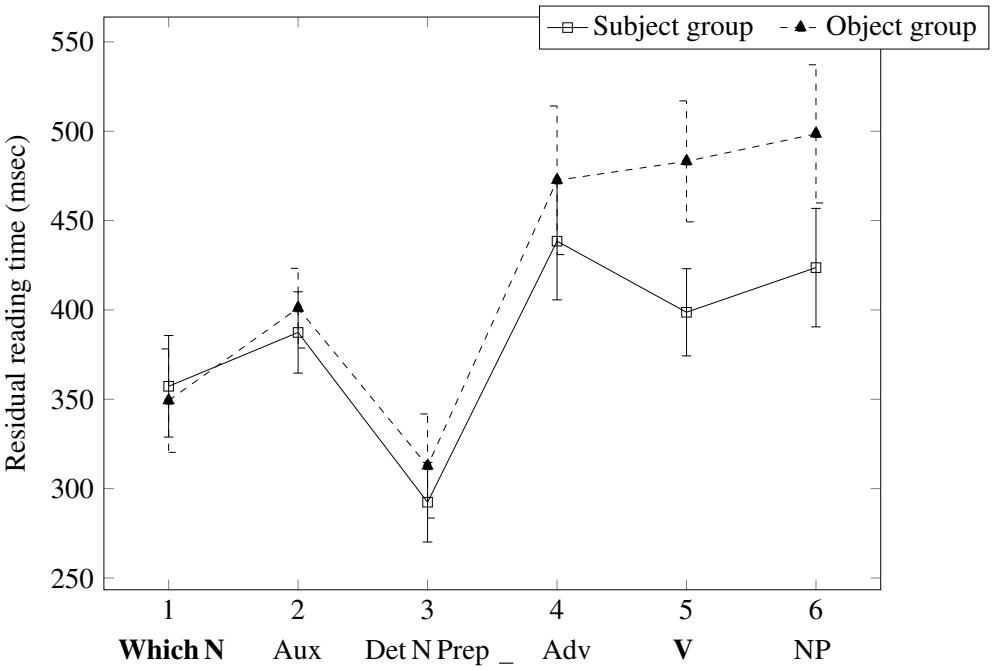


Figure 3: Mean residual reading times for all sentence regions in Block 2 (Experiment 2)

suggest that extraction from subjects is in principle construable by the parser, and their processing and acceptability is sensitive to their (syntactic, semantic and pragmatic) probability.

# 5 Experiment 3: Parasitic extraction from tensed subjects

Perhaps it can be argued that subject-internal gaps like the ones examined so far are postulated during on-line sentence processing because they can *in principle* be made licit if parasitic on a second gap. After all, such gaps are not necessarily illicit because they can still potentially be 'rescued' later in the sentence, if there is another gap downstream. In what follows we consider tense-embedded Subject Island violations because they are widely claimed to disallow any kind of amelioration, parasitic or otherwise. For example, although a subject-embedded gap in an infinitival phrase like (24a) can be made licit if parasitic on a second gap as in (24b), subject-embedded gaps in tensed phrases like (25a) remain illicit even in the presence of a second gap, as in (25b). Such contrast have been experimentally validated by Kurtzman and Crawford (1991), and Phillips (2006).

(24) a.*What did [the attempt to repair _ ] ultimately damage the car?

b. What did [the attempt to repair _ ] ultimately damage _?

c. What did [the attempt to repair the car] ultimately damage _?

(25) a.*What did [the mechanic that repaired _ ] ultimately damage the car?

b.*What did [the mechanic that repaired _ ] ultimately damage _?

c. What did [the mechanic that repaired the car] ultimately damage _?

Interestingly, Phillips (2006) found that the acceptability of (25a) is significantly lower than that of (25b), even though the parasitic gap is (allegedly) illicit. This presents a puzzle for the architectural view because if the subject-embedded gap is ungrammatical and beyond parasitic rescue, then (25a,b) should be equally unacceptable. But in the functional view expounded in §2.2 extractions like (25a) are expected to be worse than (25b) because in the former case the extracted referent is predicated by the main verb, a canonical informational focus domain. Furthermore, both (25a,b) are expected to be odd if in these particular propositions the odds of the extracted referent *what* being the patient of *repaired* happen to be particularly low. This would predict that ideal conditions the presence of tense should not hamper extraction so severely, as in counterexamples like (26).[13]

(26) a. Here is [the boy who everyone who has met _ ] thinks _ is clever.
(attributed to Janet Fodor by Engdahl (1983))

b. She is [the kind of person that everyone who meets _ ] ends up falling in love with _.
(Kayne, 1983)

c. John is [someone who everyone who meets _ ] dislikes _.
(Culicover, 1999, 179)

d. Which woman do [men who meet _ ] usually ask out _?
(attributed to Elisabet Engdahl by Pollard and Sag (1994, 226))

e. This is [a bill that the senators who objected to _ ] would probably not benefit from _.
(Chaves and Dery, 2014)

To be sure, none of the sentences in (26) are easy to process, or regarded as unquestionably acceptable by all speakers, and so the grammatical status of such parasitic gaps remains somewhat unclear. Although tensed-embedded Subject Island gaps like (25a,b) should never be licit and should never be postulated during on-line sentence processing according to Phillips (2006), we show below that such subject-embedded gaps can become as acceptable as undisputedly grammatical controls (Experiment 3) and are actively postulated during on-line sentence processing (Experiment 4), employing the same experimental paradigms as Phillips (2006).

Experiment 3 probes items like (26d,e) and shows that although they initially receive very low acceptability ratings, they quickly improve to the point of being as acceptable as uncontroversially

[13]Phillips (2006, 803,ft.6) conjectures that the acceptability of (26a–c) somehow restricted to relative clauses with a quantificational head NP, but examples like (26d,e) cast doubt on this claim. There is also a general correlation between processing cost and definite referentiality (Almor, 2000; Ariel, 2001; Epstein, 2001; Warren and Gibson, 2002), and some evidence that the presence of finite tense (a form of definite referentiality) leads to additional processing cost (Fiengo and Higginbotham, 1981; Kluender, 1992, 1998; Hofmeister et al., 2013).

grammatical controls. This sharp amelioration effect should be impossible if such gaps are grammatically and computationally illicit, as explicitly assumed by Phillips (2006, 803) and others. In this study we compare the acceptability of three types of sentences, illustrated in (27). The *Parasitic gap* items contain one gap embedded in a tensed relative inside the subject, and another gap in the complement the main verb. The *Object gap* items contain exactly one gap, corresponding to the main verb's object. In order to provide a baseline for low-acceptability items, we included in the experiment *Non-parasitic gap* items, which contain a gap located inside the tensed relative in the subject phrase, but no second gap.

(27) a. Which incident did the bystanders who witnessed _ find _ terrifying? (*Parasitic gap*)

b. Which incident did the bystanders who fled into the subway find _ terrifying? (*Object gap*)

c. Which incident did the bystanders who witnessed _ end up going into therapy? (*Non-parasitic gap*)

## 5.1 Method

### 5.1.1 Participants

We analyzed data provided by 135 participants who were recruited via AMT, using recruitment protocols identical to Experiment 1. There were 22 additional participants who reported they were not native speakers of English. Data collected from these participants were discarded.

### 5.1.2 Design and materials

There were 24 experimental items, and three versions of each experimental item as illustrated in (27). The full list is in Appendix C. The 24 experimental item triples were counterbalanced across three lists using a Latin Square design so that each participant only responded to one version of each experimental item, and were interspersed among 36 distractor items. The distractors are illustrated in (28), half of which were ungrammatical, as shown in (29). Exactly as in previous experiments, all items were pseudorandomized and interspersed with distractors.

(28) a. Which computer did the technician regard as being difficult to find parts for?

b. Which window did the previous tenant request that the insurance company fix?

c. Which machine did the regional managers ask if the employees wanted to get rid of?

d. Which song did the DJ say that the people at the club were too drunk to enjoy?

(29) a.*Which color have the designers voted that the clients would consider able?

b.*Which joke did the talkshow guests say the host would find it easy to excuse to?

c.*Which baker did the tenants ask whether the parking meters annoyed for?

d.*Which city did the cab companies determine if the parking spaces should be used?

To ensure that the propositions conveyed by relative clauses in the *Parasitic gap* items are plausible to begin with, a norming task was conducted on the declarative counterparts of the subject-embedded relative clauses. For example, the relative in (27a) was converted to the declarative clause in (30).

(30) The bystanders who witnessed this incident found it terrifying.

A group of 23 native English speakers were recruited through AMT to evaluate the plausibility of the 24 declarative counterparts of the relative clauses in the *Parasitic gap* type condition together with 15 plausible distractors such as (31a,b), 15 moderately plausible distractors like (31c,d), and 20 implausible distractors as in (31e,f). Raters were asked to use a 1 – 7 Likert scale in order to 'rate the likelihood of the situations described by the sentences'.

(31) a. The lawyer wanted to convince this jury that his client was innocent. (*plausible*)

b. A biologist identified the chemical that was killing the bees. (*plausible*)

c. The bank manager started to prepare that dessert mid-afternoon (*moderate*)

d. A comedian discovered the value of copper before it became expensive. (*moderate*)

e. The embalmer who stood up asked to wash those power lines. (*implausible*)

f. An architect cooked an ant for the monkey who texted him. (*implausible*)

Mean responses were 5.6 ($SD = 1.4$) for the experimental items, 6.45 ($SD = 0.88$) for the plausible distractors, 2.51 ($SD = 1.66$) for the moderately plausible distractors, and 1.79 ($SD = 1.41$) for the implausible distractors. Pairwise t-tests comparing the means for the experimental items against the means of the plausible distractors, as well as against the means of the moderately plausible distractors were significant (against the plausible distractors: $t = -8.91, p < 0.0001$; against the moderately plausible distractors: $t = 22.81, p < 0.0001$), indicating that the plausibility of the relative clauses in the *Parasitic gap* type items lies in between the plausibility of our 'plausible' and 'moderately plausible' distractors.

Unlike in Experiment 1, there was no plausibility norming across conditions in the present experiment. Consequently, the items are structurally and pragmatically complex in different ways across conditions, and therefore there should be no expectation that they are equally pragmatically or equally (un)acceptable either.

### 5.1.3 Procedure

The procedure was identical to that of Experiment 1.

## 5.2 Results

The mean responses for each of the three stimulus types were 4.26 ($SD = 1.77$) for *Parasitic gap* items, 5.46 ($SD = 1.56$) for *Object gap* items, and 2.67 ($SD = 1.52$) for *Non-parasitic gap* items. Note that if one adopts the same criterion as Phillips (2006, 808), then the items in the *Parasitic gaps* condition should be automatically deemed grammatical, as their mean acceptability is above the middle of the scale. Finally, the mean response for the grammatical distractors was 5.21 ($SD = 1.66$), and for the ungrammatical distractors it was 2.99 ($SD = 1.64$).

To determine whether the mean acceptability responses were statistically different from each other across the three types of sentences, three pairwise LMER comparisons with construction type as a fixed predictor were conducted. All pairs were found to be significantly different (*Parasitic gap* vs. *Object gap*: $\beta = 1.225$, $t = 14.47$, $p < 0.0001$; *Parasitic gap* vs. *Non-parasitic gap*: $\beta = -1.60$, $t = -18.9$, $p < 0.0001$; *Object gap* vs. *Non-parasitic gap*: $\beta = -2.836$, $t = -35.14$, $p < 0.0001$). In order to determine if the acceptability of the experimental items increased during the experiment, LMER models with order of experimental presentation as a fixed predictor were used, analyzing each of the three construction types separately. *Object gap* items became increasingly more acceptable ($\beta = 0.036$, $t = 4.1$, $p < 0.0001$), and so did *Parasitic gap* items, only twice as fast ($\beta = 0.062$, $t = 5.9$, $p < 0.0001$), suggesting that the acceptability ratings of the two conditions converge, given enough presentations. Note that both kinds of items were relatively complex and so the acceptability increase likely reflects the effect of comprehenders adapting to them and becoming more proficient at processing such constructions. Finally, the acceptability ratings of *Non-parasitic gap* items did change ($\beta = 0.014$, $t = 1.64$, $p = 0.1$), perhaps because such sentences described less felicitous propositions because the extracted referent had no obvious bearing on the main predicate (see §2.2). Figure 4 shows the results, according to the experimental presentation order. As in Experiment 1, no two participants saw the same presentation order and each point corresponds to the mean rating according to presentation order.

Analogously to Experiment 1, a subset of the data was formed consisting of the last four presentation orders (83 responses). The mean acceptability was 5.72 ($SD = 1.43$) for *Object gap* condition items, and 4.86 ($SD = 1.65$) for *Parasitic gap* condition items. An LMER model was run on this subset of data with gap location as a fixed predictor, and found it to be significant ($\beta = 0.73$, $t = 2.64$, $p = 0.01$). However, focusing on a smaller window consisting of the last two presentation orders (35 responses), the mean acceptability of the *Parasitic gap* condition was 4.65 ($SD = 1.66$), and the mean acceptability for the *Object gap* condition was 5.6 ($SD = 1.29$). An LMER model with construction type as a fixed predictor found no significant difference ($\beta = 0.53$, $t = 1.52$, $p = 0.148$), as Figure 5 illustrates. Given the concerns raised by such a small sample size, we adopted the same bootstrapping approach employed in Experiment 1 and sampled with replacement from the 35 total data points (17 per condition) to create 1118 data points (559 per condition). This process was repeated 100 times, and for each upsampled dataset an LMER model

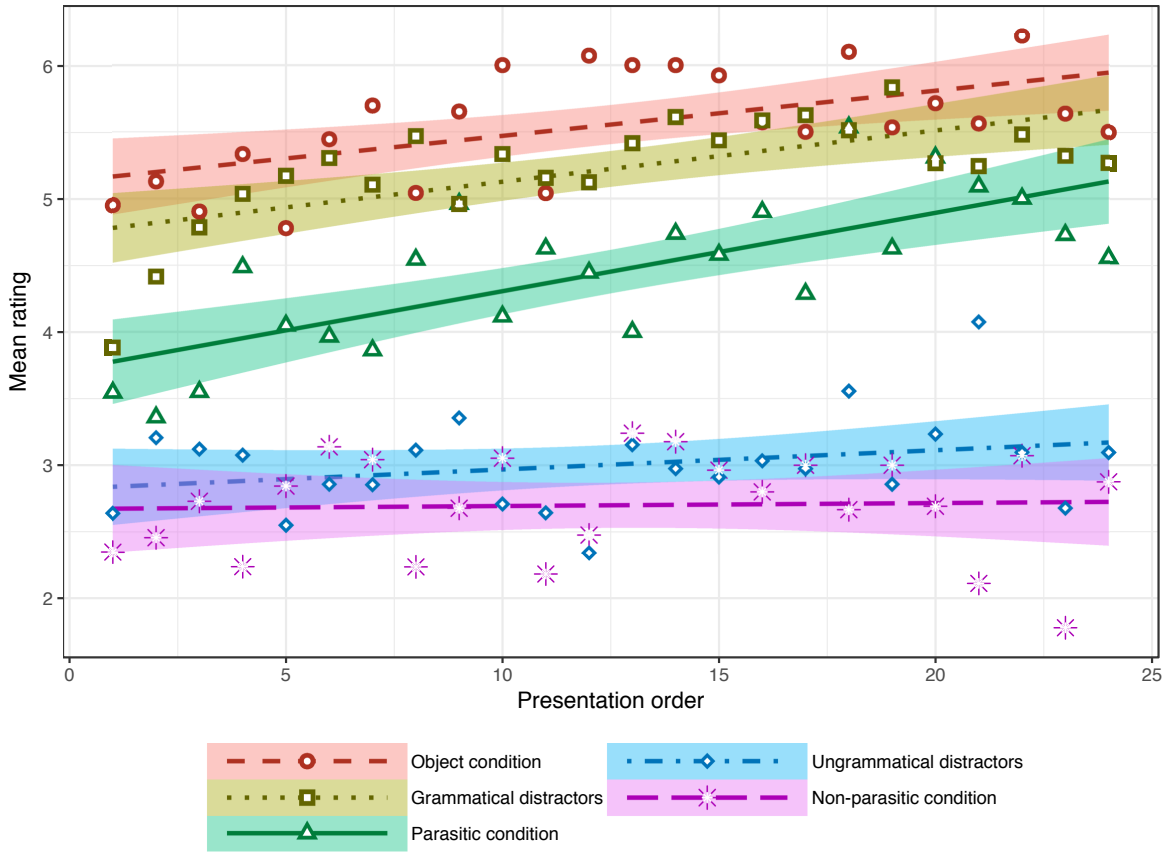


Figure 4: LMER of each item type with presentation order as a fixed predictor (Experiment 3)

was ran with gap location as a fixed predictor, allowing the random effects to have different slopes for the main factor. The results were again not significant (mean $t = 0.62$, SD $= 0.04$).

Finally, as in Experiment 1, we compared the acceptability ratings of *Parasitic gap* items that were seen by participants at the end of the experiment with an equal number of *Object gap* items seen in the beginning of the experiment, by forming a subset of our data consisting of 72 type *Parasitic gap* responses whose presentation orders were 18 or higher and of 72 type *Object gap* responses whose presentation order was either 1 or 2. The mean response for the *Parasitic* condition was 4.88 ($SD = 1.75$), and the mean response for the *Object gap* condition was 5.15 ($SD = 1.51$). Again, construction type was not significant ($\beta = 0.343$, $t = 1.35$, $p = 0.17$), suggesting that the acceptability of *Parasitic gap* items by the end of the experiment was not significantly different from the acceptability of *Object gap* items at the beginning of the experiment. The above results suggest that tense-embedded subject-internal parasitic gaps can be deemed as acceptable as uncontroversially grammatical sentences, in ideal conditions.

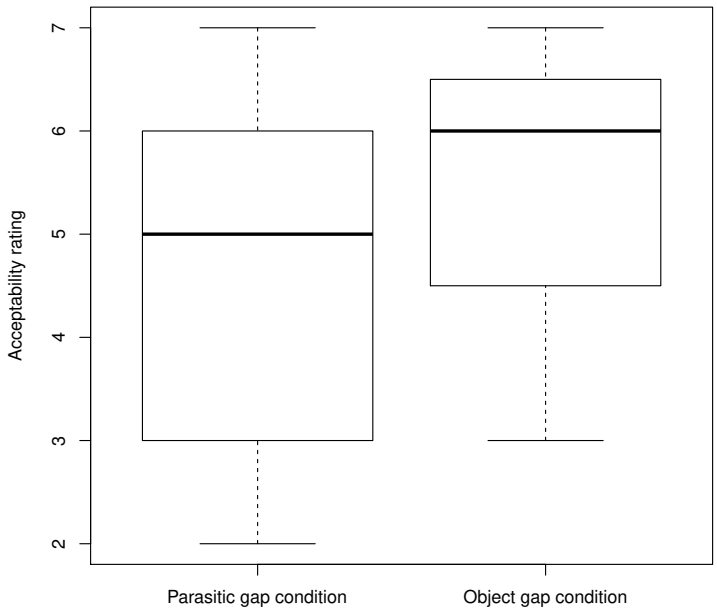


Figure 5: Subset of ratings at the end of Experiment 3

## 5.3 Discussion

The acceptability of *Parasitic gap* items increased twice as fast *Object gap* items during the experiment. We suspect these results were caused by comprehenders becoming more proficient at processing such complex sentences. Future research is needed to determine whether *Non-parasitic gap* items are truly immune to frequency effects, or simply harder to ameliorate, perhaps for the reasons outlined in §2.2. Nonetheless, our results suggest that sentences with tense-embedded parasitic subject gaps can be rated as acceptable as grammatical counterparts, given enough exposure. Our findings are therefore consistent with the view that speakers can overcome their prior expectations about where gaps are usually located if such sentences express highly plausible propositions to begin with. If the acceptability of Subject Island violations is sensitive to the probability of the filler-gap dependency in the particular sentence and context, as argued by Hofmeister et al. (2013), then it makes sense that different experiments and/or different Subject Island violation items can lead to different results. Experiment 4 provides on-line evidence that comprehenders do in fact actively postulate subject-internal gaps within such tensed relative clauses, contra Phillips (2006).

# 6 Experiment 4: Gap-filling in tensed subject phrases

Past research on gap filling in subject-embedded tensed relatives is inconclusive. Clifton and Frazier (1989) manipulated the verb inside relative clauses in a speeded acceptability experiment so that the latter could be an optionally transitive verb as in (32a) or an obligatorily intransitive verb as in (32b). The relevant subject phrase is bracketed for perspicuity.

(32) a. What did John think [the girl who always **won**] received?

b. What did John think [the girl who always **excelled**] received?

Speakers took a longer time to judge the former, suggesting the postulation of a gap inside the subject-embedded relative, but as Phillips (2006) notes, this may simply reflect readers' uncertainty over the argument structure of the optionally transitive verb. Similarly, Experiment 1 of Pickering et al. (1994) probed sentences like (33) and found a slowdown in reading time at the tensed relative clause verb, *painted*, rather than at the complement *the large mural*.

(33) I realize what [the artist who painted the large mural] ate today.

Although this finding is consistent with the formation of a long-distance dependency in which there is a gap in the subject-embedded relative clause, Pickering et al. (1994) concede that the slowdown may simply reflect overall processing load at that region of the sentence. To resolve this impasse, Phillips (2006, 807–814) conducted a semantic plausibility self-paced reading experiment which suggests that comprehenders attempt to resolve filler-gap dependencies in subject-embedded infinitival verb phrases (i.e. comprehenders postulated a gap after *expand* in (34a) when the filler phrase was compatible with the relative verb), but make no such attempt in subject-embedded tensed verb phrases (i.e. gaps were never postulated after *expanded* in (34b)).

(34) a. The school superintendent learned **which schools / which high school students** the proposal to **expand** drastically and innovatively upon the current curriculum would overburden _ during the following semester.

b. The school superintendent learned **which schools / which high school students** the proposal that **expanded** drastically and innovatively upon the current curriculum would overburden _ during the following semester.

Given these results, Phillips (2006, 803) concludes that the language processor does not postulate gaps in subject-embedded tensed structures because such gaps are grammatically illicit, parasitically or otherwise. However, the argument that parasitic gaps in subject-embedded tensed structures are illicit is undermined by the existence of counterexamples like (26) above, and by the results of our Experiment 3. We conjecture that the comprehenders in Phillips (2006) did not attempt to postulate subject-internal gaps in tensed environments like (34b) because of a flaw in the experimental items. At least 8 out of 24 items (over 30%) in the tensed and plausible conditions used in Phillips (2006) actually contained semantically infelicitous critical regions. This becomes

obvious if we focus on Phillips' relative clauses, and simply 'undo' the extraction. For example, in the plausible condition in (34b), the noun phrase becomes *the proposal that expanded the schools*, which is semantically felicitous. Unfortunately, this was not the case for several of the experimental items. In (35) we list some problematic noun phrases drawn from the critical regions of the plausible conditions used in Phillips (2006, 820).

(35) a. #The request that asked the professor ...

b. #The idea that expanded the outdated building ...

c. #The plan that prepared the extravagant reception ...

d. #The effort that battled religious groups ...

e. #The attempt that aimed the weapon ...

f. #The scheme that prepared basic foods ...

g. #The struggle that battled the deadly disease ...

h. #The plan that argued the high-profile case ...

The phrase (35a) is bizarre in the tensed condition because *request* is not a suitable agent for the verb *ask*. In contrast, the infinitival counterpart is felicitous because *to ask the question* is a complement of *request*. In other words, not only do the tensed and the non-tensed counterparts have different meanings, the former are severely deviant. The same objection can be raised for all the other phrases listed in (35). All the experimental items should have been equally felicitous in both the infinitival and the tensed conditions, but because more than half of the items in the tensed condition were odd (at least 12 were odd because the filler phrase was implausible, and at least 8 were odd as seen in (35)), it is possible that comprehenders adapted to the near-systematic oddness of the sentences in the tensed condition and simply refrained from attempting to fill gaps in that condition. If the source of extraction is defective, then the result is at least as defective. Below we show that comprehenders can in fact postulate subject-internal gaps in tensed relatives during on-line sentence processing, using the same paradigm of Phillips (2006).

## 6.1 Methods

### 6.1.1 Participants

We analyzed data provided by 30 participants who were recruited in AMT, using recruitment protocols identical to Experiment 1. All had accuracy levels of at least 80% in comprehension questions, with a mean accuracy level of 87%. There were 11 additional participants whose accuracy scores were lower than the 80% threshold. Data collected from these participants were discarded.

### 6.1.2 Design and materials

We constructed 20 pairs of experimental sentences like the two pairs shown in (36) and (37). See Appendix D for a complete list. Each pair of items differed in the semantic plausibility of the extracted phrase in region 1 as a filler for a potential gap at the end of the relative clause in region 7. For example, whereas the noun *bill* is a semantically plausible complement for *to pass*, the noun *table* is not. Region 10 contained the grammatically correct object of the verb in region 7, and region 11 corresponded to the main verb of the clause.

(36) a. Which **bill** $_1$| did $_2$| the senators $_3$| who $_4$| attended $_5$| the meeting $_6$| **to pass** $_7$| maybe $_8$| only $_9$| one statute $_{10}$| end up $_{11}$| supporting?$_{12}$| (*plausible gap at region 7*)

b. Which **table** $_1$| did $_2$| the senators $_3$| who $_4$| attended $_5$| the meeting $_6$| **to pass** $_7$| maybe $_8$| only $_9$| one statute $_{10}$| end up $_{11}$| sitting at?$_{12}$| (*implausible gap at region 7*)

(37) a. Which **door** $_1$| did $_2$| the fireman $_3$| who $_4$| smashed $_5$| a window $_6$| **to unlock** $_7$| only $_8$| just $_9$| a padlock $_{10}$| fail $_{11}$| to notice?$_{12}$| (*plausible gap at region 7*)

b. Which **rule** $_1$| did $_2$| the fireman $_3$| who $_4$| smashed $_5$| a window $_6$| **to unlock** $_7$| only $_8$| just $_9$| a padlock $_{10}$| fail $_{11}$| to follow?$_{12}$| (*implausible gap at region 7*)

Thus, the critical regions of interest were regions 7 through 11. If comprehenders attempt to resolve the filler-gap dependency at region 7, inside a subject-embedded tensed relative clause, then there should be a slowdown in reading time when the filler phrase is a semantically implausible complement for the verb at region 7. The adverbs at regions 8 and 9 were designed to delay and maximize the possibility of linking the extracted phrase to the object of the verb at region 7. Crucially, the adverbs are temporarily compatible with a parse in which they modify an upcoming main verb phrase, illustrated in (38a), or an upcoming complement, as in (38b). Although the latter is the correct parse for the adverbs in our experimental items, this is a rather infrequent and therefore dispreferred use of such adverbs, though perfectly grammatical and attested in corpora.[14]

(38) a. The senators who attended the meeting to pass the bill [**maybe only** [slept two hours]].

b. The senators who attended the meeting to pass [**maybe only** [one statute]] sat at this table.

Crucially, the verbs at region 7 are strongly transitive, and as is well-known, gaps tend to be postulated after strongly transitive verbs even if they are implausible and even before there is direct evidence for a missing object (Stowe et al., 1991). But if comprehenders postulate a gap at region

[14] E.g. examples like (i) – (iii) from COCA and online sources, validated by native speakers.

i. Eat only just as much as you think the stomach requires.

ii. (...) judges would consistently adopt and apply perhaps merely heightened scrutiny (...)

iii. (...) who himself has allegedly only one pupil (...).

7 in the implausible condition, gap-filling fails because the filler is not a plausible object for the verb. As a consequence, the presence of the adverbs in region 8 and 9 will be rather unexpected precisely because such adverbs are typically used as verb phrase modifiers, and there should be additional processing difficulty in the adverbial regions. No extra difficulty should arise in the plausible condition because the adverbs are consistent with an upcoming main verb phrase parse. Finally, note that region 7 always consisted of a 'purpose' infinitival VP adjoined to the tensed VP heading the relative. Thus, any gap postulated at region 7 crosses a total of three classic 'barriers' to extraction: modification, tense, and a subject phrase.

To ensure that there is a difference in plausibility between the plausible and implausible versions of each experimental sentence pair, a norming study was conducted in which a separate group of 53 native English speakers recruited through AMT were asked to evaluate the propositions expressed by the relative clauses of our items. For example, the two conditions in (36a) were used to produce the declaratives in (39). Participants were then asked to rate the plausibility of such sentences, using a 1–7 Likert scale, with 1 being 'very implausible', and 7 being 'very plausible'. Pairwise t-tests revealed a significant plausibility difference for all pairs (all $p$'s $< 0.05$).

(39) a. The senators attended the meeting to pass a bill. (*plausible*)

b. The senators attended the meeting to pass a table. (*implausible*)

After the norming study, the experimental sentence pairs were counterbalanced across two lists so that each participant only responded to one version of each sentence. Experimental sentences were interspersed among 50 distractor items, and the experiment was presented in a different random order for each participant. Distractors were of four types, as illustrated in (40), half of which were followed by (non-trivial) comprehension questions, also shown in (40).

(40) a. Which machine | did | the manager | of | the shop | eventually | quietly | ask | if | the employees | wanted | to get rid of? |
(Q: *Was the manager unsure whether some equipment needed to be thrown away?*)

b. Which study | did | the authors of | clearly claim | that | they | just | completed | with the help | of private funds? |
(Q: *Did the authors of the study use money from non-governmental sources?*)

c. That | was | the street | that | the witnesses | said | the suspects | placed | road blocks | at | during | the robbery. |
(Q: *Is it likely that the witnesses saw the suspects but not the road blocks?*)

d. Everyone | is | very | well | aware | of | which people | were invited | to the party | but | just | never | showed up.
(Q: *Is it true that none of the party guests failed to attend the party?*)

The first type of distractor, exemplified in (40a), consisted of a complex interrogative in which the gap is deeply embedded in a clausal complement of the matrix. The second type of distractor,

seen in (40b), consisted of a standard parasitic gap interrogative construction in which one gap is a complement of the subject nominal and the second gap is embedded in the matrix verb phrase. The third type of distractor consisted of declarative clauses with a demonstrative subject and a complex relative clause embedded in the complement phrase, as in (40c). The fourth and final type of distractor consisted of a declarative-embedded interrogative, as in (40d). As in previous experiments, the distractors were homogenous and therefore can act as additional controls.

### 6.1.3 Procedure

The procedure was identical to that of Experiment 2.

## 6.2 Results

For each region of interest, reading times less than 100 ms and more than 1,500 ms long were excluded, and data points greater or less than 2.5 standard deviations from each participants mean were replaced with these boundary values. Residual reading times were computed as in Experiment 2. An LMER model was used to test whether the residual reading times for all sentence regions were affected by the plausibility of the filler phrase. Mean residual reading times are in Figure 6. The bars represent 95% confidence intervals.

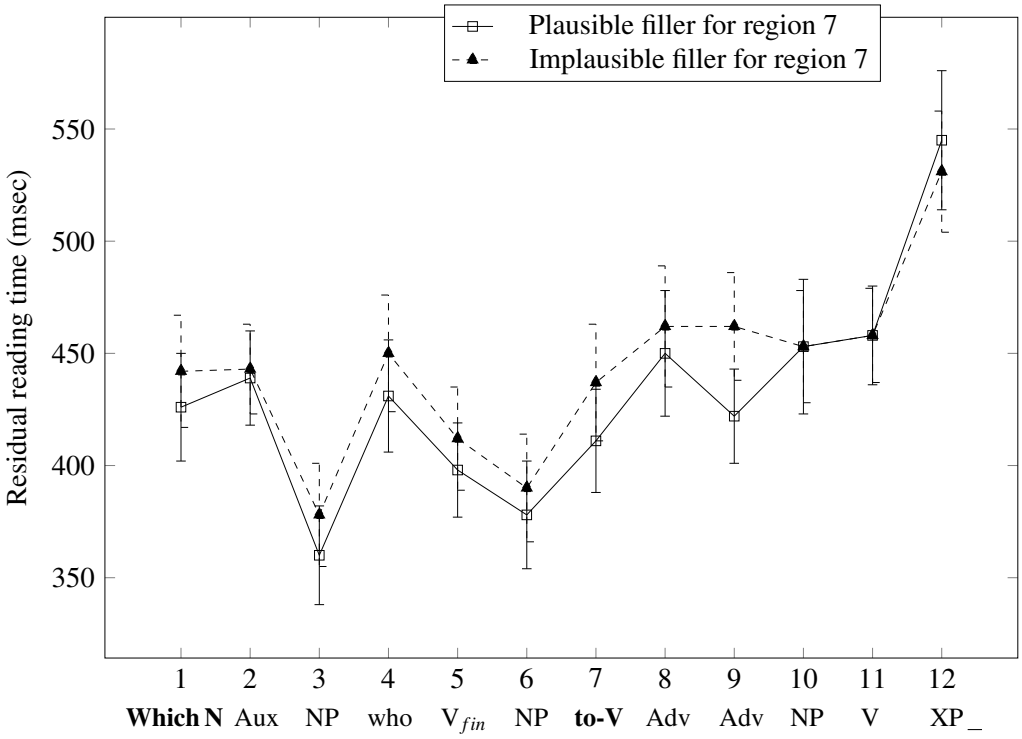


Figure 6: Mean residual reading times for all sentence regions (Experiment 4)

There was a main effect of plausibility at the verb region 7 ($\beta = -26.94, t = -2.037, p = 0.042$), indicating that participants read the infinitival verb phrase more slowly when the filler was an implausible gap-filler at that verb's region (mean 437 ms, SD = 205) than when the filler was a plausible gap-filler at that same region (mean 411 ms, SD = 229). There was no significant effect of plausibility at region 8 ($\beta = -13.63, t = -0.977, p = 0.329$), but at region 9 there was again a main effect of plausibility ($\beta = -38.35, t = -3.397, p = 0.0007$), in which participants again slowed down in the implausible filler phrase condition only (mean 462 ms, SD = 203 vs. 422 ms, SD = 183). No other regions before 7 or after 9 registered any significant effects.

### 6.3 Discussion

The results suggest that comprehenders postulated a gap at region 7, given that the items in the implausible filler phrase condition lead to a significant increase in reading time at that region. As already noted above, gaps tend to be postulated after strongly transitive verbs even if they are implausible and even before there is direct evidence for a missing object (Stowe et al., 1991). If comprehenders postulate a gap at region 7 in the implausible condition, then gap-filling fails because the filler is a semantically implausible object for the verb. And after failing to fill it because of its implausibility, comprehenders come to expect the verbs true object. However, they instead encounter adverbs that are most consistent with an upcoming main verb phrase, and therefore the slowdown at region 9 only arises in the implausible condition.

In sum, our findings suggest that comprehenders can postulate gaps in subject-embedded tensed relatives, contrary to the null effects found by Phillips (2006) in an analogous experiment. We hypothesize that the latter was caused by semantically infelicitous tensed items. The fact that active gap filling sometimes occurs in subject-embedded tensed phrases, and sometimes does not, is consistent with the hypothesis that the on-line processing of filler-gap dependencies is modulated by extra-grammatical factors, such as probabilistic information about the expected distribution of gaps, given the semantics and pragmatics of the items.

## 7 General Discussion

Following Erteschik-Shir (1981), Van Valin (1986), Kuno (1987), Takami (1992), Deane (1992) and others, we assume that extraction is in general restricted to the informational focus of the proposition. As a consequence, extraction from subject phrases should be difficult given that subjects are typically reserved for topic continuity, and subject-embedded referents are unlikely to be the informational focus of the utterance. This conclusion is independently supported by corpus data showing that attested Subject Island violations invariably involve relative clauses that describe new information rather than background information, and that complex NPs tend to be objects rather than subjects.

Our hypothesis is that if the entire proposition is highly pragmatically felicitous, then the mention of its embedded referents must be highly felicitous too, even if subject-embedded. If the mention of a given referent (subject-embedded of not) is highly felicitous, it is more likely for such a

referent to be understood as the informational focus, and thus easier for it to be extracted. This is consistent with the intuition expressed by Kluender (2004, 495), Shimojo (2002), and Chaves (2013) that Subject Island violations tend to be more acceptable when the extracted referent is relevant for the main assertion. More generally, in order for the mention of a referent to be felicitous and in accordance to the Gricean Maxims of *Quantity* ('Be brief') and *Manner* ('Avoid Prolixity'), the context must be such that the referent is not irrelevant for the proposition. As we show, because subjects tend to be topics, subject-embedded referents are rarer than object-embedded referents, even without extraction.

According to our findings, ensuring that a proposition with subject-embedded referents is highly felicitous to begin with suffices to induce the extractability of such subject-embedded referents, if their frequency is increased. Our results are most consistent with the functional approach proposed by Hofmeister et al. (2013) and Chaves (2013) in which the more syntactically, semantically, and pragmatically plausible the subject-internal gap, the more acceptable such extractions should be, especially if comprehenders have the chance to adjust to the unusual syntactic location of the gap by being exposed to multiple exemplars of such sentences. We argue that such amelioration is facilitated when the very mention of the subject-embedded referent is a highly felicitous discourse move to begin with. Such an account also explains why a subject-embedded gap often becomes acceptable 'parasitically' if co-indexed with a non-island gap: since the extracted referent is an argument of the main predicate, it is trivially relevant for the main assertion. Our account is also consistent with the findings of Polinsky et al. (2013), which suggests that Subject Island violations in transitive predicates tend to be stronger than in unaccusatives. If the subject is an agent or an actor, it initiates or controls the event, and is by default more relevant for the assertion than any subject-embedded referent. However, if the subject is not an agent or actor, then it is easier for a phrase other than the subject to be construed as relevant, including a subject-embedded referent. Thus, it is not easy to extract from subjects of transitive verbs, though not impossible; see Chomsky (2008, 160,ft.39), Jiménez–Fernández (2009), and Chaves (2013). Ultimately, Subject Island effects are likely proposition-dependent, and in particular, contingent on the degree to which the subject-embedded referent is relevant for the assertion.

The present account of Subject Island effects also requires no special assumptions to explain the behavioral phenomena. It is well-known that speakers generally attend to frequencies in linguistic input, and can adapt to changes in order to overcome the processing difficulty caused when the input is unusual and inconsistent with their prior experience. In particular, comprehenders make use of expectations about the distribution of filler-gap dependencies to efficiently prune the search space during gap detection, mitigating the processing costs associated with resolving such dependencies during on-line sentence comprehension (van Schijndel et al., 2014; Michel, 2014). We conclude that linguistic theory need not be complicated by constraints that block extraction from subjects, as the behavioral phenomena are graded, fluid and thus likely functional in nature.

# Appendices

### Appendix A: Experiment 1

1. Which animal does the mating call of {reportedly mimic the sounds of the Gray Catbird?, the Gray Catbird reportedly mimic the sounds of?}
2. Which artist does the son of {frequently collaborate with the daughter of the Governor?, the Governor frequently collaborate with the daughter of?}
3. Which athlete does the manager of {faintly resemble the agent of Tiger Woods?, Tiger Woods faintly resemble the agent of?}
4. Which celebrity does the wife of {reportedly quarrel with the fiancee of the Mayor?, the Mayor reportedly quarrel with the fiancee of?}
5. Which committee does the report of {supposedly contradict the recommendations of the experts?, the experts supposedly contradict the recommendations of?}
6. Which company does the chairman of {usually bump heads with the CEO of IBM?, IBM usually bump heads with the CEO of?}
7. Which condition does the treatment of {typically mimic the effects of eczema?, eczema typically mimic the effects of?}
8. Which country does the King of {allegedly resemble the President of Sweden?, Sweden allegedly resemble the President of?}
9. Which disease does the vaccine for {generally interact with the treatment of malaria?, malaria generally interact with the treatment of?}
10. Which inmate does the description of {plainly match the photos of the suspect?, the suspect plainly match the photos of?}
11. Which movie does the plot of {radically deviate from the script of the play?, the play radically deviate from the script of?}
12. Which nation does the budget surplus of {nearly equal the fiscal deficit of Brazil?, Brazil nearly equal the fiscal deficit of?}
13. Which politician does the son of {sometimes socialize with the ex-wife of Clint Eastwood?, Clint Eastwood sometimes socialize with the ex-wife of?}
14. Which pundit does the wife of {rarely agree with the opinions of the President?, the President rarely agree with the opinions of?}
15. Which road does the end of {actually coincide with the beginning of Route 66?, Route 66 actually coincide with the beginning of?}
16. Which singer does the brother of {regularly hang out with the daughter of Bill Clinton?, Bill Clinton regularly hang out with the daughter of?}
17. Which skirt does the color of {totally clash with the patterns of the jacket?, the jacket totally clash with the patterns of?}
18. Which song does the beat of {supposedly coincide with the melody of ”Ice, Ice Baby”?, ”Ice, Ice Baby” supposedly coincide with the melody of?}

19. Which state does the Governor of {actively oppose the senator from Vermont?, Vermont actively oppose the senator from?}
20. Which movie does the score of {clearly sound like the main theme of "The Godfather"?, "The Godfather" clearly sound like the main theme of?}
21. Which stock does the value of {often parallel the price of the dollar?, the dollar often parallel the price of?}
22. Which team does the manager of {definitely look like the coach of the Lakers?, the Lakers definitely look like the coach of?}

## Appendix B: Experiment 2

*Block 1, List 1*

1. Which animal | does | the song of | reportedly | mimic | the Gray Catbird's sounds?
2. Which artist | does | the son of | frequently | collaborate | with the Governor's daughter?
3. Which athlete | does | the manager of | clearly | resemble | Tiger Woods' agent?
4. Which tune | do | covers of | always | become | popular around Christmas?
5. Which committee | do | the findings of | supposedly | contradict | the expert's recommendations?
6. Which company | does | the chairman of | usually | disagree | with his own advisors?
7. Which condition | does | the treatment of | typically | mimic | the effects of eczema?
8. Which country | does | the King of | clearly | resemble | Sweden's President?
9. Which disease | does | the vaccine for | frequently | interact | with malaria treatments?
10. Which inmate | does | the description of | barely | match | the suspect's photos?
11. Which movie | does | the plot of | radically | deviate | from the play's script?
12. Which sculpture | do | the replicas of |traditionally | sell | for thousands of dollars?
13. Which nation | does | the budget of | nearly | equal | Brazil's fiscal deficit?
14. Which politician | does | the son of | often | socialize | with Shakira's ex-husband?
15. Which pundit | does | the wife of | rarely | support | her own husband's opinions?

*Block 2, List 1&2*

16. Which road | does | the end of | actually | coincide | with Route 66's beginning?
17. Which singer | does | the voice of | reputedly | irritate | the American Idol judges?
18. Which skirt | does | the color of | totally | clash | with the jacket's patterns?
19. Which song | does | the beat of | supposedly | mirror | the melody of 'Ice Ice Baby'?
20. Which state | does | the Governor of | actively | despise | the senator from Iowa?
21. Which stock | does | the price of | often | coincide | with the dollar's value?
22. Which team | does | the manager of | definitely | look | like the Lakers' coach?
23. Which song | does | the chorus of | almost | sound | like the national anthem?
24. Which celebrity | does | the wife of | reportedly | bicker | with the Mayor's fiancee?
25. Which company | do | the employees of | allegedly | reject | salary increases?

**Appendix C: Experiment 3**

1. Which incident did the bystanders who {witnessed find terrifying?, fled into the subway find terrifying?, witnessed end up going into therapy?}
2. Which woman do all the men who {work with want to date?, work in the corner office want to date?, work with become more productive?}
3. Which products did the customers who {ordered online not receive for three weeks?, visited the store want to buy?, ordered online not have poor credit?}
4. Which articles did the journalists who {wrote decide not to publish?, you interviewed decide not to publish?, wrote decide to quit their job?}
5. Which puzzle did the kids who {solved enjoy the most?, came to my parties enjoy the most?, solved feel very tired?}
6. Which virus did the mice that were {infected with never recover from?, in the lab never recover from?, infected with never get sick?}
7. Which client did the lawyers who {worked for end up suing?, felt discriminated end up suing?, worked for try to sue the state?}
8. Which celebrities do the people who {follow on Twitter also follow on Facebook?, are over fifty years old follow on Facebook?, follow on Twitter have college degrees?}
9. Which company did the interns who {worked for receive bonuses from?, graduated from college receive bonuses from?, worked for start non-profit organization?}
10. Which tax bill do the senators who {support probably benefit from?, cheat on taxes probably benefit from?, support probably get lots of donations?}
11. Which disease did the children who {fell victim to eventually recover from?, were malnourished recover from?, fell victim to were malnourished?}
12. Which drink did the guests who {ordered love the most?, sat on the patio love the most?, ordered sat on the patio?}
13. Which products do the customers who {buy recommend to their friends?, visit the store recommend to their friends?, buy never return to the store?}
14. Which proposal did the people who {rejected end up accepting?, were in the courthouse end up accepting?, rejected go home?}
15. Which patient did the doctors who {examined decide not to treat?, were at the private clinic decide not to treat?, examined been transferred to another hospital?}
16. Which newspaper did the editor who {criticized get fired from?, criticized the photojournalist get fired from?, criticized blackmail the mayor?}
17. Which politician do all the people who {support end up hating?, support the campaign end up hating?, support end up changing their political party?}
18. Which law did the congressman who {proposed accidentally vote against?, was arrested for fraud accidentally vote against?, proposed fired his secretary?}
19. Which movies do the people who {rent usually watch with friends?, are single usually watch with friends?, rent usually have nightmares?}

20. Which program do the children who {participate in enjoy the least?, you interview enjoy the least?, participate in sign consent form?}
21. Which suspect did the policeman who {arrested forget to handcuff?, went missing yesterday forget to handcuff?, arrested write to the newspapers?}
22. Which witness did the detectives who {questioned accuse of lying?, were fired accuse of lying?, questioned quit the police force?}
23. Which toys did the kids who {liked try to steal?, came to the party try to steal?, liked have great time?}
24. Which girl do all the boys who {dance with fall in love with?, move to my street fall in love with?, danced with leave the party early?}

**Appendix D: Experiment 4**

1a. Which greeting | did | the shoppers | who | came | to the store | to buy | sometimes | simply | a newspaper | get | the most?|
1b. Which product | did | the shoppers | who | came | to the store | to buy | sometimes | simply | a newspaper | get | the most?|
2a. Which problem | did | the fans | who | went | to the bar | to watch | only | probably | some contest | end up | having later?|
2b. Which game | did | the fans | who | went | to the bar | to watch | only | probably | some contest | end up | seeing instead?|
3a. Which complaint | did | the neighbor | who | got | a trap | to catch | basically | just | squirrels | end up | getting?|
3b. Which critter | did | the neighbor | who | got | a trap | to catch | basically | just | squirrels | end up | also ensnaring?|
4a. Which error | did | the customers | who | went | to our website | to purchase | typically | mostly | clothing | detest | the most?|
4b. Which product | did | the customers | who | went | to our website | to purchase | typically | mostly | clothing | detest | the most?|
5a. Which complaint | did | the reporter | who | hid | in the bushes | to photograph | allegedly | only | some birds | get | fired for?|
5b. Which celebrity | did | the reporter | who | hid | in the bushes | to photograph | allegedly | only | some birds | get | punched by?|
6a. Which nickname | did | the witness | who | returned | to court | to answer | probably | merely | a survey | help | reveal?|
6b. Which question | did | the witness | who | returned | to court | to answer | probably | merely | a survey | help | resolve?|
7a. Which building | did | the doctor | who | moved | to Nigeria | to cure | maybe | just | sore throats | end up | dying in?|

7b. Which epidemic | did | the doctor | who | moved | to Nigeria | to cure | maybe | just | sore throats | end up | dying from?|

8a. Which avenue | did | the defendant | who | sold | the house | to pay | perhaps | merely | the mortgage | have | to abandon?|

8b. Which fee | did | the defendant | who | sold | the house | to pay | perhaps | merely | the mortgage | have | to waive?|

9a. Which rule | did | the fireman | who | smashed | a window | to unlock | only | just | a padlock | fail | to follow?|

9b. Which door | did | the fireman | who | smashed | a window | to unlock | only | just | a padlock | fail | to notice?|

10a. Which table | did | the senators | who | attended | the meeting | to pass | maybe | only | one statute | end up | sitting at?|

10b. Which bill | did | the senators | who | attended | the meeting | to pass | maybe | only | one statute | end up | supporting?|

11a. Which candy | did | the children | who | got | a vaccination | to prevent | apparently | only | measles | become | allergic to?|

11b. Which disease | did | the children | who | got | a vaccination | to prevent | apparently | only | measles | become | immune to?|

12a. Which vehicle | did | the gardener | who | dug | a hole | to plant | probably | mostly | some seeds | cause | damage to?|

12b. Which tree | did | the gardener | who | dug | a hole | to plant | probably | mostly | some seeds | cause | damage to?|

13a. Which sport | did | the teenager | who | flew | to Hollywood | to meet | probably | just | a pen pal | fall | in love with?|

13b. Which celebrity | did | the teenager | who | flew | to Hollywood | to meet | probably | just | a pen pal | fall | in love with?|

14a. Which injury | did | the mason | who | climbed | the wall | to repair | supposedly | mostly | some bricks | end up | suffering from?|

14b. Which chimney | did | the mason | who | climbed | the wall | to repair | supposedly | mostly | some bricks | end up | fixing too?|

15a. Which inspector | did | the engineer | who | pressed | a button | to activate | only | just | one machine | complain | about?|

15b. Which computer | did | the engineer | who | pressed | a button | to activate | only | just | one machine | complain | about?|

16a. Which scam | did | the journalist | who | traveled | to Sochi | to interview | mostly | only | athletes | fall | victim to?|

16b. Which celebrity | did | the journalist | who | traveled | to Sochi | to interview | mostly | only | athletes | fall | in love with?|

17a. Which senator | did | the congressman | who | drafted | a document | to propose | maybe | simply | a tax break | anger | the most?|

17b. Which plan | did | the congressman | who | drafted | a document | to propose | maybe | simply | a tax break | oppose | the most?|

18a. Which fee | did | the subscribers | who | use | the Internet | to download | mostly | only | music | hate | the most?|

18b. Which movies | did | the subscribers | who | use | the Internet | to download | mostly | only | music | watch | at home?|

19a. Which mistake | did | the kids | who | were | at the store | to buy | perhaps | simply | candy | cause | their parents | to make?|

19b. Which toy | did | the kids | who | were | at the store | to buy | perhaps | simply | candy | cause | some damage | to?|

20a. Which substance | did | the policeman | who | exited | the car | to confront | presumably | only | a juvenile | find | in the sidewalk?|

20b. Which suspect | did | the policeman | who | exited | the car | to confront | presumably | only | a juvenile | point | a gun at?|

# References


Abeillé, Anne, Hemforth, Barbara, Winckel, Elodie and Gibson, Edward. 2018. A construction-conflict explanation of the subject-island constraint, 31th Annual CUNY Conference on Human Sentence Processing.

Almor, Amit. 2000. Constraints and mechanisms in theories of anaphor processing. In Matthew W. Crocker, Martin J. Pickering and Charles Clifton (eds.), *Architectures and mechanisms for language processing*, pages 341–354, Cambridge, MA: MIT Press.

Altmann, Gerry T. M. and Kamide, Yuki. 1999. Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition* 73, 247– 264.

Amy, Gérard and Noziet, Georges. 1978. Memory requirements and local ambiguities of parsing strategies. *Journal of Psycholinguistic Research* 20(3), 233–250.

Arai, Manabu and Keller, Frank. 2013. The use of verb-specific information for prediction in sentence processing. *Language and Cognitive Processes* 4(28), 525–560.

Ariel, Mira. 2001. Accessibility theory: an overview. In T. Sanders, J. Schilperoord and W. Spooren (eds.), *Text Representation: Linguistic and Psycholinguistic Aspects*, pages 29–87, Amsterdam: John Benjamins.

Bates, Douglas, Maechler, Martin, Bolker, Ben and Walker, Stevern. 2014. lme4: Linear mixed-effects models using 'Eigen' and S4, R package version 1.1-7.

Bayer, Josef and Salzmann, Martin. 2013. That-trace effects and resumption: How improper movement can be repaired. In P. Brandt and E. Fuß (eds.), *Repairs - The Added Value of Being Wrong*, pages 275 – 333, Berlin: de Gruyter.

Boeckx, Cedric. 2003. *Islands and Chains. Resumption as Stranding*. Benjamins, Amsterdam.

Boston, Marisa Ferrara, Hale, John, Kliegl, Reinhold, Patil, Umesh and Vasishth, Shravan. 2008. Parsing costs as predictors of reading difficulty: An evaluation using the Potsdam Sentence Corpus. *Journal of Eye Movement Research* 2, 1–12.

Chafe, Wallace L. 1994. *Discourse, consciousness, and time: The flow and displacement of conscious experience in speaking and writing*. Chicago, University Chicago Press.

Chaves, Rui P. 2012. On the grammar of extraction and coordination. *Natural Language and Linguistic Theory* 30(2), 465–512.

Chaves, Rui P. 2013. An expectation-based account of subject islands and parasitism. *Journal of Linguistics* 2(49), 285–327.

Chaves, Rui P. 2018. Freezing as a probabilistic phenomenon. In Jutta Hartmann, Marion Knecht, Andreas Konietzko and Susanne Winkler (eds.), *Freezing - Theoretical Approaches and Empirical Domains*, De Gruyter.

Chaves, Rui P. and Dery, Jeruen E. 2014. Which subject islands will the acceptability of improve with repeated exposure? In R. E. Santana-LaBarge (ed.), *31st West Coast Conference on Formal Linguistics*, pages 96–106, Somerville, MA: Cascadilla Proceedings Project.

Chen, Evan, Gibson, Edward and Wolf, Florian. 2005. Online syntactic storage costs in sentence comprehension. *Journal of Memory and Language* 52, 144–169.

Chesi, Cristiano and Bianchi, Valentina. 2014. Subject islands, reconstruction, and the flow of the computation. *Linguistic Inquiry* 45(4), 525–569.

Chomsky, Noam. 1973. Conditions on transformations. In Stephen Anderson and Paul Kiparsky (eds.), *A Festschrift for Morris Halle*, pages 232–286, New York: Holt, Reinhart & Winston.

Chomsky, Noam. 1977. *Essays on form and interpretation*. New York: North-Holland.

Chomsky, Noam. 1995. *The Minimalist Program*. The MIT Press. Cambridge. MA.

Chomsky, Noam. 2008. On Phases. In Robert Freidin, David Michaels, Carlos P. Otero and Maria Luisa Zubizarreta (eds.), *Foundational Issues in Linguistic Theory: Essays in Honor of Jean-Roger Vergnaud*, pages 133–165, Cambridge, MA, USA: MIT Press.

Chung, Sandra and McCloskey, James. 1983. On the interpretation of certain island facts in GPSG. *Linguistic Inquiry* 14, 703–714.

Clark, Herbert H. and Wasow, Thomas. 1998. Repeating words in spontaneous speech. *Cognitive Psychology* 37, 201–242.

Clausen, David R. 2011. Informativity and Acceptability of Complex Subject Islands, poster presented at the 24th Annual CUNY Sentence Processing Conference, Stanford, CA.

Clifton, Charles E. Jr. and Frazier, Lyn. 1989. Comprehending sentences with long distance dependencies. In Michael K. Tanenhaus and Gregory N. Carlson (eds.), *Linguistic structure in language processing*, pages 273–317, Dordrecht: Kluwer.

Crain, Stephen and Fodor, Janet D. 1985. How can grammars help parsers? In David Dowty, Lauri Kartunnen and Arnold M. Zwicky (eds.), *Natural Language Parsing: Psycholinguistic, Computational, and Theoretical Perspectives*, pages 94–128, Cambridge, UK: Cambridge University Press.

Crawford, Jean. 2011. Using Syntactic Satiation Effects to Investigate Subject Islands. In *WCCFL 29 Proceedings*, University of Arizona.

Creel, Sarah C., Aslin, Richard N. and Tanenhaus, Michael K. 2008. Heeding the voice of experience: The role of talker variation in lexical access. *Cognition* 106, 633–664.

Culicover, Peter W. 1999. *Syntactic Nuts: Hard Cases in Syntax*. Volume One of Foundations of Syntax, Oxford University Press.

Culicover, Peter W. 2013. *Grammar & Complexity*. Oxford University Press.

Davies, Mark. 2008. The Corpus of Contemporary American English: 450 million words, 1990-present, available online at http://corpus.byu.edu/coca/.

Deane, Paul D. 1992. *Grammar in Mind and Brain*. Mouton de Gruyter.

DeLong, Katherine A., Urbach, Thomas P. and Kutas, Marta. 2005. Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience* 8(8), 1117–1121.

Demberg, Vera and Keller, Frank. 2008. Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. *Cognition* 109(2), 193–210.

Do, Monica L. and Kaiser, Elsi. 2017. The Relationship between Syntactic Satiation and Syntactic Priming: A First Look. *Frontiers in Psychology* 8, 1851.

Drummond, Alex. 2013. Ibex 0.3.7 Manual, [http://spellout.net/latest_ibex_manual.pdf].

Eady, Stephen J. and Fodor, Janet D. 1981. Is center embedding a source of processing difficulty?, presented at the Linguistics Society of America Annual Meeting.

Engdahl, Elisabet. 1983. Parasitic gaps. *Linguistics and Philosophy* 6, 3–34.

Enochson, Kelly and Culbertson, Jennifer. 2015. Collecting Psycholinguistic Response Time Data Using Amazon Mechanical Turk. *PLoS ONE* 10(3).

Epstein, Richard. 2001. The definite article, accessibility, and the construction of discourse referents. *Cognitive Linguistics* (12), 333–378.

Erteschik-Shir, Nomi. 1981. More on extractability from quasi-NPs. *Linguistic Inquiry* 12, 665–670.

Erteschik-Shir, Nomi and Lappin, Shalom. 1979. Dominance and the functional explanation of island phenomena. *Theoretical Linguistics* 6, 41–86.

Farmer, Thomas A., Fine, Alex B., Yan, Shaorong, Cheimariou, Spyridoula and Jaeger, Florian T. 2014. Error-Driven Adaptation of Higher-Level Expectations During Natural Reading. In *Proceedings of the 36th Annual Meeting of the Cognitive Science Society*, Austin, TX: Cognitive Science Society.

Federmeier, Kara D. and Kutas, Marta. 1999. A rose by any other name: Long-term memory structure and sentence processing. *Journal of Memory and Language* 41, 469–495.

Ferreira, Fernanda. 1991. Effects of length and syntactic complexity on initiation times for prepared utterances. *Journal of Memory and Language* 30(2), 210–233.

Fiengo, Robert and Higginbotham, James. 1981. Opacity in NP. *Linguistic Analysis* 7(4), 395–421.

Fine, Alex B., Jaeger, Florian T., Farmer, Thomas A. and Qian, Ting. 2013. Rapid expectation adaptation during syntactic comprehension. *PLoS ONE* 8(10).

Fine, Alex B. and Jaeger, T. Florian. 2013. Evidence for implicit learning in syntactic comprehension. *Cognitive Science* 37(3), 578–591.

Fine, Alex B., Qian, Ting, Jaeger, Florian T. and Jacobs, Robert A. 2010. Syntactic Adaptation in Language Comprehension. In *Proceedings of ACL Workshop on Cognitive Modeling and Computational Linguistics*, pages 18–26.

Frazier, Lyn. 1987. Syntactic processing: evidence from Dutch. *Natural Language and Linguistic Theory* 5, 519–559.

Fukuda, Shin, Nakao, Chizuru, Omaki, Akira and Polinsky, Maria. 2018. Revisiting subject-object asymmetry: subextraction in Japanese. In *Proceedings of the 10th Workshop on Altaic Formal Linguistics*, MIT Working Papers in Linguistics, to appear.

Futrell, Richard. 2012. *Processing Effects of the Expectation of Informativity*. Ma thesis, Stanford.

Gibson, Edward. 2006. The interaction of top-down and bottom-up statistics in the resolution of syntactic category ambiguity. *Journal of Memory and Language* 54, 363–388.

Gibson, Edward, Piantadosi, Steve and Fedorenko, Kristina. 2011. Using Mechanical Turk to obtain and analyze English acceptability judgments. *Language and Linguistics Compass* 5(8), 509–524.

Goldberg, Adele E. 2006. *Constructions at Work: the nature of generalization in Language*. Oxford: Oxford University Press.

Goldberg, Adele E. 2013. Backgrounded constituents cannot be extracted. In Jon Sprouse and Norbert Hornstein (eds.), *Experimental Syntax and Island Effects*, pages 221–238, Cambridge University Press.

Goodall, Grant. 2011. Syntactic satiation and the inversion effect in English and Spanish *wh-* questions. *Syntax* 14, 29–47.

Haegeman, Liliane, Ángel L. Jiménez-Fernández and Radford, Andrew. 2014. Deconstructing the subject condition in terms of cumulative constraint violation. *The Linguistic Review* 31(1), 73–150.

Hiramatsu, Kazuko. 2000. *Accessing Linguistic Competence: Evidence from children's and adults' acceptability judgments*. Ph.D. thesis, University of Connecticut.

Hofmeister, Philip, Casasanto, Laura Staum and Sag, Ivan A. 2013. Islands in the grammar? Standards of evidence. In Jon Sprouse and Norbert Hornstein (eds.), *Experimental Syntax and islands effects*, pages 42 – 63, Cambridge: Cambridge University Press.

Hofmeister, Philip and Sag, Ivan A. 2010. Cognitive Constraints and Island Effects. *Language* 86(2), 366–415.

Huang, Cheng-Teh James. 1982. *Logical relations in Chinese and the theory of grammar*. Ph.d dissertation, MIT, Cambridge, Massachusetts.

Huddleston, Rodney D., Pullum, Geoffrey and Peterson, Peter. 2002. *The Cambridge Grammar of the English Language*, Chapter 12: Relative clause constructions and unbounded dependencies, pages 1033–1096. Cambridge University Press.

Ivanova, Iva, Pickering, Martin, Branigan, Holly, F McLean, Janet and Costa, Albert. 2011. The comprehension of anomalous sentences: Evidence from structural priming. *Cognition* 122, 193–209.

Jackendoff, Ray S. 2002. *Foundations of Language: Brain, meaning, grammar, evolution*. Oxford University Press.

Jiménez–Fernández, Ángel. 2009. On the Composite Nature of Subject Islands: A phase-based approach. *SKY Journal of Linguistics* 22, 91–138.

Jurka, Johannes, Nakao, Chizuru and Omaki, Akira. 2011. It's not the end of the CED as we know It: revisiting German and Japanese subject islands. In Mary Byram Washburn, Katherine McKinney-Bock, Erika Varis, Ann Sawyer and Barbara Tomaszewicz (eds.), *Proceedings of the 28th West Coast Conference on Formal Linguistics*, pages 124 – 132, Somerville, MA: Cascadilla Proceedings Project.

Just, Marcel Adam, Carpenter, Patricia A. and Wooley, Jacqueline D. 1982. Paradigms and Processes in Reading Comprehension. *Journal of Experimental Psychology: General* 111(2), 228–238.

Kamide, Yuki, Scheepers, Christoph and Altmann, Gerry T. M. 2003. Integration of syntactic and semantic information in predictive processing: Crosslinguistic evidence from German and English. *Journal of Psycholinguistic Research* 32(1), 37–55.

Kaschak, Michael P. 2006. What this construction needs is generalized. *Memory & Cognition* 34(2), 368–379.

Kayne, Richard. 1981. ECP Extensions. *Linguistic Inquiry* 12, 93–133.

Kayne, Richard. 1983. Connectedness. *Linguistic Inquiry* 14, 223–249.

Kehler, Andrew. 2002. *Coherence, Reference, and the Theory of Grammar*. Stanford: CSLI Publications.

Keller, Frank. 2003. A Probabilistic Parser as a Model of Global Processing Difficulty. In Richard Alterman and David Kirsh (eds.), *Proceedings of the 25th Annual Conference of the Cognitive Science Society*,

pages 646–651.

Kluender, Robert. 1992. Deriving island constraints from principles of predication. In Helen Goodluck and Michael Rochemont (eds.), *Island Constraints: Theory, Acquisition and Processing*, pages 223–258, Dordrecht: Kluwer.

Kluender, Robert. 1998. On the distinction between strong islands and weak islands: a processing perspective. In Peter W. Culicover and Louise McNally (eds.), *Syntax and Semantics 29: The Limits of Syntax*, pages 241–279, New York, NY: Academic Press.

Kluender, Robert. 2004. Are subject islands subject to a processing account? In B. Schmeiser, V. Chand, A. Kelleher and A. Rodriguez (eds.), *In Proceedings of the WCCFL 23*, pages 101–125, Somerville, MA: Cascadilla Press.

Kluender, Robert and Kutas, Marta. 1993. Subjacency as a processing phenomenon. *Language and Cognitive Processes* 8, 573–633.

Kubota, Yusuke and Lee, Jungmee. 2015. The Coordinate Structure Constraint as a discourse-oriented principle: Further evidence from Japanese and Korean. *Language* 91(3), 642–675.

Kuno, Susumo. 1972. Functional Sentence Perspective: A case study from Japanese and English. *Linguistic Inquiry* 3, 299–320.

Kuno, Susumu. 1987. *Functional syntax – anaphora, discourse and empathy*. The University of Chicago Press, Chicago and London.

Kuno, Susumu and Takami, Ken-Ichi. 1993. *Grammar and discourse principles: functional syntax and GB theory*. Chicago & London: University of Chicago Press.

Kurtzman, Howeard S. and Crawford, Loren Frost. 1991. Processing parasitic gaps. In Tim Sherer (ed.), *Proceedings of the 21st Annual Meeting of the North Eastern Linguistics Society*, pages 217–231, Amherst, MA: LSA Publications.

Kutas, Marta and Hillyard, Steven A. 1984. Brain potentials during reading reflect word expectancy and semantic association. *Nature* 5947(307), 161–163.

Kuznetsova, Alexandra, Brockhoff, Per Bruun and Christensen, Rune Haubo Bojesen. 2015. Package lmerTest. Technical Report.

Kynette, Donna and Kemper, Susan. 1986. Aging and the loss of grammatical forms: a cross-sectional study of language performance. *Language and Communication* 6, 65–72.

Lambrecht, Knud. 1988. There was a farmer had a dog: Syntactic amalgams revisited. In *Proceedings of the Fourteenth Annual Meeting of the Berkeley Linguistics Society*, pages 319–339.

Lambrecht, Knud. 1994. *Information Structure and Sentence Form: Topic, focus, and the mental representation of discourse referents*. Cambridge University Press.

Lambrecht, Knud. 2000. When subjects behave like objects: An analysis of the merging of S and O in sentence-focus constructions across languages. *Studies in Language* 24(3), 611–682.

Larson-Hall, Jenniver and Herrington, Richard. 2010. Improving Data Analysis in Second Language Acquisition by Utilizing Modern Developments in Applied Statistics. *Applied Linguistics* 31, 368–390.

Lasnik, Howard and Saito, Mamoru. 1992. *Move α: Conditions on its application and output*. MIT Press.

Lau, Ellen, Stroud, Clare, Plesch, Silke and Phillips, Colin. 2006. The role of structural prediction in rapid syntactic analysis. *Brain and language* 1(98), 74–88.

Lau, Jey Han, Clark, Alexander and Lappin, Shalom. 2015. Unsupervised Prediction of Acceptability Judgements. In *Proceedings of the 53rd Annual Conference of the Association of Computational Linguistics, Beijing*, pages 1618–1628.

Levine, Robert D. and Hukari, Thomas E. 2006. *The unity of unbounded dependency constructions*. CSLI

Publications, Leland Stanford Junior University.

Levine, Robert D., Hukari, Thomas E. and Calcagno, Michael. 2001. Parasitic gaps in English: some overlooked cases and their theoretical implications. In Peter Culicover and Paul M. Postal (eds.), *Parasitic Gaps*, pages 181–222, Cambridge, MA: MIT Press.

Levine, Robert D. and Sag, Ivan A. 2003. Some empirical issues in the grammar of extraction. In Stefan Müller (ed.), *Proceedings of the HPSG-2003 Conference, Michigan State University, East Lansing*, pages 236–256, Stanford: CSLI Publications.

Levy, Roger. 2008. Expectation-based syntactic comprehension. *Cognition* 3(106), 1126–1177.

Levy, Roger and Andrew, Galen. 2006. Tregex and Tsurgeon: tools for querying and manipulating tree data structures. In *5th International Conference on Language Resources and Evaluation (LREC 2006)*.

Levy, Roger, Fedorenko, Evelina, Breen, Mara and Gibson, Ted. 2012. The processing of extraposed structures in English. *Cognition* 12(1), 12–36.

Levy, Roger and Keller, Frank. 2013. Expectation and locality effects in German verb-final structures. *Journal of Memory and Language* 2(68), 199–222.

Luka, Barbara and Barsalou, Lawrence. 2005. Structural facilitation: Mere exposure effects for grammatical acceptability as evidence for syntactic priming in comprehension. *Journal of Memory and Language* 52, 436–459.

Mak, Willem M., Vonk, Wietske and Schriefers, Herbert. 2008. Discourse structure and relative clause processing. *Memory & Cognition* 36, 170–181.

Marcus, Mitchell, Santorini, Beatrice and Marcinkiewicz, Mary Ann. 1993. Building a large annotated corpus of English: The Penn Treebank. *Computational Linguistics* 19, 313 – 330.

Matchin, William, Almeida, Diogo, Sprouse, Jon and Hickok, Gregory. 2018. Subject island violations involve increased semantic processing, but not increased verbal working memory resources: evidence from fMRI, poster presented at the 18th Conference on Human Sentence Processing, UC Davies, CA.

McCawley, James D. 1981. The syntax and semantics of English relative clauses. *Lingua* 53, 99–149.

Melnick, Robin, Jaeger, T. Florian and Wasow, Thomas. 2011. Speakers employ fine-grained probabilistic knowledge, 85th Annual Meeting of the Linguistic Society of America. Pittsburgh, PA.

Menn, Lise. 1974. Assertions not made by the main clause of a sentence. *Studies in the Linguistic Sciences (University of Illinois)* 4(1), 132–143.

Michaelis, Laura and Francis, Hartwell. 2007. Lexical subjects and the conflation strategy. In N. Hedberg and R. Zacharski (eds.), *Topics in the Grammar-Pragmatics Interface: Papers in honor of Jeanette K. Gundel*, Amsterdam: John Benjamins.

Michel, Daniel. 2014. *Individual Cognitive Measures and Working Memory Accounts of Syntactic Island Phenomena*. Doctoral dissertation, University of California, San Diego, CA.

Müller, Gereon. 2011. *Constraints on Displacement. A phase-based approach*, volume 7 of *Language Faulty and Beyond*. Amsterdam: Benjamins.

Munro, Robert, Bethard, Steven, Kuperman, Victor, Lai, Vicky, Melnick, Robin, Potts, Christopher C., Schnoebelen, Tyler and Tily, Harry. 2010. Crowdsourcing and language studies: the new generation of linguistic data. In *Proceedings of the NAACL-2010 Workshop on Creating Speech and Language Data with Amazons Mechanical Turk*, pages 122–130.

Ni, Weijia, Crain, Stephen and Shankweiler, Donald. 1996. Sidestepping garden paths: Assessing the contributions of syntax, semantics and plausibility in resolving ambiguities. *Language & Cognitive Processes* 11(3), 283–334.

Nunes, Jairo. 2004. *Linearization of Chains and Sideward Movement*. Cambridge, Mass.: The MIT Press.

Nunes, Jairo and Uriagereka, Juan. 2000. Cyclicity and extraction domains. *Syntax* 3, 20–43.

Petten, Cyma Van and Kutas, Marta. 1990. Interactions between sentence context and word frequency in event-related brain potentials. *Memory & Cognition* 18(4), 380–393.

Phillips, Colin. 2006. The real-time status of island phenomena. *Language* 82, 795–823.

Phillips, Colin. 2013. Some arguments and non-arguments for reductionist accounts of syntactic phenomena. *Language and Cognitive Processes* 28, 156–187.

Phillips, Colin, Kazanina, Nina and Abada, Shani H. 2005. ERP effects of the processing of syntactic long-distance dependencies. *Cognitive Brain Research* 22(3), 407 – 428.

Phillips, Colin, Wagers, Matthew W. and Lau, Ellen F. 2011. Grammatical Illusions and Selective Fallibility in Real-Time Language Comprehension. In Jeffrey Runner (ed.), *Experiments at the Interfaces. Syntax & Semantics, vol. 37*, pages 147–180, Emerald Group Publishing Limited.

Pickering, Martin and Barry, Guy. 1991. Sentence Processing Without Empty Categories. *Language and Cognitive Processes* 6, 229–259.

Pickering, Martin J., Barton, Stephen and Shillcock, Richard. 1994. Unbounded dependencies, island constraints and processing complexity. In Charles E. Jr. Clifton, Lyn Frazier and Keith Rayner (eds.), *Perspectives on sentence processing*, pages 199 – 224, London: Lawrence Erlbaum.

Polinsky, Maria, Gallo, Carlos Gómez, Graff, Peter, Kravtchenko, Ekaterina, Morgan, Adam Milton and Sturgeon, Anne. 2013. Subject islands are different. In Jon Sprouse and Norbert Hornstein (eds.), *Experimental Syntax and islands effects*, Cambridge: Cambridge University Press.

Pollard, Carl and Sag, Ivan A. 1994. *Head-driven phrase structure grammar*. Chicago: University of Chicago Press and Stanford: CSLI.

Rackowski, Andrea and Richards, Norvin. 2005. Phase Edge and Extraction. *Linguistic Inquiry* 36, 565–599.

Rizzi, Luigi. 2007. On Some Properties of Criterial Freezing. In *CISCL Working Papers on Language and Cognition 1*, pages 145–158.

Roark, Brian, Bachrach, Asaf, Cardenas, Carlos and Pallier, Christopher. 2009. Deriving lexical and syntactic expectation-based measures for psycholinguistic modeling via incremental top-down parsing. In *Proceedings of the 2009 conference on empirical methods in natural language processing*, pages 324–333.

Roland, Douglas, Mauner, Gail, O'Meara, Carolyn and Yun, Hongoak. 2012. Discourse expectations and relative clause processing. *Journal of Memory and Language* 66, 479 – 508.

Ross, John R. 1967. *Constraints on Variables in Syntax*. Ph.d. dissertation, MIT, Cambridge, Massachusetts. [Published in 1986 as *Infinite Syntax!* Norwood, NJ: Ablex Publishing].

Sabel, Joachim. 2002. A minimalist analysis of syntactic islands. *The Linguistic Review* 19, 271 – 315.

Sag, Ivan A., Hofmeister, Philip and Snider, Neal. 2007. Processing complexity in subjacency violations: the complex noun phrase constraint. In *Proceedings of the 43rd Annual Meeting of the Chicago Linguistic Society. Chicago: CLS*.

Santorini, Beatrice. 2007. (Un?)expected movement, University of Pennsylvania. `http://www.ling.upenn.edu/~beatrice/examples/movement.html`. Accessed: Jan 2 2012.

Shimojo, Mitsuaki. 2002. Functional theories of island phenomena: The case of Japanese. *Studies in Language* 26, 67–123.

Smith, Nathaniel J. and Levy, Roger. 2008. Optimal processing times in reading: A formal model and empirical investigation. In B. C. Love, K. McRae and V. M. Sloutsky (eds.), *Proceedings of the thirtieth annual conference of the Cognitive Science Society*, pages 595–600.

Snyder, William. 2000. An experimental investigation of syntactic satiation effects. *Linguistic Inquiry* 31, 575–582.

Snyder, William. 2017. On the nature of syntactic satiation. Unpublished ms.

Sprouse, Jon. 2007. Continuous Acceptability, Categorical Grammaticality, and Experimental Syntax. *Biolinguistics* 1, 118 – 129.

Sprouse, Jon. 2009. Revisiting Satiation: Evidence for an equalization response strategy. *Linguistic Inquiry* 40(1), 329 – 341.

Sprouse, Jon. 2011. A validation of Amazon Mechanical Turk for the collection of acceptability judgments in linguistic theory. *Behavior Research Methods* 43, 155–167.

Sprouse, Jon, Caponigro, Ivano, Greco, Ciro and Cecchetto, Carlo. 2015. Experimental syntax and the variation of island effects in English and Italian. *Natural Language & Linguistic Theory* 34(1), 307–344.

Staub, Adrian and Clifton, Charles E. Jr. 2006. Syntactic prediction in language comprehension: Evidence from *either... or*. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 32(2), 425–436.

Stepanov, Arthur. 2007. The end of CED? Minimalism and extraction domains. *Syntax* 10(1), 80–126.

Stowe, Laurie A. 1986. Parsing WH-constructions: Evidence for on-line gap location. *Language and Cognitive Processes* 1, 227–245.

Stowe, Laurie A., Tanenhaus, Michael K. and Carlson, Gregory N. 1991. Filling gaps online - use of lexical and semantic information in sentence processing. *Language and speech* 34, 319–340.

Sussman, Rachel S. and Sedivy, Julie C. 2003. The time-course of processing syntactic dependencies: Evidence from eye movements during spoken narratives. *Language and Cognitive Processes* 18, 143–163.

Tabor, Whitney, Juliano, C and Tanenhaus, MK. 1997. Parsing in a dynamical system: An attractor-based account of the interaction of lexical and structural constraints in sentence processing. *Language and Cognitive Processes* 2–3(12), 211–271.

Takahashi, Daiko. 1994. *Minimality of movement*. Phd thesis, University of Connecticut.

Takami, Ken-ichi. 1992. *Preposition stranding: from syntactic to functional analyses*. Berlin: Mouton de Gruyter.

Tollan, Rebecca and Heller, Daphna. 2016. Elvis Presley on an island: *wh* dependency formation inside complex NPs. In *Proceedings of NELS 46*, page pp2.

Trueswell, John C. and Tanenhaus, Michael K. 1994. Semantic Influences on Parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language* 33(3), 258–318.

Tsiamtsiouris, Jim and Cairns, Helen Smith. 2009. Effects of syntactic complexity and sentence-structure priming on speech initiation time in adults who stutter. *Journal of Speech, Language, and Hearing Research* 52(6), 1623–1639.

van Schijndel, Marten, Schuler, William and Culicover, Peter W. 2014. Frequency Effects in the Processing of Unbounded Dependencies. In *Proceedings of CogSci 2014*, Quebec, Quebec, Canada: Cognitive Science Society.

Van Valin, Robert D. Jr. 1986. Pragmatics, island phenomena, and linguistic competence. In *Papers from the Parasession on Pragmatics and Grammatical Theory*, volume 22(2), pages 223 – 233, CLS.

Van Valin, Robert D. Jr. 1995. Toward a functionalist account of so-called extraction constraints. In B. Devriendt, L Goossens and J. van der Auwera (eds.), *Complex structures: A functionalist perspective*, pages 26–60, Berlin: Mouton de Gruyter.

Wagers, Matt and Phillips, Colin. 2009. Multiple dependencies and the role of the grammar in real-time comprehension. *Journal of Linguistics* 45, 395–433.

Wanner, Eric and Maratsos, Michael. 1978. An ATN approach to comprehension. In Morris Hale, Joan Bres-

nan and George A. Miller (eds.), *Linguistic theory and psychological reality*, pages 119–161, Cambridge, MA: MIT Press.

Warren, Tessa and Gibson, Edward. 2002. The influence of referential processing on sentence complexity. *Cognition* 85, 79–112.

Warren, Tessa and Gibson, Edward. 2005. Effects of NP type and in reading cleft sentences in English. *Language and Cognitive Processes* 20, 751–767.

Wilcox, Rand. 2005. *Introduction to Robust Estimation and Hypothesis Testing*. San Francisco: Elsevier.