Rui P. Chaves

Freezing as a probabilistic phenomenon

Abstract: This paper shows that freezing effects are graded rather than categorical, and that different kinds of freezing are not equally strong. Building on Hofmeister et al. (2015: 470), I argue that freezing effects are at least in part caused by their extremely unusual structure, with two disparate *foci* governed by the same verb. By being inconsistent with comprehenders' expectations about the distribution of gaps, such constructions likely create a processing conflict between what is expected and the actual input. Experiment 1 suggests that such expectations are malleable, given that the oddness of extracting from an extraposed phrase disappears by virtue of making such constructions as likely as their non-extraposed counterparts. Experiments 2 and 3 suggest that the oddness created by crossing extraposition and extraction paths also disappears, but at a much lower rate. I propose that the latter constructions are more improbable and therefore worse than the former because (a) they are preempted by simpler and more likely alternative (local) parses (Fodor 1978) in which the point of retrieval and integration does not coincide with the point of reanalysis (Hofmeister et al. 2015), (b) involve crossing non-local dependencies (which are independently known to be more difficult than non-crossing dependencies (Fodor 1978), and therefore bound to be rarer), and (c) have disparate *foci* and therefore atypical pragmatic requirements (Huck and Na 1990; Bolinger 1992).

Keywords: Extraposition, extraction, islands, frequency, adaptation

1 Introduction

Ross (1967: 305) first noticed that leftward extraction (1a) and extraposition (1b) cause low acceptability when they interact, as seen in (2), a phenomenon known as *freezing*. In (2a) there is extraction from an extraposed PP, in (2b) there is extraction from an extraposed NP, and in (2c) an extraction from a PP crossed with direct object extraposition.

(1) a. Who$_j$ did you [give [a picture of _$_j$] [to Robin]]?
b. Did you [give _$_i$ [to Robin] [a picture of my brother]$_i$]?

Rui P. Chaves: Linguistics Department, University at Buffalo, SUNY

(2) a.*Who$_j$ did you [give a picture _$_i$] [to Robin] [of _$_j$]$_i$?
b.*Who$_j$ did you [give _$_i$ [to John] [a picture of _$_j$]$_i$]?
c.*Who$_j$ did you [give _$_i$ [to _$_j$] [a picture of my brother]$_i$]?

Most accounts of freezing involve theory-dependent assumptions (e.g. Wexler and Culicover 1980; Takahashi 1994; Rizzi 2007, among others) which usually predict that any movement out of *any* moved phrase is necessarily illicit. This is not quite so. For example, (3a) shows that a clause moved rightwards does not block leftward sub-extraction, (3b) exhibits complement extraposition from an extracted *wh*-phrase, and (3c) contains extraction from an extracted *wh*-phrase. It remains unclear how the full range of facts can be explained via syntactic constraints, without stipulation.

(3) a. What$_i$ did he explain _$_j$ to Mary [that she should write _ $_i$]$_j$?
b. [How many videos _$_i$]$_j$ are there _$_j$ on the web [of Mitt Romney getting booed]$_i$?
c. [Which handout]$_j$ can't you remember [how many COPIES OF _ $_j$]$_i$ you have to print _ $_i$?

The structure of the paper is as follows. Section 2 discusses non-configurational accounts freezing phenomena, and provides evidence that freezing effects like (2a) are created by independently motivated phonological phrasing constraints, following Huck and Na (1990). Freezing effects like (2b,c) require a different explanation, however. Section 3 describes three sentence acceptability experiments which show that, although the acceptability of such freezing constructions are initially low, speakers gradually come to regard such constructions as being significantly more acceptable. These results suggest that freezing effects are transient and at least in part plausibly caused by the fact that constructions with multiple *foci* are extremely unusual. By making such constructions more frequent, comprehenders can adapt their expectations and come to regard freezing violations as less severe. The reported experiments also show that the amelioration rates caused by increased frequency are much more robust for (2b) than for (2c). Finally, Section 4 describes a probabilistic model of the behavioral phenomena.

2 Non-syntactic accounts

Huck and Na (1990) offer a simple explanation for why sentences like (2a) have low acceptability. They begin by noting that such examples improve with contrastive stress and appropriate contextualization, as in (4).

(4) a. Okay, you saw a picture yesterday, but just whom did you see a picture yesterday OF _ ?
b. Here's an article in the Tribune by Trevor, of all people; he's someone I'd expect to read a story in the paper ABOUT _ .
c. I know Alger found letters in the file TO Chambers, certainly, but I'm not sure I can remember whom he found letters in the files FROM _ . (Huck and Na 1990: 66)

Huck and Na (1990) argue that this kind of amelioration is analogous to the one observed in pronouns (Zwicky 1982, 1986), as illustrated in (5). These examples show that unstressed pronouns like *it* and *me* cannot be separated from the verb, because the former are unable to project their own phonological phrase.

(5) a. Mia told to Noel *[it] / [that joke] yesterday.
b. Mia told [it] / [that joke] to Noel yesterday.
c. She called up *[me] / [the janitor] about the fire.
d. She called [me] / [the janitor] up about the fire.

For Huck and Na (1990), the same is true for unstressed prepositions: when an unstressed stranded proposition is separated from its selecting head by another phrase, oddness ensues for prosodic reasons. Thus, the oddness of (2a) is due to independently motivated phonological phrasing constraints rather than syntax proper. Apparent counterexamples like (6a), where the stranded P is itself fronted, and no amount of stress can ameliorate the sentence, pose no problem. Note that *to* in (6a) is an argument marking preposition, and therefore it is semantically defective. As such, it cannot be contrasted with anything or instantiate the required pragmatic function of topicalized expressions. Note that acceptability does in fact improve if the stranded preposition is semantically richer, as seen in (6b).

(6) a.*Who$_i$ do you think that [TO/to $\_{}_i$]$_j$, John gave a book $\_{}_j$. (Postal 1972)
b. This is [a bridge]$_i$ that I think [UNDERNEATH $\_{}_i$]$_j$, Robin would never park $\_{}_j$.

In the remainder of this paper I focus on (2b,c), which cannot be accounted for by prosodic phrasing alone. Fodor (1978: 457) notes that (2c) has a syntactically highly probable temporary alternative parse in which *to* combines with the NP *a picture of my brother*. The existence of this local ambiguity likely disrupts parsing, especially as it occurs in a portion of the sentence that contains two gaps in close succession. Indeed, constructions with two independent gaps in close proximity are licit, but not trivial to process, as seen in (7), specially if the extraction paths cross (Fodor 1978), as in (7b).

(7) a. This is a problem which$_i$ John$_j$ is difficult to talk to _$_j$ about _$_i$.
b. Who$_j$ can't you remember which papers$_i$ you sent copies of _$_i$ to _$_j$?

Hofmeister et al. (2015: 477) make a similar observation, by noting that constructions like (2c) must cause increased processing effort since the point of retrieval and integration coincides with the point of reanalysis. The existence of a preferential alternative parse that is locally licit but globally illicit can in turn lead to a 'digging-in' effect (Ferreira and Henderson 1991, 1993; Tabor and Hutchins 2004), in which the more committed the parser becomes to a syntactic parse, the harder it is to backtrack and reanalyze the input. The net effect of these factors is that the correct parse of (2c) is less probable and therefore harder to identify than that of (2b), which suffers from none of these problems, and is regarded to be more acceptable than (2c) by Fodor (1978: 453) and others.[1]

Finally, there may also be problems caused by the use of preposition stranding. Gries (2004) provides corpus data suggesting that P stranding tends to be used when the processing cost of the utterance is not already high, whereas PP extraction tends to be used otherwise. For example, it should be easier to process PP extractions and NP extractions because in the former there is more information about the syntactic function of the fronted phrase, by virtue of the presence of the preposition. In contrast, in NP extraction the fronted phrase can have virtually any syntactic function. Indeed, the observation that PP extraction from Subject Islands is more acceptable than NP extraction goes back to Ross (1967). Given that (2b,c) involve two different types of displacement, that would favor the use of PP extraction, but this would require using the pronoun

1 The *Clause Non-Final Incomplete Constituent Constraint* (Kuno 1973: 130) prohibited extraction out of phrases in a clause non-final position, and may be best explained by essentially the same kinds of garden-path-like processing problems created by a mid-sentence gap, specially given the number of counterexamples found in the literature (Jackendoff and Culicover 1972; Hukari and Levine 1991; Fodor 1992).

whom, which many speakers generally deem to be marked or awkward. Register effects and prescriptive pressures may further complicate matters.

Finally, Huck and Na (1990) and Bolinger (1992) also conjecture that freezing effects may be in part due to a pragmatic conflict created by extraposition and extraction: *wh*-movement has extracted a phrase leftward, focusing interest on that expression, while at the same time extraposition has moved a constituent rightward, focusing interest on that constituent as well. Note also that objects tend to be extraposed when they are discourse new, and even more so when they are heavy (Wasow 2002: 71). Therefore, the theme phrase *a picture of John* in (2c) is strongly biased to be discourse new, but this clashes with the fact that an entirely different entity, the recipient, is leftward extracted, and therefore is the *de facto* new information that the open proposition is about. No such mismatch exists in (2a) or (2b), in contrast, where the extraposed theme is more directly linked to the entity targeted by leftward extraction.

In this work I examine the possibility that the extreme low frequency of freezing constructions plays a role in the freezing effects in (2b,c), drawing inspiration from Hofmeister et al. (2015: 470), which conjectures that extraposed constituents are more difficult to process when they appear in environments that are more unexpected. Levy et al. (2012) provides evidence that the processing difficulty associated with extraposed relatives depends on how expected such constructions are: sentences in which a relative clause is highly expected facilitates the comprehension of an extraposed relative. In addition, there is also broader evidence that processing difficulty is affected by expectations about lexical (Kutas and Hillyard 1984; Altmann and Kamide 1999; Metzing and Brennan 2003; DeLong et al. 2005; Creel et al. 2008; Arai and Keller 2013; DeLong et al. 2005; Van Berkum et al. 2005; Gibson 2006; Kutas and Federmeier 2011; Levy and Keller 2013), syntactic (Demberg and Keller 2008; Ferreira and Clifton 1986; McRae et al. 1998), semantic (Federmeier and Kutas 1999; Altmann and Kamide 1999; Kamide et al. 2003), and pragmatic (Ni et al. 1996) information. By using their statistical experience with their language to generate predictions about upcoming syntactic structure, comprehenders are able to efficiently and robustly process complex and ambiguous linguistic input. As an analogy, consider the garden-path sentence (8), which contains a temporary ambiguity between a grammatical parse in which *plans* is a noun, and an ungrammatical parse in which *plans* is a verb. Since the sequence 'N $plans_V$ to' is many orders of magnitude more frequent than 'N $plans_N$ to', the verbal parse is preferred.[2]

(8) The government plans to raise taxes were defeated.

Crucially, the nominal parse is not only much more (locally) likely than the verbal parse, it also does not create any syntactic problems until the end the sentence, which makes the source of the misparse difficult to identify. Hence, the grammatical parse is likely to be preempted. In this work I explore the possibility that freezing effects are similarly sensitive to probabilistic information.

3 Experimental evidence

In what follows I provide experimental evidence that increased frequency of freezing violations can ameliorate freezing effects. These results are related to what Hiramatsu (2000), Francom (2009), Kravtchenko et al. (2009), and Chaves and Dery (2014), found for Subject Island violations, usually regarded as another type of freezing effect. If sentences with freezing island violations were ungrammatical, and if grammars cannot construct representations for ungrammatical representations – as explicitly argued by Phillips (2006: 803), Sprouse (2007), Wagers and Phillips (2009), and Phillips (2013) – then it should be impossible to improve their acceptability by simply increasing their frequency. If the language processor cannot construct ungrammatical structures, then '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' (Sprouse 2007: 123).

3.1 Experiment 1

This experiment focuses on classic freezing effects like (9b,d,f), and shows that the oddness attributed to such constructions can be made to vanish by virtue of repeated exposition. In other words, by making (9a,c,e) as frequent as (9b,d,f) the difference in acceptability steadily disappears.

2 For example, in the 520,000,000 word Corpus of Contemporary American English (COCA; Davies, 2008-), the sequence '`[nn*] plans.[v*] to`' has a frequency of 1744 whereas the sequence '`[nn*] plans.[nn*] to`' is much rarer, with a frequency of 8.

(9) a. Who_i did you give [a picture of $\_{}_i$] to Robin?
b.*Who_i did you give $\_{}_j$ to Robin [a picture of $\_{}_i]_j$?
(Wexler and Culicover 1980)
c. Who_i did you see [some beautiful pictures of $\_{}_i$] yesterday?
d.*Who_i did you see $\_{}_j$ yesterday [some beautiful pictures of $\_{}_i]_j$?
(Johnson 1985: 74)
e. What_i did you give [a book about $\_{}_i$] to John?
f.*What_i did you give $\_{}_j$ to John [a book about $\_{}_i]_j$?
(Lasnik and Saito 1992: 103)

3.1.1 Methods

A total of 70 self-reported native English speakers with IP addresses originating from the United States were recruited through Amazon.com's Mechanical Turk (AMT) crowdsourcing marketplace.[3] The task was open to anyone, but participants were asked to report if they happened to be native speakers. Participants were instructed to rate how natural each sentence was, by giving it a number from 1 (very unnatural) to 7 (very natural). The experimental items consisted of 20 pairs of sentences, a sample of which is shown in (10). The full set of items is in Appendix A.

(10) a. Which cake did you serve Mark several slices of? (*In situ*)
b. Which cake did you serve to Mark several slices of? (*Ex situ*)
c. Which problem did you write Robin several emails about? (*In situ*)
d. Which problem did you write to Robin several emails about? (*Ex situ*)

In the *in situ* (control) condition, the verb phrase consists of a double object construction where the verb is followed by the direct object and then followed by the sentence-final indirect object containing the gap (e.g. [*serve* [*Mark*] [*several slices of* $\_{}_i$]]). In the *ex situ* condition, however, the verb phrase consists of a prepositional indirect-object construction in which the object containing the gap is extraposed over the oblique (e.g. [*serve* $\_{}_j$ [*to Mark*] [*several slices of* $\_{}_i]_j$]), constituting a freezing violation analogous to (9d). Hence, the acceptability of the latter should be consistently lower than that of their *in situ* counterparts.

3 For evidence that linguistic data obtained via AMT parallels data obtained in the laboratory see Melnick et al. (2011), Gibson et al. (2011), and Sprouse (2011).

The experimental items were pseudorandomized and counterbalanced across two lists using a Latin Square design, and interspersed with 40 distractor sentences. This way, no participant saw both versions of the same item, and no two participants saw the items in the same order. There were six kinds of distractor sentence, exemplified in (11), half of which were ungrammatical. In the actual stimuli, no diacritic '*' was present, of course.

(11) a. Which library did you donate several books about Oprah to?
b. Which houses did the roofs of get damaged by the explosion?
c. Which individual did you see Robert with?
d.*Which school did you compare Ann to attend some classes at?
e.*Which restaurant did you order multiple dishes to Larry about?
f.*Which windows did you strain James of?

3.1.2 Results

The mean acceptability rating for the *in situ* condition was 5.76 (SD = 1.35), and for the *ex situ* condition was 4.51 (SD = 1.73). The mean rating for the grammatical distractors was 5.54 (SD = 1.77), and for the ungrammatical distractors it was 2.26 (SD = 1.49).

Linear mixed-effect regression (LMER) models were fitted using the `lme4` package (version 1.1–7) (Bates et al. 2014) in `R` (version 3.1.2), and the intercept was adjusted by item type, subjects, and lists, in order to account for random effects. An LMER model with sentence type and presentation order as fixed factors and items, subjects and lists as random factors confirmed that *in situ* items were rated significantly higher than *ex situ* items ($t = 9.848$, $p < 0.001$), and revealed that acceptability ratings improved as the experiment progressed ($t = 5.621$, $p < 0.001$). No interactions were detected between item type and presentation order ($t = -0.947, p = 0.34$). LMER models for each item type further confirmed that both *in situ* items ($t = 4.85, p < 0.001$) and *ex situ* items ($t = 5.38, p < 0.001$) improved with presentation order. Separate models just for distractors detected no change in acceptability as the experiment progressed for ungrammatical distractors ($t = 1.07$, $p = 0.28$), but an increase of acceptability occurred in the case of grammatical distractors ($t = 4.77$, $p < 0.001$). Figure 1 depicts the change in acceptability as a function of presentation order for all stimuli. Since no two participants saw the experimental items in the same order, each dot corresponds to the average across different items in each condition. Moreover, the increase in acceptability judgements during the experiment cannot be attributed to the particular order in which the items were presented.

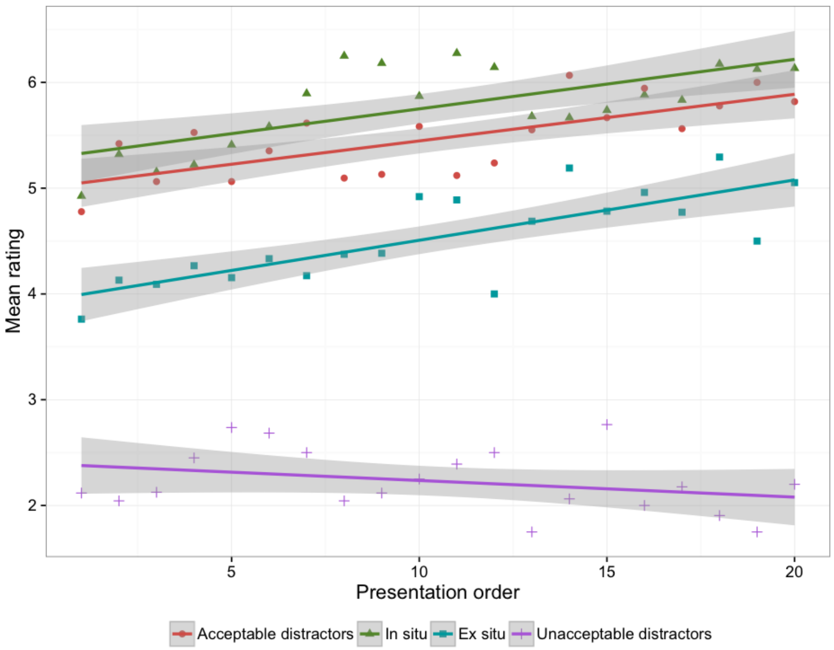

Fig. 1. LMER of each item type with presentation order as a fixed predictor

Next, the acceptability ratings of 60 responses in the *in situ* condition with presentation order of 1, 2, and 3 were compared against the acceptability ratings of 60 responses in the *ex situ* condition with presentation orders of 18, 19, and 20. A linear mixed-effect regression model with sentence type as a fixed predictor (allowing the intercept to be adjusted by items, subjects, and lists) revealed that the acceptability of the *in situ* items in the beginning of the experiment was not significantly different from the acceptability of the *ex situ* items by the end of the experiment ($t = 1.123$, $p = 0.2633$).

3.1.3 Discussion

The experimental findings suggest that the acceptability of freezing violations standardly regarded as illicit (Wexler and Culicover 1980; Johnson 1985; Lasnik and Saito 1992) disappears by merely increasing their frequency. The gradual increase of acceptability over the course of only 10 sentences in the *ex situ* condition may reflect comprehender's adaptation to highly unusual constructions, by prompting them to revise their expectations about the syntactic distribution of English gaps. Given enough exposure, the results suggest that both conditions would eventually converge towards the upper end of the scale, and become equally likely and acceptable.

The present interpretation of the behavioral evidence is not unlike that of Kim et al. (2011), for which the acceptability cline in category mismatches in Verb Phrase Ellipsis (VPE) constructions is argued to correlate with how consistent the prior heuristics are with the particular input: highly acceptable VPE constructions are those consistent with comprehenders' parsing heuristics, and have syntactic analyses that are found comparatively early during sentence comprehension. Conversely, less acceptable VPE constructions involve more processing as well as more processing conflicts because they are inconsistent with heuristics that normally aid sentence processing. According to surprisal theory (Hale 2001; Levy 2008), more cognitive effort is required to process input that is less expected. The relation between surprisal and processing effort has been experimentally validated (Boston et al. 2008; Demberg and Keller 2008; Roark et al. 2009; Smith and Levy 2008), as has the correlation between sentence acceptability and probability (Keller 2003; Lau et al. 2015; Manning 2003).

3.2 Experiment 2

Experiment 2 focuses on freezing effects like those in (12), where the extraposition of the direct object putatively blocks the leftward extraction from the oblique complement. If the oddness of (12b) and (12d) is due to grammar, then their acceptability should not improve with repeated exposition, let alone approach the acceptability of their *in situ* counterparts in (12a,c).

(12) a. Who$_i$ does this sentence appear to be fine to _$_i$?
b.*Who$_i$ does this sentence appear _$_j$ to _$_i$ [to be fine]$_j$?
(Langendoen and Pullum 1977)
c. Who$_i$ did you give [a picture of Sandy] to _$_i$?
d.*Who$_i$ did you give _$_j$ to _$_i$ [a picture of Sandy]$_j$?
(Wexler and Culicover 1980)

As we shall see, the acceptability of sentences like (12d) improves only marginally with repeated exposure, suggesting that the freezing effects examined in Experiment 2 (e.g. **Who did you give to a book about Sandy?*) are more severe than those examined in Experiment 1 (e.g. **Who did you give a book about to John?*), and therefore unlikely to be cause by the exact same factors.

3.2.1 Methods

A set of different 76 participants were recruited through AMT, using the same methodology as in Experiment 1. The experimental items consisted of 20 pairs of sentences listed in Appendix B, a sample of which is given in (13). In the *in situ* condition, the gap is located immediately after a sentence-final stranded preposition, whereas in the *ex situ* condition the direct object has been extraposed over the stranded preposition. The latter is isomorphic to (12d).

(13) a. Who did you promise a sum of $1,000 to? (*In situ*)
b. Who did you promise to a sum of $1,000? (*Ex situ*)
c. Who did you forward a copy of the contract to? (*In situ*)
d. Who did you forward to a copy of the contract? (*Ex situ*)

As in Experiment 1, the item pairs were pseudorandomized, counterbalanced across two lists, and interspersed with 40 distractor sentences. There were various types of distractor sentence, some with *who/what* phrases, others containing *which*. As before, half of the distractors were grammatical, half were ungrammatical, illustrated in (14).

(14) a. What did you call back your boss about?
b. Which restaurant did you ask me to book a reservation for?
c. Who did you recommend several songs by Rihanna to?
d.*What did you look for a neighbor at?
e.*Which road did you drive me to go to various places with?
f.*Who did you visit several friends of mine?

3.2.2 Results

The mean acceptability of the *in situ* items was 6.42 (SD = 0.88) and the mean for *ex situ* items was 3.17 (SD = 1.64). An LMER model with item type and presentation order as fixed factors confirmed that sentences in the *in situ* condition were rated significantly higher than sentences in the *ex situ* condition ($t = 40.94$, $p < 0.001$), and revealed that acceptability ratings improved as the experiment progressed ($t = 2.73$, $p < 0.01$). No interactions were detected between item type and presentation order ($t = 0.19, p = 0.84$). LMER models for each item type further confirmed that both *in situ* items ($t = 3.39, p < 0.001$) and *ex situ* items ($t = 2.91, p < 0.01$) improved with presentation order, as in Figure 2. Again, ungrammatical distractors showed no improvement.

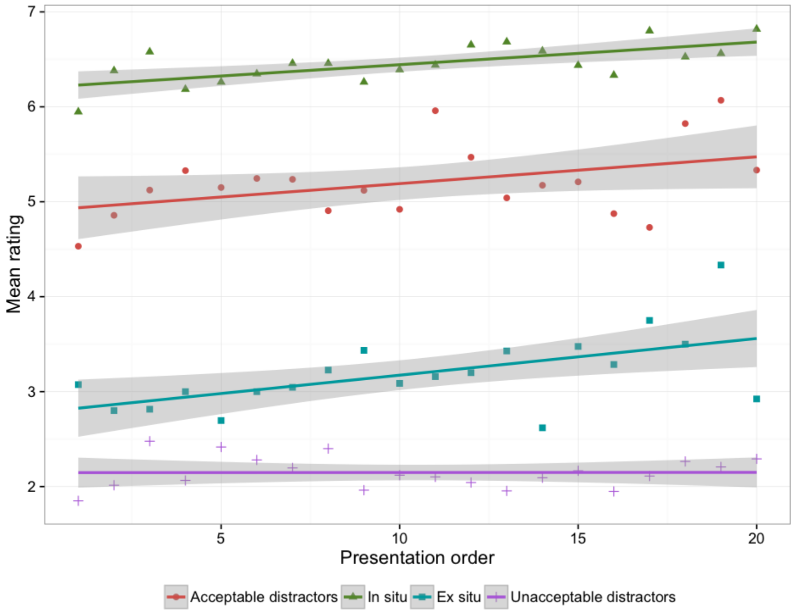

Fig. 2. LMER of each item type with presentation order as a fixed predictor

3.2.3 Discussion

The results of Experiment 2 suggest that this type of freezing effect is difficult to ameliorate with frequency alone. Any theory of freezing effects should explain this fact. The mean acceptability of the freezing violation items approaches the middle of the scale, presumably where the threshold for acceptability lies, but never quite crosses it. As discussed in §2, it is possible that freezing violations like (12d) do not exhibit strong amelioration effects because they are more unlikely and unexpected. First, comprehenders may overlook the possibility that there is a gap immediately after the mid-sentence preposition because the preposition is adjacent to an NP. Second, extraposition and extraction target the same sentence region, which is dispreferred as it likely creates processing difficulty that hampers and preempts the correct parse. Third, extraposed phrases usually describe new information, but the extraction targets the non-extraposed phrase. This conflict created by extraposition and extraction may prevent speakers from viewing such constructions as fully felicitous, without a suitable contextualization for the double *foci*. Moreover, the discourse contexts in which such multiple *foci* would are felicitous are arguably rare, and therefore unlikely, all else being equal.

3.3 Experiment 3

If the sentence-medial stranded preposition in the *ex situ* condition in Experiment 2 mislead comprehenders into assuming that the preposition would combine with the following NP, then the low acceptability of the *ex situ* items in Experiment 2 might be in part due to difficulty detecting the correct grammatical parse. Experiment 3 examined if participants would rate *ex situ* items more favorably if a small pause were to be inserted at the gap site, thus more overtly indicating the location of the gap as in (15).

(15) Who did you owe to // a debt of millions of dollars?

Drawing from work on silent reading by Quinn et al. (2000), Fodor (2002) and various others, the items in Experiment 2 were modified to include prosodic cues that signaled the 'brief pause', in (15). The present experiment should be regarded as a preliminary study, since speech stimuli should ideally be use to probe the effect of a pause in sentence acceptability.

3.3.1 Methods

A different set of 76 participants from those that participated in the previous experiments was recruited through AMT. All enrollment criteria and participant instructions were identical to that of Experiment 1. The experimental items consisted of the same 20 pairs of items used in Experiment 2, counterbalanced across two lists, pseudorandomized, and mixed with distractors, exactly as in previous experiments. However, the items contained the symbol '❚❚' as shown in (16), which participants were told sigaled 'a brief pause'.

(16) a. Who ❚❚ did you promise a sum of $1,000 to? (*In situ*)
b. Who did you promise to ❚❚ a sum of $1,000? (*Ex situ*)

The break in (16a) is grammatical, though somewhat marked. The goal of this manipulation is to allow *ex situ* items to benefit from some prosodic information, and to make *in situ* items less prototypical. Hence, the acceptability between *in situ* items and *ex situ* items should be less extreme than in Experiment 2.

3.3.2 Results

The mean acceptability for the *in situ* condition was 5.06 (SD = 1.86), and for the *ex situ* condition 3.58 (SD = 1.73). The means for the ungrammatical and grammatical distractors were 2.77 (SD = 1.72) and 4.38 (SD = 1.97) respectively. An LMER model with sentence type and presentation order as fixed factors, and items, subjects and lists as random factors confirmed that *in situ* items were rated significantly higher than items in the *ex situ* condition ($t = 28.08$, $p < 0.001$), and revealed that acceptability ratings improved as the experiment progressed ($t = 2.4$, $p = 0.01$). No interactions were detected between item type and presentation order ($t = -0.734, p = 0.46$). Contrary to Experiment 2, however, separate LMER analysis revealed that *ex situ* items improved faster during the experiment ($t = 4.37, p < 0.001$) than *in situ* items did ($t = 3.17$, $p = 0.001$).

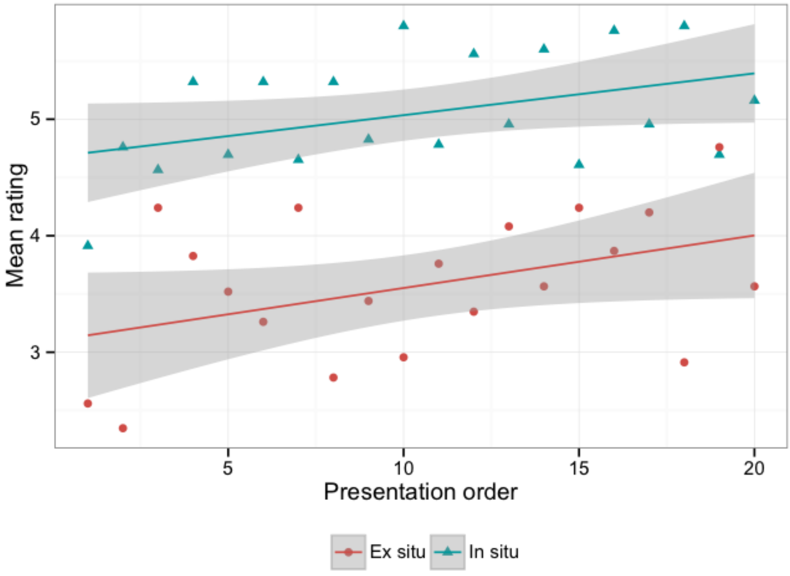

Fig. 3. LMER of each item type with presentation order as a fixed predictor

A regression model with sentence type as a fixed predictor revealed that the acceptability ratings of 48 responses in the *in situ* condition whose presentation order was either 1 or 2 was not significant from the acceptability ratings of 48 responses in the *ex situ* condition whose presentation orders were 19 or 20 ($t = -0.45$, $p = 0.64$), but broader selections of presentation order windows yield significant mean differences. For example, the acceptability ratings of 60 responses in the *in situ* condition whose presentation order was either 1, 2, or 3 was significantly different ($t = -2.07$, $p = 0.04$) from that of 60 responses in the *ex situ* condition whose presentation orders were 18, 19, or 20.

Although the acceptability of *ex situ* items by the end of the experiment only tended to approach the acceptability of the *in situ* items at the beginning of the experiment, it is nonetheless clear that the acceptability of *ex situ* items gradually approached and crossed the midpoint of the acceptability scale, which raises questions with regard to their ungrammaticality status.

3.3.3 Discussion

The results from the above three experiments indicate that freezing effects like those in (9), repeated below in (17a), are stronger and harder to ameliorate those in (12), repeated in (17b).

(17) a.*Who$_i$ did you give _$_j$ to _$_i$ [a picture of Sandy]$_j$?
b.*Who$_i$ did you give _$_j$ to Robin [a picture of _$_i$]$_j$?
(Wexler and Culicover 1980)

Although the prosodic cue had a mild ameliorative effect, and increase of acceptability was not as strong as the one observed in Experiment 1, the *ex situ* condition exhibited a stronger amelioration effect than the *in situ* condition. This is unexpected if such freezing violations are impossible for the parser to construct and prime. Secondly, it is also possible that if comprehenders had been exposed to more experimental items than the acceptability of (17a) would more clearly converge into the upper range of the scale. Thirdly, it is also possible that with appropriate contextualization, some of the oddness created by the presence of the double *foci* can be circumvented. Further research is necessary in order to more directly compare the acceptability of freezing violations with and without prosodic cues, using actual auditory stimuli. It is possible that the prosodic break at the gap site in the *ex situ* condition would have been more effective in circumventing the freezing effect if the experimental materials consisted of actual speech rather than written stimuli.

As discussed in §2, there are important differences between the two types of freezing violation in (17) that may explain the contrast in amelioration rates. In (17a) there is a highly likely alternative syntactic parse in which *to* combines with the following NP *a picture of Sandy*, thus creating additional processing difficulty, two gaps in close succession, crossing displacement dependencies, and an information structural conflict created by extraposition and extraction of completely different entities. As a consequence, comprehenders may be less likely to regard such constructions as fully felicitous, and more reluctant to revise their expectations about the distribution of extraction and extraposition.

4 A probabilistic model

The experimental results reported in this paper are consistent with the hypothesis that comprehenders make use of expectations about the syntactic distribution of filler-gap dependencies to efficiently prune the search space during gap detection, and mitigate the processing costs associated with resolving such dependencies during on-line sentence comprehension. The more syntactically, semantically, and pragmatically unlikely the position of a gap, the harder for the language processor to overcome the conflict between the expected structure and the actual input, and the harder it is to reanalyze the structure.

Indeed, freezing violations like (2) are not attested in corpora. No single occurrence is found in the 520,000,000 word COCA corpus, for example. As discussed in §2 such constructions may be odd not only because they involve two *foci* that require very peculiar contexts, but also because they contain temporary structural ambiguities which are more likely to be resolved incorrectly and to persist until the end of the sentence. The rarity of dual *foci* constructions causes them to be unexpected, and therefore more likely to be preempted by any extant (local) alternative parses. These factors plausibly conspire to hamper the production and comprehension of freezing constructions.

If sentence processing is guided by probabilistic information about the distribution of gaps, then it is also likely that such expectations are malleable and can be changed to reflect variations in the input, just like other types of syntactic expectation have been shown to be malleable (Fine et al. 2010; Kamide and Mitchell 1997; Fine and Jaeger 2013; Farmer et al. 2014). In ideal conditions, comprehenders can adapt their prior syntactic expectations to match those in the current context. In what follows I sketch a simple model that can predict the frequency-based behavioral effects detected by the experiments in §3.

Sentence processing proceeds incrementally, which means that as each word is processed, speakers determine the most likely syntactic structure given the current discourse context, and predict with varying degrees of certainty what the remainder of the sentence should consist of. For example, suppose the observed input is w_1 = *Lisa*, w_2 = *said*. The most likely partial tree t consistent with the input is $[_{\text{S}}[_{\text{NP}}\textit{Lisa}][_{\text{VP}}[_{\text{V}}\textit{said}][\,...\,]]]$, where the ellipsis indicates the node that upcoming input is expected to be. Only parse trees consistent with the input and the grammar are permitted, and among those, the parse tree with highest probability given by the grammar and the input are preferred to those with lower probability. We can define the most likely tree $\hat{t}$ for the input $w_1 ... w_n$ to be the most probable tree from the set of trees T_G that a grammar G licenses for the string $w_1 ... w_n$, as seen in (18).

(18) $\hat{t} = \arg \max_{t \in T_G(w_1 ... w_n)} P(t)$

There are many ways to estimate what the most probable tree is or what the most probable continuation is, and there are many ways to analyze extraction and extraposition. Before addressing the former problem, I'll discuss the latter. I start by assuming that each tree node is a set of attribute-value pairs $\{a_1 : v_1, \ldots, a_n : v_n\}$, where a is an attribute and v is its value, encoding part-of-speech, case, valence, etc. Simplifying somewhat the HPSG framework of Sag et al. (2003), a nominal like *Lisa* corresponds to (19), which I abbreviate as 'NP' for convenience.

(19) {PHON : /lisə/, POS : *n*, VALENCE : ⟨ ⟩, NUM : *sing*, PER : *3rd*, GEN : *fem*, CASE : *nom*, SLASH : {}, EXTRA : {}, SEMANTICS : *lisa*}

Crucially, I follow Kim and Sag (2005) and others in representing leftward displacement dependencies with the attribute SLASH, and rightward displacement dependencies with the attribute EXTRA. In this framework, any given local constituent tree is of the form $[^{\tau}_{\alpha}\ \beta_1 ... \beta_n]$ where τ is the type of construction, α is the mother and $\beta_1 ... \beta_n$ are its local daughters. For example, in Figure 4 the node that combines the subject and the verbal phrase is $[^{\textit{head-subj-cxt}}_{\mathrm{S}}\ \mathrm{NP}\ \mathrm{VP}]$. Here, attributes other than SLASH and EXTRA are omitted due to space limitations.[4]

The label VP is merely an abbreviation for a verbal expression with one element listed in VALENCE, and S is verbal constituent with no elements listed in VALENCE. For illustration, the V *gave* in Figure 4 is shown in full in 20a, the VP *gave a book* is shown in (20b), and the S node is in (20c). Phrase-structure rules are responsible for matching and ordering the elements in VALENCE with the

4 Although only SLASH and EXTRA are show in this discussion, the model formalized below takes into consideration *all* attribute-value pairs, from *all* nodes in the local tree, including semantics and phonology. Thus, information about phonological phrasing too can influence the model's behavior. Moreover, if we augment our attribute-value sets with an attribute dedicated to information structure, then information about *foci* can be factored in and potentially influence the probability of a given tree, in a given context.

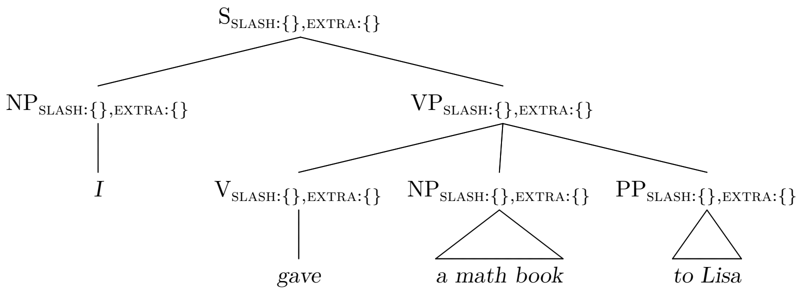

Fig. 4. A clause without extraposition or extraction

sisters of the head, combining phonological information, composing the semantic representations, and projecting the correct information to the mother node.

(20) a. {PHON : /geɪv/, POS : v, VALENCE : ⟨NP, NP, PP⟩, SLASH : {}, EXTRA : {}, SEMANTICS : $\lambda z.\lambda y.\lambda x.give'(x, y, z)$}

b. {PHON : /geɪv ʌ mæθ bʊk tu lisə/, POS : v, VALENCE : ⟨NP⟩, SLASH : {}, EXTRA : {}, SEMANTICS : $\lambda x.\exists y(book(y) \wedge about(y, math') \wedge give'(x, y, lisa'))$}

c. {PHON : /ɑɪ geɪv ʌ mæθ bʊk tu lisə/, POS : v, VALENCE : ⟨ ⟩, SLASH : {}, EXTRA:{}, SEMANTICS : $\exists y(book(y) \wedge about(y, math') \wedge give'(speaker', y, lisa'))$}

However, in a topicalization sentence like the one in Figure 5 the verbal nodes in the extraction path bear the specification [SLASH : {NP^y}], except when the extraction is terminated, at the matrix node $[^{head\text{-}filler\text{-}cxt}_{\mathrm{S}_{\text{SLASH}:\{\}}}\ \mathrm{NP}^y\ \mathrm{S}_{\text{SLASH}:\{\text{NP}^y\}}]$, where the first daughter is identified with the element in the SLASH value of the clause.

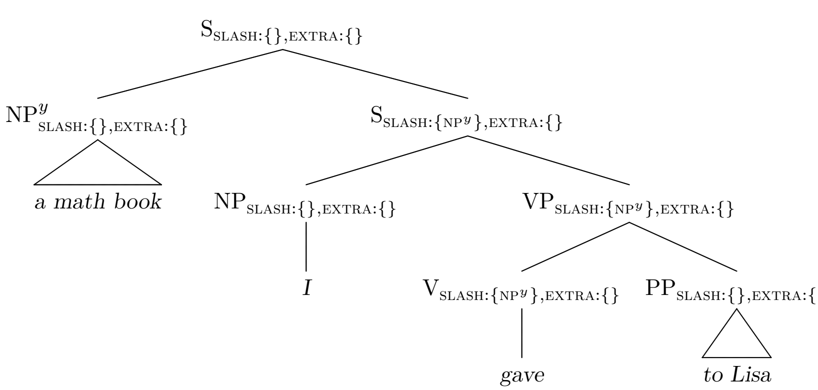

Fig. 5. A clause with extraction

Similarly, in an extraposition construction it is up to the attribute EXTRA to allow an object to be realized to the right of its canonical position. In the case seen in Figure 6 the node $\left[^{\textit{head-extraposed-cxt}}_{\mathrm{VP}_{\textsc{extra}:\{\}}}\ \mathrm{VP}_{\textsc{extra}:\{\textsc{np}^y\}}\ \mathrm{NP}^y\right]$ allows the the verb phrase to combine with the a right-dislocated object in EXTRA. When both extraposition and extraction occur in the same clause, both attributes SLASH and EXTRA have non-empty values which propagate in the structure and are linked to their filler phrases independently.

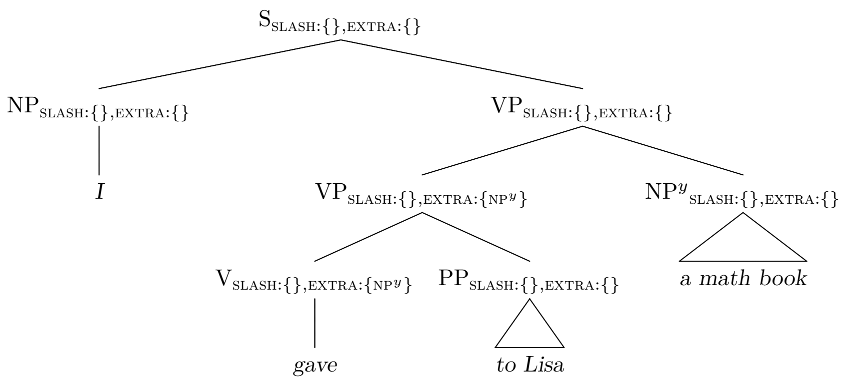

Fig. 6. A clause with extraposition

We can now return to (18), and define the probability of parse t as the product of the probability of the constituent trees c that t is composed of:

(21) $P(t) \;= \prod_{c \,\lhd\, t} P(c)$

Finally, the probability $P(c)$ of any given local tree c can be defined in terms of the probability of its features. As a consequence, the probability of the same verbal node having non-empty values for SLASH and EXTRA simultaneously (as in *Who did you give to a picture of Sandy?*) will be exceedingly low, given that these constructions do not occur. In contrast, the probability of a verb's direct object being extraposed (i.e. of bearing [SLASH : { }] and [EXTRA : {NP}]) and the probability of a verb's direct object containing a leftward extraction (i.e. of bearing [SLASH : {NP}] and [EXTRA : { }]) are not as low because these constructions do in fact occur, as the attestation samples in (22) and (23) show.

(22) a. Webb approved the sale _ to Iraq [of military transport helicopters] (...) (COCA: 1995 MAG)

b. And what was the value _ to Michelangelo [of being part of that]?
(COCA: 2008 SPOK)

c. Just two weeks ago, Britain stopped a shipment _ to Iraq [of devices that could be used to trigger nuclear weapons].
(COCA: 1990 SPOK)

(23) a. (...) this was something James didn't seem to have [a problem with _].
(COCA: 2007 FIC)

b. Others, we're going to have to find [some housing for _].
(COCA: 1994 SPOK)

c. There was one last question my editor was dying to know [the answer to _].
(COCA: 2004 NEWS)

Our statistical model should therefore predict that any given node has an extremely low probability of having non-empty values for SLASH and EXTRA simultaneously (simply because these do not occur) and that the independent probabilities of having non-empty values for SLASH and EXTRA are higher (given that the latter do in fact occur). Such a model should be flexible enough to allow the former to become more likely if non-empty values for SLASH and EXTRA are repeatedly observed simultaneously, for the same verb, as in the experiments reported above. As the increase in acceptability due to repeated exposure is linear, one plausible choice is the log-linear model in (24).[5]

(24) $$P([^{\tau}_{\alpha}\,\beta_1...\beta_n]) = \frac{\exp\left(\sum_{a:v} w_{a:v} \times f_{a:v}([^{\tau}_{\alpha}\,\beta_1...\beta_n])\right)}{\sum_{c\,\in\,T_\tau} \exp\left(\sum_{a:v} w_{a:v} \times f_{a:v}(c)\right)}$$

Here, $w_{a:v}$ is an $\mathbb{R}$-valued weight for a given attribute-value combination $a{:}v$ (including joints of attribute-value combinations, as described below), $f_{a:v}(c)$ is the indicator function yielding 1 if $a{:}v$ is present in the given local tree c and 0 otherwise, and T_τ is the set of local trees of type τ in the treebank. In HPSG, the values of SLASH and EXTRA can only be of one of two types: *empty-set* and *non-empty-set*. In in (24), the former is coded as 0 and the latter as 1.

For example, suppose that the local constituent tree c we are evaluating specifies that SLASH and EXTRA both have non-empty values. In that case, the model will take into consideration at least three weighted factors: one for non-

5 See Miyao and Tsujii (2002,2008) for more discussion about this class of models and their computational implementation. Alternatively, such a model may be recast in terms of Data-Oriented Parsing, such as Linadarki (2006).

empty SLASH, one for non-empty EXTRA and a third one for the joint non-empty pair SLASH and EXTRA. More formally, the numerator of (24) will be: $\exp(w_{\text{SLASH:1}} + w_{\text{EXTRA:1}} + w_{\text{SLASH:1\&EXTRA:1}} + \ldots)$, where the weights for attributes other than the mother's SLASH and EXTRA are not shown for space limitations. Although the two first weights have moderate positive values, the third is strongly negative (since such gap distributions are extremely rare), and therefore the model yields a minuscule non-zero probability for such gap distributions. Consequently, such local constituent trees will be heavily dispreferred if there are more likely competing alternatives, and very difficult to consider when comprehenders realize the parse is incorrect and attempt a reanalysis. However, if tree structures simultaneously containing non-empty values for SLASH and EXTRA are made more likely, then the weight $w_{\text{SLASH:1\&EXTRA:1}}$ increases, and so does the probability of the double-extraction analysis. Since the model is linear, a linear increase in the frequency of such structures leads to a linear increase in their probability. And as the latter parses becomes more likely, they require less cognitive effort to compute than before, and their acceptability improves (perhaps as a function of both the probability and the required processing effort).

If on the other hand the constituent tree c we are evaluating specifies that SLASH is empty but EXTRA is not, then the numerator of (24) will instead be $\exp(w_{\text{SLASH:0}} + w_{\text{EXTRA:1}} + w_{\text{SLASH:0\&EXTRA:1}} + \ldots)$, all of which are positive weights, as intended. Consequently, such structures will be more likely, and therefore easier to process, harder to preempt, and ultimately deemed more acceptable.

Consequently, frequency alone can significantly ameliorate freezing effects as long as there are no additional factors hampering or preempting the correct parse; for example, no additional processing difficulty incurred by crossing extraction and extraposition, converging on the same sentence region, no competition from highly likely and alternative parses caused by sentence-medial P stranding, and no informational-structural conflicts due to multiple *foci*.

It is possible that the kinds of expectations that we are concerned with here are straighforwardly created by extra-grammatical heuristic parsing rules drawn from frequently occurring patterns, deployed during on-line sentence comprehension. There is evidence that speakers resort to such heuristics. For example, given the lexical input '*who did …*' with a high-falling intonation, speakers of English create the expectation of an $\text{S}_{\text{SLASH:}\{\text{NP}^{y}\},\text{EXTRA:}\{\}}$ constituent in which y is co-referential with the *wh*-phrase. Often times this heuristic will aid language processor, by correctly pre-activating the right linguistic structures and rules, but sometimes it will not, as in continuations where there is no extraction whatsoever, e.g. '*who did that?*'. Although such heuristics involve linguistic information, they need not be part of the grammar *per se*. As Kroch (2001: 722) notes, "There is no doubt, however, that human beings like other animals, track the frequencies of events in their environment, including the frequency of linguistic events.” For other examples of similar heuristics likely deployed during sentence processing see Sag and Wasow (2011, 2015), including some that involve visual and gestural information, social knowledge, style, and genre. For an overview about how this kind of expectation can represented in a Bayesian framework see Manning (2003).

5 Conclusion

Speakers use probabilistic information as an heuristic to predict upcoming linguistic input and aid the processing of complex and ambiguous utterances. Such heuristics are adaptive in order to overcome variation and unexpected patterns. Hofmeister et al. (2015) propose that the unacceptability of freezing constructions is caused by their unanticipated syntactic structure, which creates processing difficulty. The experimental results reported in this paper are consistent with this hypothesis, since freezing effects can be ameliorated simply by making such structures more frequent. Following Fodor (1978), Huck and Na (1990), Bolinger (1992), Hofmeister et al. (2015), I argue that freezing effects are strongest when the correct syntactic analysis is extremely unlikely due to independent factors, such as the existence of alternative syntactic parses, crossing extraction pathways, which likely cause additional processing difficulty and make the correct parse construction unlikely. Extraposition freezing constructions are argued to be rare – and therefore unexpected – because two *foci* are governed by the same verb, and as such may only be pragmatically felicitous in peculiar contexts.

Acknowledgment: I am very grateful to the organizers of the Freezing workshop – Andreas Konietzko, Marion Knecht, Jutta Hartmann, and Susanne Winkler – as well as to the University of Tübingen, for their kind invitation and support. I am also very thankful to the workshop’s audience, two anonymous reviewers, and to both John Hale and John Pate for their comments and suggestions.

References

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

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

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

Bolinger, Dwight. 1992. The role of accent in extraposition and focus. *Studies in Language* 16. 265–324.

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

Chaves, Rui P. & Jeruen E. Dery. 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*, 96–106. Somerville, MA: Cascadilla Proceedings Project.

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

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

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

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

Farmer, Thomas A., Alex B. Fine, Shaorong Yan, Spyridoula Cheimariou & Florian T. Jaeger. 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. & Marta Kutas. 1999. A rose by any other name: Long-term memory structure and sentence processing. *Journal of Memory and Language* 41. 469–495.

Ferreira, Fernanda & Charles E. Jr. Clifton. 1986. The independence of syntactic processing. *Journal of Memory and Language* 25. 348–368.

Ferreira, Fernanda & John M. Henderson. 1991. Recovery from misanalyses of garden-path sentences. *Journal of Memory and Language* 31. 725–745.

Ferreira, Fernanda & John M. Henderson. 1993. Basic reading processes during syntactic analysis and reanalysis. *Canadian Journal of Psychology: Special Issue on Reading and Language Processing* 47. 247–275.

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

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

Fodor, Janet Dean. 1978. Parsing strategies and constraints on transformations. *Linguistic Inquiry* 9. 427–473.

Fodor, Janet Dean. 1992. Islands, learnability and the lexicon. In Helen Good luck and Michael Rochemont (eds.), *Island constraints: theory, acquisition and processing*, 109–180. Dordrecht: Kluwer.

Fodor, Janet Dean. 2002. Prosodic disambiguation in silent reading. In M. Hirotani (ed.), *NELS 32*, 113–132. Amherst, MA: GLSA Publications.

Francom, Jerid. 2009. *Experimental syntax: exploring the effect of repeated exposure to anomalous syntactic structure - evidence from rating and reading tasks*. University of Arizona, Ph.d. thesis.

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. 363–338.

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

Gries, Stefan. 2004. Preposition stranding in English: predicting speakers' behavior. In Vida Samiian (ed.), *Proceedings of the Western Conference on Linguistics*. Vol. 12, 230–241. California State University, Fresno, CA.

Hale, John. 2001. A probabilistic Earley parser as a psycholinguistic model. In *Proceedings of NAACL-2001, Pittsburg, PA*, 159–166, ACL.

Hiramatsu, Kazuko. 2000. *Accessing linguistic competence: evidence from children's and adults' acceptability judgments*. University of Connecticut, Ph.D. thesis.

Hofmeister, Philip, Peter Culicover & Susanne Winkler. 2015. Effects of processing on the acceptability of 'frozen' extraposed constituents. *Syntax* 18(4). 464–483.

Huck, Geoffrey J. & Younghee Na. 1990. Extraposition and focus. *Language* 66. 51–77.

Hukari, Thomas & Robert Levine. 1991. On the disunity of unbounded dependency constructions. *Natural Language and Linguistic Theory* 9. 97–144.

Jackendoff, Ray & Peter Culicover. 1972. A reconsideration of dative movement. *Foundations of Language* 7. 397–412.

Johnson, Kyle. 1985. *A case for movement*. MIT, Doctoral dissertation.

Kamide, Yuki & Don C. Mitchell. 1997. Relative clause attachment: non-determinism in Japanese parsing. *Journal of Psycholinguistic Research* 26. 247–254.

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

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*, 646–651.

Kim, Christina S., Gregory M. Kobele, Jeffrey T. Runner & John T. Hale. 2011. The acceptability cline VP Ellipsis. *Syntax* 14(4). 318 – 354.

Kim, Jong-Bok & Ivan A. Sag. 2005. English object extraposition: a constraint-based approach. In Stefan Müller (ed.), *Proceedings of the 12th International Conference on Head-Driven Phrase Structure Grammar, Department of Informatics, University of Lisbon*, 192–212. Stanford: CSLI Publications. http://cslipublications.stanford.edu/HPSG/6/.

Kravtchenko, Ekaterina, Maria Polinsky & Ming Xiang. 2009. Are all subject islands created equal? March 26–28, poster presented at the 22nd CUNY conference on Human Sentence Processing. Davis, CA.

Kroch, Anthony S. 2001. Syntactic change. In M. Baltin and C. Collins (Eds.), *Handbook of contemporary syntactic theory*, 699–729. Oxford: Blackwell.

Kuno, Susumu. 1973. *The structure of the Japanese language*. Cambridge, MA: The MIT Press.

Kutas, Marta & K. D. Federmeier. 2011. Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology* 62(1). 621–647.

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

Langendoen, Terence D. & Geoffrey K. Pullum. 1977. Preposition stranding in English: A problem and a mystery. In Samuel E. Fox, Woodford A. Beach and Shulamith Philosoph (eds.), *CLS Book of Squibs/Cumulative Index 1968-1977*, 64–65. Chicago, IL: Linguistic Society.

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

Lau, Jey Han, Alexander Clark & Shalom Lappin. 2015. Unsupervised prediction of acceptability judgements. In *Proceedings of the 53rd Annual Conference of the Association of Computational Linguistics, Beijing*, 1618–1628.

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

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

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

Linadarki, Evita. 2006. *Linguistic and Statistical Extensions of Data Oriented Parsing*, University of Essex, PhD Dissertation.

Manning, Christopher D. 2003. Probabilistic syntax. In Rens Bod, Jennifer Hay and Stefanie Jannedy (eds.), *Probabilistic linguistics*, 289 – 341. MIT Press.

McRae, Ken, Micheal J. Spivey-Knowlton & Micheal K. Tanenhaus. 1998. Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension. *Journal of Memory and Language* 38(3). 283–312.

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

Metzing, Charles & Susan E. Brennan. 2003. When conceptual pacts are broken: Partner-specific effects on the comprehension of referring expressions. *Journal of Memory and Language* 49, 201–213.

Miyao, Yusuke & Junichi Tsujii. 2002. Maximum entropy estimation for feature forests. In *Proceedings of the Human Language Technology Conference*.

Miyao, Yusuke & Junichi Tsujii. 2008. Feature forest models for probabilistic HPSG parsing. *Computational Linguistics* 34(1).35-80.

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

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.

Postal, Paul. 1972. On some rules that are not successive cyclic. *Linguistic Inquiry* 3. 211–222.

Quinn, Deirdre, Hala Abdelghany & Janet Dean Fodor. 2000. More evidence of implicit prosody in silent reading: French, English and Arabic relative clauses. Poster presented at the 13th Conference on Human Sentence Processing, La Jolla, CA.

Rizzi, Luigi. 2007. On some properties of criterial freezing. In *CISCL Working Papers on Language and Cognition 1*, 145–158.

Roark, Brian, Asaf Bachrach, Carlos Cardenas & Christopher Pallier. 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*, 324–333.

Ross, John R. 1967. *Constraints on Variables in Syntax*. Cambridge, MA: MIT doctoral dissertation.

Sag, Ivan A. & Thomas Wasow. 2011. Performance-compatible competence grammar. In Robert D. Borsley and Kersti Börjars (eds) *Non-transformational syntax: formal and explicit models of grammar*, 359–377. Wiley-Blackwell.

Sag, Ivan A. & Thomas Wasow. 2015. Flexible processing and the design of grammar, *Journal of Psycholinguistic Research* 44. 47–63.

Sag, Ivan A., Thomas Wasow & Emily M. Bender. 2003. *Syntactic theory: a formal introduction*. Stanford: CSLI Publications.

Smith, Nathaniel J. & Roger Levy. 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*, 595–600.

Sprouse, Jon. 2007. Continuous acceptability, categorical grammaticality, and experimental syntax. *Biolinguistics* 1. 118 – 129.

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

Tabor, Whitney & Sean Hutchins. 2004. Evidence for self-organized sentence processing: digging-in effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 30(2). 431–450.

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

Van Berkum, Jos J. A., Colin M. Brown, Pienie Zwitserlood, Valesca Kooijman & Peter Hagoort. 2005. Anticipating upcoming words in discourse: Evidence from ERPs and reading times. *Learning and Memory* 31(3). 443–467.

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

Wasow, Thomas. 2002. *Postverbal behavior*. CSLI Publications.

Wexler, Kenneth & Peter Culicover. 1980. *Formal principles of language acquisition*. Cambridge, MA: MIT Press.

Zwicky, Arnold. 1982. Stranded to and phonological phrasing in English. *Linguistics* 20, 3–37.

Zwicky, Arnold M. 1986. The unaccented pronoun constraint in English. In Arnold M. Zwicky (ed.), *Interfaces, volume 32 of Ohio State University Working Papers in Linguistics*, 100–114, Ohio State University Department of Linguistics.

Appendix A

1. Which cake did you serve (to) Mark several slices of?
2. Which cigarettes did you toss (to) Sean several packages of?
3. Which coffee did you sell (to) Roger several blends of?
4. Which documents did you forward (to) Jake various copies of?
5. Which house did you pay (to) the IRS too much tax on?

6. Which lands did you lease (to) Scott multiple parcels of?
7. Which logos did you hand (to) Frank several drawings of?
8. Which building did you (to) grant Sam too much access to?
9. Which machine did you ship (to) Quinn several parts of?
10. Which mistake did you assign (to) Lee too much blame for?
11. Which plant did you offer (to) Pam several varieties of?
12. Which problem did you write (to) Robin several emails about?
13. Which products did you send (to) John some samples of?
14. Which provisions did you bring (to) Carl abundant supplies of?
15. Which riddle did you read (to) Ben several versions of?
16. Which stocks did you owe (to) Kim some dividends from?
17. Which student did you pass (to) Bree several notes about?
18. Which theory did you show (to) Mia some basic principles of?
19. Which topic did you lend (to) Doug several books about?
20. Which wine did you slide (to) Rose several glasses of?

Appendix B

1. Who did you promise (to) a sum of $1,000 (to)?
2. Who did you forward (to) a copy of the contract (to)?
3. Who did you owe (to) a debt of gratitude (to)?
4. Who did you sell (to) your share of the company (to)?
5. Who did you send (to) several letters of apologies (to)?
6. Who did you award (to) the custody of the children (to)?
7. Who did you feed (to) a handful of roasted peanuts (to)?
8. Who did you show (to) the letter from the IRS (to)?
9. Who did you teach (to) the basics of poker (to)?
10. Who did you throw (to) the bouquet of flowers (to)?
11. Who did you disclose (to) the breach of security (to)?
12. Who did you give (to) a box of Belgian chocolates (to)?
13. Who did you grant (to) full access to the pool (to)?
14. Who did you mention (to) the cost of the damage (to)?
15. Who did you offer (to) a bribe of $25,000 (to)?
16. Who did you owe (to) a debt of millions of dollars (to)?
17. Who did you pass (to) several notes about me (to)?
18. Who did you pay (to) a bribe of $2,000 (to)?
19. Who did you serve (to) a slice of frosted cake (to)?
20. Who did you toss (to) the box of matches (to)?