COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science 45 (2021) e12927

ISSN: 1551-6709 online
DOI: 10.1111/cogs.12927

Quantifying Structural and Non-structural Expectations in Relative Clause Processing

Zhong Chen,[a] John T. Hale[b]

[a]*Department of Modern Languages and Cultures, Rochester Institute of Technology*
[b]*Department of Linguistics, University of Georgia*

Received 30 November 2019; received in revised form 16 October 2020; accepted 29 October 2020

Abstract

Information-theoretic complexity metrics, such as Surprisal (Hale, 2001; Levy, 2008) and Entropy Reduction (Hale, 2003), are linking hypotheses that bridge theorized expectations about sentences and observed processing difficulty in comprehension. These expectations can be viewed as syntactic derivations constrained by a grammar. However, this expectation-based view is not limited to syntactic information alone. The present study combines structural and non-structural information in unified models of word-by-word sentence processing difficulty. Using probabilistic minimalist grammars (Stabler, 1997), we extend expectation-based models to include frequency information about noun phrase animacy. Entropy reductions derived from these grammars faithfully reflect the asymmetry between subject and object relatives (Staub, 2010; Staub, Dillon, & Clifton, 2017), as well as the effect of animacy on the measured difficulty profile (Lowder & Gordon, 2012; Traxler, Morris, & Seely, 2002). Visualizing probability distributions on the remaining alternatives at particular parser states allows us to explore new, linguistically plausible interpretations for the observed processing asymmetries, including the way that expectations about the relativized argument influence the processing of particular types of relative clauses (Wagers & Pendleton, 2016).

Keywords: Sentence processing; Relative clause; Animacy; Entropy reduction; Surprisal

1. Introduction

There is growing evidence in psycholinguistics suggesting that sentence comprehension is predictive. Each new word introduces information that helps readers to shape expectations about the rest of the sentence (Hale, 2006). Computational models of sentence processing based on this idea reflect syntactic constraints and structural frequencies about

Correspondence should be sent to Zhong Chen, Eastman Building 3179, 18 Lomb Memorial Drive, Rochester Institute of Technology, Rochester, NY 14623. E-mail: z.chen@rit.edu

grammatical category, phrasal hierarchy, as well as local and non-local dependencies (Demberg, Keller, & Koller, 2013; Hale, 2001, 2003; Levy, 2008).

This information that comprehenders work with can be divided into two kinds: "structural" information that is rightfully viewed as part of the grammar and "non-structural" information that may be extragrammatical. Agreement affixes are a classic example of the first category. A verb that fails to agree with a preceding noun (in the appropriate structural relationship) is simply not expected. The animacy status of a noun phrase (NP) —whether its referent is alive or not—exemplifies another kind of information. The power of animacy over sentence processing is evident in languages like Chinese and German (Li, Bates, & MacWhinney, 1993; MacWhinney, Bates, & Kliegl, 1984). Section 3.2 reviews a body of evidence suggesting this non-structural property plays an important role in English as well, a language where it arguably is not grammaticalized.

The agenda of this study is to combine both types of information in unified models of word-by-word sentence processing difficulty. Its empirical domain is relative clauses (RCs). Since the 1970s RCs have challenged cognitive scientists to recognize the filler–gap relationship in processing models (Kaplan, 1972; Wanner & Maratsos, 1978). Subsequent research spotlighted the role of animacy in the comprehension of RCs (Gennari & MacDonald, 2008; Traxler, Morris, & Seely, 2002). These two considerations have led us to define a formal grammar that combines animacy with a movement analysis of fillers and gaps. By making this grammar probabilistic, we join a consensus across cognitive science that frequency exerts an influence on the processing not just of individual words but larger units as well, even RC subtypes. This phenomenon has been characterized using information-theoretical complexity metrics such as surprisal (for a review, see Hale, 2016). But the nature of its interaction with non-structural cues remains to be seen. This open question—how to combine independently acknowledged explanatory factors—is a long-standing goal in cognitive science and the central concern of this paper.[1]

Specifically, we compare quantitative reading time predictions made by surprisal and an alternative complexity metric, entropy reduction (Hale, 2003), against human data from reading RCs. Surprisal quantifies each incoming word's expectedness, whereas entropy reduction links the reduction of comprehenders' uncertainty to their disambiguation effort. Both metrics accommodate the non-structural animacy feature, in addition to those traditionally thought of as structural, for example, verb transitivity. By examining syntactic derivations that are still "in play" at the target words, we identify linguistically plausible explanations for the effects of reading difficulties in sentence comprehension.

The rest of this paper is organized as follows. Section 2 surveys theoretical accounts of RC processing, including a detailed introduction of the surprisal and entropy reduction hypotheses. Section 3 reviews key empirical evidence in the literature, in particular, the SUBJECT ADVANTAGE and the ANIMACY EFFECT. Section 4 describes the procedure for making word-by-word processing difficulty predictions. In Section 5, we show that entropy reduction does a better job than surprisal in deriving the subject–object asymmetry throughout the RC region (Staub, 2010; Staub, Dillon, & Clifton, 2017). Its quantitative predictions are also compatible with processing effects observed in experiments with various animacy manipulations (Lowder & Gordon, 2012; Traxler et al., 2002). We discuss the results and future improvements in Sections 6 and 7, before concluding the paper in Section 8.

2. Principles proposed to account for RC processing

Relative clauses have been a perennial topic of interest within the cognitive science of language. Their particular syntactic structure and distinctive processing profile has fascinated psychologists, linguists, and computer scientists alike (see, e.g., Wanner & Maratsos, 1978). A pair of RC examples in English is shown below:

(1) a. Subject relatives (SRs).

The reporter$_i$ who e_i attacked the senator left the room.

b. Object relatives (ORs).

The reporter$_i$ who the senator attacked e_i left the room.

In (1a), the NP "the reporter" serves as the RC head. It is not only the subject of the matrix clause, but also the agent of the embedded verb "attacked." Relativizing "the reporter" from the underlying subject position within the RC is notated with a gap, symbolized by the empty category e_i. In (1b), the same head noun is co-indexed with an empty element at the embedded object position following the verb. The indices in (1) indirectly suggest which interpretation the construction receives. For example, the logical object of "attacked" is identical to the matrix subject "reporter" in ORs.

The processing of RCs provides an ideal test case for general sentence processing principles, which we briefly review in this section. A large selection of literature documents a robust finding that SRs are easier to process than ORs, known as the subject advantage. This processing asymmetry has been observed in a variety of different measures, including: reading times (Grodner & Gibson, 2005; King & Just, 1991), eye-tracking (Staub, 2010; Traxler et al., 2002), ERPs (King & Kutas, 1995), fMRI (Just, Carpenter, Keller, Eddy, & Thulborn, 1996), and PET (Stromswold, Caplan, Alpert, & Rauch, 1996). It has also been found in languages other than English, including those where RCs appear before the head noun (Lin, 2008) and a language that allows both post-nominal and pre-nominal RCs (Wagers, Borja, & Chung, 2018).

Among the processing principles advanced as candidate explanations for the universal subject advantage pattern, recent studies have appealed particularly to working-memory theories and expectation-based accounts. In general, the former explain the pattern in terms of reduced memory load in SRs compared to ORs, whereas the latter suggest that readers have higher structural expectations for the more frequent SRs. In this section, we discuss several representative proposals that fall under those two classes of sentence processing principle.

2.1. *Working-memory theories*

Gibson and colleagues (Gibson, 1998, 2000; Grodner & Gibson, 2005; Warren & Gibson, 2002) have sought to relate the comprehension difficulty of RCs to working memory processes. In particular, under the Dependency Locality Theory (DLT; Gibson, 2000), the sentence-initial head noun is kept in the working memory until reading the embedded verb. In ORs, the longer distance between the head noun and the embedded verb results in extra storage and memory retrieval costs.

Lewis and colleagues (Lewis & Vasishth, 2005; Lewis, Vasishth, & Van Dyke, 2006) have emphasized the role of memory retrieval in a model that essentially applies Anderson et al.'s (2004) ACT-R cognitive architecture to the special domain of sentence processing.[2] Their model has greater coherence with other cognitive theories and acknowledges the locality effect such that constructions involving long-distance dependencies bring heavier working memory load and cause longer processing delays. Their model is also more explicit in addressing the interference effect than earlier memory-based proposals, such as the HOLD hypothesis (Wanner & Maratsos, 1978) and the DLT. Processing English ORs imposes additional memory burden, because the RC head and the embedded subject compete in the memory retrieval process at the embedded verb. This kind of inhibitory interference in building subject–verb dependencies is similarity-based, since both candidates are NPs. Evidence for the similarity-based interference has also been reported in processing reflexive-antecedent dependencies and on other processing cues, such as number, gender, and animacy (Jäger, Engelmann, & Vasishth, 2017).

However, the predictions made by working-memory theories are at times inconsistent with the observed comprehension difficulty profile in pre-nominal RCs. For example, the subject advantage has been reported in Chinese (Lin & Bever, 2006),[3] Japanese (Ishizuka, Nakatani, & Gibson, 2003; Nakamura & Miyamoto, 2013), and Korean (Kwon, Kluender, Kutas, & Polinsky, 2013; Kwon, Lee, Gordon, Kluender, & Polinsky, 2010), despite the fact that the embedded verb is farther away from the head noun in SRs than in ORs in those languages.

2.2. *Expectation-based accounts*

The difficulty of processing RCs has also been explained in terms of readers' expectations regarding alternative syntactic constructions. Perhaps the most direct way that reflects the expectations on SRs and ORs is their relative frequencies in a corpus sample. Readers have more experience with SRs because they are more common in almost all languages, including English (Roland, Dick, & Elman, 2007).

2.2.1. *Surprisal*

One instantiation of the idea that connects sentence processing difficulties and readers' expectations is surprisal (Hale, 2001; Levy, 2008), which has been proven useful across many methodologies, for example, eye-tracking (Boston, Hale, Kliegl, Patil, & Vasishth, 2008; Demberg & Keller, 2008), ERPs (Frank, Otten, Galli, & Vigliocco, 2015), and fMRI (Brennan, Stabler, Van Wagenen, Luh, & Hale, 2016; Hale, Lutz, Luh, & Brennan, 2015; Henderson, Choi, Lowder, & Ferreira, 2016). Proposed to reflect the degree of "surprise" when reading each word from left to right, surprisal is defined below in (2) as the log-ratio between the forward probabilities of string prefixes before and after the *n*th word in a sentence.

$$Surprisal(w_n) = \log_2 \frac{P(w_0 \ldots w_{n-1})}{P(w_0 \ldots w_n)}. \tag{2}$$

The key idea of surprisal is that these expectations are conditioned on the left context. This contrasts with the Tuning Hypothesis (Mitchell, Cuetos, Corley, & Brysbaert, 1995), which crystallized the idea that human comprehension is tuned to unconditional construction frequency. However, the assumption is similar: rare constructions like ORs involve a derivation that must use a low-probability rule. At a certain point this derivation is forced by the left context. This suffices to derive the subject advantage. But in many cases surprisal—in combination with reasonable grammars—predicts effort on the wrong word. For example, it fails to capture the processing difficulty profile in RCs and in other long-distance syntactic dependencies (Levy & Gibson, 2013).

2.2.2. *Entropy Reduction*

Hale (2003, 2006) revived Wilson and Carroll's (1954) Entropy Reduction idea and applied it to the analysis of sentence processing asymmetries. In this scenario, the information-theoretic notion, Entropy, is defined below in (2). The random variable X might take values that are derivations of a probabilistic grammar. One could further specialize X to reflect derivations proceeding from various grammatical categories, for example, noun phrase (NP), verb phrase (VP), sentence (S), etc. The expression $H(S)$ reflects the average uncertainty of guessing any derivation generated by a grammar with the start symbol "S."

$$H(X) = -\sum_{x \in X} p(x) \log_2 p(x). \tag{3}$$

Like probability, entropy can be conditional. Letting $w_1 w_2 \ldots w_n$ stand for an n-word initial substring of a sentence generated by a grammar whose start symbols is S, the notation $H(S|w_1 w_2 \ldots w_n)$ denotes the conditional entropy of just those S-derivations whose first n-words are $w_1 w_2 \ldots w_n$. This is the uncertainty about the rest of the sentence, which can be calculated using standard techniques from computational linguistics such as chart parsing (Bar-Hillel, Perles, & Shamir, 1964; Nederhof & Satta, 2008).[4] By abbreviating $H(S|w_1 w_2 \ldots w_n)$ with H_n, formula (4) defines the complexity metric ER_n as the difference between conditional entropies before and after w_n, a particular word at a particular position in a sentence.

$$ER_n = \begin{pmatrix} H_{n-1} - H_n & \text{if this difference is positive} \\ 0 & \text{otherwise} \end{pmatrix}. \tag{4}$$

The background assumption of entropy reduction is that human sentence comprehension is making progress toward a disambiguated parser state. Unless the sentence is globally ambiguous, readers' uncertainty will be brought to zero at the end of the utterance. Intuitively, sentence-medial disambiguations occur whenever the uncertainty about the sentence's structure goes down. This is because readers' "beliefs" in various syntactic alternatives take on a more concentrated probability distribution (Jurafsky, 1996). In contrast, a new word could also cause the structural uncertainty to increase locally, which results in a negative entropy reduction value. It happens when the remaining syntactic alternatives become more disorganized, such that there exist many equiprobable syntactic expectations, or the new word opens up more derivations than it closes down. In those cases, the comprehender has made no progress toward the goal of a unique reading and any negative value calculated for those transitions is only considered as zero (Hale, 2016, p. 405). The entropy reduction hypothesis generalizes the idea that "flipping the preferred interpretation" of a sentence prefix leads to delays in sentence processing (Narayanan & Jurafsky, 2002), except that the "flip" only counts if the reader moves toward a less-confused state of mind. It therefore differs from surprisal in terms of how the structural frequency information is reflected in processing difficulties.[5]

The entropy reduction hypothesis has been applied to both naturalistic text stimuli (Frank, 2013; Wu, Bachrach, Cardenas, & Schuler, 2010) and controlled experimental materials (Linzen & Jaeger, 2015). A number of studies have also shown that it correlates with measured reading times and neural signals related to sentence-medial ambiguities, including those observed in processing RCs (Chen, Jäger, & Hale, 2012; Frank, 2013; Lowder, Choi, Ferreira, & Henderson, 2018; Nelson, Dehaene, Pallier, & Hale, 2017; Yun, Chen, Hunter, Whitman, & Hale, 2015). In the next section, we review some of the most crucial experimental findings in RC processing and discuss in detail the predictions made by different processing principles.

3. Experimental observations during RC comprehension

3.1. The subject–object asymmetry

The subject advantage in English RCs is robust, making SRs easier to comprehend than ORs. However, one of the major differences among previous studies is the regions where the additional difficulty sets in for ORs. Investigating the loci of the subject advantage is especially important in evaluating different processing theories.

In a self-paced reading experiment, Grodner and Gibson (2005) explored whether the distance between the head noun and the embedded verb is a determinant of reading difficulty in RCs. They found that ORs were read slower than SRs at the RC verb, but not at the RC-internal NP, as shown in Table 1. Working-memory theories, such as the DLT, attribute this result to the additional integration cost at the embedded OR verb due to its preceding subject NP. In contrast, expectation-based surprisal was unable to account for this finding and instead predicted an early processing difficulty at the onset of the NP,

Table 1
Mean reading times (ms) at the same word suggest a preference of SRs at the embedded verb, as highlighted in bold (Grodner & Gibson, 2005, Table A1). The column follows the word order in ORs from left to right

Type	*who*	Det	Noun	RC Verb
SR	349.8	334.3	384.0	**354.8**
OR	343	348.1	357.6	**422.0**

namely the determiner (Hale, 2001; Levy, 2008). This is because readers have a higher expectation for the more frequent SRs upon seeing the relative pronoun "who." While anticipating a verb as the next word, the determiner will come as a surprise and slows down the comprehension. Levy (2008) therefore explained the late effect in ORs as the processing spillover from the preceding embedded NP. The entropy reduction hypothesis, on the other hand, faithfully derives the reading difficulty at the embedded OR verb by focusing on the processing cost of disambiguation (Hale, 2003). In ORs, the uncertainty about the rest of the sentence is greatly reduced by the embedded verb as comprehenders become certain that there will be no recursive modification after.

Staub (2010) tested the same sort of stimuli as Grodner and Gibson (2005) in an eye-tracking study. His Experiment 1 replicated the processing slow-down at the embedded OR verb in all three reading time measures, that is, first fixation duration, gaze duration, and go-past time. Importantly, the effect of clause type also appeared to be significant early at the NP, which is consistent with surprisal's prediction. The subject advantage was therefore statistically significant throughout the RC region. In addition, readers were more likely to look back when they read the NP in ORs, as evidenced by the higher probability of regressive saccades in Table 2. Using a complement clause as the baseline, Experiment 2 of the same paper further showed that the penalty in processing ORs was larger, and presented in more eye-tracking dependent measures, on the RC subject than on the verb.

The findings in Staub (2010) were replicated in a more recent eye-tracking study with higher statistical power (Staub et al., 2017). The penalty in ORs was quantitatively greater on the subject than on the RC verb, driven primarily by regressive eye movements. Other experimental paradigms, such as the maze task (Forster, Guerrera, & Elliot,

Table 2
The subject advantage, highlighted in bold, was found throughout the RC region in Experiment 1 of Staub (2010), with a larger effect at the NP, primarily driven by regressive eye movements out of ORs

	Type	*who/that*	Det	Noun	RC Verb
Go-past time (ms)	SR	283	**272**	**375**	**333**
	OR	303	**382**	**459**	**420**
Regression proportion	SR	0.12	**0.12**	**0.16**	0.17
	OR	0.11	**0.36**	**0.40**	0.15

2009) and the self-guided reading (Hatfield & Artos, 2016), were also able to obtain the effect at the embedded NP. Staub and colleagues argue that both experience-based expectations and the memory retrieval processes play a role in processing RCs.[6] Their argument is in a way similar to the "two-factor" proposal by Demberg and Keller (2009) to account for patterns in reading time data for naturally occurring RCs, where the role of readers' expectation (more specifically, surprisal) is "somewhat curtailed" (Hale, 2016, p. 404).[7] As demonstrated in Section 5.1, the alternative expectation-based complexity metric, entropy reduction, derives a prediction that more closely matches the difficulty profile observed by Staub et al. (2017). The result distinguishes itself from surprisal, which has unfortunately failed to predict any increased reading difficulty at the embedded verb in ORs.

3.2. The role of animacy in RC processing

Previous research has found that the reading difficulty of RCs is modulated by the animacy status of the RC head and the embedded NP. Evidence supporting the so-called animacy effect is crosslinguistic (English: Traxler et al., 2002; Traxler, Williams, Blozis, & Morris, 2005, Dutch: Mak, Vonk, & Schriefers, 2002, 2006; Chinese: Hsiao & MacDonald, 2016; Wu, Kaiser, & Andersen, 2012). Although RCs with two animate NPs have been commonly tested in empirical studies, they are not the most frequent type in natural language usage (Roland et al., 2007). For example, ORs tend to modify an inanimate head, because more verbs take inanimate than animate objects.

Traxler et al. (2002) tested the processing of RCs by manipulating the animacy of NPs involved. In two of the four conditions in (5), the head noun was animate while the embedded noun was inanimate. In the other two conditions, the head noun was inanimate while the noun in the RC was animate. Eye-tracking results in Table 3 show that ORs were read faster if they followed an inanimate head than an animate head.[8] The subject advantage, which we define here as the difference between SRs and ORs in the RC region, was smaller when the RC head was inanimate. Gennari and MacDonald (2008) have argued that the animacy effect can be seen as frequency-driven. ORs with inanimate heads are easier because objects are more likely to be inanimate. On the contrary, ORs modifying an animate head contradict the comprehenders' language experience. Traxler et al. (2005) tested the same set of sentences among three groups of participants with

Table 3
Eye-tracking results on RCs with various animacy patterns (Traxler et al., 2002). The subject advantage is defined as the numerical difference between SRs and ORs headed by an NP with the same animacy status

Type	Head NP	Embedded NP	Quasi-first-pass (ms)	Subject Advantage
SR	+anim	−anim	747	153
OR	+anim	−anim	900	
SR	−anim	+anim	752	44
OR	−anim	+anim	796	

different working memory capacity. A reduced subject advantage was found across the board in RCs with an inanimate head. Participants with high working memory capacity, in particular, read ORs faster if the head noun was inanimate.

(5) The four conditions tested in Experiment 3 of Traxler et al. (2002)

a. SR, head NP: +animate, embedded NP: −animate

The director that watched the movie received a prize.

b. OR, head NP: +animate, embedded NP: −animate

The director that the movie pleased received a prize.

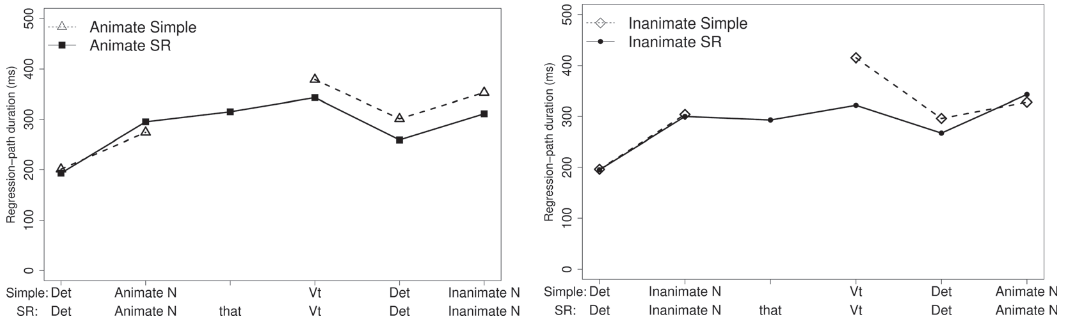
c. SR, head NP: −animate, embedded NP: +animate

The movie that pleased the director received a prize.

d. OR, head NP: −animate, embedded NP: +animate.

The movie that the director watched received a prize.

Table 3 also suggests that, contrary to ORs, the reading time was largely unchanged in SRs, regardless of the head noun's animacy status. This result has allowed Lowder and Gordon (2012) to investigate the relationship between the NP animacy and the syntactic structure in an eye-tracking experiment using sentences in (6), where one of the SR conditions had an animate head and the other had an inanimate head. Two simple sentence conditions were also created by dropping the relative pronoun "that." The subject–verb dependency was therefore within the same clause in (6b) and (6d), but crossed one clausal boundary in (6a) and (6c). Fig. 1 illustrates the mean regression-path durations across



(a) Sentences with an animate subject (b) Sentences with an inanimate subject

Fig. 1. Mean regression path durations in reading the four conditions in (6) (Lowder & Gordon, 2012). For easier comparison with our modeling results, words are labeled by abstract categories, for example, "Vt" for transitive verbs.

the four conditions. In Fig. 1b the less-expected inanimate subject led to a significant delay at the matrix verb when the subject–verb dependency was formed in simple sentences. In comparison, at the embedded verb, SRs were read much faster. The same effect at the verb was not found in Fig. 1a when the sentence-initial NP was animate. As a result, the animacy-driven effect at the verb was mitigated when the subject and the verb belonged to different clauses, in this case, the matrix clause and the RC.

(6) The conditions of Experiment 2 in Lowder and Gordon (2012)

a. SR, +animate subject

The cowboy [**that** concealed the pistol] was known to be unreliable.

b. Simple sentence, +animate subject

The cowboy concealed the pistol last night in the saloon.

c. SR, −animate subject

The pistol [**that** injured the cowboy] was known to be unreliable.

d. Simple sentence, −animate subject

The pistol injured the cowboy last night in the saloon.

Lastly, Wagers and Pendleton (2016) tested whether different animacy patterns trigger different expectations and in turn modulate the difficulty of RCs. In two self-paced reading experiments using a filled-gap design (Crain & Fodor, 1985; Stowe, 1986), readers were more likely to predict a subject gap given an animate argument. If an NP filled the gap position at a later time, a reanalysis was needed. No predicative linkage was found between an inanimate head noun and the subject gap. Based on statistical estimates and cloze results, Wagers and Pendleton hypothesized that inanimate NP fillers are more "equivocal" and do not generate a specific syntactic expectation. In the present work, we seek to provide a more detailed explanation in Section 6 by visualizing the parser states conditioned on NP fillers with different animacy values.

4. From grammar to processing difficulty predictions

This section lays out the general procedure behind the modeling results reported next in Section 5. We used probabilistic grammars to formalize both structural and non-structural factors that figure prominently in the psycholinguistic literature discussed above in Section 3. By computing probability distributions at sentence-internal positions and updating these distributions with information from successive words, we used information-theoretic complexity metrics (see Section 2.2) to derive predictions about word-by-word comprehension difficulty.

4.1. Formal grammar as a route to cognitively plausible explanation

An important aspect of this project is its linguistic interpretability. To achieve this, we used formal grammars that specify exactly which analyses the model is entertaining. This "interpretability-for-free" property contrasts starkly with extensional approaches based solely on model outputs. The extensional approach is typified by research that uses deep-learning neural nets to achieve very high language model performance on held-out text corpora. These high performance levels typically come at the cost of decreased interpretability. Such neural nets are notoriously prone to learning irrelevant correlations (Ettinger, 2020; Futrell et al., 2019; Kuncoro et al., 2018; van Schijndel, Mueller, & Linzen, 2019). To avoid this problem and focus our investigation more squarely on structural constraints like locality in Grodner and Gibson (2005) and non-structural factors such as animacy in Traxler et al. (2002), we instead proceed with an explicit grammar whose generalization ability rests upon well-chosen syntactic analyses.

In particular, we used Minimalist Grammars (MGs; Stabler, 1997) to express the syntactic relationship between filler and gap that is manifested in an RC. MGs are a transformational grammar formalism that adopts ideas from the Minimalist Program (Chomsky, 1995). Stabler's formalization involves two generalized transformations: merge and move. Merge is a binary rule, analogous to ordinary context-free grammar rules or function application in categorial grammars (Berwick & Epstein, 1995). Move, on the other hand, is unary and non-concatenative. In our case, it explicitly relates gaps to fillers in their surface positions. Using a translator written by Guillaumin (2005), MGs can be converted to equivalent Multiple Context Free Grammars (MCFGs; Seki, Matsumura, Fujii, & Kasami, 1991) whose derivations exhibit rich non-local dependencies, including crossing dependencies, that we find in natural language (Clark, 2014). This level of complexity is what Stabler identifies as the "hidden consensus" (Stabler, 2013) around the explanatory role of formal grammar in human sentence processing.[9] In virtue of being mildly context-sensitive, conditional entropies from probabilistic MCFGs can be calculated exactly using software like the Cornell Conditional Probability Calculator (CCPC).[10] Hunter (to appear) provides an engaging overview of how these sorts of expressive grammars can contribute to sentence processing and experimental syntax research, advocating for MGs in particular as part of an integrated account of language and cognition.[11]

4.2. Assigning weights to grammar rules

Ever since the pioneering Tuning Hypothesis (Mitchell et al., 1995), it has been recognized that human comprehenders' expectations reflect corpus statistics. Seeking to zoom in on specific, theoretically relevant conditions (e.g., ditransitive verb embedded in an OR), our model relies on treebanks to estimate the probability of particular grammar rules. The CCPC tool then calculates the probability of entire derivations (Chen, Hunter, Yun, & Hale, 2014).[12]

Table 4 is an example of 10 derivations with highest probabilities generated by a hypothetical grammar fragment. Although in this table each probability number seems to be associated with a grammar-produced "string,"for example, "he matter –ed," they are in fact the probability of a syntactic derivation. Therefore, it is possible that an ambiguous "string" shows up more than once in this kind of list in which it corresponds to multiple derivations.

As described earlier, the quantities that the entropy reduction hypothesis advances as a cognitive model are conditional entropies. These values reflect uncertainties about every analysis for every grammatically plausible sequence of words that can follow a given string prefix. To compute these conditional entropies, the CCPC tool uses chart parsing to recover probabilistic "intersection" grammars conditioned on each string prefix of the target sentence (Nederhof & Satta, 2008). An intersection grammar derives all and only the derivations that are consistent with the initial string prefix. It implicitly defines comprehenders' expectations about how the sentence would continue. Given the string prefix, the conditional entropy models the degree of confusion that a reader is in at that point in the sentence. Comparing the conditional entropies before and after a new word, any decrease quantifies the disambiguation work that, ideally, could have been done at that word. The CCPC tool also records the "remainder set" of syntactic alternatives from the intersection grammar to get an intuitive picture of how uncertainties are reduced during parsing. For example, given an intersection grammar conditioned on the one-word string prefix "David," it outputs all derivations that are still "in play," with the most probable ones listed in Table 5. They share the same highlighted string prefix, in contrast to those generated by a non-intersection grammar in Table 4. These syntactic alternatives are essential in understanding how entropy reduction values are calculated. More importantly, the remainder sets shed light on linguistic interpretations for the processing effect observed at a given word, especially when the effect is driven by disambiguation.

Table 4
Ten example derivations with highest probabilities generated by a hypothetical grammar fragment. They are equivalent to the remaining syntactic alternatives conditioned on a null string prefix

Probability	Syntactic Alternatives
0.00379689	*he matter -ed*
0.00315572	*David matter -ed*
0.00236679	*Sally matter -ed*
0.00221223	*they matter -ed*
0.00221223	*I matter -ed*
0.00144425	*the treat be -s clever*
0.00129323	*the treat be -s young*
0.00129323	*the treat be -s right*
0.00129323	*the treat be -s poor*
0.00099117	*the treat be -s strange*

Table 5
Ten derivations generated by a hypothetical "intersection" grammar with highest conditional probabilities given the one-word string prefix, "David"

Conditional Probability	Remaining Alternatives
0.04162610	***David*** *matter -ed*
0.00815858	***David*** *doesn't matter*
0.00607872	***David*** *matter -s*
0.00594986	***David*** *matter*
0.00263401	***David*** *pay -ed for the treat*
0.00178260	***David*** *sell -ed the treat*
0.00177774	***David*** *tell -ed the treat*
0.00173645	***David*** *get -ed the treat*
0.00173645	***David*** *leave -ed the treat*
0.00167712	***David*** *have -ed pay -en for the treat*

5. Modeling the processing of RCs

In this section, we lay out our modeling work in detail. Following the procedure introduced in Section 4, we calculated full entropy values for the parser states during incremental comprehension. The reductions of uncertainty degree at the target words are consistent with two major processing effects, namely the subject advantage and the animacy effect. We show that the entropy reduction predictions are better aligned with processing difficulty patterns related to sentence-medial disambiguation, whereas the surprisal metric is more sensitive to reading time delays driven by low frequencies.

5.1. Deriving the subject advantage

The parser's degree of uncertainty at any time hinges on grammatical alternatives given the sentence prefix. Therefore, it is necessary to identify non-trivial linguistic parameters that generate alternative derivations to compete with the target structure. This is particularly important for post-nominal RCs, because the remainder set of syntactic derivations after the relative pronoun or complementizer amount to a collection of variations on the ultimate, winning RC structure. The parameters we examined crosscut distinctions between different construction types and are shown in Table 6. We wrote a grammar fragment to cover all of these parameters, each of which had been used in previous experiments on RC processing.[13] When the human parser makes different choices concerning those parameters, it essentially considers various syntactic derivations as necessary complements to the syntactic treatment of the target RC itself.

Most transformational syntax literature on post-nominal RCs focuses on the *wh*-movement of NP elements, namely, how the object "reporter" in (7a) moves to the sentence-initial position, with a RC attached to its right in (7b). We adopted the promotion analysis (Bianchi, 2002; Kayne, 1994; Vergnaud, 1985) such that both the relative pronoun "who" and the head noun "reporter" are promoted to the left edge of RC.[14]

Table 6
The linguistic parameters incorporated in the grammar fragment of modeling the subject-object asymmetry in relative clauses

Parameter	Representative Study
Full NP vs. pronoun	Warren and Gibson (2002)
Ditransitive verb	Grodner and Gibson (2005)
Reduced relative clause	Staub (2010)
Phrasal verb	Staub et al. (2017)
PP adjunct	Staub et al. (2017)

(7) a. Simple sentence/Underlying Form of OR

The senator attacked the reporter.

b. Object relative clause

The reporter$_i$ [who the senator attacked e_i] left the room.

As an example, Fig. A1 in Appendix A illustrates the derivation of a post-nominal OR. The terminal nodes in this tree diagram are abstract syntactic categories. This use of abstract categories as formatives (rather than words) focuses the modeling on syntactic decisions. The complementizer "that" promotes the underlying object to its left and forms a relativizable element "Noun that" with the *wh* feature. The *wh*-element-marked "DP_1" will eventually move to the left of RC, leaving behind a co-indexed trace "t_1" at the embedded object position.

As described in Section 4, the CCPC tool transforms MG derivations into multiple context-free rules in the sense of Seki et al. (1991). To estimate the probability of those grammar rules, we used the pattern matching tool Tregex (Levy & Andrew, 2006) to obtain attestation counts of constructions from the Wall Street Journal and Brown portions of Penn Treebank 3 (Marcus, Santorini, & Marcinkiewicz, 1993). Table 7 lists the key findings of our corpus study.[15] It shows that among other things SRs indeed occur much more frequently than ORs in English.[16]

This grammar fragment, weighted by the construction frequencies discussed above, allowed us to make reading time predictions at the same word for SRs and ORs using the two complexity metrics of sentence processing. The critical RC region comprised the embedded NP and verb following the complementizer *that*.[17] As discussed in Section 3.1, Grodner and Gibson (2005) found a subject advantage at the RC verb. Eye-tracking experiments by Staub and colleagues provided evidence for processing delays at both the NP and the verb within an OR, and that the former induces greater difficulty (Staub, 2010; Staub et al., 2017).

To fairly compare the two complexity metrics under the same modeling process, Table 8 reports reading time predictions made by both on the subject–object asymmetry in English RCs. In accord with previous works (Hale, 2001; Levy, 2008), our own calculation of word-by-word surprisals predicts that ORs are read slower at the embedded NP,

Table 7
The attestation counts and examples of relevant constructions in Penn Treebank 3. The underline in the RC examples marks the extraction site

Construction	Example	Count
SR	*the forces that __ threatened her*	2,564
OR with relative pronoun	*the parts* ***which*** *he wrote __*	377
OR without relative pronoun	*the teenagers I interviewed __*	1,202
NP extracted from PP	*the life they believe in __*	123
NP with RC modifier	*the data* ***we seek*** *__*	6,869
NP without RC modifier	*the data*	94,216
pronoun as subject	***they*** *miss the hot cereal*	27,801
NP as subject	***the method*** *changed*	13,639
pronoun as object	*he chose* ***me***	2,954
NP as object	*he got* ***the approval***	7,347
pronoun as embedded subject	*the designation* ***she*** *liked __*	643
NP as embedded subject	*the product which* ***the SEC*** *approved __*	117
OR with transitive verb	*the teenagers I* ***interviewed*** *__*	837
OR with ditransitive verb	*a set I* ***gave*** *__ to the Salvation Army*	91
PP adjunct to intransitive verb	*it all began* ***on an autumn afternoon***	8,129
PP adjunct to transitive verb	*Charlie ate some supper* ***in the kitchen***	7,490
PP complement	*it may refer* ***to a specific person***	2,946

Table 8
Surprisal and entropy reduction predict the reading difficulty at each word in RCs, with the loci of subject advantage highlighted in bold

	Type	that	Det	Noun	RC verb
Surprisal	SR	4.54	**0.51**	0	0.84
	OR	4.54	**5.84**	0	0.14
Entropy Reduction	SR	0	**0.68**	0	**1.87**
	OR	0	**1.32**	0	**2.07**

but not at the RC verb. Although it is possible to impose different degrees of parallelism available to a parser (Boston, Hale, Vasishth, & Kliegl, 2011), no qualitative difference was found in the surprisal results when the calculation was restricted to consider only 3, 5, or 10 syntactic analyses with highest probabilities, as shown in Table B1 in Appendix B. The entropy reduction hypothesis, however, is linked to readers' uncertainty over all remaining structural derivations at any given point. Comparing the predictions for SRs and ORs at the same word, the entropy reduction results better fit to the eye-tracking data in Staub et al. (2017), such that SRs are easier to process throughout the RC region, with a more apparent advantage over ORs at the embedded NP.

5.2. Modeling the animacy effect

The animacy of NP has been identified as a determinant of RC difficulty in previous experimental investigations. In this work, we offer an information-theoretic interpretation of the animacy-related processing patterns. It was done by specifying the animacy feature on a handful of rules in our unlexicalized grammar fragment. This method is equivalent to propagating diacritical marks of animacy on the lexical head to higher nodes in a lexicalized grammar. In addition, we separated RC verbs in the grammar by their extraction sites. For example, "V-SR" and "V-OR" stand for the embedded verbs in SRs and ORs, respectively. This is a case of grandparent annotation in the sense of Johnson (1998) to ensure that fine-grained probabilistic information is captured at the grammar weighting stage, especially when the frequency distribution of animacy interacts with the RC type. Using only one category for RC verbs, for example, "V-RC", could obscure this sort of distributional difference.

A similar but finer grained corpus search added weights to the grammar rules. We used the animacy-annotated Switchboard corpus of Conversational American English (Godfrey, Holliman, & McDaniel, 1992).[18] The annotation for animacy distinctions in the parsed part of the Switchboard corpus is based on a hierarchy of 10 classes: human, organizations, animal, automata, vehicles, place, time, other physical objects, abstract entities, and those describing heterogeneous groups of entities (Zaenen et al., 2004). Similar to Bowman and Chopra (2012), we adopted Zaenen et al.'s binary classification by treating the first five classes as animate and the rest as inanimate. It is of course an open question as to how to divide the 10 classes into two sets. For example, one may suggest that vehicles and machines should instead be treated as inanimate. However, we do not think that the results of our modeling strongly depend on those borderline categories in the animacy hierarchy. The key frequency patterns estimated from the corpus counts, for example, the animacy of head noun, remain largely unchanged, even if we only consider human and animal as animate, and the other eight classes inanimate.[19]

We quantified the animacy distribution in the matrix clause and in different types of RC. The corpus counts in Table 9 confirm the pattern reported by Roland et al. (2007),[20] such that animate NPs are much more likely to be modified by an SR, while inanimate NPs tend to head an OR. Pronouns were again treated separately from full NPs and were found to be frequently used within the RCs. Matrix clauses have more inanimate than animate objects. They are also more likely to have animate subjects, regardless of the verb transitivity.

As discussed in Section 3.2, Traxler et al. (2002) found that ORs with inanimate heads were read faster than those with animate heads. As a result, the subject advantage was much attenuated in RCs with an inanimate head. Our reading difficulty predictions are consistent with these empirical results. Both the surprisal and entropy reduction results in Table 10 suggest that changing the animacy status of the head noun reverses the subject advantage pattern. Comparing the predictions made by the two complexity metrics, the total entropy reduction value for ORs is slightly larger given an animate head, whereas the much inflated surprisal grasps the substantially increased reading time for this type of sentence, as empirically evidenced in Table 3.

Table 9
The counts of relevant constructions in the animacy-annotated Switchboard corpus were used to estimate the probabilities of animacy-specified grammar rules

Construction	Example	Count
SR with animate head	***students*** *who __ have graduated*	499
OR with animate head	***people*** *that I know __*	133
SR with inanimate head	***things*** *that __ help you*	727
OR with inanimate head	***problems*** *that he had __*	1,074
SR with animate object	*friends that __ have* ***kids***	59
SR with inanimate object	*the one that __ opened the* ***floodgates***	124
SR with pronoun object	*anybody who __ bills* ***you***	82
OR with animate subject	*the one that my* ***parents*** *took __*	1,060
OR with inanimate subject	*deductibles that our* ***insurance*** *doesn't cover __*	131
OR with pronoun subject	*engineers who* ***they*** *know __*	1,020
animate subject with intransitive verb	*people only* ***go*** *forward*	11,709
animate subject with transitive verb	*my husband* ***broke*** *the coffee pot*	10,896
inanimate subject with intransitive verb	*things* ***improved***	4,847
inanimate subject with transitive verb	*Aspen* ***got*** *three feet*	4,888
animate object	*I had a* ***friend***	4,660
inanimate object	*they review* ***movies***	14,914
pronoun object	*they destroyed* ***it***	7,130

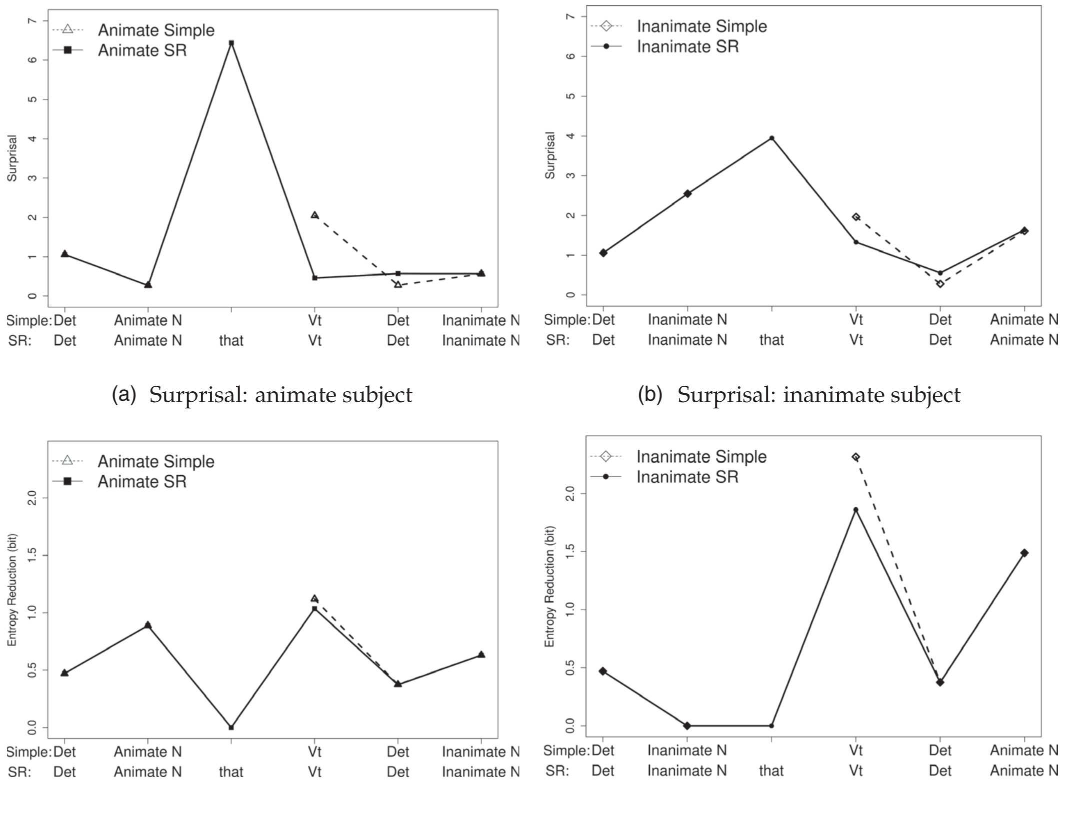
Table 10
The reading time predictions made by both complexity metrics suggest that the animacy of head nouns affects subject advantage, defined as the numerical difference between SRs and ORs

	Type	Head NP	Embedded NP		Subject Advantage
Surprisal	SR	+anim	−anim	1.61	**10.66**
	OR	+anim	−anim	12.27	
	SR	−anim	+anim	3.49	**−1.17**
	OR	−anim	+anim	2.32	
Entropy Reduction	SR	+anim	−anim	2.04	**1.01**
	OR	+anim	−anim	3.05	
	SR	−anim	+anim	3.72	**−0.73**
	OR	−anim	+anim	2.99	

The animacy feature not only plays a role in the subject–object processing asymmetry, but also interacts with clausal structures. Lowder and Gordon (2012) reported that the preference for animate subjects was less pronounced if the subject and the verb were separated by a clausal boundary around the complementizer "that." Their experimental results in Fig. 1 show that given an animate subject there was no difference at the verb in processing RCs and simple sentences. The reading time difference between simple sentences and RCs became larger with a sentence-initial inanimate subject. Based on the same animacy-specified weighted grammar, the entropy reduction model faithfully derives this processing effect. As illustrated in Fig. 2d, the less-expected inanimate subject causes a heightened reduction of uncertainty at the verb, suggesting additional processing delays when the verb is in the matrix clause than in the RC. This effect of sentence type is predicted to be less significant in sentences starting with an animate subject in Fig. 2c. The surprisal model, on the other hand, is not able to capture such pattern at the verb. Contrary to the experimental observation, it instead predicts the longest processing delay at the complementizer, as shown in Figs. 2a and b.

6. Discussion of the predictions

The entropy reduction hypothesis has shown success in predicting the comprehension difficulty profile of many constructions, including RCs in this work. However, it is not



(a) Surprisal: animate subject (b) Surprisal: inanimate subject

(c) Entropy Reduction: animate subject (d) Entropy Reduction: inanimate subject

Fig. 2. The entropy reduction hypothesis predicts more precisely that the processing difficulty driven by the inanimate subject is reduced by the clausal boundary, consistent with the findings in Lowder and Gordon (2012).

always straightforward to obtain meaningful linguistic explanations on how the degree of uncertainty is computed at critical sentence-medial parser states. As discussed in Section 4, when calculating the full entropy value for each string prefix in the sentence, the CCPC tool records all viable syntactic alternatives that are still "in play." By illustrating how structural uncertainties fluctuate in transitions from one word to the next, we characterize the disambiguation work in linguistically meaningful terms.

6.1. What derives the subject advantage throughout the RC?

Evidence from self-paced reading experiments suggests a penalty on the verb within an OR (Grodner & Gibson, 2005). Comparing the two RC types, the subject advantage there, although a small one, was replicated in subsequent eye-tracking studies (Staub, 2010; Staub et al., 2017). As shown in Section 5.1, entropy reduction has had better success in predicting such an effect, whereas modeling efforts using surprisal, including our own calculation, have all failed to do so, even when the parser considers a large number of parallel syntactic derivations.

The present study offers an explanation of the comprehension difficulty on the embedded OR verb based on readers' expectation of choosing a particular type of verb. In Fig. 3, the two boxes illustrate the parser states before and after reading in the target verb, which has been mostly transitive in the experimental stimuli of previous studies. Each parser state is represented by a set of syntactic remainders sharing the same string prefix in bold. The degree of uncertainty about the rest of the sentence is high after the embedded subject NP, because there exist multiple alternatives to the transitive verb "Vt" as the next word. In particular, the sentence could continue with either a phrasal verb "Vph" ("The reporter that the senator laughed at"), or a ditransitive verb "Vdi" ("The reporter that the senator introduced to the President ..."). Remaining parses with recursive center embeddings are also possible before the embedded verb sets in (Hale, 2003), for example, in sentences like "The reporter that the senator [who lost the re-election] attacked filed a civil case."[21] The succeeding transitive verb disconfirms all those possibilities, resulting in a drop of entropy value that simulates the processing difficulty in the form of increased reading time.

Although the early entropy reduction model in Hale (2003) captured the extra processing delay at the OR verb, there was no significant difference between SRs and ORs at the

Prob	Remainder
0.217	***Det Noun that Det Noun*** *Vt Vt Det Noun*
0.094	***Det Noun that Det Noun*** *Vt Vt Pronoun*
0.077	***Det Noun that Det Noun*** *Vt Vi Prep Det Noun*
0.058	***Det Noun that Det Noun*** *Vt Vi*
0.032	***Det Noun that Det Noun*** *Vt Vt Det Noun Prep Det Noun*
0.029	***Det Noun that Det Noun*** *Vph Prep Vt Det Noun*
0.025	***Det Noun that Det Noun*** *Vt Vph Prep Det Noun*
0.024	***Det Noun that Det Noun*** *Vdi Prep Det Noun Vt Det Noun*
0.016	***Det Noun that Det Noun*** *Vt Prep Det Noun Vt Det Noun*
0.010	***Det Noun that Det Noun*** *Vt Vdi Pronoun Prep Det Noun*
...	...
entropy = 6.287	

$\xrightarrow{\text{ER}=2.07}$

Prob	Remainder
0.320	***Det Noun that Det Noun Vt*** *Vt Det Noun*
0.138	***Det Noun that Det Noun Vt*** *Vt Pronoun*
0.114	***Det Noun that Det Noun Vt*** *Vi Prep Det Noun*
0.085	***Det Noun that Det Noun Vt*** *Vi*
0.048	***Det Noun that Det Noun Vt*** *Vt Det Noun Prep Det Noun*
0.037	***Det Noun that Det Noun Vt*** *Vph Prep Det Noun*
0.024	***Det Noun that Det Noun Vt*** *Prep Det Noun Vt Det Noun*
0.021	***Det Noun that Det Noun Vt*** *Vdi Pronoun Prep Det Noun*
0.021	***Det Noun that Det Noun Vt*** *Vt Pronoun Prep Det Noun*
0.017	***Det Noun that Det Noun Vt*** *Vi Prep Det Noun Prep Det Noun*
...	...
entropy = 4.213	

Fig. 3. The transition between the parser states before and after the embedded verb in ORs. Only the top 10 syntactic remainders are listed in each box.

embedded NP. This is different from what surprisal predicts at this word and is somewhat inconsistent with the eye-tracking data where the subject advantage was found throughout the RC region. By adopting a grammar formalism that makes the filler–gap relationship explicit, our calculation of word-by-word entropy reduction predicts a larger subject advantage at the embedded NP, more precisely, at the determiner before the noun. When the complementizer "that" signals an upcoming relativized structure, OR alternatives are much less expected because of their lower frequencies. To illustrate the parser state at that moment, only one of the top 10 remaining derivations in the left box of Fig. 4 is object-extracted. Because OR derivations are ranked relatively lower than SR derivations among possible alternatives given the head noun and the complementizer, they will be disambiguated to a greater degree when the parser is forced to integrate the upcoming determiner. In other words, more disambiguation work will have to happen before those low-probability derivations reach the top of the list at the subsequent parser state after the transition.

Lastly, although our discussion has so far focused on the RC region, expectation-based surprisal and entropy reduction are not able to account for the increased processing difficulty on the matrix verb following an OR. Such an effect has been reported in a number of previous studies with different explanations. Staub et al. (2017), for example, argue that the extra cost on the matrix verb is neither due to spillover processing from the preceding OR verb (Grodner & Gibson, 2005) nor the higher retrieval interference (Van Dyke, 2007; Van Dyke & Lewis, 2003) but rather the parser's engagement in serial execution of memory retrievals (McElree, 2006; McElree, Foraker, & Dyer, 2003) and the difficulty of rapidly switching between different syntactic or thematic roles (MacWhinney & Pleh, 1988; Sheldon, 1974).

6.2. What derives the animacy effect?

Previous modeling exercises on RC processing have only taken readers' structural knowledge into consideration. This leaves out non-structural factors like animacy which also play a crucial role in human sentence comprehension. We addressed this issue directly by extending grammar-based models and showing that they indeed predict observed effects in sentence processing experiments. In particular, the results of Traxler

Prob	Remainder
0.104	***Det Noun that*** *Vt Det Noun Vt Det Noun*
0.045	***Det Noun that*** *Vt Pronoun Vt Det Noun*
0.045	***Det Noun that*** *Vt Det Noun Vt Pronoun*
0.037	***Det Noun that*** *Vt Det Noun Vi Prep Det Noun*
0.037	***Det Noun that*** *Vi Prep Det Noun Vt Det Noun*
0.034	***Det Noun that*** *Pronoun Vt Vt Det Noun*
0.028	***Det Noun that*** *Vi Vt Det Noun*
0.028	***Det Noun that*** *Vt Det Noun Vi*
0.019	***Det Noun that*** *Vt Pronoun Vt Pronoun*
0.016	***Det Noun that*** *Vi Prep Det Noun Vt Pronoun*
...	...
entropy = 7.609	

ER=1.32

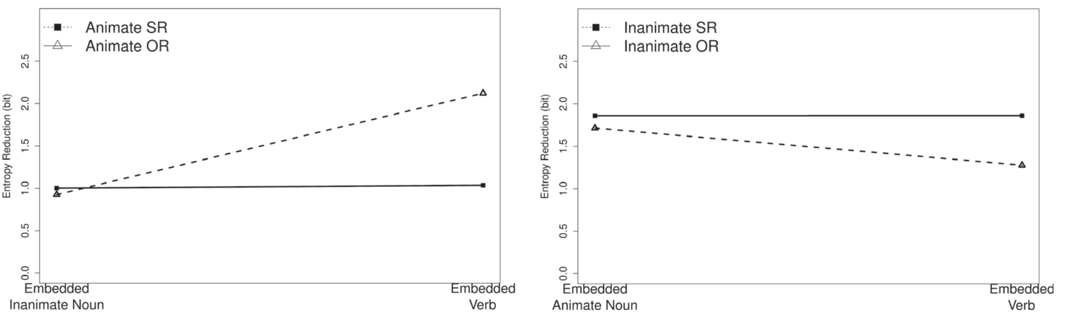
Prob	Remainder
0.217	***Det Noun that Det*** *Noun Vt Vt Det Noun*
0.094	***Det Noun that Det*** *Noun Vt Vt Pronoun*
0.077	***Det Noun that Det*** *Noun Vt Vi Prep Det Noun*
0.058	***Det Noun that Det*** *Noun Vt Vi*
0.032	***Det Noun that Det*** *Noun Vt Vt Det Noun Prep Det Noun*
0.029	***Det Noun that Det*** *Noun Vph Prep Vt Det Noun*
0.025	***Det Noun that Det*** *Noun Vt Vph Prep Det Noun*
0.024	***Det Noun that Det*** *Noun Vdi Prep Det Noun Vt Det Noun*
0.016	***Det Noun that Det*** *Noun Vt Prep Det Noun Vt Det Noun*
0.010	***Det Noun that Det*** *Noun Vt Vdi Pronoun Prep Det Noun*
...	...
entropy = 6.287	

Fig. 4. The transition between the parser states before and after reading in the determiner of the embedded NP in ORs.

et al. (2002, 2005) have shown that SRs are equally or even more difficult to process than ORs when the head noun is inanimate. The total surprisal and entropy reduction values we calculated in Table 10 are in line with the observed animacy effect in processing RCs. The surprisal results, in particular, exhibit a close correspondence with the construction frequencies in Table 9. The large surprisal for ORs following an animate head directly reflects their frequency status, as they are least popular among the four RC constructions with different animacy patterns. As for the entropy reduction predictions, the neutralized subject advantage is primarily due to the additional difficulty of processing SRs modifying an inanimate head than an animate head. The total entropy reduction values for ORs are largely on the same level, regardless of the animacy pattern.

We further examined the entropy reduction results at the word level to locate the exact position(s) where SRs were predicted to be harder than ORs. Comparing the four types of RC with different animacy manipulations in Fig. 5, we can understand the absence of subject advantage by reference to the contrastive predictions at the embedded RC verb. When the head noun is inanimate, more uncertainties are reduced in SRs than in ORs at the embedded verb. The CCPC tool visualizes the parser states before and after this critical position in the form of sorted lists of "in play" remainders. In Fig. 6a, the degree of uncertainty remains high given an inanimate head prefix because SRs and ORs (highlighted in blue) are both still considered as possible alternatives. This has led to a more balanced frequency distribution with no dominating remainders, for example, there are no derivations with an over 10% probability. Reading in the following verb disambiguates the sentence to a greater degree. In Fig. 6b, when it is time to form the subject–verb dependency within ORs, readers have already expected the next word to be a transitive verb because there is only one other variant (starting with a ditransitive verb) among the top 10 alternatives. Less disambiguation work is therefore computed for the transition to the next parser state in ORs with an inanimate head.

As Lowder and Gordon (2012) have demonstrated, the clausal boundary introduced by the complementizer "that" alleviates the processing difficulty of integrating a verb with a



(a) RCs with an animate head noun (b) RCs with an inanimate head noun

Fig. 5. Entropy reduction predicts no subject advantage for RCs with an inanimate head due to the greater processing difficulty at the embedded verb in SRs than in ORs.

Prob	Remainder
0.047	***Det InaniN that*** *Vt Det InaniN Vi*
0.047	***Det InaniN that*** *Det AniN Vt Vi*
0.046	***Det InaniN that*** *Pronoun Vt Vi*
0.035	***Det InaniN that*** *Vt Pronoun Vi*
0.034	***Det InaniN that*** *Vi Vi*
0.025	***Det InaniN that*** *Vt Det AniN Vi*
0.016	***Det InaniN that*** *Vt Det InaniN Vt Det InaniN*
0.016	***Det InaniN that*** *Det AniN Vt Vt Det InaniN*
0.016	***Det InaniN that*** *Pronoun Vt Vt Det InaniN*
0.012	***Det InaniN that*** *Vt Pronoun Vt Det InaniN*
...	...
entropy = 9.092	

ER=1.86 →

Prob	Remainder
0.124	***Det InaniN that Vt*** *Det InaniN Vi*
0.092	***Det InaniN that Vt*** *Pronoun Vi*
0.065	***Det InaniN that Vt*** *Det AniN Vi*
0.043	***Det InaniN that Vt*** *Det InaniN Vt Det InaniN*
0.032	***Det InaniN that Vt*** *Pronoun Vt Det InaniN*
0.027	***Det InaniN that Vt*** *Det InaniN Vdi Det AniN to Det InaniN*
0.025	***Det InaniN that Vt*** *Det InaniN Vdi Det InaniN to Det InaniN*
0.023	***Det InaniN that Vt*** *Det InaniN Vt Det AniN*
0.022	***Det InaniN that Vt*** *Det AniN Vt Det InaniN*
0.022	***Det InaniN that Vt*** *Det InaniN Vi in Det InaniN*
...	...
entropy = 7.231	

(a) SRs with an inanimate head

Prob	Remainder
0.196	***Det InaniN that Det AniN*** *Vt Vi*
0.068	***Det InaniN that Det AniN*** *Vt Vt Det InaniN*
0.043	***Det InaniN that Det AniN*** *Vt Vdi Det AniN to Det InaniN*
0.039	***Det InaniN that Det AniN*** *Vt Vdi Det InaniN to Det InaniN*
0.036	***Det InaniN that Det AniN*** *Vt in Det InaniN Vi*
0.036	***Det InaniN that Det AniN*** *Vt Vt Det AniN*
0.035	***Det InaniN that Det AniN*** *Vt Vi in Det InaniN*
0.033	***Det InaniN that Det AniN*** *Vdi to Det InaniN Vi*
0.024	***Det InaniN that Det AniN*** *Vt Vt Pronoun*
0.013	***Det InaniN that Det AniN*** *Vt in Det AniN Vi*
...	...
entropy = 7.379	

ER=1.28 →

Prob	Remainder
0.248	***Det InaniN that Det AniN Vt*** *Vi*
0.085	***Det InaniN that Det AniN Vt*** *Vt Det InaniN*
0.055	***Det InaniN that Det AniN Vt*** *Vdi Det AniN to Det InaniN*
0.049	***Det InaniN that Det AniN Vt*** *Vdi Det InaniN to Det InaniN*
0.046	***Det InaniN that Det AniN Vt*** *in Det InaniN Vi*
0.045	***Det InaniN that Det AniN Vt*** *Vt Det AniN*
0.044	***Det InaniN that Det AniN Vt*** *Vi in Det InaniN*
0.030	***Det InaniN that Det AniN Vt*** *Vt Pronoun*
0.016	***Det InaniN that Det AniN Vt*** *in Det AniN Vi*
0.016	***Det InaniN that Det AniN Vt*** *in Det InaniN Vt Det InaniN*
...	...
entropy = 6.102	

(b) ORs with an inanimate head

Fig. 6. The parser states before and after the transition to the embedded verb in SRs and ORs with an inanimate head.

less-expected inanimate subject. The entropy reduction predictions in Fig. 2 more closely approximate this effect than the calculated surprisal values. The CCPC tool is also able to offer an explanation. In Fig. 7, the string prefix with an inanimate subject is very ambiguous because the transitivity of the forthcoming matrix verb is still undetermined. At the matrix verb, the frequency distribution of syntactic alternatives becomes more concentrated. The top three remainders in the right box, in particular, have a total probability of more than 80%. Transitioning from a highly uncertain parser state to a much certain one results in a 2.32 bits of entropy reduction, larger than the one calculated at the embedded verb in an SR in Fig. 6a.

Prob	Remainder
0.304	***Det InaniN*** *Vi*
0.105	***Det InaniN*** *Vt Det InaniN*
0.067	***Det InaniN*** *Vdi Det AniN to Det InaniN*
0.061	***Det InaniN*** *Vdi Det InaniN to Det InaniN*
0.055	***Det InaniN*** *Vt Det AniN*
0.054	***Det InaniN*** *Vi in Det InaniN*
0.037	***Det InaniN*** *Vt Pronoun*
0.019	***Det InaniN*** *Vi in Det AniN*
0.012	***Det InaniN*** *Vdi Det AniN to Det AniN*
0.011	***Det InaniN*** *Vdi Det InaniN to Det AniN*
...	...
entropy = 5.374	

ER=2.32 →

Prob	Remainder
0.424	***Det InaniN Vt*** *Det InaniN*
0.223	***Det InaniN Vt*** *Det AniN*
0.148	***Det InaniN Vt*** *Pronoun*
0.043	***Det InaniN Vt*** *Det InaniN in Det InaniN*
0.023	***Det InaniN Vt*** *Det AniN in Det InaniN*
0.015	***Det InaniN Vt*** *Det InaniN in Det AniN*
0.015	***Det InaniN Vt*** *Pronoun in Det InaniN*
0.008	***Det InaniN Vt*** *Det AniN in Det AniN*
0.006	***Det InaniN Vt*** *Det InaniN who Vt Det InaniN*
0.006	***Det InaniN Vt*** *Det InaniN who Det AniN Vt*
...	...
entropy = 3.058	

Fig. 7. The string prefix of an inanimate subject allows a variety of alternatives. The succeeding main verb greatly disambiguates it.

Looking beyond the major effect at the verb, the results of Lowder and Gordon (2012) suggest neither particular speed-up nor slow-down in terms of fixation durations at the complementizer. In Figs. 2a and b, surprisal predicts the largest processing delay at "that," quite the contrary. The slow-down is even larger in the condition with an animate subject. We can interpret the surprisal result as a direct reflection of frequencies, that is, RCs are less frequent than simple sentences. Corpus counts in Table 9 also suggest that more RCs modify an inanimate noun than an animate noun. On the other hand, the entropy reduction predicts that no disambiguation work is done at the word "that."This is because the uncertainty about the rest of the sentence remains high, even after the complementizer winnows down the possible alternatives to RCs. Both the SR and OR alternatives are still in play, as shown in the left box of Fig. 6a. Here, the different predictions made by the two complexity metrics serve as a good example to echo the conceptual difference between them, such that entropy reduction takes one step away from the frequency-driven surprisal and instead focuses on the internal uncertainty and disambiguation of parser states (Hale, 2016).

Visualizing the parser state at the complementizer, when relativized structures are expected, provides additional evidence in support of Wagers and Pendleton (2016). They reported that comprehenders have a greater likelihood to encode a subject gap after animate fillers, whereas inanimate fillers are rather unlikely to generate a specific syntactic expectation on where the gap locates. In Table 11, we calculated the probabilities of two RC types conditioned on a head noun with different animacy status. The results are compatible with corpus counts in Table 9 which we used to assign weights to the animacy-specified grammar. Inanimate head nouns are indeed more "equivocal" in generating a syntactic expectation. The parser state at the word "that" in Fig. 8a more straightforwardly illustrates the higher expectation of positing a subject gap after an animate head NP than after an inanimate one. Among the top 10 possible remainders, only two lower ranking constructions are object-extracted. This is in contrast to the parser state given an inanimate head, repeated here as Fig. 8b. Whether the sentence continues as an SR or an OR has yet to be determined at that point since the top 10 derivations are equally split between the two RC types.

7. Toward discourse-informed predictions

To model the effect of animacy in incremental sentence comprehension, we adopted a simple binary classification of NP animacy, namely animate versus inanimate, along the

Table 11
Probabilities of RC derivations conditioned on head NPs with different animacy values

	$P(RC \mid Head)$	
	SR	OR
Animate head	0.887	0.113
Inanimate head	0.526	0.474

Prob	Remainder
0.106	***Det AniN that*** *Vt Det InaniN Vi*
0.078	***Det AniN that*** *Vt Pronoun Vi*
0.056	***Det AniN that*** *Vt Det AniN Vi*
0.055	***Det AniN that*** *Vi Vi*
0.026	***Det AniN that*** *Vt Det InaniN Vt Det InaniN*
0.020	***Det AniN that*** *Vt Pronoun Vt Det InaniN*
0.019	***Det AniN that*** *Det AniN Vt Vi*
0.018	***Det AniN that*** *Pronoun Vt Vi*
0.017	***Det AniN that*** *Vt Det InaniN Vdi Det AniN to Det InaniN*
0.015	***Det AniN that*** *Vt Det InaniN Vdi Det InaniN to Det InaniN*
...	...
entropy = 7.929	

(a) Animate head NP prefix

Prob	Remainder
0.047	***Det InaniN that*** *Vt Det InaniN Vi*
0.047	***Det InaniN that*** *Det AniN Vt Vi*
0.046	***Det InaniN that*** *Pronoun Vt Vi*
0.035	***Det InaniN that*** *Vt Pronoun Vi*
0.034	***Det InaniN that*** *Vi Vi*
0.025	***Det InaniN that*** *Vt Det AniN Vi*
0.016	***Det InaniN that*** *Vt Det InaniN Vt Det InaniN*
0.016	***Det InaniN that*** *Det AniN Vt Vt Det InaniN*
0.016	***Det InaniN that*** *Pronoun Vt Vt Det InaniN*
0.012	***Det InaniN that*** *Vt Pronoun Vt Det InaniN*
...	...
entropy = 9.092	

(b) Inanimate head NP prefix

Fig. 8. The parser states at the complementizer suggest that an animate head is more likely to predict an SR than an inanimate head. OR constructions are highlighted in blue.

lines of MacWhinney et al. (1984) and Traxler et al. (2002). However, this simplification should not be misconstrued as a theoretical claim. Typologists such as Silverstein (1976) and Dixon (1979) have argued that grammatical processes in different languages are sensitive to the relative degree of animacy of the NPs involved. In particular, they have proposed that continuous categories ranging from most animate to least animate can be ranked under an animacy hierarchy like the one in (8).[22]

(8) The Animacy Hierarchy (Dixon, 1979, p. 85)

1st Person Pronoun > 2nd Person Pronoun > 3rd Person Pronoun > Proper Noun > Human Common Noun > Animate Common Noun > Inanimate Common Noun

Macdonald, Brandt, Theakston, Lieven, and Serratrice (2020) have recently reported that the binary "lexical animacy" of head-nouns facilitates children's interpretation of English RCs. As RCs unfold, children are also sensitive to the "perceptual animacy" along a continuum, such that lexically inanimate entities are conceptualized as more or less animate given contextual cues like motion. We can define these hierarchical or continuous cognitive representations of animacy in finer-grained grammars by specifying the person feature of pronouns and making additional classification of common nouns and verbs depending on whether they support the semantic reversibility of the RC.

The present work has also set up a quantitative framework to discuss the broader role of discourse in sentence comprehension. Previous literature has started to address whether this kind of non-structural constraint serves as a potential explanation for supposed processing universals. Roland, Mauner, O'Meara, and Yun (2012), for instance, have argued for the importance of discourse information in processing RCs and in sentence comprehension in general. They have proposed an account to explain the inverse processing pattern found in RCs with pronouns (Reali & Christiansen, 2007), as compared to those without. They claim that the distributional frequency of word chunks in pronominal RCs (the pronominal status) is in fact a result of discourse influence (the informational status). For example, ORs tend to have a discourse-old pronoun referent as the embedded subject whereas SRs tend to have a discourse-new referent as the embedded object. Readers therefore have "discourse expectations" on the RC structure. A violation of expectations will lead to a higher degree of comprehension difficulty. The background theory behind this discourse expectation idea is the corpus study of spoken conventional English in Fox and Thompson (1990). They have shown that ORs are more likely to contain a pronoun rather than a full NP because the embedded subject refers back to a referent in the ongoing discourse, a process known as "grounding." The "discourse-new/old" idea of Roland et al. (2012) is also compatible with the topichood hypothesis proposed for processing Dutch RCs (Mak, Vonk, & Schriefers, 2002). The conclusion was that there exists a limitation on the referent of the discourse-old NP/pronoun. It could not be any individual mentioned in the context, but rather the topic of the ongoing discussion.[23] All the studies mentioned above are rooted in GIVENNESS, which is similar to the priming effect in functional linguistics (Givón, 1983).

Recent work has also started to integrate world knowledge and linguistic experience in models of online sentence comprehension, including both surprisal (Venhuizen, Crocker, & Brouwer, 2019a) and entropy reduction (Venhuizen, Crocker, & Brouwer, 2019b). The modeling framework described in this paper is flexible enough to include discourse information as an ingredient, for example, by quantifying the expectation of ongoing discourse referents as a degree of uncertainty. The complexity metric will consider the co-occurrence between the type of NP and its discourse status (e.g., topic, mention, or neutral) in the corpus, especially when it is applied with narrative and conversational text stimuli. This co-occurrence information could be encoded as more subcategorized grammar rules, in a way similar to how we treated the animacy information. An alternative solution would define the discourse constraint as a penalty score during incremental parsing, that is, as long as there is no violation of discourse expectation, there will be no extra burden in comprehension. Otherwise, a penalty will be added, which will result in longer processing delays.

8. Conclusion

In this paper, we discuss the model of incremental sentence comprehension based on readers' expectations. As is well known, both structural and non-structural information affect the processing difficulty profile of relativized structures. By elaborating an expectation-based model with the animacy status of NPs, we formalize the amount of information contributed by each incoming word in reducing both structural and non-structural uncertainties.

Using a software that combines a chart parser with a calculator of derivations' conditional probabilities, we demonstrate how word-by-word entropy reductions faithfully reflect two prominent processing patterns: the asymmetry between subject and object relatives as well as the effect of head-noun animacy. With an expressive grammar that directly defines the syntactic movement in relativization, our model derives the additional processing difficulty in object-extracted relatives. Importantly, the entropy reduction prediction confirms that the ORs are harder to process than SRs throughout the RC region, with a more pronounced effect at the embedded NP (Staub, 2010; Staub et al., 2017). In comparison, neither surprisal (Hale, 2001; Levy, 2008) nor earlier entropy reduction (Hale, 2003) calculations have been able to correctly capture the loci of subject advantage. By encoding the frequency distribution of NP animacy in weighted grammars, our model accounts for the processing difficulty discrepancies among conditions with various animacy manipulations (Traxler et al., 2002). In particular, based on a completely parallel parser (i.e., full entropy), the model predicts that ORs are less difficult than SRs when the head nouns are inanimate. Entropy reduction also successfully derives the processing burden when a less-preferred inanimate subject is integrated with the main verb. The eye-tracking data in Lowder and Gordon (2012) suggest that there exists an interaction between the syntactic structure, for example, the clausal boundary created by “that”, and the animacy of subject argument. Entropy reduction predictions at the verb are more in line with this observation than the surprisal results.

To explain how these predictions are made, we calculate the probability of alternative completions of initial substrings. Enumerating syntactic alternatives that are in play at a given point essentially visualizes the content of a ranked parallel parser state with additional features in the formal grammar fragment, such as animate or inanimate. This feature brings into relief the linguistic interpretation of the predicted difficulty contrasts. For example, we show that SRs with an inanimate head are actually harder to process than their OR counterparts. This is because the parser is highly uncertain about choosing an SR or OR as the remainder given an inanimate head (Wagers & Pendleton, 2016). More disambiguation would have to be completed later within the RC region. This is certainly a finer-grained explanation than simply arguing that SRs headed by inanimate nouns are less frequent, which is directly reflected in the surprisal metric. Similarly, we look into possible remaining alternatives conditioned by the string prefix of an inanimate subject. We find that the parser still allows a variety of remainders while more disambiguation is done later at the main verb than at the verb embedded in an SR. This could result in different levels of processing difficulty at the verb between a simple sentence and a subject relative, as suggested by Lowder and Gordon (2012).

Acknowledgments

The early stages of this work were supported by Doctoral Fellowship DD056-A-12 from Chiang Ching-kuo Foundation to Z.C. and by National Science Foundation Award 0741666 to J.T.H. We thank T. Florian Jaeger for the animacy-annotated Switchboard corpus.

Open Research badges

This article has earned Open Data and Open Materials badges. Data and materials are available at https://osf.io/g97bc/.

Notes

1. We combine explanatory factors, for example, from psycholinguistics and linguistics, into a single model that derives quantitative predictions. Recent work demonstrates that this goal is achievable. For instance, Brasoveanu and Dotlačil (2020) integrate discourse semantics with a theory of memory. The work reported in this paper aspires to the same sort of integration. It combines a particular aspect of noun meaning with syntax and frequency-sensitive processing.
2. See Vasishth, Nicenboim, Engelmann, and Burchert (2019) for a recent review.
3. Early work by Hsiao and Gibson (2003) reported an OR preference in Chinese. However, Vasishth, Chen, Li, and Guo (2013) were unable to replicate this result using the same experimental stimuli. In a random effect Bayesian meta-analysis of 15 studies, they showed that the overall evidence for the subject advantage is strong in Chinese, with an approximate posterior probability of 70% – 80%. Indeed, in self-paced reading and eye-tracking experiments with stimuli that controlled local ambiguities, Chinese SRs were read faster (Jäger, Chen, Li, Lin, & Vasishth, 2015; Wu, Kaiser, & Vasishth, 2018).
4. Jurafsky and Martin (2008) offered a tutorial presentation of chart parsing in their natural language processing textbook.
5. Blachman (1968) has clarified the difference between surprisal and entropy reduction on a mathematical level. For instance, entropy reductions are additive, whereas surprisals are not. For a brief review, see Appendix C in Yun, Chen, Hunter, Whitman, and Hale (2015).
6. Two studies of Russian RCs also support an integration of the memory-based and expectation-based theories, as these two independent sources of processing difficulty might affect RC comprehension at different stages (Levy, Fedorenko, & Gibson, 2013; Price & Witzel, 2017).
7. To our knowledge, there has not been a quantitative apportionment of explanatory burden across these two factors within the Demberg and Keller (2009) framework for this particular phenomenon. Another option would be to combine those two factors into a single principle that is different from entropy reduction. This theoretical point has received considerable attention in the literature. However, to date no such explanation is available for all of the processing patterns summarized in Section 3.1. See Lewis et al. (2006, the last bullet in Box 3), Campanelli, Van Dyke, and Marton (2018), and Futrell, Gibson, and Levy (2020), *inter alia*.
8. For consistency, Table 3 only includes Traxler et al.'s quasi-first-pass results, a term similar to the go-past measure reported by Staub (2010) in Table 2. The processing pattern driven by animacy was similar in other dependent measures, including first-pass, first-pass regression, and total time.
9. MGs are in the same complexity class as Tree Adjoining Grammar (Joshi, Levy, & Takahashi, 1975) and have a dependency interpretation (Boston, Hale, & Kuhlmann, 2010).

10. The CCPC program generalizes the Simple Language Generator of Rohde (1999) and is freely available at: https://github.com/timhunter/ccpc.
11. Graf, Monette, and Zhang (2017) have incorporated syntactic notions of locality in minimalist parsing, which makes encouraging predictions about the processing behavior of RCs. Similarly, by evaluating its performance on various word order and processing asymmetries, De Santo (2020) has demonstrated MGs' value as "a transparent, interpretable link between structural representations and off-line processing behavior."
12. Section 4 of Yun et al. (2015) explains the grammar weighting procedure in a tutorial fashion.
13. Like Linzen and Jaeger (2015), our grammar fragment was small enough for exhaustive parsing. Approximation techniques, such as beam search for *full entropy* (Roark, Bachrach, Cardenas, & Pallier, 2009) or future surprisal for *single-step entropy* (van Schijndel & Schuler, 2017), were not necessary.
14. This is in contrast to the two-step adjunction analysis (Chomsky, 1977, 1986, 1995). First, the *wh*-movement deposits the relative pronoun "who" from the underlying object position to the leftmost edge of the constituent, "the senator attacked," and forms a complementizer phrase (CP), that is, "who the senator attacked." Second, this CP is adjoined to "the reporter" as a postmodifier. Hale (2006) reported that entropy reduction predictions based on the adjunction analysis were inconsistent with experimental observations.
15. The grammars and corpus search queries are available at https://osf.io/g97bc/.
16. The adjuncthood of PPs is not clearly annotated in the Penn Treebank. We therefore estimated the rate of PP adjunct (73.4%) versus PP complement (26.6%) based on the distribution of function tags (see Section 2.2 of the Penn Treebank bracketing guidelines in Bies et al., 1995). We considered PPs with -DTV, -CLR, and -PUT tags as complements, and those tagged with -BNF, -DIR, -LOC, -MNR, and -PRP as adjuncts.
17. For simplicity, our grammar did not differentiate between relative pronouns, for example, "who," "which," and the complementizer "that."
18. We chose not to use the Switchboard corpus in modeling the subject advantage effect because of its telephone-conversation genre. Unlike the Wall Street Journal and Brown corpora, it does not contain enough samples of many constructions derived by the syntactic parameters which we planned to explore.
19. The counts for the first four constructions in Table 9 will be 418, 93, 804, and 1,114 respectively, if one adopts this alternative classification of animacy. The frequency distribution of head noun animacy remains the same.
20. Roland et al. (2007) estimated the frequency of NP animacy in RCs by hand-coding a random sample of 100 RCs each from the Brown and Switchboard corpora.
21. Derivations with recursive center embeddings are not shown in the parser state illustration because of their lower probabilities.
22. Croft (2003) calls it "Extended Animacy Hierarchy" and argues that it involves three distinct but related functional dimensions: person hierarchy, referentiality hierarchy, and semantic animacy hierarchy.

23. It is also possible that the pronouns' lexical properties and their frequencies in the relevant syntactic contexts affect the processing of RCs (Fedorenko, Piantadosi, & Gibson, 2012; Gibson, Tily, & Fedorenko, 2013).

References

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological Review*, *111*(4), 1036–1060.

Bar-Hillel, Y., Perles, M., & Shamir, E. (1964). On formal properties of simple phrase structure grammars. In Y. Bar-Hillel (Ed.), *Language and information: Selected essays on their theory and application* (pp. 116–150). Reading, MA: Addison-Wesley.

Berwick, R. C., & Epstein, S. D. (1995). On the convergence of 'minimalist' syntax and categorial grammar. In A. Nijholt, G. Scollo, & R. Steetkamp (Eds.), *Algebraic methods in language processing 1995: Proceedings of the Twente workshop on language technology 10, jointly held with the first algebraic methodology and software technology (AMAST) workshop on language processing* (pp. 143–148). Enschede, The Netherlands: Universiteit Twente.

Bianchi, V. (2002). Headed relative clauses in generative syntax. *Part II. Glot International*, *6*(8), 1–13.

Bies, A., Ferguson, M., Katz, K., MacIntyre, R., Tredinnick, V., Kim, G., Schasberger, B. (1995). *Bracketing guidelines for Treebank II style Penn Treebank project*. (Technical Report). Philadelphia, PA: CIS, University of Pennsylvania.

Blachman, N. (1968). The amount of information that y gives about x. *IEEE Transactions on Information Theory*, *14*(1), 27–31.

Boston, M. F., Hale, J., Kliegl, R., Patil, U., & Vasishth, S. (2008). Parsing costs as predictors of reading difficulty: An evaluation using the potsdam sentence corpus. *Journal of Eye Movement Research*, *2*(1), 1–12.

Boston, M. F., Hale, J., & Kuhlmann, M. (2010). Dependency structures derived from minimalist grammars. In C. Ebert, G. Jäger, & J. Michaelis (Eds.), *The mathematics of language 10/11* (Vol. 6149, pp. 1–12). Heidelberg: Springer.

Boston, M. F., Hale, J., Vasishth, S., & Kliegl, R. (2011). Parallel processing and sentence comprehension difficulty. *Language and Cognitive Processes*, *26*(3), 301–349.

Bowman, S. R., & Chopra, H. (2012). Automatic animacy classification. In *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies: Student Research Workshop* (pp. 7–10). Montreal, Canada.

Brasoveanu, A., & Dotlačil, J. (2020). *Computational cognitive modeling and linguistic theory*. Dordrecht: Springer.

Brennan, J. R., Stabler, E., Van Wagenen, S. E., Luh, W.-M., & Hale, J. (2016). Abstract linguistic structure correlates with temporal activity during naturalistic comprehension. *Brain and Language*, *157*, 81–94.

Campanelli, L., Van Dyke, J. A., & Marton, K. (2018). The modulatory effect of expectations on memory retrieval during sentence comprehension. In T. T. Rogers, M. Rau, X. Zhu, & C. W. Kalish (Eds.), *Proceedings of the 40th annual conference of the Cognitive Science Society* (pp. 1434–1439). Austin, TX: Cognitive Science Society.

Chen, Z., Hunter, T., Yun, J., & Hale, J. (2014). Modeling sentence processing difficulty with a conditional probability calculator. In P. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds.) *Proceedings of the 36th annual meeting of the Cognitive Science Society* (pp. 1856–1857). Austin, TX: Cognitive Science Society.

Chen, Z., Jäger, L., & Hale, J. (2012). Uncertainty reduction as a predictor of reading difficulty in Chinese relative clauses. In Y.-O. Biq & L. Chen (Eds.), *Proceedings of the 13th international symposium on Chinese languages and linguistics (IsCLL-13)* (pp. 245–261). Taipei, Taiwan: National Taiwan Normal University.

Chomsky, N. (1977). On wh-movement. In P. Culicover, T. Wasow, & A. Akmajian (Eds.), *Formal syntax* (pp. 71–132). New York: Academic.

Chomsky, N. (1986). *Knowledge of language: Its nature, origin and use*. New York, NY: Praeger.

Chomsky, N. (1995). *The minimalist program*. Cambridge, MA: MIT Press.

Clark, A. (2014). *An introduction to multiple context free grammars for linguists*. Available at: https://alexc17.github.io/papers/mcfgsforlinguists.pdf. Accessed December 9, 2020.

Crain, S., & Fodor, J. D. (1985). How can grammars help parsers? In D. Dowty, L. Kartunnen, & A. M. Zwicky (Eds.), *Natural language parsing: Psycholinguistic, computational, and theoretical perspectives* (pp. 94–128). Cambridge: Cambridge University Press.

Croft, W. (2003). *Typology and universals*. Cambridge, UK: Cambridge University Press.

De Santo, A. (2020). Structure and memory: A computational model of storage, gradience, and priming (Unpublished doctoral dissertation). State University of New York at Stony Brook.

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

Demberg, V., & Keller, F. (2009). A computational model of prediction in human parsing: Unifying locality and surprisal effects. In N. Taatgen & H. van Rijn (Eds.), *Proceedings of the 29th meeting of the Cognitive Science Society* (pp. 1888–1893). Amsterdam, The Netherlands: Cognitive Science Society.

Demberg, V., Keller, F., & Koller, A. (2013). Incremental, predictive parsing with psycholinguistically motivated tree-adjoining grammar. *Computational Linguistics*, *39*(4), 1025–1066.

Dixon, R. M. (1979). Ergativity. *Language*, 59–138.

Ettinger, A. (2020). What BERT is not: Lessons from a new suite of psycholinguistic diagnostics for language models. *Transactions of the Association for Computational Linguistics*, *8*, 34–48.

Fedorenko, E., Piantadosi, S., & Gibson, E. (2012). Processing relative clauses in supportive contexts. *Cognitive Science*, *36*(3), 471–497.

Forster, K. I., Guerrera, C., & Elliot, L. (2009). The maze task: Measuring forced incremental sentence processing time. *Behavior Research Methods*, *41*(1), 163–171.

Fox, B. A., & Thompson, S. A. (1990). A discourse explanation of the grammar of relative clauses in English conversation. *Language*, 297–316.

Frank, S. L. (2013). Uncertainty reduction as a measure of cognitive load in sentence comprehension. *Topics in Cognitive Science*, *5*(3), 475–494.

Frank, S. L., Otten, L. J., Galli, G., & Vigliocco, G. (2015). The ERP response to the amount of information conveyed by words in sentences. *Brain and Language*, *140*, 1–11.

Futrell, R., Gibson, E., & Levy, R. P. (2020). Lossy-context surprisal: An information-theoretic model of memory effects in sentence processing. *Cognitive Science*, *44*(3), e12814.

Futrell, R., Wilcox, E., Morita, T., Qian, P., Ballesteros, M., & Levy, R. (2019). Neural language models as psycholinguistic subjects: Representations of syntactic state. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT), Volume 1 (Long and Short Papers, pp. 32–42). Minneapolis, MN.

Gennari, S. P., & MacDonald, M. C. (2008). Semantic indeterminacy in object relative clauses. *Journal of Memory and Language*, *58*(2), 161–187.

Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, *68*, 1–76.

Gibson, E. (2000). Dependency locality theory: A distance-based theory of linguistic complexity. In A. Marantz, Y. Miyashita, & W. O'Neil (Eds.), *Image, language, brain: Papers from the first mind articulation project symposium* (pp. 95–126). Cambridge, MA: MIT Press.

Gibson, E., Tily, H., & Fedorenko, E. (2013). The processing complexity of English relative clauses. In M. Sanz, I. Laka, & M. K. Tanenhaus (Eds.), *Language down the garden path: The cognitive and biological basis for linguistic structure* (pp. 149–173). Oxford: Oxford University Press.

Givón, T. (1983). *Topic continuity in discourse: A quantitative cross-language study* (Vol. *3*). Amsterdam, the Netherlands: John Benjamins.

Godfrey, J., Holliman, E., & McDaniel, J. (1992). Switchboard: Telephone speech corpus for research and development. In Proceedings of ICASSP-92: 1992 IEEE international conference on acoustics, speech, and signal processing (Vol. 1, pp. 517–520). San Francisco, CA.

Graf, T., Monette, J., & Zhang, C. (2017). Relative clauses as a benchmark for minimalist parsing. *Journal of Language Modelling*, *5*(1), 57–106.

Grodner, D., & Gibson, E. (2005). Consequences of the serial nature of linguistic input for sentenial complexity. *Cognitive Science*, *29*(2), 261–290.

Guillaumin, M. (2005). Conversations between mildly context-sensitive grammars (Internship Report). Ecole Normale Superieure and UCLA.

Hale, J. (2001). A probabilistic Earley parser as a psycholinguistic model. In Proceedings of the second meeting of the North American Chapter of the Association for Computational Linguistics on Language technologies (NAACL-HLT) (pp. 1–8).

Hale, J. (2003). The information conveyed by words in sentences. *Journal of Psycholinguistic Research*, *32* (2), 101–123.

Hale, J. (2006). Uncertainty about the rest of the sentence. *Cognitive Science*, *30*(4), 643–672.

Hale, J. (2016). Information-theoretical complexity metrics. *Language and Linguistics Compass*, *10*(9), 397–412.

Hale, J., Lutz, D., Luh, W.-M., & Brennan, J. (2015). Modeling fMRI time courses with linguistic structure at various grain sizes. In Proceedings of the 6th workshop on cognitive modeling and computational linguistics (pp. 89–97). Denver, CO.

Hatfield, H., & Artos, T. (2016). The locus of processing for object relative clauses and the impact of methodology. *Language, Cognition and Neuroscience*, *31*(2), 190–195.

Henderson, J. M., Choi, W., Lowder, M. W., & Ferreira, F. (2016). Language structure in the brain: A fixation-related fMRI study of syntactic surprisal in reading. *NeuroImage*, *132*, 293–300.

Hsiao, F., & Gibson, E. (2003). Processing relative clauses in Chinese. *Cognition*, *90*, 3–27.

Hsiao, Y., & MacDonald, M. C. (2016). Production predicts comprehension: Animacy effects in Mandarin relative clause processing. *Journal of Memory and Language*, *89*, 87–109.

Hunter, T. (to appear). Formal methods in experimental syntax. In J. Sprouse (Ed.), *The Oxford handbook of experimental syntax*. Oxford: Oxford University Press. Retrieved from https://timhunter.humspace.ucla.edu/parsing/expt-syntax-handbook.pdf

Ishizuka, T., Nakatani, K., & Gibson, E. (2003). Relative clause extraction complexity in Japanese. Poster presented at the 16th annual CUNY conference on human sentence processing. Boston, MA.

Jäger, L., Chen, Z., Li, Q., Lin, C.-J.-C., & Vasishth, S. (2015). The subject-relative advantage in Chinese: Evidence for expectation-based processing. *Journal of Memory and Language*, *79–80*, 97–120.

Jäger, L., Engelmann, F., & Vasishth, S. (2017). Similarity-based interference in sentence comprehension: Literature review and Bayesian meta-analysis. *Journal of Memory and Language*, *94*, 316–339.

Johnson, M. (1998). PCFG models of linguistic tree representations. *Computational Linguistics*, *24*, 613–632.

Joshi, A. K., Levy, L. S., & Takahashi, M. (1975). Tree adjunct grammars. *Journal of Computer and System Sciences*, *10*(1), 136–163.

Jurafsky, D. (1996). A probabilistic model of lexical and syntactic access and disambiguation. *Cognition*, *20*, 137–194.

Jurafsky, D., & Martin, J. H. (2008). *Speech and language processing: An introduction to natural language processing, computational linguistics and speech recognition* (2nd ed.). Upper Saddle River, NJ: Prentice Hall.

Just, M. A., Carpenter, P., Keller, T., Eddy, W., & Thulborn, K. (1996). Brain activation modulated by sentence comprehension. *Science*, *274*(5284), 114.

Kaplan, R. M. (1972). Augmented transition networks as psychological models of sentence comprehension. *Artificial Intelligence*, *3*, 77–100.

Kayne, R. S. (1994). *The antisymmetry of syntax*. Cambridge, MA: MIT Press.

King, J., & Just, M. A. (1991). Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language*, *30*(5), 580–602.

King, J., & Kutas, M. (1995). Who did what and when? Using word- and clause-level ERPs to monitor working memory usage in reading. *Journal of Cognitive Neuroscience*, *7*(3), 376–395.

Kuncoro, A., Dyer, C., Hale, J., Yogatama, D., Clark, S., & Blunsom, P. (2018). LSTMs can learn syntax-sensitive dependencies well, but modeling structure makes them better. In Proceedings of the 56th annual meeting of the Association for Computational Linguistics (ACL) Volume 1 (Long Papers) (pp. 1426–1436). Melbourne, Australia: Association for Computational Linguistics.

Kwon, N., Kluender, R., Kutas, M., & Polinsky, M. (2013). Subject/object processing asymmetries in Korean relative clauses: Evidence from ERP data. *Language*, *89*, 537–585.

Kwon, N., Lee, Y., Gordon, P. C., Kluender, R., & Polinsky, M. (2010). Cognitive and linguistic factors affecting subject/object asymmetry: An eye-tracking study of pre-nominal relative clauses in Korean. *Language*, *86*(3), 546–582.

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

Levy, R., & Andrew, G. (2006). Tregex and Tsurgeon: Tools for querying and manipulating tree data structures. In N. Calzolari, K. Choukri, A. Gangemi, B. Maegaard, J. Mariani, J. Odijk, & D. Tapias (Eds.), *Proceedings of the 5th international conference on language resources and evaluation (LREC 06)* (pp. 2231–2234). Genoa, Italy: European Language Resources Association (ELRA).

Levy, R., Fedorenko, E., & Gibson, E. (2013). The syntactic complexity of Russian relative clauses. *Journal of Memory and Language*, *69*(4), 461–495.

Levy, R., & Gibson, E. (2013). Surprisal, the PDC, and the primary locus of processing difficulty in relative clauses. *Frontiers in Psychology*, *4*, 229.

Lewis, R. L., & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science*, *29*, 1–45.

Lewis, R. L., Vasishth, S., & Van Dyke, J. A. (2006). Computational principles of working memory in sentence comprehension. *Trends in Cognitive Science*, *10*(10), 447–454.

Li, P., Bates, E., & MacWhinney, B. (1993). Processing a language without inflections: A reaction time study of sentence interpretation in Chinese. *Journal of Memory and Language*, *32*, 169–192.

Lin, C.-J.-C. (2008). The processing foundation of head-final relative clauses. *Language and Linguistics*, *9*(4), 813–839.

Lin, C.-J.-C., & Bever, T. G. (2006). Subject preference in the processing of relative clauses in Chinese. In D. Baumer, D. Montero, & M. Scanlon (Eds.), *Proceedings of the 25th West Coast conference on formal linguistics* (pp. 254–260). Somerville, MA: Cascadilla Proceedings Project.

Linzen, T., & Jaeger, T. F. (2015). Uncertainty and expectation in sentence processing: Evidence from subcategorization distributions. *Cognitive Science*, *40*(6), 1382–1411.

Lowder, M. W., Choi, W., Ferreira, F., & Henderson, J. M. (2018). Lexical predictability during natural reading: Effects of surprisal and entropy reduction. *Cognitive Science*, *42*, 1166–1183.

Lowder, M. W., & Gordon, P. C. (2012). The pistol that injured the cowboy: Difficulty with inanimate subject–verb integration is reduced by structural separation. *Journal of Memory and Language*, *66*(4), 819–832.

Macdonald, R., Brandt, S., Theakston, A., Lieven, E., & Serratrice, L. (2020). The role of animacy in children's interpretation of relative clauses in English: Evidence from sentence–picture matching and eye movements. *Cognitive Science*, *44*(8), e12874.

MacWhinney, B., Bates, E., & Kliegl, R. (1984). Cue validity and sentence interpretation in English, German and Italian. *Journal of Verbal Learning and Verbal Behavior*, *23*, 127–150.

MacWhinney, B., & Pleh, C. (1988). The processing of restrictive relative clauses in Hungarian. *Cognition*, *29*(2), 95–141.

Mak, W., Vonk, W., & Schriefers, H. (2002). The influence of animacy on relative clause processing. *Journal of Memory and Language*, *47*(1), 50–68.

Mak, W., Vonk, W., & Schriefers, H. (2006). Animacy in processing relative clauses: The hikers that rocks crush. *Journal of Memory and Language*, *54*, 466–490.

Marcus, M. P., Santorini, B., & Marcinkiewicz, M. A. (1993). Building a large annotated corpus of English: The Penn treebank. *Computational Linguistics*, *19*(2), 313–330.

McElree, B. (2006). Accessing recent events. *Psychology of Learning and Motivation*, *46*, 155–200.

McElree, B., Foraker, S., & Dyer, L. (2003). Memory structures that subserve sentence comprehension. *Journal of Memory and Language*, *48*(1), 67–91.

Mitchell, D. C., Cuetos, F., Corley, M. M. B., & Brysbaert, M. (1995). Exposure-based models of human parsing: Evidence for the use of coarse-grained (nonlexical) statistical records. *Journal of Psycholinguistic Research*, *24*, 469–488.

Nakamura, M., & Miyamoto, E. T. (2013). The object before subject bias and the processing of double-gap relative clauses in Japanese. *Language and Cognitive Processes*, *28*(3), 303–334.

Narayanan, S., & Jurafsky, D. (2002). A Bayesian model predicts human parse preference and reading time in sentence processing. In T. G. Dietterich, S. Becker, & Z. Ghahramani (Eds.), *Advances in neural information processing systems* (Vol. 14, pp. 59–65). Cambridge, MA: MIT Press.

Nederhof, M.-J., & Satta, G. (2008). Computation of distances for regular and context-free probabilistic languages. *Theoretical Computer Science*, *395*, 235–254.

Nelson, M., Dehaene, S., Pallier, C., & Hale, J. (2017). Entropy reduction correlates with temporal lobe activity. In Proceedings of the 7th workshop on cognitive modeling and computational linguistics (CMCL 2017) (pp. 1–10). Valencia, Spain.

Price, I. K., & Witzel, J. (2017). Sources of relative clause processing difficulty: Evidence from Russian. *Journal of Memory and Language*, *97*, 208–244.

Reali, F., & Christiansen, M. H. (2007). Processing of relative clauses is made easier by frequency of occurrence. *Journal of Memory and Language*, *57*(1), 1–23.

Roark, B., Bachrach, A., Cardenas, C., & Pallier, C. (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 (Vol. 1, pp. 324–333).

Rohde, D. (1999). The simple language generator: Encoding complex languages with simple grammars (Tech. Rep.). Pittsburgh, PA: Department of Computer Science, Carnegie Mellon University.

Roland, D., Dick, F., & Elman, J. L. (2007). Frequency of basic English grammatical structures: A corpus analysis. *Journal of Memory and Language*, *57*, 348–379.

Roland, D., Mauner, G., O'Meara, C., & Yun, H. (2012). Discourse expectations and relative clause processing. *Journal of Memory and Language*, *66*(3), 479–508.

Seki, H., Matsumura, T., Fujii, M., & Kasami, T. (1991). On multiple context-free grammars. *Theoretical Computer Science*, *88*(2), 191–229.

Sheldon, A. (1974). On the role of parallel function in the acquisition of relative clauses in English. *Journal of Verbal Learning and Verbal Behavior*, *13*, 272–281.

Silverstein, M. (1976). Hierarchy of features and ergativity. In R. M. W. Dixon (Ed.), *Grammatical categories in Australian languages* (pp. 112–171). Canberra: Australian Institute of Aboriginal Studies.

Stabler, E. (1997). Derivational minimalism. In C. Retoré (Ed.), *Logical aspects of computational linguistics* (pp. 68–95). New York: Springer-Verlag.

Stabler, E. (2013). The epicenter of linguistic behavior. In M. Sanz, I. Laka, & M. K. Tanenhaus (Eds.), *Language down the garden path: The cognitive and biological basis of linguistic structures* (pp. 316–323). New York: Oxford University Press.

Staub, A. (2010). Eye movements and processing difficulty in object relative clauses. *Cognition*, *116*(1), 71–86.

Staub, A., Dillon, B., & Clifton, C. (2017). The matrix verb as a source of comprehension difficulty in object relative sentences. *Cognitive Science*, *41*, 1353–1376.

Stowe, L. A. (1986). Parsing wh-constructions: Evidence for on-line gap location. *Language and Cognitive Processes*, *1*(3), 227–245.

Stromswold, K., Caplan, D., Alpert, N., & Rauch, S. (1996). Localization of syntactic comprehension by positron emission tomography. *Brain and Language*, *52*, 452–473.

Traxler, M., Morris, R. K., & Seely, R. E. (2002). Processing subject and object relative clauses: Evidence from eye movements. *Journal of Memory and Language*, *47*, 69–90.

Traxler, M., Williams, R., Blozis, S., & Morris, R. (2005). Working memory, animacy, and verb class in the processing of relative clauses. *Journal of Memory and Language*, *53*, 204–224.

Van Dyke, J. A. (2007). Interference effects from grammatically unavailable constituents during sentence processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(2), 407.

Van Dyke, J. A., & Lewis, R. L. (2003). Distinguishing effects of structure and decay on attachment and repair: A cue-based parsing account of recovery from misanalysed ambiguities. *Journal of Memory and Language*, *49*(3), 285–316.

van Schijndel, M., Mueller, A., & Linzen, T. (2019). Quantity doesn't buy quality syntax with neural language models. In *Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (EMNLP-IJCNLP)* (pp. 5831–5837). Hong Kong, China.

van Schijndel, M., & Schuler, W. (2017). Approximations of predictive entropy correlate with reading times. In G. Gunzelmann, A. Howes, T. Tenbrink, & E. Davelaar (Eds.), *Proceedings of the 39th annual meeting of the Cognitive Science Society* (pp. 1260–1265). London: Cognitive Science Society.

Vasishth, S., Chen, Z., Li, Q., & Guo, G.-L. (2013). Processing Chinese relative clauses: Evidence for the subject-relative advantage. *PLoS One*, *8*(10), e77006.

Vasishth, S., Nicenboim, B., Engelmann, F., & Burchert, F. (2019). Computational models of retrieval processes in sentence processing. *Trends in Cognitive Sciences*, *23*(11), 968–982.

Venhuizen, N. J., Crocker, M. W., & Brouwer, H. (2019a). Expectation-based comprehension: Modeling the interaction of world knowledge and linguistic experience. *Discourse Processes*, *56*(3), 229–255.

Venhuizen, N. J., Crocker, M. W., & Brouwer, H. (2019b). Semantic entropy in language comprehension. *Entropy*, *21*(12), 1159.

Vergnaud, J.-R. (1985). *Dépendances et Niveaux de Représentation en Syntaxe (Vol. 13)*. Amsterdam, the Netherlands: John Benjamins Publishing.

Wagers, M. W., Borja, M. F., & Chung, S. (2018). Grammatical licensing and relative clause parsing in a flexible word-order language. *Cognition*, *178*, 207–221.

Wagers, M. W., & Pendleton, E. (2016). Structuring expectation: Licensing animacy in relative clause comprehension. In K.-M. Kim, P. Umbal, T. Block, Q. Chan, T. Cheng, K. Finney, M. Katz, & S. Nickel-Thompson (Eds.), *Proceedings of the 33rd West Coast Conference on Formal Linguistics (WCCFL)* (pp. 29–46). Somerville, MA: Cascadilla Proceedings Project.

Wanner, E., & Maratsos, M. (1978). An ATN approach in comprehension. In M. Halle, J. Bresnan, & G. Miller (Eds.), *Linguistic theory and psychological reality* (pp. 119–161). Cambridge, MA: MIT Press.

Warren, T., & Gibson, E. (2002). The influence of referential processing on sentence complexity. *Cognition*, *85*(1), 79–112.

Wilson, K., & Carroll, J. B. (1954). Applications of entropy measures to problems of sequential structure. In C. E. Osgood & T. A. Sebeok (Eds.), *Psycholinguistics: A survey of theory and research* (pp. 103–110). Bloomington, IN: Indiana University Press.

Wu, F., Kaiser, E., & Andersen, E. (2012). Animacy effects in Chinese relative clause processing. *Language and Cognitive Processes*, *27*(10), 1489–1524.

Wu, F., Kaiser, E., & Vasishth, S. (2018). Effects of early cues on the processing of Chinese relative clauses: Evidence for experience-based theories. *Cognitive Science*, *42*, 1101–1133.

Wu, S., Bachrach, A., Cardenas, C., & Schuler, W. (2010). Complexity metrics in an incremental right-corner parser. In *Proceedings of the 48th annual meeting of the Association for Computational Linguistics (ACL-2010)* (pp. 1189–1198). Association for Computational Linguistics.

Yun, J., Chen, Z., Hunter, T., Whitman, J., & Hale, J. (2015). Uncertainty in processing relative clauses across East Asian languages. *Journal of East Asian Linguistics*, *24*(2), 113–148.

Zaenen, A., Carletta, J., Garretson, G., Bresnan, J., Koontz-Garboden, A., Nikitina, T., . . . Wasow, T. (2004). Animacy encoding in English: Why and how. In *Proceedings of the 2004 ACL Workshop on Discourse Annotation* (pp. 118–125). Association for Computational Linguistics.

Appendix A: The tree diagram of English object relative clauses

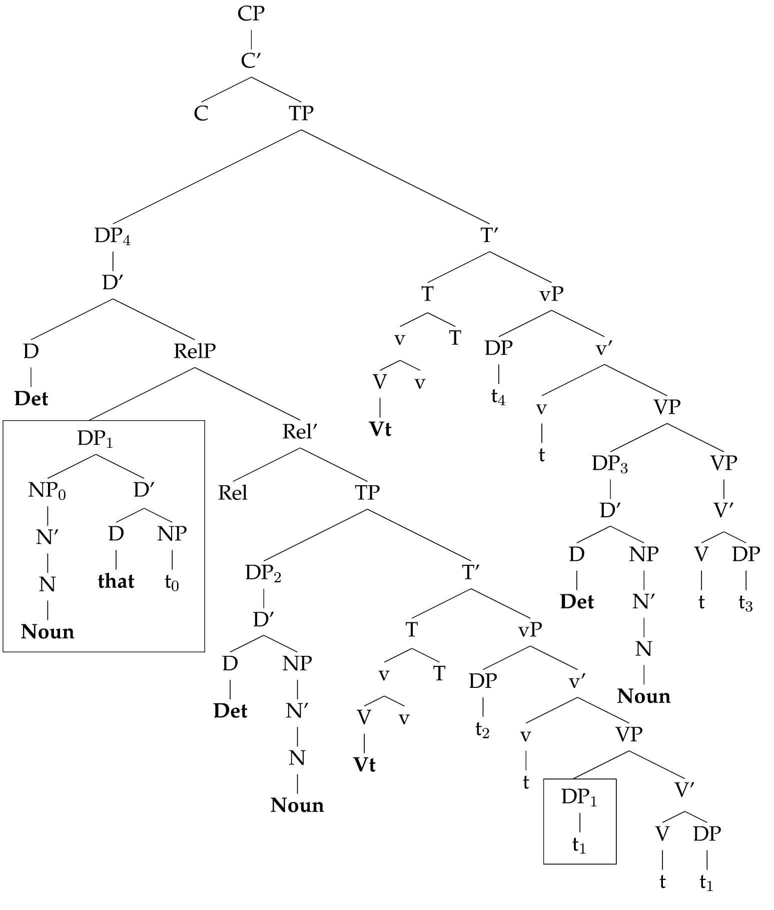


Fig A1. The tree diagram of English ORs derived by the grammar fragment under the promotion analysis (Kayne, 1994). Abstract categories at the terminal nodes correspond to the surface string "Det Noun [that Det Noun Vt] Vt Det Noun."

Appendix B: Word-by-word surprisal predictions for RC processing

Table B1.
Word-by-word surprisals of RC processing, based on all derivations with a probability equal to or greater than 10^{-5}, or on only a subset of derivations (e.g., 3, 5, or 10) with highest probability. The results predict the subject advantage at the determiner, but not at the RC verb

Derivations	Type	*that*	Det	Noun	RC verb
$k = 3$	SR	5.91	0.06	0	0
	OR	5.91	4.25	0	0
$k = 5$	SR	5.78	0.17	0	0.01
	OR	5.78	4.38	0	0
$k = 10$	SR	5.39	0.28	0	0.22
	OR	5.39	4.76	0	0.14
$p \geq 10^{-5}$	SR	4.54	0.51	0	0.84
	OR	4.54	5.84	0	0.14