To appear, Proceedings of WCCFL33 (2015)

Structuring Expectation: Licensing Animacy in Relative Clause Comprehension

Matthew W. Wagers & Emily Pendleton

1. Introduction

1.1. Animacy and the predictive encoding of relative clauses

Broadly speaking, subject relative clauses are easier to understand than object relative clauses (Wanner & Maratsos, 1978, Ford, 1983, Holmes & O'Regan, 1981, among many others). This can be demonstrated by a variety of methods that measure the speed and accuracy with which sentences containing relative clauses are processed. As an example, compare (1a), which contains a subject relative clause (SRC), to (1b), which contains an object relative clause (ORC).

(1) (a) *Subject relative clause,*
The lobbyist [that ___ quoted the journalist on the radio] lost her job.
(b) *Object relative clause*
The lobbyist [that the journalist quoted ___ on the radio] lost her job.

The SRC advantage is very robust cross-linguistically. Even in cases where non-subject relatives may ultimately be easier to process, the 'fingerprints' of a subject advantage can be detected (Clemens et al., 2015, Borja, Chung & Wagers, 2015).

However, the SRC advantage can be neutralized under a variety of conditions. For example, if the subject of an ORC is a quantified expression or a pronoun, then the asymmetry is substantially reduced (Gordon, Hendrick & Johnson, 2001, 2004). Another way to neutralize the SRC advantage is to change the animacy of the relativized argument (Mak, Vonk & Schriefers, 2002, Traxler, Morris & Seely, 2005, Gennari & MacDonald, 2008, among others). (2) illustrates a contrast that should be less severe than (1), because the relativized argument in (2), 'report', is inanimate.

(2) (a) *Subject relative clause, inanimate filler*
The report [that ___ quoted the journalist in the introduction] was not well-known.
(b) *Object relative clause, inanimate filler*
The report [that the journalist quoted ___ in the introduction] was not well-known.

Accounts of the SRC advantage - and for why lexical and phrasal factors like animacy, semantic type, or pronominality moderate it - appeal to a combination of predictive and integrative mechanisms (Staub, 2010, Gordon & Lowder, 2012, Lowder & Gordon, 2014).

Explanations based on predictive mechanisms generally turn on the notion that comprehenders (implicitly) entertain expectations about how a sentence will continue and that processing can be more difficult if those expectations are proven wrong. Explanations based on integrative mechanisms

---

[1] Matthew Wagers, University of California, Santa Cruz, mwagers@ucsc.edu. Emily Pendleton, University of California, Santa Cruz. The authors thank the audience members of WCCFL33, where parts of this research were presented as a component of the first author's keynote address; and attendees of the 27th Annual CUNY Conference on Human Sentence Processing (2014, The Ohio State University, Columbus, OH), where parts of this research were presented as the poster "Animacy and the Filled Gap Effect." MW acknowledges the support of NSF BCS #1251429. Supplemental materials - including item files, data sets, analysis scripts, corpus query and result files - may be accessed from the first author's web site (http://people.ucsc.edu/~mwagers).

usually link loci of difficulty to instances of composition between non-adjacent constituents. This difficulty is determined in large part by how the constituents are encoded in memory and how and when they are retrieved. In principle predictive and integrative explanations need not be independent, because knowledge of how the representation must be extended to form a grammatical utterance can be used to optimize how working memory should be allocated or managed - i.e., what should be maintained versus stored, how features are encoded, how cues are assembled to retrieved stored items (cf. Ericsson & Kintsch, 1995).

Gennari & MacDonald (2008, 2009) argued that the cause of the animacy effect in relative clause processing was related to a predictive mechanism, and one rooted in a tight relationship between production and comprehension. The idea is that the presence of an animate relativized argument induces expectations of a subject gap more strongly than an inanimate relativized argument. In a sentence completion task, they show that producers tend to complete relative clauses with inanimate arguments as object relatives (65% of the time) compared to animates, which they complete as subject relatives 90% of the time. Thus, animate arguments engender a prediction that is foiled when the actual string should be analyzed as an object relative clause, but inanimate argument engender a prediction that is always compatible with the incoming string. In a self-paced reading experiment, Gennari & MacDonald (2008) find that processing difficulty emerges as early as the relative clause subject. Thus, they argue, there must be a role for prediction in explaining the asymmetry, since they verb has yet to be encountered. Lowder & Gordon (2014) find a similar effect in eye-tracking, but with more tightly controlled stimuli. However, they find that the effect is only robust if the relativized argument and relative clause subject are not closely related by real-world knowledge. For example, ORCs relating *mayors* and *senators* do not lead to substantial difficulty on the relative clause subject, but ORCs relating *waiters* and *senators* do.

This latter finding makes it clear that the line separating prediction from integration is not always clear. In particular, if the co-occurrence of two argument DPs allows the comprehender to generate reasonable expectations about likely predicates or predicate features, then prediction can plausibly feed integration even before the predicate is encountered (Poeppel & Monahan, 2011, Omaki et al. 2015). One limitation of previous studies is that they either compare ORCs directly to SRCs (e.g., Traxler, Morris & Seely, 2005), or compare ORC subjects that follow different kinds of relativized arguments (Lowder & Gordon, 2014). But if integration is not limited to the verb itself, such designs make it difficult to disentangle prediction and integration. A more informative comparison would include relative clause subjects under contrastive levels of syntactic prediction, so that there is a baseline of *maximal prediction* against which to measure the effects of animacy.

In this paper, we would like to explore the specific contribution that prediction makes to the processing of relative clauses, and understand how it is a function of animacy. To do so, we used a *filled-gap design* (Crain & Fodor, 1985, Stowe, 1986, Wagers & Phillips, 2014) in which we compare two types of relative clause subjects: ones that are only potentially predicted, depending on the animacy of the relativized arguments; and ones that are *guaranteed*, independent of animacy by virtue of principles of syntactic connectivity.

In brief, we find that relativized animate arguments are predictively linked to the highest subject position of a relative clause, and no such predictive linkage is made for relativized inanimates.

1.2. The filled-gap design

The filled-gap design was introduced by Crain & Fodor (1985) and Stowe (1986) as a way of questioning how comprehenders "undo" the dependencies created by movement. For the rest of the paper, we will adopt the theory-neutral parlance of "fillers" to refer to the constituents displaced in a long-distance dependency and "gaps" to refer to their site of canonical thematic interpretation, i.e., where they were displaced from. For example in (3), the filler, in boldface, is 'who' and the gap, indicated by an underscore, is the object position of the embedded question.

(3) My brother wanted to know **who** Ruth will bring __ home at Christmas

As comprehenders build incremental syntactic representations, how could they know where the gaps are? Broadly speaking, this is an ambiguity resolution question - at least for gaps inside VP, where the comprehender must arbitrate between positing a gap or analyzing the VP as intransitive (Fodor, 1978).

In some early proposals (Jackendoff & Culicover, 1971, Wanner & Maratsos, 1978), gaps were only posited when parsing the VP otherwise 'failed' somehow. However, as word-by-word processing measures became increasingly practical, it also became clear that gaps were posited eagerly and as a first-resort - a generalization reflected in the 'Active Filler Strategy' (Frazier, 1987). Why parsers should be so organized as to prioritize linking fillers to gaps seems to be an instantiation of a broader principle: incremental linguistic representations should maximize the satisfaction of obligatory grammatical principles and constraints (Aoshima et al., 2004, Wagers & Phillips, 2009). A representation that includes a filler but not a gap is deficient in several respects: the filler lacks a thematic role, its position in the phrase structure is otherwise unlicensed, etc.

One of the key pieces of evidence in favor of an Active Filler Strategy parser comes from the filled-gap paradigm. We will illustrate the logic of this paradigm using examples from Stowe (1986). Consider again (3), and compare it to (4). Notice that now the gap is the complement to a preposition, and the embedded object position is filled by a pronoun. Stowe (1986) reasoned that if comprehenders predictively linked the filler with a gap in object position, then they should be surprised to encounter *us*. And that is exactly what was found: a "filled-gap effect".

(4) My brother wanted to know **who** Ruth will bring *us* home to ___ at Christmas.

To measure surprise, an appropriate baseline must be established. Stowe (1986) accomplished this by comparing embedded *wh*-questions to embedded *if*-questions, as in (5).

(5) My brother wanted to know if Ruth will bring *us* home to Mom at Christmas.

To test for a filled-gap effect, reading times to *us* in (4) were compared with reading times to *us* in (5), where there are no expectations for gaps to foil.

Stowe (1986) concluded that comprehenders do not treat subject gaps like non-subject gaps, because there was no significant filled-gap effect at the position of the highest embedded subject, 'Ruth'. And various arguments seemed to justify this conclusion - for example, a ban on string-vacuous movement seemed compatible with Stowe's data; and the fact that visible subjects are obligatory in English finite clauses mean that there is not the same ambiguity problem as with VP-internal gaps. However one problem with interpreting the lack of a filled-gap effect at that position is that it occurs immediately after the filler. It seems reasonable to suppose that predictively encoding a gap in any position should not be an instantaneous mental process, but one that takes some time. Therefore, the lack of a filled-gap effect for phrases in subject position could simply reflect the fact that the representation had not yet been extended -- not that it isn't extended in principle. Lee (2004) showed that if a brief adjunct was inserted after the filler, then a filled gap effect could be detected in subject position. As the comparison between (6a) and (6b) illustrates, Lee also introduced another innovation: instead of a baseline sentence with no filler-gap dependency, the baseline sentence incorporated a filler-gap dependency with a non-DP gap - as is created by pied-piping.

(6) a. That is the **cult which**, in the early eighties, *Elaine* inspired many friends to make a deep commitment to $\_\_\__{DP}$
b. That is the **cult to which**, in the early eighties, *Elaine* inspired many friends to make a deep commitment $\_\_\__{PP}$

The use of pied-piping provides an excellent control, since it holds almost every other fact about the sentence constant except the exact nature of the prediction: in both (6a) and (6b), a gap is predicted; but only in (6a) is it a DP gap (but cf. §3.2). Accordingly, demonstrating a filled-gap effect in subject position, at *Elaine*, is strong evidence that the parser has registered an expectation for a gap in that position. In what follows, we will use the design principles of Lee (2004) to test the effect of animacy of relative clause processing.

2. Reading time evidence for predictively-encoded gaps

We conducted two self-paced reading experiments to test the idea that only animate fillers cause the comprehender to predictively encode a subject gap. Both experiments used a filled-gap design to contrast conditions in which an expectation was grammatically legal, with a baseline condition, in which an expectation is grammatically impossible. The first experiment will be described in detail, followed by a briefer report for the second.

2.1. Design, Materials, Participants and Methods

Experimental materials followed a 2×2 design which crossed the factors of ANIMACY (*inanimate, animate*) and FILLER category (*NP, PP*). The contrast between NP and PP fillers is critical to the filled-gap design, as it allows us to test for foiled expectation at NP argument positions. The contrast between levels of animacy allows us to ask whether the expectation for gaps is a function of the animacy of the filler. 31 item sets were created which realized the experimental design. An example set is given in (7) for both levels of FILLER. The italicized text indicates the material in the string which was modified across conditions. Here, the animate argument is illustrated in the string and its inanimate version is given inside curly braces. The critical region, the embedded relative clause subject, is indicated in bold font, and the relative string position of the intended gap site is indicated with an underscore.

(7) a. FILLER:*NP*
The scholar looked to *his aging mentor who*, { *the controversial text which* }
only recently, **the academic community** owed much of their findings to $\_\__{NP}$.
b. FILLER:*PP*
The scholar looked to *his aging mentor* to *whom*, { *the controversial text to which* }
only recently, **the academic community** owed much of their findings $\_\__{PP}$.

The filler phrase was always three words in length, and of the form *Det – Mod – N.* Animate head nouns were always accompanied by the relative pronoun *who* (or *to whom* when pied-piped); inanimate conditions always included the relative pronoun *which* (or *to which*). A variety of prepositional phrases were used, chosen to maximize naturalness of pied-piping in the *PP* conditions: *for* (10 items); *to* (7); *on* (6); *about* (4); *with, of, around, under* (1 each). Some were selected by the relative clause verb (as in example 7), while others were less tightly constrained by the predicate, like locative or benefactive PPs.

The interrupting RC-initial adjunct phrase was always 1 to 3 words in length, and the number of words in the embedding sentence was adjusted accordingly so that the ordinal position of the critical region was identical across sentences. The critical region was always three words long and consisted of determiner and an NP with noun or adjective modifier. We did not completely control for the animacy of the relative clause subject, but they overwhelmingly referred to individual humans or collections of humans (as in 'the academic community').

Each sentence was followed by a *yes/no* comprehension question. The referents or events probed by each question were approximately counterbalanced so that that they might equally be introduced by phrases in either the matrix sentence or the relative clause. There were 71 filler sentences varied in structure in complexity. The full set of materials, including comprehension questions and filler sentences, may be accessed at the first author's website (see footnote 1).

30 undergraduates were recruited from the UC Santa Cruz Department of Linguistics subject pool, and they received course credit for their participation. The experiment was controlled by Linger (Version 2.94; Rohde, 2003)

2.2. Analysis

Before analyzing the target experimental data, we calculated each participant's median reading time and average accuracy on comprehension questions (expressed as a logit). These values were transformed into *z*-scores. Participants whose scaled reading time was greater than 2, or whose scaled accuracy was less than 2, were set aside. This policy identified two participants, one for slowness; and another for inaccuracy. Thus, 28 data files entered the full analysis. Finally, we also trimmed reading times at either end of the RT spectrum, removing the top and bottom 0.5 percentile (removing 1% of observations overall). The original range of reading times spanned 63 ms – 12,063 ms, but the trimmed range spanned 143 ms – 1895 ms.

The general statistical analysis we employed was to submit the dependent variable – either reading times in milliseconds, or question-answering correctness – to a linear mixed-effects model with ANIMACY and FILLER both as fixed effects and as nested effects under participant and item (*lme4*, Version 1.1; Bates, Maechler, Bolker & Walker, 2014). In the case of reading times, we identified residuals whose absolute normally-scaled value exceeded 2 and re-fit the model without those observations (Baayen & Milin, 2010). For targeted pairwise comparisons, we followed the same protocol, but entered only the relevant fixed effect. In the case of question-answering correctness, we used the binomial link function. To report *p*-values associated with the model *t*-score, we used the Kenward-Roger approximation for degrees of freedom, calculated by *pbkrtest,* (Version 0.4; Halekoh & Højsgaard, 2014). Because error rates were relatively high, we did not exclude trials with incorrectly answered questions for the analysis reported here. However, we did examine the data with only correct trials, and found no qualitative differences in the patterns reported below.

Analysis scripts and data files may be accessed at the first author's website.

2.3. Results

In brief: we find that there is a filled-gap effect at the relative clause subject – but only for animate filler phrases. Figure 1 shows average reading time per word and condition, with the critical region indicated by a box. When the filler is an animate NP, readers slow down during the subject phrase in comparison to the PP baseline. However, when the filler is an inanimate NP, there is no corresponding slow-down. Thus, readers seem to expect a subject gap just when the filler is animate.

Critical region. The filled-gap effect may vary in its onset and timing in ways that are not well understood (Wagers & Phillips, 2014). Therefore, on each trial we analyzed the cumulative reading time for the relative clause subject by summing the reading time for the determiner, modifier and noun. Participants read the subject region, on average, 97 milliseconds slower when the filler was an animate NP (compared to an animate PP); but only 1 millisecond slower when the filler was an inanimate NP (compared to an inanimate PP). Table 1 reports these averages, as well as a summary of the regression. The effect of FILLER was significant, and the interaction between ANIMACY and FILLER was marginally significant. The source of this interaction was explored in two pairwise comparisons, which revealed that the contrast between *NP* and *PP* conditions was significant for *Animate* sentences (t = 2.5) but not for *Inanimate* sentences (t = -0.2).

Verb and VP regions. The one-word region immediately following the subject was either a lexical verb, tense or a modal. There, reading times were elevated both for *Animate* conditions and *NP* conditions, but the effects were small and unreliable (*t*s = 1.5, 1.6, respectively). The one-word region that came three words after the subject head varied in its syntactic category across items, but it was generally a word inside the direct object. In this region, reading times for *PP* conditions were slower compared to *NP* conditions by about 18 ms, a marginally significant effect ($t = -1.7$, $p < .10$).

Pre-critical regions. We examined reading times one word before the critical region, to test for potential baseline effects. This region was always part of the RC-initial adjunct phrase. There was a slight elevation for *PP* conditions (20 ms) and *Animate* conditions (18 ms), but these differences were not significant (*t*s = -1.7, 1.6 respectively). We do not report the analysis of regions preceding the single-word *pre-critical* region, because of the variability in the words in each position. The word-by-word averages reported in Figure 1 do not suggest any serious baseline issues.

Accuracy. Average accuracy on the comprehension questions was 78%, and did not vary substantially or reliably by condition (range: 77% - 79%; standard error by-participants, per-condition: 3%). However, it was much lower than accuracy on filler sentences (94%).

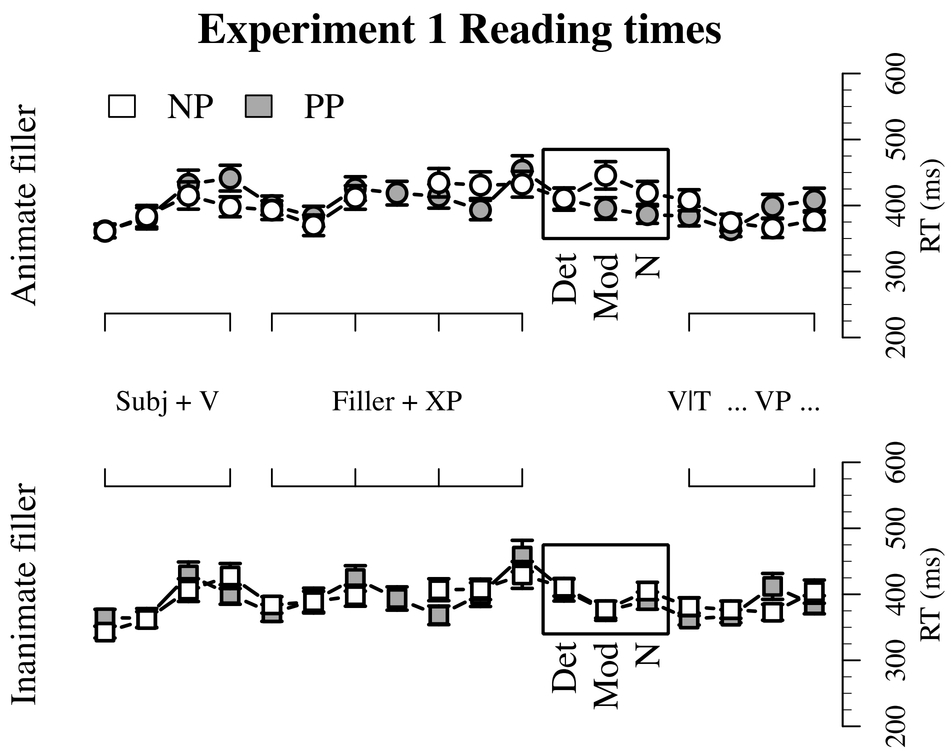

Figure 1 **Only animate fillers create an expectation for a subject gap** Average word-by-word reading times are given in milliseconds, with standard errors per word and condition. The critical region is indicated by a box. Only for animate fillers (top panel) is there a significant slow-down for NP fillers (open symbols) compared to the PP filler baseline.

	Cumulative Reading Time (ms)		
	NP	*PP*	Filled Gap Effect RT(*NP*) - RT(*PP*)
Animate	1277 (36)	1180 (33)	97 (49)
Inanimate	1179 (28)	1178 (28)	1 (40)

Model coefficients			
fixed effect		t-*score*	
ANIMACY	17 (23)	0.8	
FILLER	53 (25)	2.1	*
ANIM. × FILLER	84 (45)	1.9	•
Intercept	1147 (47)	24	

Table 1 **Cumulative reading times in subject position: summary and model** *Left:* cumulative reading times for the relative clause subject (*Det - Mod - N* summed per trial) and FILLER contrast for each level of ANIMACY. The outlined cell indicates the crucial filled-gap effect for *Animates* only. *Right:* fixed effect coefficients from the linear mixed-effects model of reading time on the experimental factors. *p:* .01 < * < .05 • < .10. *Both*: standard error in parentheses.

2.4. Experiment 2: Replication and extension

Our first experiment revealed a selective subject filled gap effect, one which invites the conclusion that only animate fillers are predictively linked to a gap in subject position. Before accepting this conclusion, however, we conducted a replication experiment to test the generality of the results in a different sample from the same participant population. We also varied the items slightly, in response to the following potential criticism. Suppose it were the case that the 1-3 word pre-subject phrases simply did not always provide enough time for comprehenders to link the filler and the gap. In such a case, what we observed might be that comprehenders always link the filler to the first available gap position, but it takes longer for comprehenders to link *inanimate* fillers (perhaps in general). This state-of-affairs would not merely be a variant of our initial conclusion, because it does not rely on differential expectations. Instead it relies on assuming qualitatively comparable processes that have differential finishing-time distributions. In our follow-up, therefore, we extended the pre-subject phrases to lengthen the time between filler and subject (=first potential gap site). These new phrases were more heterogeneous: on average, they were 4 words in length, but varied between 2-6 words. An example is given in (8).

(8) a. FILLER:*NP*
The executive introduced *the target buyers who*, { *the redesigned car which* } according to his consultants, **the insurance branch** should tailor their ads to $\_\__{NP}$ …
b. FILLER:*PP*
The executive introduced *the target buyers to who*, { *the redesigned car to which* } according to his consultants, **the insurance branch** should tailor their ads $\_\__{PP}$ …

There were 31 item sets and 32 participants. 2 item sets were removed because of exceptionally low accuracy on comprehension questions (19 and 22%, respectively). 2 participants were set aside because of low performance (69% and 62%). Overall, question answering accuracy was 86%, with participants doing better on non-target trials (92%) than target experimental trials (73%). For experimental trials, there was no significant variation by condition. Figure 2 shows reading times for the critical region, and surrounding words.

Critical region. At the determiner of the relative clause subject, there was a significant interaction between ANIMACY and FILLER ($p < .05$). This interaction was explored in two pairwise comparisons on the factor FILLER, which revealed that it was driven by two effects: a (marginal) slow-down for *NP* fillers in *Animate* condition (26 ms, $t(50) = 1.8$, $p < .10$), and a non-significant speed-up for *PP* fillers in the *Inanimate* condition (-16 ms, $t(50) = -1.4$). A further comparison showed that there was no effect of ANIMACY on the *PP* conditions ($t(50) = 1.0$). There were no further significant effects at either the modifier or the head noun. In a cumulative analysis, there was a marginal slow-down for *Animates* (83 ms, $t(33) = 1.9$, $p < .10$), but the interaction between FILLER and ANIMACY did not reach significance ($t(33) = 1.50$). Pairwise comparisons revealed that the effect of ANIMACY was only found in the *NP* conditions, such that *Animate* filler conditions were read significantly slower than *Inanimate* filler conditions (79 ms, $t(35) = 2.4$, $p < .05$).

Precritical region. At the word preceding the subject (which included a comma), *PP* conditions were read significantly slower than *NP* conditions (29 ms, $t(37)= -2.1$, $p < .05$). At the preceding word, *Animate* conditions were read significantly slower than *Inanimate* conditions (20 ms, $t(39)=2.5$, $p < .05$). There were no significant effects at the preceding word. Given the heterogeneity in lengths, we did not consider pre-critical regions.

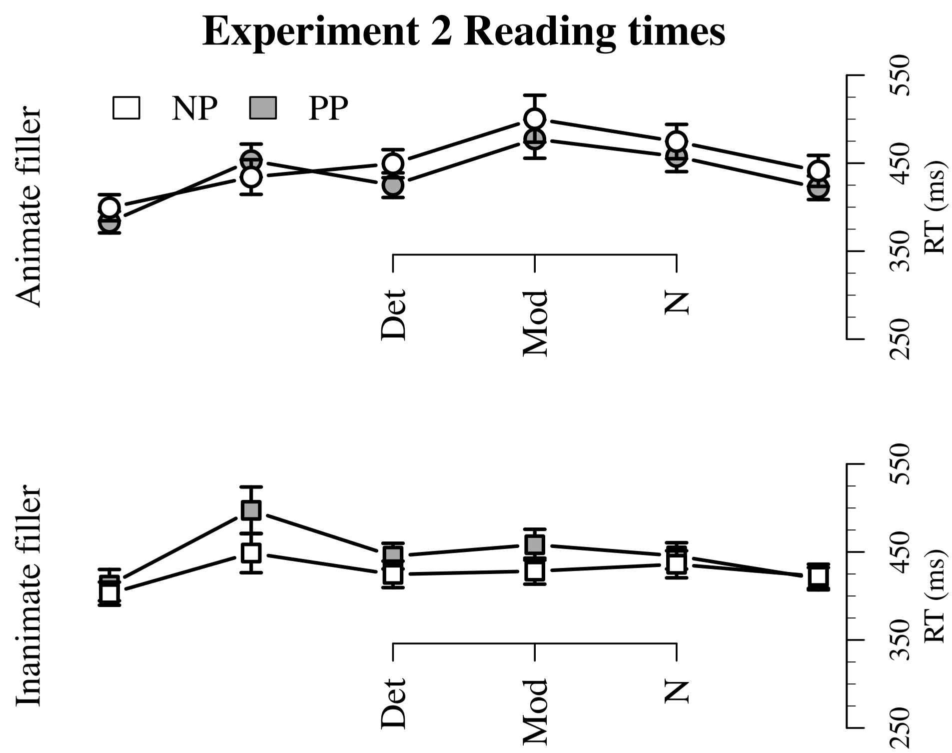

Figure 2 **Replication data set: animate NP fillers cause the most difficulty**
Average word-by-word reading times are given in milliseconds, with standard errors per-word and condition.

2.5. The experiments - summing up

The results of the second experiment were consistent with the first. When the filler phrase was animate, participants were slowed down at the relative clause subject phrase. In the first experiment, there was a clear, reliable slow-down with respect to the corresponding *PP* condition. In the second experiment, the slowdown was most evident by comparison to the corresponding *Inanimate* condition. Overall, though, the pattern of mean reading times was similar, and it was likely the case that the second experiment was underpowered. A post-hoc estimate of power in the *Animate NP-PP* pairwise comparison suggests we achieved power (1-β) of 0.43 in Experiment 2, compared to 0.84 in Experiment (G*Power Version 3.1; Faul et al. 2009).

Nonetheless, taking the two experiments together, it seems clear that relative clauses with animate fillers engender disruption as early as the subject phrase. Moreover, this disruption is not present when the fillers are animate but require linkage to a *PP* gap, as in the case of pied-piping. That finding indicates that the difficulty is probably related to dependency construction per se and not merely the presence of two NPs with overlapping features. Therefore, we conclude that comprehenders predictively encode a representation with a subject gap and that they do so with greater likelihood when the filler is animate.

Two other effects seems worth commenting on. First: in both experiments, we observed consistent patterns in the RC-initial adjunct. In particular, this phrase was slightly harder to read. It was also harder when the filler was animate. We conjecture that these effects might reflect either encoding interference - as in the case of PPs - or maintenance - as in the case of animates. It is harder to tie them to dependency formation, given that no dependency can be formed in these adjuncts (except, possibly, in cases of parasitic gaps for clausal adjuncts). Given the design of the experiment, it is impossible to say much more.

The second effect worth noting, from Experiment 1, is the increased reading times for PP conditions in the immediate post-verbal regions inside the relative clause. This effect gives some assurance that comprehenders were truly processing the pied-piped sentences to some degree of depth. Combined with the lack of variation in accuracy, it suggests that comprehenders did not fail in some special way when the filler was pied-piped, even though such constructions are often considered unnatural (see Wagers & Phillips, 2014, for similar evidence).

3. Discussion

3.1. Candidate accounts for the animacy asymmetry

We began this investigation with a question about *why* relative clauses formed on an inanimate object are easier to understand than relative clauses formed on an animate object. This generalization has been well-known for some time now. It has not been difficult to conceive of a general, functionally-oriented, explanation for the asymmetry between the two kinds of relative clauses. Something along the lines of: animate relative clauses tend to be relatively rarer as object relative clauses, because, all else equal, animate arguments tend to be mapped to a subject position in the syntactic structure. But what are the real-time incremental processes that take place, such that difficulty accrues for the animates? We have offered very direct evidence that relativized animate objects trigger reanalyses more often than relativized inanimate objects. That evidence came in the form of the selective filled gap effect, an index of foiled expectations. Thus, the path from string to the analysis is a more complex one for object relative clauses with an animate filler phrase.

To be concrete, we suppose the following:

(9) **Why relativized objects make comprehension harder if they are animate**

(i) Relativized arguments, *Rel*, predictively and incrementally extend comprehenders' partial syntactic representation to include a gap.

(ii) There is some probability, *s*, that this gap is inserted into the subject position.

(iii) s_{anim}, p(*InsertSubjGap* | *Rel*=ANIM) > s_{inanim}, p(*InsertSubjGap* | *Rel*=INANIM)

(iv) The chance, *r*, that reanalysis will be required is directly proportional *s*.

(v) $r_{anim} \geq r_{inanim}$

(vi) Reanalysis exacts some cost.

Assumption (i) amounts to the Active Filler Strategy (Frazier, 1987). Assumption (ii) could be added as a parameter of that theory, and it is most generally stated as a distribution over possible syntactic positions. It is an interesting question how this distribution is determined. Under a strictly syntactic view of locality, the probability of insertion into subject position should be maximal (or highest) because the subject position is the closest potential gap site in terms of dominance relations. Although not thinking exactly in terms of incrementality, it was filler-gap locality that Keenan & Comrie (1977) proffered to explain the Accessibility Hierarchy.

Under another construal of locality - for example, the one advocated in Aoshima et al. (2004) or Wagers (2014) - the position of the gap should be the first one that can be tested against the input. Call this construal "diagnostic locality" (cf. Chater, Crocker & Pickering, 1998). What counts as *the first one* thus depends at least on how syntactic structure is mapped onto word order. In a head-initial, verb-medial language like English, the gap would be preferentially posited in subject position. In Chamorro, a verb-initial language with more flexible word order, the position of the gap could be more flexible - but the presence of subject-verb agreement probably leads the gap to also be preferentially posited in the subject position (Borja, Chung & Wagers, 2015). However, in a verb-final language like Japanese, more deeply embedded syntactic positions precede less deeply embedded ones. Consequently, the gap is posited in embedded object positions before it is posited in higher positions (Aoshima et al., 2004).

But neither syntactic locality or diagnostic locality alone seemingly can account for (iii), that fact that probability of insertion depends on the animacy of the relativized argument. Where do the values for s_{anim} and s_{inanim} come from? It is this question which will occupy the rest of this paper.

Before preceding, however, a word about assumptions (9)(iv)-(vi) is necessary. We have (softly) framed the theory in terms of a single-path parser (Lewis, 2000). A single-path parser pursues one interpretation at a time, and this is a feature which we believe is most congenial to the limited focus of attention that characterizes working memory (McElree et al. 2003; see Lewis, Vasishth & Van Dyke, 2006, for a broader defense) as well as findings indicating that syntactic ambiguity resolution is probably non-competitive (Clifton & Staub, 2008; cf. Levy, 2008). Thus the assumptions (9)(iv)-(vi) are stated in terms of reanalysis. However none of the following arguments, it seems to us, depend on the difficulty stemming strictly from reanalysis. If multiple parses are represented simultaneously, then we can restate (9)(iv)-(vi) in terms of resource-allocation. Namely, difficulty is engendered by shifting the distribution of probability/confidence over a set of multiple parses (Hale, 2001, 2006, Levy, 2008) when the crucial information arrives that a phrase occupies the subject position.

3.2. How filler animacy affects the position of the gap: take 1

A natural hypothesis is that *s*, the probability of insertion, is directly related to the language comprehenders' experience. That is to say, that p(*InsertSubjGap* | Rel_j) ~ p(GAP=*Subj* | Rel_j). We'll refer to this latter quantity as the experiential probability, or p_{exp}. As we've already mentioned, previous experimental studies have found a relationship between animacy and relative clause type both in production experiments (Gennari & MacDonald, 2008) and in natural productions (Roland, Dick & Elman, 2007). Roland, Dick & Elman (2007) examined 100 relative clauses and estimated p_{exp} in two corpora (Switchboard and Brown). In Switchboard p_{exp} was 91% for animates and 31% for inanimates, In Brown, it was 75% and 47%, respectively. Thus it seems clear that the animacy of the filler NP can be used to make a reasonable prediction about the upcoming gap site.

We attempted to validate these estimates by using a different data source: the New York Times subsection of the English Gigaword corpus (Second edition; Graff et al., 2005), a version of which was parsed using the Stanford parser by Pranav Anand (p.c.). It is important to note that we are not necessarily looking to find quantitatively similar numbers as previous estimates, but to test the correlation between animacy and gap position. The actual values obtained will depend in important ways on the exact class of structures targeted, the genre of text, etc. Using *TGrep2* (Version 1.15; Rohde, 2005), we searched for instances of finite, unreduced relative clauses with the visible complementizes/relative pronouns *that, which,* and *who(m)*. We restricted our attention to these cases because they correspond most closely to the materials used in our experiment. Relative clauses were automatically classified as having a (highest) subject gap if there was an NP dominated by the same S that dominated the (highest) relative clause VP. All queries and result files may be downloaded from the first author's web site.

	Relativizer				
Gap	*that*	*which*	*who*	*whom*	
=*Subj*	41,276	85,643	140,319	1°	267,239
≠*Subj*	6,988	5,696	1,308	1,053	15,045
	48,264	91,339	141,627	1,054	282,284

p_{exp}(GAP=*Subj* | ...)*		*Surprisal*(GAP≠*Subj*)	
FILLER=*PP*	0.005†	.01	*baseline*
FILLER=*NP*	0.954	4.4	
... & *Rel* = "who"	0.995†	7.6	*> abrupt*
... & *Rel* = "which"	0.945	4.2	*< abrupt*

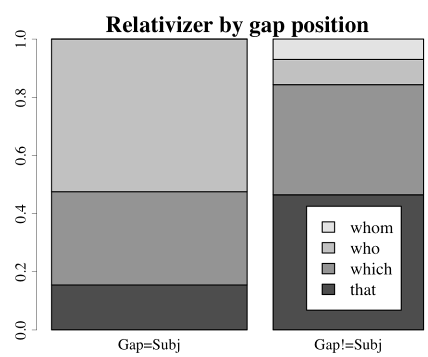

Table 2 **The animate relative pronoun 'who' strongly predicts a subject gap** *Top left*: raw counts of (highest) subject and non-subject relative clauses, by relativizer. Double lines highlight the relative pronouns used in our experiment. *Bottom left*: estimates of p_{exp}, conditional probability of a subject gap, and surprisal values for encountering the subject inside the relative clause. *Right*: proportional breakdown of relativizers according to gap type. *Notes:* °This token contained a true subject gap, and thus probably a hyper-corrective use of 'whom'. †By hypothesis, see discussion in text; *p* was constrained to [.005, .995]. *Our query had a significant false positive rate for non-subject relative clauses, since the corpus annotation often assimilated subject phrases with other pre-verbal NPs (like "this week"). Inspecting a sample of non-subject RCs (n=100) revealed 16% false positives, compared with 1% in the case of subject RCs. *p*(GAP=*Subj*) was calculated after adjusting for the false positive rate, so it is slightly higher than the raw counts suggest.

As with previous authors, we find that SRCs predominate: 95% of our tokens contained the gap in highest subject position. The severity of this asymmetry is heightened compared to, e.g., Roland, Dick & Elman (2007), because we did not consider non-finite relative clauses or relative clauses without a visible complementizer or relative pronoun. Omitting the complementizer or relative pronoun is possible when the gap is not in highest subject position, and these are precisely the kind of object relative clauses that are most common. Omission is also possible in conjunction with a subject gap and deletion of BE ("whiz deletion"), but we didn't consider those types either. However, we were most interested in the cases which characterize our experimental stimuli, which are outlined in the table above. If one examines comparable types from Roland, Dick & Elman (2007)'s Table 7 and a comparable genre (Wall Street Journal), the subject gap rate is similarly high: 86%.

What we can see from an initial analysis is what accords with our intuitions: the presence of the animate relative pronoun 'who' is highly informative (in the context of the already low uncertainty). 'Which', on the other hand, is less discriminative. One way to illustrate this is to calculate the *Surprisal* - or negative log probability - of the non-subject analysis. If the relative pronoun is *who*, encountering a phrase in subject position provides 7.6 bits of information; whereas if the relative pronoun is *which*, the subject phrase provides 4.2 bits. A completely uninformative word would provide 0 bits of information - and that is the case for PP fillers, by hypothesis. It is not exactly true that pied-piping absolutely guarantees a subject in the highest position, as subject-verb inversion can be triggered in appropriate environments, cf (10).

(10) This class is characterized by the verb [$_{RC}$ *under which* is embedded **the S containing the cyclic subject __**]. Aissen & Hankamer (1972)

But the logic of the experiment holds as long as the animacy of the pied-piped filler does not affect subject-verb inversion. No such correlation has been observed, but it is an good question for future investigation. See Aissen & Hankamer (1972) for more details.

So far we have linked our estimate of p_{exp} to the relativizer itself and whether it is *who*, essentially. Of course, the comprehender need not rely on the relative pronoun alone, and can also make use of the real-world knowledge about the referent or kind denoted by the filler. To estimate the conditional probability of a subject relative clause, given the actual filler phrase (and its relativizer), we conducted two further analyses.

First, we inspected a small sample of subject and non-subject relative clauses (n=100 each), and computed p(GAP=*Subj*) both for animates and inanimates. Then, to obtain a sharper and less-biased estimate, we randomly selected 1000 tokens each from subject and non-subject relative clauses in our data base. We extracted the filler NPs and submitted them to workers on Amazon Mechanical Turk to classify as human or non-human. (For full details of this experiment, consult the supplemental materials on the web.) We then computed the conditional probabilities for either filler type.

Table 3 gives the raw counts for four subcategories in the annotation experiment. The probability of an animate (=human) filler conditioned on the gap type was based directly from these counts, and then the probability of gap type conditioned on animacy was calculated via Bayes Theorem. The Mechanical Turk estimates were based on 38 annotators and they are in good agreement with our by-hand estimates, which are given in brackets in the table. The likelihood of a subject gap, given an animate filler phrase, is 98%, whereas the likelihood of a non-subject gap, given an animate filler phrase, is 90%. In terms of Surprisal, this is a difference at the subject phrase of about 2.3 bits: the subject phrase is predicted to be more informative (~difficult) when the filler phrase is animate.

	Filler category (counts)			
GAP	*Human*	*Non-human*	*Institu-tion*	*Not sure*
=Subj	530	344	66	51
≠Subj	232	670	50	36

	Probabilities			
GAP	p(ANIM|GAP)	p(ANIM)	$p(GAP=Subj \mid ...)$	
=Subj	0.61 [0.59]	0.589 [0.567]	ANIM	0.98 [0.98]
≠Subj	0.26 [0.23]		INANIM	0.90 [0.90]

Table 3 **Two estimates of how strongly filler animacy predicts gap type** *Left*: counts of isolated DPs, classified by what relative clause type they were extracted from and how they were classified by annotators. DPs classified as institutions (like companies or charities) and DPs about which the annotator was unsure were not included in probability estimates. *Right*: Conditional probability of an animate filler, given the gap type (along the rows); the baseline probability of an animate, and the conditional probability of a subject gap given an animacy value.

3.3. How an RC-initial adjunct affects the position of the gap

The patterns that emerge from the initial corpus investigation are consistent with previous findings, and lend plausibility to the idea that the parameter *s* in our theory of relative clause processing - the probability of *predictively encoding the relative clause analysis with a gap in the subject position,* given features of the filler phrase - is straightforwardly related to the experienced probability of a subject gap occurring after a filler with those features, as compiled from experience and here estimated from a corpus. In this section, however, we consider one important way in which our experimental stimuli differed from the types of tokens we examined in the corpus: the presence of an RC-initial adjunct phrase.

RC-initial adjunct phrases can be found in naturally-occurring tokens. For example, (11) is a recent example from the New York Times. This sentence contains an (appositive) relative clause, a gap in subject position, and an adjunct phrase which precedes the TP.

(11) Dr. Doeleman had planned to spend the night working out new techniques to point the telescope, [$_{\text{RC}}$ which **among its other problems** ___ was afflicted by a persistent and annoying electrical hum]. Overbye, D. (2015, June 8). Black Hole Hunters. *The New York Times.* Retrieved from http://nyti.ms/1MiSFBR.

Although the gap position is marked here after the adjunct, it is impossible to tell from this single example whether the adjunct phrase is higher or lower in the tree than the subject. But, this is a convention for writing the string, and a harmless one for present purposes: from the perspective of the parser, the subject gap analysis can be confirmed or rejected only when the material in T/V is encountered or when an actual subject phrase in encountered.

We wondered whether we could find many such examples of RC-initial adjuncts, and whether the distribution of gaps among such examples differed. We turned again to the parsed New York Times subsection of Gigaword, and collected examples of relative clauses with non-subject phrases that preceded the verb. We found 3488 examples of relative clauses that contained an RC-initial adjunct phrase. Only one of them contained a non-subject gap, which is given in (12).

(12) That resignation, [$_{\text{RC}}$ which **at the time** Grace attributed ___ to differences of "style and philosophy" with senior management], marked the beginning of an extraordinary period of corporate infighting... `nyt_eng_199504.tgrep2:33748`

The remainder of the results contained subject gaps, like (13).

(13) Paul Anderson, [$_{\text{RC}}$ who **because of his picture on a box of Wheaties** ___ was known as the world's strongest man to a generation of children growing up in the late 50s], was the last American to win Olympic gold in 1956. `nyt_eng_199407.tgrep2:37630`

At this point we might tendentiously conclude that the odds of finding a non-subject gap, after a relative clause with an initial adjunct, were an infinitesimal 3487:1 against - animate or no. However, some care is needed. Firstly, because of the parsing errors in the corpus, there were some false positives: sentences identified as containing an RC-initial adjunct which were mis-parsed (usually the false-positives contained genitive filler phrases with an PP modifier). We examined a random sample of 100 tokens, and found that 87 of those tokens contained the target adjunct. If we adjusted our count accordingly, this would result in an odds estimate of 3033:1 against - still not very encouraging for object gaps. We also estimated *p*(ANIM) from this sample and found that it was not substantially different from the sample without an initial adjunct: 63% (cf. 58% in Table 3).

A more interesting issue is that the vast majority of RCs with 'qualified' adjuncts were appositive RCs, as in (11)-(13). In our hand-classified random sample, 84 of 87 relative clauses seemed like appositive RCs; and cursory examination of the other hits made it clear this was probably generally true. It is hard to say whether this is a fact about the environments in which those adjuncts are most natural, or whether this is a stylistic fact about the genre of newswire. We also did not exclude appositive relative clauses from our initial corpus search. Returning to those data, we found that about 67% of our subject relative clauses and 54% of non-subject relative clauses were set apart by commas – a reasonable if imperfect classifier for appositive RCs. Note that, if we exclude the appositives, it does not substantially change our baseline estimate of an RC having a highest subject gap – it becomes about 93%.

Given our ignorance of the genre-contingency of appositives, we decided to turn to another method: a Cloze completion task. We asked workers on Mechanical Turk to complete stem sentences taken from the actual items we used in the experiment. To illustrate, consider example (7), again, modified as (14) below. Task participants were asked to provide a completion to sentences that included the filler (14a), the filler and associated relative pronoun (14b), or the filler, relative pronoun and adjunct (14c). Stem type was crossed with animacy of the filler/relative pronoun in a 2 × 4 design (the 4th level of stem type, with a subject phrase in place, is not shown here; for complete details, see supplemental materials). Between 103-119 completions were provided for each condition. The chance of a subject gap completion is given in the right margin, according to filler type. In the case of "Filler Only" stems, the non-subject gap rate is also provided in brackets because most completions did not include a relative clause at all. Thus the sum of these two numbers is the overall relative clause completion rate.

(14)	**Cloze Completion Task**		***p*(GAP=*Subj* | *Stem*)**	
	Stem Types		*ANIM*	*INANIM*
	a.	*Filler Only*	*19%*	*10%*
		The scholar looked to *his aging mentor ...*	[*1%*]	[*10%*]
	b.	*Filler + Relative Pronoun*	*99%*	*89%*
		The scholar looked to *his aging mentor who ...*		
	c.	*Filler + Relative Pronoun + Adjunct*	*99%*	*94%*
		The scholar looked to *his aging mentor who, only recently*		

The first finding is general convergence between our corpus investigation and this completion task. Including a relative clause initial adjunct increases the rate of subject gap completions for inanimates. However, based on these results, we can see that there is still a difference in informativity based on the animacy of the filler: surprisal at the subject phrase is predicted to be 6.6 bits for animates, but only 4.1 bits for inanimates. While average surprisal is higher, the contrast is about the same as found for animacy+relativizer alone. Table 4 provides a summary of our estimates of *p*(GAP=*Subj*) and the corresponding measure Surprisal(GAP≠*Subj*).

	Relative pronoun alone (*from corpus*)		… & filler animacy (*from corpus*)		… & filler animacy & adjunct (*from Cloze*)	
	p(GAP=*Subj*)	*Surprisal* (GAP≠*Subj*)	p	*Surprisal*	p	*Surprisal*
ANIM (*who*)	0.995	7.6	0.98	5.6	0.99	6.6
INANIM (*which*)	0.945	4.2	0.90	3.3	0.94	4.1
$Surprisal_{ANIM} - Surprisal_{INANIM}$		*3.4*		*2.3*		*2.5*

Table 4 **All sources of information considered predict differential difficulty for object relatives if the relative pronoun is animate *who*.** The first two columns show our corpus-based estimates from the NYT subsection of Gigaword. The third column is based on the Cloze task designed from our actual experimental materials. Surprisal is measured in bits.

3.4. How filler animacy affects the position of the gap: take 2

At this point, let us step out of the weeds and see what we've accomplished. We've considered several sources of information to assess what kind of expectations they are capable of establishing in principle. What we found is that subject gaps are always more likely for animates compared to inanimates, regardless of whether we consider the animacy of the (visible) relative pronoun in isolation or in conjunction with real-world knowledge about the animacy of filler phrase. This attaches plausibility to the idea that experience with relative clauses of the type under consideration can correctly influence a comprehender's predictively encoded representations.

But now we must confront the issue of baselines. This was difficult to do meaningfully in the case of our corpus investigations, given that we restricted our attention to one data source (newswire). And, in general, we don't want to advocate for the use of corpus estimates as literal stand-ins for mental probabilities. Instead we'd rather think of the corpus values as one source of data, among others, to a mechanism of generalization (cf. Roland, Dick & Elman, 2007, on this point). On the one hand, the behavioral evidence from Experiments 1 and 2 strongly impels us to the conclusion that the comprehender doesn't predictively insert subject gaps when the filler is inanimate. On the other, the completion data – using identical materials – shows that individuals overwhelmingly *produce* subject gaps, especially when the RC-initial adjunct is present.

In essence, what we've found is a strong mismatch between the incremental behavior of the comprehender and the cumulative behavior of the producer. In a way that we didn't anticipate, the inclusion of the adjuncts made this mismatch all the more striking. If we return to our theory of why animate relativized objects makes comprehension harder, iterated as (15) below, then what we really need to account for is clause (15iii'): why s_{inanim} is not only smaller than s_{anim}, but how it could basically be zero. This will not fall out obviously from experiential probability alone, since relativized inanimate subjects are abundant in general and, in particular, they overwhelmingly predominate for our sentence tokens, which include an RC-initial adjunct.

(15) **Why relativized objects make comprehension harder if they are animate (take 2)**
- (i) Relativized arguments, *Rel*, predictively and incrementally extend comprehenders' partial syntactic representation to include a gap.
- (ii) There is some probability, *s*, that this gap is inserted into the subject position.
- (iii) s_{anim}, p(*InsertSubjGap* | *Rel*=ANIM) > s_{inanim}, p(*InsertSubjGap* | *Rel*=INANIM)
- **(iii′)** $\boldsymbol{s_{inanim} \sim 0}$
- (iv)… The chance, *r*, that reanalysis/reallocation will be required depends on *s*. …

What proposal could take the place of the idea that *s* derives directly from p_{exp}?

One possibility is that comprehenders actually do not pay attention to many of the details that characterize our stimuli, and that their estimates of p(GAP=*Subj*|REL=*Anim*) reflect a broader and more abstract generalization about relative clauses, about A-bar dependencies, or about subject phrases in general. If we think about a general probability statement, like p(GAP=*Subj* | X & Y & Z), we can ask if any valid contingencies can participate in a language user's estimates of what's going to come next; or if there are restrictions on what can be X, Y, Z, etc. Let us suppose that the comprehender's behavior is sensitive to contingencies stated over what we'll call "grammatically-active" features. *Animacy* is a plausible grammatically-active feature because it restricts the choice of (relative) pronoun, it participates in verb-argument selection restrictions, etc. Moreover, it is not only grammatically-active but it is relevant to the prediction of interest, that is, a prediction about how an argument should be linked to a syntactic position. Possibly relevant to this point is the fact that, when comprehenders spontaneously generated relative clauses in the Cloze task (as in condition 14a), they did so at comparable rates when the stem ended in either an animate or inanimate DP (about 20% of completions). However, for animates, 95% of the relative clauses they formed had a gap in the highest subject position; for inanimates, there was an even split between highest subject relatives and others.

Whether or not an XP is preposed must be, by our hypothesis, orthogonal to how the argument-to-position linking is made. Language users, correspondingly, store no information about whatever contingencies happen to hold by virtue of an XP being preposed. In a sense, then, this is a generalized version of the Argument Structure Hypothesis (Tutunjian & Boland, 2008). According to that hypothesis, adjunct attachment is not sensitive to co-occurence frequency of predicate/adjunct pairings while argument attachment is. In our experiments, the situation is somewhat inverted: the fact than an adjunct exists in a particular position occasions no change in a predictively-coded attachment based on animacy. Of course, the question arises of *why* preposed XPs should so strongly co-occur with subject gaps. One possibility is that there is a locality pressure in incremental planning, that is somehow not operative in comprehension. Another possibility is that, given the size of the preposed constituent, a (transitive) clause with a subject gap can be more felicitously phrased prosodically than one with a pronounced subject phrase. The best explanation would be one in which a conspiracy of factors filters out the non-subject gaps, a conspiracy whose exact character we must leave for future research.

We'd like to finally and briefly articulate an alternative hypothesis: namely that the probability of an analysis, appropriately restricted, is *not* exclusively what drives comprehension processes. We want to say, specifically, that comprehenders won't always favor analyses that are incrementally most likely. Instead, it is the expected utility of these incremental analyses that guides comprehenders' decisions (Chater, Crocker & Pickering, 1998). Expected utility is determined not only by the probability of an analysis being correct, but what "preferences" are satisfied by virtue of the analysis being correct. Suppose that optimizing well-formedness is one such 'preference' (cf. Merlo & Stevenson, 2000, Aoshima, Phillips & Weinberg, 2004). Moreover, suppose that grammatical status is gradiently affected by argument alignment, such that sentences with inanimate subjects are, in fact, somewhat less grammatical than structurally-parallel sentences with animate subjects. If this were true, then (holding conditional probability constant) an incremental analysis that linked an animate argument to the subject position would satisfy more preferences than one that linked an inanimate.

We don't want to push this latter hypothesis too strongly or get too far beyond the data we've offered. But it provides a basis for thinking about when and why comprehenders predictively extend incremental representations: they do so to license elements in the string that need to be licensed. In English, inanimate arguments can clearly occupy the subject position, even if an animate argument occupies a less prominent position. But, there are languages that more strongly penalize such pairings. For example, Chamorro categorically bans inanimate DPs from occupying the subject position of a transitive clause if its object is an animate DP (Chung, 2012, Clothier-Goldschmidt, 2015). Silverstein (1976) observed a broader typological generalization about the association of person/animacy features and grammatical role, which Aissen (1999) explained in Optimality Theoretic syntax. The core of that proposal is that grammars encode prominence scales, like: *Subject* > *Object* and *Animate* > *Inanimate*, and that these scales are harmonically aligned to generate a set of constraints like:

(16) (a) **Subject/Inanimate* >> **Subject/Animate*
(b) **Object/Animate* >> **Object/Inanimate*

These constraints are differently ordered in different languages (Prince & Smolensky, 1993), but we might expect them to be weakly active in English (Bresnan, Dingare & Manning, 2001). An interesting possibility is that the alignment of these scales is one source of prediction in incremental linguistic structure building, and that comprehenders attempt to maximize expected utility of their parsing decisions by maximizing well-formedness (cf. Smolensky, 2006).

3.5. Conclusion

In this paper we have argued that the difficulty of non-subject relative clauses is modulated by animacy because different animacy values trigger different expectations. When the filler is animate, the comprehender's incremental syntactic representation is predictively extended to encode a subject gap. The relative clause subject is incompatible with this representation, and triggers reanalysis, repair or reallocation of resources. However, when the filler in inanimate, no subject gap is predictively encoded and therefore the relative clause subject is processed with no special difficulty. Thus we affirm the basic account of Gennari & MacDonald (2008), although we have framed it in terms of syntactic, rather than semantic, representations. Because our experimental design contrasts possible subject gap sentences (*NP* fillers) with impossible subject gap sentences (*PP* fillers), our argument is less liable to the concern that any animacy effect on the relative clause subject is due to an attempted integration of the two arguments. This does not mean to say that features of the subject play no role in the cumulative difficulty of the relative clause, because clearly they may (Traxler et al. 2005, Lowder & Gordon, 2014) - however, we can now pinpoint a role for prediction.

One open question concerns the symmetry of prediction, i.e., do inanimate *NP* fillers generate a specific syntactic expectation? The maximize well-formedness hypothesis suggests they would (at least we grounded it in harmonic alignment). On the other hand, statistical estimates - by Gennari & MacDonald (2008), Roland, Dick & Elman (2007), and our cloze results (*filler*-only condition) - seem to show that while animates are strongly affiliated with subject gaps, inanimates are more equivocal. Consequently, if the grammatically-active contingencies hypothesis were correct, then we might expect no specific expectation to be generated by inanimates. This prediction is a more complicated one to test, because of how the intervening material between filler and a potential object position could re-shape the comprehender's expectations.

A final question - or a confession - concerns the fact that Lee (2004)'s paper on the subject filled-gap effect apparently used inanimates (or non-individuated human groups) for the filler, as in (6), 'cult'. This would be straightforwardly problematic for our claims, except that Lee's critical region was inside a demonstrative cleft, and not a restrictive relative clause. For such clefts there are many arguments that the dependency with the clefted constituent is mediated between the matrix subject — i.e., *it, that*, *this* — and the gap inside the cleft, which has been extraposed (Hedberg, 2000). Thus the apparent incompatibility of our findings with Lee (2004) may be resolved. At least it points the way to future inquiry, one which broadens to related grammatical domains the traditional concerns of complexity that have been so deeply explored with relative clauses.

References

Aissen, Judith. 1999. Markedness and subject choice in Optimality Theory. *Natural Language and Linguistic Theory, 17*, 673-711.

Aissen, Judith, & Hankamer, Jorge (1972). Shifty Subjects: A Conspiracy in Syntax? *Linguistic Inquiry*, *3*(4), 501–504.

Aoshima, Sachiko, Phillips, Colin, & Weinberg, Amy (2004). Processing filler-gap dependencies in a head-final language. *Journal of Memory and Language*, *51,* 23–54.

Baayen, R. Harald, & Milin, Petar (2010). Analyzing reaction times. *International Journal of Psychological Research, 3(2)*, 12–28.

Bates, Douglas, Maechler, Martin, Bolker, Ben, & Walker, Steven (2014). lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-7. Retrieved from http://CRAN.R-project.org/package=lme4.

Borja, Manuel F., Chung, Sandra, & Wagers, Matthew (2015). Constituent order and parser control processes in Chamorro. In Amber Camp et al. (Eds.) *Proceedings of AFLA 21.*

Bresnan, Joan, Dingare, Shipra, & Manning, Christopher D. (2001). Soft constraints mirror hard constraints: Voice and person in English and Lummi. In *Proceedings of the LFG01 Conference* (pp. 13-32).

Chater, Nick, Crocker, Matthew, & Pickering, Martin. (1998). The rational analysis of inquiry: the case of parsing. In Nick Chater, Matthew Crocker & Martin J. Pickering (Eds.), *Rational models of cognition.* Oxford: Oxford University Press.

Chung, Sandra (2012). On reaching agreement late. In Andrea Beltrama, Tasos Chatzikonstantinou, Jackson L. Lee, Mike Pham and Diane Rak (Eds.). *Proceedings from the Annual Meeting of the Chicago Linguistic Society, Volume 48* (pp. 169-90). Chicago, IL: Chicago Linguistic Society.

Clifton, Charles, & Staub, Adrian (2008). Parallelism and Competition in Syntactic Ambiguity Resolution.

Language and Linguistics Compass, *2*, 234-50. doi:10.1111/j.1749-818X.2008.00055.x.
Clothier-Goldschmidt, Scarlett (2015). *The distribution and processing of referential expressions: evidence from English and Chamorro* (Master's thesis). University of California, Santa Cruz.
Crain, Steven, & Fodor, Janet D. (1985). How can grammars help parsers? In David Dowty, Lauri Kartunnen, & Arnold Zwicky (Eds.), *Natural language parsing* (pp. 94– 128). Cambridge: Cambridge University Press.
Ericsson, K. A., & Kintsch, W. (1995). Long-Term Working Memory. *Psychological Review*, *102*(2), 211–45. doi:10.1037/0033-295X.102.2.211.
Faul, Franz, Erdfelder, Edgar, Buchner, Axel, & Lang, Albert-Georg (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*, 1149-60.
Fodor, Janet D. (1978). Parsing Strategies and Constraints on Transformations. *Linguistic Inquiry*, *9*(3), 427–73.
Ford, Marilyn. (1983). A method for obtaining measures of local parsing complexity throughout sentences. *Journal of Verbal Learning and Verbal Behavior, 22*, 203–18.
Frazier, Lyn (1987). Syntactic processing: Evidence from Dutch. *Natural Language and Linguistic Theory*, *5*, 519–60.
Gennari, Silvia P., & MacDonald, Maryellen C. (2008). Semantic indeterminacy in object relative clauses. *Journal of Memory and Language, 58*, 161–87.
Gennari, Silvia P., & MacDonald, Maryellen C. (2009). Linking production and comprehension processes: The case of relative clauses. *Cognition, 111,* 1–23.
Gordon, Peter C., Hendrick, Randall, & Johnson, Marcus (2001). Memory interference during language processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 1411–1423.
Gordon, Peter C., Hendrick, Randall, & Johnson, Marcus (2004). Effects of noun phrase type on sentence complexity. *Journal of Memory and Language, 51*, 97–114.
Gordon, Peter C., & Lowder, Matthew W. (2012). Complex sentence processing: A review of theoretical perspectives on the comprehension of relative clauses. *Language and Linguistics Compass, 6,* 403–415.
Graff, David, Kong, Jungbo, Chen, Ke, & Maeda, Kazuaki (2005). English Gigaword Second Edition LDC2005T12. Philadelphia: Linguistic Data Consortium. Retrieved from https://catalog.ldc.upenn.edu/LDC2005T12.
Hale, John (2001). A probabilistic Earley parser as a psycholinguistic model. In *Proceedings of the 2nd meeting of the North American Chapter of the Association for Computational Linguistics* (pp. 159–66). Retrieved from: http://www.aclweb.org/anthology/N /N01/N01-1021.pdf.
Hale, John (2006). Uncertainty about the rest of the sentence. *Cognitive Science, 30*, 609–42.
Hedberg, Nancy (2000). The referential status of clefts. *Language, 76*(4), 891-920.
Halekoh, Ulrich, & Højsgaard, Søren (2014). A Kenward-Roger Approximation and Parametric Bootstrap Methods for Tests in Linear Mixed Models - The R Package pbkrtest. *Journal of Statistical Software, 59(9)*, 1-30. Retrieved from http://www.jstatsoft.org/v59/i09/.
Holmes, V. M., & O'Regan, J. K. (1981). Eye fixation patterns during the reading of relative-clause sentences. *Journal of Verbal Learning and Verbal Behavior, 20*, 417–30.
Jackendoff, Ray, & Culicover, Peter (1971). A reconsideration of dative movements. *Foundations of Language, 7,* 307-412.
Keenan, Edward L., & Comrie, Bernard. (1977). Noun phrase accessibility and universal grammar. *Linguistic Inquiry*, *8*, 63-99.
Lee, Ming-Wei (2004). Another look at the role of empty categories in sentence processing (and grammar). Journal of Psycholinguistic Research, *33*, 51–73.
Levy, Roger (2008). Expectation-based syntactic comprehension. *Cognition*, *106*(3), 1126–77. doi:10.1016/ j.cognition.2007.05.006.
Lewis, Richard L. (2000). Specifying architectures for language processing: process, control, and memory in parsing and interpretation. In Matthew W. Crocker et al. (Eds.), *Architectures and Mechanisms for Language Processing*. Cambridge, UK: Cambridge University Press.
Lewis, Richard L., Vasishth, Shravan, & Van Dyke, Julie A. (2006). Computational principles of working memory in sentence comprehension. *Trends in Cognitive Sciences*, *10*(10), 447–54. doi:10.1016/j.tics.2006.08.007.
Lowder, Matthew W., & Gordon, Peter C. (2014). Effects of animacy and noun-phrase relatedness on the processing of complex sentences. *Memory and Cognition, 42,* 794-805. doi:10.3758/s13421-013-0393-7.
Mak, Willem M., Vonk, Wietske, & Schriefers, Herbert (2002). The influence of animacy on relative clause processing. *Journal of Memory and Language, 47,* 50–68.
Merlo, Paola, & Stevenson, Suzanne (2000). Lexical syntax and parsing architecture. In Matthew W. Crocker et al. (Eds.), *Architectures and Mechanisms for Language Processing*. Cambridge, UK: Cambridge University Press.
Omaki, Akira, Lau, Ellen F., Davidson White, Imogen, Dakan, Myles L., Apple, Aaron, & Phillips, Colin (2015). Hyper-active gap filling. *Frontiers in Psychology*, *06*, 1–18. doi:10.3389/fpsyg.2015.00384.
Poeppel, David, & Monahan, Philip J. (2011). Feedforward and feedback in speech perception: Revisiting analysis by synthesis. *Language and Cognitive Processes, 26(7),* 935–51. doi:10.1080/01690965.2010.493301.

Prince, Alan, Smolensky, Paul. (1993). Optimality Theory: Constraint Interaction in Generative Grammar. *RuCCS Technical Report No. 2,* Piscataway, NJ: Rutgers University Center for Cognitive Science.

Roland, Douglas, Dick, Frederic, & Elman, Jeffrey L. (2007). Frequency of basic English grammatical structures: A corpus analysis. *Journal of Memory and Language*, *57*(3), 348–379. doi:10.1016/j.jml.2007.03.002.

Rohde, Douglas T. (2003). Linger [Computer software]. http://tedlab.mit.edu/~dr/Linger/.

Rohde, Douglas T. (2005). TGrep2 [Computer software]. http://tedlab.mit.edu/~dr/Tgrep2/.

Silverstein, Michael. (1976). Hierarchy of Features and Ergativity. In R. M. W. Dixon (Ed.), *Grammatical Categories in Australian Languages* (pp. 112-71). Canberra: Australian Institute of Aboriginal Studies.

Staub, Adrian (2010). Eye movements and processing difficulty in object relative clauses. *Cognition, 116,* 71–86.

Smolensky, Paul. (2006). Harmony in Linguistic Cognition. *Cognitive Science, 30,* 779-801.

Stowe, Laurie (1986). Parsing wh–constructions: Evidence for on–line gap location. *Language and Cognitive Processes, 1*, 227–46.

Traxler, Matthew J., Williams, Rihana S., Blozis, Shelley A., & Morris, Robin K. (2005). Working memory, animacy, and verb class in the processing of relative clauses. *Journal of Memory and Language, 53,* 204–24.

Tutunjian, Damon, & Boland, Julie (2008). Do we need a distinction between arguments and adjuncts? evidence from psycholinguistic studies of comprehension. *Language and Linguistics Compass, 2*(4)*,* 631-46. doi: 10.1111/j.1749-818x.2008.00071.x.

Wagers, Matthew W., & Phillips, Colin (2014). Going the distance: Memory and control processes in active dependency construction. *The Quarterly Journal of Experimental Psychology*, *67*(7), 1274-1304.

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

Wanner, Eric, & Maratsos, Michael (1978). An ATN approach to comprehension. In Morris Halle, Joan Bresnan, & George A. Miller (Eds.), *Linguistic theory and psychological reality* (pp. 119–61). Cambridge, MA: MIT Press.