The Syntactic Complexity of Russian Relative Clauses

Roger Levy\textsuperscript{a}, Evelina Fedorenko\textsuperscript{b}, Ted Gibson\textsuperscript{b}

\textsuperscript{a} Department of Linguistics, UC San Diego, 9500 Gilman Drive #0108, La Jolla, CA 92093-0108, USA
\textsuperscript{b} Department of Brain \& Cognitive Sciences Massachusetts Institute of Technology 77 Massachusetts Ave, Cambridge, MA 02139, USA

Abstract

Although syntactic complexity has been investigated across dozens of studies, the available data still greatly underdetermine relevant theories of processing difficulty. Memory-based and expectation-based theories make opposite predictions regarding fine-grained time course of processing difficulty in syntactically constrained contexts, and each class of theory receives support from results on some constructions in some languages. Here we report four self-paced reading experiments on the online comprehension of Russian relative clauses together with related corpus studies, taking advantage of Russian’s flexible word order to disentangle predictions of competing theories. We find support for key predictions of memory-based theories in reading times at RC verbs, and for key predictions of expectation-based theories in processing difficulty at RC-initial accusative noun phrase (NP) objects, which corpus data suggest should be highly unexpected. These results suggest that a complete theory of syntactic complexity must integrate insights from both expectation-based and memory-based theories.

1. Introduction

Human language is distinctive among the communicative systems found in nature in its infinite expressivity. To a first approximation, every utterance that a comprehender hears is one that they have never heard before. The comprehender must thus deploy finitely-represented knowledge of language in real time to analyze the utterance. A crucial aspect of this knowledge is that of syntax, which allows a comprehender to recover the meaningful relationships between words arranged in sequences that may never have previously been encountered. The cognitive difficulty incurred in the deployment of syntactic knowledge is, however, differential and localized: not all sentences are equally easy to comprehend successfully, nor are all parts of a given sentence equally easy, and in many cases the difficulty of a given sentence is attributable to its specific syntactic properties. One key part of the central problem of sentence comprehension can thus be stated as follows: what major cognitive constraints govern the deployment of syntactic knowledge to achieve understanding in real time?

It has long been known that one major cognitive constraint in the deployment of syntactic knowledge is that humans cannot simultaneously pursue all possible analyses of an input string (partial or complete) in a cost-free way. Hence extensive work has been done on
the problem of SYNTACTIC AMBIGUITY RESOLUTION, where a local ambiguity of syntactic interpretation is subsequently resolved through the influence of one or more information sources (Bever, 1970; Ferreira and Clifton, 1986; Frazier and Fodor, 1978; MacDonald et al., 1994; Mitchell, 1994; Tanenhaus et al., 1995; Tanenhaus and Trueswell, 1995; Trueswell et al., 1994; among many others). For example, the first three words of sentence (1) are ambiguous between readings in which the defendant is the agent or the patient of the verb "examined":

(1) The defendant examined by the lawyer turned out to be guilty.

For native English speakers there is measurable processing difficulty during comprehension of the rest of the sentence, which rules out the agentive reading. Though there remains disagreement regarding precise empirical details in syntactic ambiguity resolution, most notably how quickly non-syntactic information sources can be utilized, whether more than one analysis can ever be simultaneously entertained (e.g., Clifton Jr. et al., 2003), and the extent to which globally incoherent analyses are considered (Taboer et al., 2004), considerable evidence has also accumulated demonstrating humans' abilities in this area, and probability theory has emerged as a powerful formal framework for describing the cognitive constraints relevant in ambiguity resolution (Jurafsky, 1996).

This paper reports the results of experiments whose goal was to shed further light on cases of processing difficulty which do not seem to arise from ambiguity in the analysis of a partial input string; we will use the term SYNTACTIC COMPLEXITY to describe such cases (Gibson, 1998, 2000; Lewis, 1996; Miller and Chomsky, 1963; Yngve, 1960, inter alia). There is less agreement in the field on the nature of the cognitive constraints underlying syntactic complexity; one hope is that theories of syntactic complexity in locally unambiguous contexts may be able to subsume theories of ambiguity resolution and thus lead to a more parsimonious and satisfactory theory overall (Clifton and Frazier, 1989; Gibson, 1991, 1998; Grodner et al., 2002; Hale, 2001, 2003, 2006; Levy, 2008). One of the most-studied cases of syntactic complexity is the asymmetry in processing difficulty between English SUBJECT-EXTRACTED and OBJECT-EXTRACTED transitive relative clauses (RCs) as in (2) below, in which both the head noun phrase (NP; the reporter in 2) and the RC-internal NP (the subject in an object-extracted RC, or the object in a subject-extracted RC; the senator in (2)) are animate, definite, and full.

(2) a. The reporter who attacked the senator hoped for a story. (Subject-extracted RC; SRC)
   b. The reporter who the senator attacked hoped for a story. (Object-extracted RC; ORC)

A wide range of experimental studies (Wanner and Maratsos, 1978; Ford, 1983; King and Just, 1991; Grodner and Gibson, 2005; Gordon et al., 2001; Traxler et al., 2002, inter alia) have demonstrated that comprehension difficulty is differential for these cases: the object-extracted RC (ORC; (2b)) is more difficult than the subject-extracted RC (SRC; (2a)). These studies have also demonstrated that processing difficulty is localized: the locus of greatest processing difficulty is at the ORC verb (Grodner and Gibson, 2005). More recently, the results of
Staub (2010) suggest that the onset of the subject NP in ORCs may also be a locus of some processing difficulty (a point we will return to in Section 3.3). Hence ORCs of the type seen in (2b) are more complex than the SRCs of the type seen in (2a), and the measurable processing difficulty associated with that complexity is localizable to two different regions within the RC. Relative clauses have played a particularly prominent role in the study of online syntactic comprehension, partly because they exemplify one of the formally most complex corners of natural language syntax and play a key role in how language achieves its full richness of expressive capacity, and partly because, as in the example above, they have been a rich source of syntactic complexity results in which ambiguity avoidance plays no apparent role. The English SRC/ORC processing asymmetry of (2) serves as an effective touchstone for describing the wide variety of theories of syntactic complexity prominent in the literature today and upon which the new research reported in this paper will bear.

1.1. Memory versus expectations as foundations of syntactic complexity

In broad strokes, two prominent classes of theory regarding the key cognitive constraint determining syntactic complexity can be identified: theories based on MEMORY LIMITATIONS and theories based on EXPECTATIONS (see Gibson and Wu, in press, for a similar summary). One such theory based on memory limitations is the DEPENDENCY LOCALITY THEORY (DLT, also called SYNTACTIC PREDICTION LOCALITY THEORY: Gibson, 1998, 2000), according to which the key operations in syntactic comprehension are STORAGE and RETRIEVAL of potential elements in structural dependency relationships within a sentence, and INTEGRATION of a retrieved preceding element into a structural dependency relation with the current input. On this theory, the resources involved in retrieval, integration, and maintenance of stored-element representations are limited. Thus dependency integrations are more difficult when more elements need to be integrated simultaneously, and when the retrieved elements have greater linear distance from the integration site. The DLT successfully predicts the English SRC/ORC processing difficulty asymmetry: the most integration-intensive word in either RC of (2) is the ORC verb attacked, with which both the preceding subject and object NPs must simultaneously be integrated; no other word in (2a) or (2b) involves more than one simultaneous integration (see Gibson, 1998, 2000 for further details).

A closely related theory is the DECAY/REACTIVATION/RETRIEVAL-INTERFERENCE theory of Lewis and Vasishth (2005; see also Vasishth and Lewis, 2006; Lewis et al., 2006). In this theory, the representation of a sentence in real-time comprehension is an incrementally extended syntactic structure; similar to DLT, the theory’s processing bottleneck is retrieval of a preceding syntactic element or elements from this structure, with which the current input word must be integrated. Once an element is stored in memory, its activation level begins to decay, so that greater linear distance between a dependent and its governor generally increases the difficulty of the dependency integration, as in DLT. A distinguishing feature of the decay/reactivation theory, however, is that when elements of the incremental structure are accessed intermediately, they are reactivated, counteracting decay. This reactivation means that additional intervening constituents can under some circumstances facilitate rather than hinder an integration spanning long linear distances (Vasishth and Lewis, 2006). Countervailing against the facilitatory effect of reactivation, however, is SIMILARITY-BASED
Interference (SBI; Gordon et al., 2001, 2004; Lewis and Vasishth, 2003; Lewis et al., 2006; McElree, 2000; McElree et al., 2003; Van Dyke and McElree, 2006): because memory is content-addressable and retrieval involves cue-based competition among stored syntactic elements, retrieval is more difficult and error-prone when the preceding context contains other elements featurally similar to the retrieval target. For (2), decay/activation and other SBI theories predict the English SRC/ORC processing difficulty asymmetry because in the ORC case, both reporter and senator need to be retrieved at the RC verb and associated with their appropriate semantic roles, but they interfere with one another due to their similarity (e.g., both are animate, singular, and definite). Both DLT and decay/reactivation/retrieval-interference theories make fine-grained predictions regarding the processing difficulty of each word in a sentence.

In expectation-based theories of syntactic complexity, in contrast, the key constraining factor is not memory but rather experience and/or generalization: structures with which individuals have more direct experience, or which they infer to be likely in a particular context given their linguistic and world knowledge, are easier to process in comprehension. In word-order frequency theories, surface orderings of word classes which occur more frequently in the input are hypothesized to be favored and thus easier to process during comprehension (Bever, 1970; MacDonald and Christiansen, 2002). In such theories, the greater processing difficulty of the ORC in (2) would be attributed to the fact that its surface word order, Object-Subject-Verb (reporter-senator-attacked in 2b), is rare in English, whereas the SRC has the ubiquitous surface word order Subject-Verb-Object. Such theories are closely related to the Tuning Hypothesis (Mitchell et al., 1995), which posits that coarse-grained structural statistics are tracked in linguistic input and used to make decisions in online comprehension, though the Tuning Hypothesis has historically been framed with respect to problems of ambiguity resolution rather than with problems of syntactic complexity. The predictions of word-order frequency theories regarding where difficulty will be observed, however, are relatively coarse-grained, not word-by-word.

Another expectation-based theory is surprisal, according to which comprehenders maintain and update fine-grained expectations regarding upcoming input at multiple levels of linguistic structure (including but not limited to syntax), and the difficulty of processing an input in the context in which it appears is monotonically decreasing in its conditional probability (Hale, 2001; Levy, 2008; Smith and Levy, 2008). Surprisal is in some ways like a word-by-word instantiation of the word-order theory outlined above, but does not commit to the stance that expectations are based on superficial sequences of word categories. Rather, in many models instantiating surprisal theory rich syntactic context is taken into account (Boston et al., 2008, 2011; Demberg and Keller, 2008; Hale, 2001; Levy, 2008; Roark et al., 2009; Smith and Levy, 2008), so that, for example, RC-internal word order expectations might in principle be completely different from the expectations arising in independent clauses with superficially similar word order, depending on the grammatical properties of RCs in the language in question. Surprisal can account for the overall difference in English SRC/ORC comprehension difficulty because among transitive RCs whose head noun and RC NP are both full, definite NPs, SRCs are much more common than ORCs (Hale, 2001; Levy, ...
hence, the overall surprisal of the ORC is higher than that of the SRC. Surprisal is less effective, however, at predicting where processing difficulty in ORCs is localized: it predicts that the processing penalty is paid at the onset of the RC NP, which disconfirms the possibility that the RC is subject-extracted (see discussion in Grodner and Gibson, 2005 and Levy, 2008). As mentioned earlier, the results of Staub (2010) suggest that there is in fact a processing cost at this point, but the bulk of experimental data point to the RC verb as the primary locus of ORC processing difficulty (Gordon et al., 2001; Grodner and Gibson, 2003; Staub, 2010).

A third expectation-based theory is the Entropy Reduction Hypothesis (ERH; Hale, 2003, 2006). In the ERH, the entropy (Shannon, 1948; Cover and Thomas, 1991) of the distribution of possible structural completions of the sentence at any point in incremental processing is a quantity of fundamental interest; processing difficulty is posited to ensue when a word causes a large drop in this entropy. According to the analysis of Hale (2003), the ERH successfully localizes processing difficulty at the verb of English ORCs: the point immediately following any common noun is high-entropy because common nouns are often recursively postmodified; the possibility of recursive postmodification yields a high-entropy distribution over sentence continuations. The RC verb rules out this infinity of possible NP postmodifications, and thus yields a large drop in entropy. In SRCs, in contrast, the verb follows the word who, which does not admit the possibility of recursive postmodification, so that the drop in entropy induced by the RC verb is much smaller. Expectation-based theories gain some additional degree of support from studies indicating that more frequent types of ORCs are in fact easier to process (Gordon et al., 2001, 2004; Gennari and MacDonald, 2009; Reali and Christiansen, 2007; Traxler et al., 2002; Warren and Gibson, 2002). As one particularly striking example, Reali and Christiansen (2007) found that among English RCs with pronominal RC NPs, ORCs (such as the woman who you called) are actually more frequent than SRCs (such as the woman who called you). Reali and Christiansen also found that among RCs of this type it is ORCs, not SRCs, that are read more quickly.

Although memory- and expectation-based approaches are aligned in predicting the general pattern of English ORCs being more difficult than SRCs, they differ in their specific empirical predictions regarding word-by-word processing difficulty, both for RCs and other constructions. These differences can perhaps be cast into sharpest relief when syntactically constrained contexts are considered: cases where the preceding context of a sentence sets up an expectation that some syntactic category X will be encountered in upcoming input, but precisely when X will appear and what word will instantiate it remain unknown until it is encountered. In these cases, expectation-based and memory-based theories make close to opposite predictions regarding the effect of processing difficulty of X as a function of the number of X’s preceding dependents. For memory-based theories, the more material appearing in the input before X is encountered, the greater the burden placed on memory and hence the harder X should be to process when it is encountered. For expectation-based theories, in contrast, additional material can on average only help the comprehender sharpen her expectations regarding the location and identity of X; this additional material should thus in general facilitate processing.
of X when it appears. English relative clauses are such a syntactically constrained context: once the initiation of the RC is cued by the relative pronoun, the comprehender knows that an RC verb must appear (Grodner and Gibson, 2005; Levy, 2008). In the SRC, this verb appears immediately after the RC onset (Figure 1a), at which point the comprehender was as yet uncertain as to whether the RC is subject-extracted and has seen only one of the arguments of the RC verb. In the ORC, in contrast, this verb appears after the comprehender knows that the RC is not subject-extracted and has seen two arguments of the RC verb (Figure 1b). Thus the comprehender should have a stronger expectation in the ORC that the verb will appear when it does in fact appear, and should have sharper expectations regarding the identity of this verb; but at the same time needs to perform more memory retrieval operations upon encountering the verb, and these retrieval operations may be more difficult than in the SRC case. In the case of English RCs as seen in (2), the observed pattern of processing difficulty matches the predictions of memory-based theories such as DLT and SBI.

However, results from empirical investigation of other syntactically-constrained contexts conform in many cases with the predictions of expectation-based theories, not with those of memory-based theories. As one example, Vaisishth and Lewis (2006) used self-paced reading to study online comprehension of Hindi object-extracted relative clauses, as in (3) below:

(3) a. Vo kaagaz jisko us larke-ne dekhaa bahut puraanaa thaa.  
that paper which that boy-saw very old was  
“The paper that that boy saw was very old.”

b. Vo kaagaz jisko us larke-ne mez-ke piiche gire.hue dekhaa bahut  
that paper which that boy-saw table-erg behind fallen saw very

1In surprisal, this average benefit of additional preceding material can be mathematically proven: it is equivalent to the well-known proof that conditionalizing a random variable X on another random variable Y can never decrease entropy (Cover and Thomas, 1991, chapter 2): \( H(X|Y) \leq H(X) \).

2Additional preceding material that is not dependent on X can also create a similar contrast. For example, contrast the English ORC the farmer who the girl with the telescope spotted with the ORC the farmer who the girl spotted; the former example has an additional PP, with the telescope, preceding the RC verb. Even though this preceding PP is dependent on the NP subject, not on the RC verb, the semantic content of the PP nevertheless provides a useful clue as to the RC verb’s identity, and so expectation-based theories predict that the PP gives a processing benefit at the RC verb (assuming that adding the PP would in fact increase the likelihood that the RC continues with spotted, as measurable, for example, in a Cloze study). Nevertheless, in this paper we focus on experimental manipulations involving the preceding dependents of a verb, because they elicit a stronger contrast in the predictions of expectation-based and memory-based theories: on memory-based theories such as DLT and SBI, the predicted difficulty increase from additional preceding dependents is greater than from additional preceding non-dependent material, because the additional preverbal dependents not only interfere, they must also themselves be integrated with the verb.
The RC verb *dekhaa* ("saw") has more preceding dependents in [3b] than in [3a], thus memory-based theories predict greater integration difficulty. However, Vasishth and Lewis found that reading times at the RC verb were faster, not slower, in [3b] than in [3a]. This finding is difficult to reconcile with the DLT: additional preverbal dependents apparently reduce, rather than increase, processing difficulty at the verb. For decay/reactivation this finding can be explained as the additional intervening constituents reactivating the prediction for a clause-final verb set up by the RC onset and the RC-initial subject NP (Vasishth and Lewis, 2006). For the ERH, this finding is predicted because the additional preverbal dependents in [3b] render the uncertainty immediately before the RC-final verb regarding how the RC might be completed lower than in [3a]. The same considerations hold for surprisal as for the ERH; furthermore, the extra preverbal dependents may give the comprehender additional predictive benefit regarding the identity of the RC verb (e.g., there are fewer eventualities that might hold with "paper" as object and "fallen behind a table" as a secondary predicate than with "paper" as object alone). Similar patterns of results have been found in comprehension of verb-final main clauses in German (Konieczny, 2000; Konieczny and Döring, 2003) and Japanese (Miyamoto and Nakamura, 2003; Uno and Garnsey, 2008; Nakatani and Gibson, 2010); there has also been one report of such effects in English main verbs following subject-modifying relative clauses (Jaeger et al., 2008).

1.2. Syntactic complexity theories specific to relative clauses

The work we present here also bears on theories of syntactic complexity specific to relative clauses—specifically, differences in processing complexity of RCs of different extraction types. Perspective Shift (MacWhinney and Peh, 1998) proposes that the English SRC/ORC asymmetry seen in [2] may arise from a processing penalty specific to cases where the grammatical roles of the head noun in the main and relative clauses differ (cf. Gibson et al., 2005). Universal Structural Asymmetry theories (Lin and Bever, 2006; O’Grady, 1997) propose that SRCs should always be easier to comprehend than ORCs due to the higher structural position and thus greater accessibility of the SRC’s extraction site.

1.3. Grammatical properties of languages and disentangling theories

As described in the foregoing discussion, the syntactic complexity in comprehension of verbs in general and RC verbs in particular is an area of considerable theoretical interest in which much empirical data are available and yet fail in many cases to distinguish conclusively among competing theories. The starting point for the new studies presented in this paper is the observation that a number of potentially crucial grammatical properties tend to be confounded in the available data: the word order preference of the language investigated, the morphological richness of the language, and the construction type investigated. The clearest cases supporting memory-based theories come from studies of SRCs and ORCs in English (Wanner and Maratsos, 1978; Ford, 1983; King and Just, 1991; Gordon et al., 2001;
Grodner and Gibson, 2005; Fedorenko et al., 2011), French (Holmes and O’Regan, 1981; Frauenfelder et al., 1980; Cohen and Mehler, 1996), and Chinese (Hsiao and Gibson, 2003, Gibson and Wu, in press; though see Lin and Bever, 2006; Chen et al., 2010). These studies focus on relative clauses and involve languages with relatively rigid word order, predominantly SVO, and with sparse morphological marking of grammatical roles. The clearest cases supporting expectation-based theories come from studies of Hindi (Vasishth, 2002; Vasishth and Lewis, 2006), German (Konieczny, 2000; Konieczny and Döring, 2003; Levy and Keller, 2012), and Japanese (Miyamoto and Nakamura, 2003; Ishizuka et al., 2003; Uneo and Garnsey, 2008; Nakatani and Gibson, 2010), languages with predominantly verb-final word order, relatively greater flexibility of non-verbal constituent ordering, and rich systems of morphological case explicitly marking the grammatical role of the main dependents of the verb. Furthermore, with the exception of Vasishth and Lewis’s study of Hindi RCs and Levy and Keller’s study of German RCs, the cases supporting expectation-based theories did not investigate relative clauses.

Here we report data from the comprehension of relative clauses in Russian, which are attractive in several respects given the current theoretical and empirical landscape. Like Chinese, English, and French but unlike Hindi, German, Korean, and Japanese, the predominant word order in Russian is SVO; its relative clauses appear postnominally as in English, French, Hindi, and German. Unlike the SVO languages mentioned above, however, Russian also has a rich morphological case system which marks the grammatical roles of verbal dependents, similar to Hindi, German and Japanese but unlike English and Chinese. Furthermore, Russian word order is freer than any of the above languages: although SVO is the predominant word order, all permutations of major clausal constituents are in fact permissible. This word order freedom allows us a flexibility of experimental design unavailable in these other languages: we can completely disentangle what material intervenes between an RC onset and the RC verb both from RC extraction type and from the inventory of clausal constituents encountered within the RC as whole. As a result, studying the online comprehension of Russian relative clauses may allow us to discriminate among competing theories of syntactic complexity more effectively than has been possible thus far. In Experiment 1 we use this flexibility to tease apart the contributions of extraction type and RC word order to Russian RC syntactic complexity. In Experiment 2 we use it to parametrically vary the number of clausal constituents intervening between the head noun and the RC verb; we also compare the effects of NP argument interveners and NP adjunct interveners on RC verb processing difficulty.

The studies reported here are to our knowledge the first work on the online comprehension of Russian relative clauses. However, since the first presentation of Experiments 1b and 2a Maria Polinsky has communicated to us closely related work carried out on comprehension of SRCs and ORCs with scrambled and default word order (i.e., the same experimental design we use in Experiment 1) among monolingual child, monolingual adult, heritage child, and heritage adult speakers of Russian, testing speakers’ ability to match reversible action pictures to SRC and ORC descriptions (Polinsky, 2011). Among all but heritage adult speakers, comprehension was high across the board; among heritage adult speakers, SRC comprehension was highly accurate but ORC comprehension was at chance.
2. Experiment 1

In this experiment we use the word order flexibility of Russian to disentangle effects of extraction type from effects of word order and dependency locality on RC syntactic complexity. All RCs are subject-modifying; hence any effects of word order on RC processing difficulty could not be accounted for purely by perspective-shift or universal structural asymmetry theories. Furthermore, our manipulation of word order will have some power to discriminate expectation-based theories—specifically surprisal and potentially word-order frequency theories—from memory-based theories. We cross extraction type (SRC versus ORC) with whether the RC-internal word order is default (VO in SRCs, SV in ORCs) or scrambled (OV in SRCs, VS in ORCs) with respect to Russian’s canonical SVO main-clause word order, as in (4) below.

(4) a. [SRC, DEFAULT]
   Slesar kotoryj udaril elektrika poterjal terpenie v spore.
   Repairman who.NOM hit electrician.ACC lost patience in argument.
   “The repairman who hit the electrician lost his patience in the argument.”

b. [SRC, SCRAMBLED]
   Slesar kotoryj elektrika udaril poterjal terpenie v spore.
   Repairman who.NOM electrician.ACC hit lost patience in argument.
   “The repairman who hit the electrician lost his patience in the argument.”

c. [ORC, DEFAULT]
   Slesar kotorogo elektrik udaril poterjal terpenie v spore.
   Repairman who.ACC electrician.NOM hit lost patience in argument.
   “The repairman whom the electrician hit lost his patience in the argument.”

d. [ORC, SCRAMBLED]
   Slesar kotorogo udaril elektrik poterjal terpenie v spore.
   Repairman who.ACC hit electrician.NOM lost patience in argument.
   “The repairman whom the electrician hit lost his patience in the argument.”

This design can also be interpreted as crossing RC extraction type with the locality (proximity) of the RC verb with respect to the relative pronoun (LOCAL in [4a] and [4d] versus NON-LOCAL in [4b] and [4c]). This disentangling of verb-relative pronoun locality from RC extraction type would not be possible in languages like English with more fixed word order (e.g., Grodner and Gibson, 2005). With this design, however, comes a second difference between English and Russian: the Russian relative pronouns used in (4) explicitly encode the grammatical function of the head NP through case marking, whereas English *that* and

---

4We refer to ORCs with SV and VS internal word orders as “default” and “scrambled” respectively for purposes of characterizing our experimental design on the logic that if the default independent-clause word order of Russian is SVO, then SV word order is obtained for an ORC by extracting the object out of a default-order transitive clause. See Section 2.1, however, for a caveat as to how much should be read into the “default/scrambled” theoretically.
who do not. In Experiment 1a we restrict our attention to these cases; in Experiment 1b we also investigate the situation where a case-syncretized relative pronoun, chto, is used to introduce the RC.

2.1. Information structure and word order in Russian relative clauses

Although Russian is frequently described as a language with “free” word order, it is widely recognized among linguists of Russian that there is strong functional motivation for the choices of different word orders in different context. Although a comprehensive review of the literature on this issue is far beyond the scope of the present paper, here we briefly describe issues most relevant to the present studies; the reader is referred to Krylova and Khavronina (1988), King (1995), and Bailyn (2011) for influential accounts with further references. All prominent accounts ascribe at least some degree of word order variability to information-structural considerations. Perhaps the most widely recognized characterization is the bipartite division of every Russian sentence into theme and rheme—loosely speaking, that which the sentence is about and the new information conveyed (these terms roughly correspond to topic and focus in much of both the generative and functional linguistics literatures). On the influential account of Krylova and Khavronina (1988), for example, SVO is the “default” order of simple transitive sentences in Russian, and it is generally agreed that among sentences with the most common intonational contour (so-called “non-emotive” sentences), deviations from the default word order require a context in which SVO would not satisfy the principle of the theme entirely preceding the rheme. For example,

(5) Elektrika udaril slesar
electrician.ACC hit repairman.NOM
“The repairman hit the electrician.”

would be inappropriate in a null context, but would be appropriate as, for example, an answer to the question, “Who hit the electrician?”, which would render elektrika and udaril as part of the theme, with slesar the rheme.

Unfortunately, the vast majority of available literature in this area deals with word order in independent clauses, leaving it far less clear how such theories relate to the ordering of words within the Russian relative clause. The recent review of Bailyn (2011), for example, spends an entire chapter on word order variation in independent clauses; on the topic of subordinate-clause word order, all that is said is, “... because of the tight connection between discourse structure and word order, subordinate clauses may show less word order variation than main clauses.” Nor is this atypical of the literature. It is not even clear a priori whether the notion of “default” word order exists for Russian relative clauses. Our corpus analysis and reading-time studies will turn out to be consistent with the hypothesis that for transitive subject-extracted RCs, VO would best be considered the “default”; but our results will turn out to be less clear regarding the possibility of a “default” order for object-extracted RCs. Regarding the relationship with theme-rheme or topic-focus structure, it has been informally suggested to us that the right edge of the RC may be associated with focus (Maria Polinsky,
p.c.), but this issue does not seem to have been written about extensively. In our reading-time studies, the key effects are located either before the right edge of the RC (Experiment 1a) or before the comprehender could know she has reached the right edge (Experiment 1b), rendering the potential role of any such right-edge/focus association unclear.

Finally, we should make a brief remark regarding the role of information structure for online processing-difficulty effects in studies such as ours. As will become clear in our experimental results, differing word orders do induce differing levels of processing difficulty; most notably, SRCs with OV order will turn out to be read more slowly than SRCs with VO order, or than ORCs with either order. It seems quite plausible that information structure plays a role in this result: OV order in SRCs is rare, and may, for example, be natural only when the embedded object NP is discourse-given. If this is the case, then OV order in SRCs would be all the more unexpected in the null contexts in which we present our experimental sentences. An information-structure-based explanation of the reading-time result would require a linking theory between the discourse context (in this case null), the word order encountered, and comprehension difficulty. Among the theories we have considered, memory-based and universal structural asymmetry theories do not obviously present prospects for such a linking. Expectation-based theories such as surprisal do: the discourse context is literally part of the probabilistic conditioning context, and unexpected word orders are surprising, directly giving rise to processing difficulty when evidence of the unexpected word order is encountered.

2.2. Predictions of different theories of syntactic complexity

We now describe the predictions of each type of theory for reading of these sentences. Perspective-shift and universal structural asymmetry theories predict a main effect of extraction type, with greater difficulty for ORCs than for SRCs. Memory-based theories predict an interaction between extraction type and word-order canonicity, or equivalently a main effect of locality, with greatest difficulty in the SRC scrambled and ORC default word order conditions. This differential difficulty effect should appear at the RC verb.

As a part of determining the predictions of expectation-based theories, we conducted corpus searches to tabulate frequencies of each of the four types of Russian relative clauses in the Russian Dependency Treebank, a collection of texts hand-annotated for dependency structure (Boguslavsky et al., 2000, 2002). Searches were carried out with the Tregex tool (Levy and Andrew, 2006; search patterns given in Appendix A). The results are shown in Table 1.

These results reveal several patterns relevant to our study. We do not find dramatic differences between general SRC and ORC frequencies; the strongest such difference is a 2.2:1 ratio among RCs with case-marked relative pronouns and full RC NPs. (For comparison, the ratio found for English by Roland et al., 2007, in the parsed Brown Corpus is 7.6:1.) Finer-grained inspection, however, revealed that the external distribution of RCs with case-syncretized relative pronouns differed across RC extraction type: whereas SRCs occurred in otherwise-typical contexts, e.g.:
Table 1: Results of Experiment 1 corpus search for frequencies of Russian subject- and object-extracted RCs in default & scrambled word order. See text for further discussion of fine-grained differences between distributions of case-marked and case-syncretized SRCs and ORCs.

<table>
<thead>
<tr>
<th>RC Word Order</th>
<th>kotoryj/kotorogo (Case-MARKED)</th>
<th>chto (Case-SYNCRETIZED)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All constituents</td>
<td>Full NPs only</td>
</tr>
<tr>
<td>SRC, VO</td>
<td>154</td>
<td>147</td>
</tr>
<tr>
<td>SRC, OV</td>
<td>9</td>
<td>4</td>
</tr>
<tr>
<td>ORC, VS</td>
<td>42</td>
<td>41</td>
</tr>
<tr>
<td>ORC, SV</td>
<td>74</td>
<td>29</td>
</tr>
</tbody>
</table>

(6) ... toj chistoj nezhnosti, chto beregla menja v detstve... 
... that pure tenderness, that protected me.ACC in childhood...

ORCs were, with only one exception, of the type use, chto “everything that”, edinstvennoe, chto “the only thing that”, poslednee, chto “the last thing that”—that is, the head noun was almost invariably inanimate and extremely semantically light. If comprehenders track fine-grained co-occurrences of this sort, then for contexts of the type seen in (4) they should interpret chto as a strong indicator that the upcoming RC will be subject-extracted; if, in contrast, they track coarser-grained statistics they may treat chto as a marker of an RC onset that is nevertheless relatively neutral to the RC’s extraction type (Mitchell et al., 1995). We will return to this issue in Experiment 1b.

Among SRCs with both relative pronoun types there is a strong preference for canonical, local word order (VO); among ORCs with both relative pronoun types, although both VS and SV orders inside the RC are seen, the preference is for the more canonical, less local ordering (SV). Similar to the findings of Reali and Christiansen (2007) for SRC/ORC relative frequency, however, we find that these statistics are further dependent on the type of RC NP. However, there is a general tendency for pronominal RC-internal subjects and objects to appear earlier in the RC; when the search is restricted to only full RC NPs, the VS order is preferred over the SV order.

At the level of the RC as a whole, then, there is a clear prediction made by word-order frequency theories for SRCs: local word order (SVO) will be easier than non-local word order (SOV). The predictions for ORCs are less clear, and depend on the granularity at which word order statistics are computed. If we aggregate across all RC NP types (OSV and OVS), the prediction is that default word-order ORCs will be easier than scrambled ORCs; if we consider only RCs with full NPs, however (the type we use in our materials), the prediction is that scrambled ORCs should be easier than default word-order ORCs. At the level of the entire RC, surprisal makes difficulty predictions similar to those of word-order theories; but surprisal and the ERH also make more fine-grained predictions about word-by-word processing difficulty that are worth elaborating further. For surprisal, the first place where the difference in processing difficulty between SRCs and ORCs (favoring SRCs) could show up is the relative pronoun when it is case-marked, since the case marking
indicates extraction type. Additionally, RCs are a syntactically constrained context, and more pre-verbal information regarding RC verb location and identity is available in non-local configurations (where the verb is at the end of the RC; SOV and OSV) than in local configurations (where the verb is RC-medial; SVO and OVS); and since surprisal does not assess costs for the representation storage, or retrieval memory of additional preverbal dependents, it predicts a processing advantage for non-local over local configurations at the RC verb. Finally, if comprehenders track fine-grained syntactic event frequencies then under surprisal they should find ORCs with case-syncretized relative pronouns more difficult than SRCs with case-syncretized relative pronouns. For the ERH, although one must be cautious in overstating the confidence of one’s predictions in the absence of an explicit probabilistic grammatical model, since Russian allows recursive postmodification of NPs one might reasonably expect the same processing advantage at the RC verb for local over non-local configurations as was argued for English RCs by Hale (2003). The ERH makes no obvious predictions regarding effects of RC extraction type or interactions with relative pronoun case marking.

2.3. Experiment 1a

For this experiment we constructed sentence frames on the basis of noun pairs \(\langle N_1, N_2 \rangle\) whose positioning in the sentence was interchangeable, and included both (A) variants where \(N_1\) was the head noun and \(N_2\) the RC NP, and (B) variants where \(N_2\) was the head noun and \(N_1\) the RC NP. Since it is arbitrary which noun in the pair is considered \(N_1\) and which \(N_2\), we collapse across this manipulation in all analysis of data; but including this manipulation ensures that possible differences in event plausibility are not confounded with RC extraction type. We also included a three-word prepositional phrase (PP) at the end of the RC in all conditions, so that each RC in this experiment consisted of a one-word NP, a one-word verb, and a three-word PP. The NP and verb always were the first two words in the RC, appearing in the order determined by experimental condition. The RC-final PP prevents the RC-final comma from falling on the RC NP or verb, and also gives some hope of determining whether any results arising at the main-clause verb reflect spillover or processing difficulty at the main verb itself. (The possible role of spillover in influencing RTs on the RC verb is addressed in Experiment 2.) A sample item in its eight conditions can be seen in (7) below (note that the English translation depends only on the extraction type and the A/B variant).

(7) a. [varA, SRC, DEFAULT]
Slesar, kotoryj udaril elektrika so vsego rasmaza, ushel
Repairman, who.NOM hit electrician.ACC with all strength, went
domoj s sinjakom pod glazom.
home with bruise under eye.
“The repairman, who hit the electrician with all his strength, went home with a bruise under his eye.”

b. [varA, SRC, SCRAMBLED]
Slesar, kotoryj elektrika udaril so vsego rasmaza, ushel
Repairman, who.NOM electrician.ACC hit with all strength, went
domoj s sinjakom pod glazom.
home with bruise under eye.

c. [varA, ORC, DEFAULT]
Slesar, kotorogo elektrik udaril so vsego rasmaza, ushel
Repairman, whom.ACC electrician.NOM hit with all strength, went
domoj s sinjakom pod glazom.
home with bruise under eye.
“The repairman, whom the electrician hit with all his strength, went home with
a bruise under his eye.”

d. [varA, ORC, SCRAMBLED]
Slesar, kotorogo udaril elektrik so vsego rasmaza, ushel
Repairman, whom.ACC hit electrician.NOM with all strength, went
domoj s sinjakom pod glazom.
home with bruise under eye.

“Elektrik, kotoryj udaril elektrik, so vsego rasmaza, ushel
Electrician, who.NOM hit repairman.ACC with all strength, went
domoj s sinjakom pod glazom.
home with bruise under eye.
“The electrician, who hit the repairman with all his strength, went home with
a bruise under his eye.”

e. [varB, SRC, DEFAULT]
Elektrik, kotoryj udaril slesara so vsego rasmaza, ushel
Electrician, who.NOM repairman.ACC hit with all strength, went
domoj s sinjakom pod glazom.
home with bruise under eye.

“Elektrik, kotoryj udaril slesara so vsego rasmaza, ushel
Electrician, who.NOM repairman.ACC hit with all strength, went
domoj s sinjakom pod glazom.
home with bruise under eye.
“The electrician, whom the repairman hit with all his strength, went home with
a bruise under his eye.”

f. [varB, SRC, SCRAMBLED]
Elektrik, kotoryj udaril slesara so vsego rasmaza, ushel
Electrician, who.NOM repairman.ACC hit with all strength, went
domoj s sinjakom pod glazom.
home with bruise under eye.

“Elektrik, kotoryj udaril slesara so vsego rasmaza, ushel
Electrician, who.NOM repairman.ACC hit with all strength, went
domoj s sinjakom pod glazom.
home with bruise under eye.
“The electrician, whom the repairman hit with all his strength, went home with
a bruise under his eye.”

g. [varB, ORC, DEFAULT]
Elektrik, kotorogo slesar udaril so vsego rasmaza, ushel
Electrician, whom.ACC repairman.NOM hit with all strength, went
domoj s sinjakom pod glazom.
home with bruise under eye.

“Elektrik, kotorogo slesar udaril so vsego rasmaza, ushel
Electrician, whom.ACC repairman.NOM hit with all strength, went
domoj s sinjakom pod glazom.
home with bruise under eye.
“The electrician, whom the repairman hit with all his strength, went home with
a bruise under his eye.”

h. [varB, ORC, SCRAMBLED]
Elektrik, kotorogo udaril slesar so vsego rasmaza, ushel
Electrician, whom.ACC hit repairman.NOM with all strength, went
domoj s sinjakom pod glazom.
home with bruise under eye.
2.3.1. Participants

Sixteen native Russian speakers living in or visiting the United States participated in this experiment at the University of California at San Diego for cash compensation. None had arrived in the United States before age 13, and all reported that they continue to use Russian on a regular basis and consider it the language they are most comfortable with.

2.3.2. Materials

Thirty-two items (listed in full in Appendix C) were constructed following the pattern of [7]. Each participant saw only one of the eight conditions of each item according to a Latin square design. These experimental stimuli were interleaved with 20 items from an unrelated experiment and 52 random fillers such that no two experimental sentences were seen consecutively.

2.3.3. Procedure

Sentences were presented to participants in a non-cumulative word-by-word moving-window self-paced procedure on a PC laptop computer running the Linger software (Rohde, 2005). Each trial began with a series of dashes displayed on the computer screen in place of the words in the sentence. The first press of the space bar revealed the first word in the sentence, and each subsequent press of the space bar revealed the next word in the sentence and masked the previous word. Punctuation was displayed together with the word preceding it. The times between button presses were recorded to the nearest millisecond. Each sentence was followed by a yes-or-no comprehension question probing the participant’s understanding of the content of the sentence. Written instructions in Russian were given at the outset of the experiment.

2.3.4. Results

Statistical analysis procedures. We used “mixed-effects”, sometimes called “multi-level” or “hierarchical”, models for all analyses. For reading-time data we used linear mixed-effects (LME; Baayen et al., 2008) models, and for question-answering data we used logistic mixed-effects models (Jaeger, 2008). All our predictive factors were dichotomous, and we centered them by coding one level of the factor as -0.5 and the other as 0.5, rendering lower-order effects interpretable as in standard ANOVAs even when higher-order effects were included. Our fixed-effects model structure always reflected the factorial structure of our experiment; and we always used “maximal” random-effects structure for our theoretically critical variables—that is, by-participant and by-item random effects with the same specification as for our fixed effects. Using maximal random-effects structure means that our analyses make the same assumptions about participant- and item-specific sensitivities to experimental conditions as in traditional ANOVAs, and ensures that the analyses are not anti-conservative with respect to the question of whether our data suggest that the effects of theoretical interest would generalize to new participants and items (Barr et al., in revision). We present p values computed by treating the t statistic resulting from LME analysis as approximately normally distributed (justified for datasets of our size; Baayen et al., 2008). Analyses were carried out using R’s lme4 package (Bates and Sarkar, 2008).
Table 2: QA accuracy for Experiment 1a

<table>
<thead>
<tr>
<th></th>
<th>Default</th>
<th>Scrambled</th>
</tr>
</thead>
<tbody>
<tr>
<td>SRC</td>
<td>0.91 (0.19)</td>
<td>0.82 (0.17)</td>
</tr>
<tr>
<td>ORC</td>
<td>0.87 (0.20)</td>
<td>0.85 (0.15)</td>
</tr>
</tbody>
</table>

Raw reading times were analyzed as follows, unless otherwise specified. Recordings in any region above 5000ms or below 100ms were discarded, means and standard deviations were then computed for each region in each condition, and any measurement more than four standard deviations above the mean was discarded. These procedures resulted in loss of 0.80% of data. The remaining measurements were then subjected to mixed-effects analyses; data from both correctly-answered and incorrectly-answered trials were included. Error bars in graphs represent standard errors of by-subject means. In-text descriptions of reading-time results are limited to regions of theoretical interest.

Comprehension accuracy. Table 2 shows comprehension accuracy as a function of experimental condition; no main effects or interactions are statistically significant. We see a numerical but insignificant interaction with highest accuracy in the SRC default-order and ORC scrambled conditions. This pattern could equivalently be viewed as a main effect of locality, with higher accuracy in the local conditions.

Reading times. We treated each of the first nine words of the sentence as its own region; an example is given in (8) below:

(8) Slesar, kotoryj udaril elektrika so vsego rasmaxa, ushel domoj
    Subj, RelPro RCVerb RCNP RCfinal1 RCfinal2 RCfinal3 MatrixVerb Spillover
    ... 
    ... 

Figure 2 shows average reading times for each of these nine regions of analysis. The interactions between RC type and scrambling are significant at the RC Verb ($p<0.001$), the RC NP ($p<0.05$), the first two words of the RCfinal region (both $p<0.01$), and the matrix verb ($p<0.025$). Pairwise comparisons indicate that nominative RC NPs in ORCs are read faster RC-initially than postverbally ($p<0.025$) but that accusative RC NPs in SRCs are not ($p=0.296$). At the RC verb we see a large interaction between extraction type and word order, with faster RTs in the SRC default and ORC scrambled conditions (the two local conditions). We also see this same interaction throughout the RC-final PP and onto the main-clause verb, with the exception of the final word in the RC, which shows a related numeric pattern (non-scrambled SRC RTs lowest) but without significant differences across condition.

We also conducted an aggregated analysis of mean per-region RTs starting at the relative pronoun and ending at the main verb. This analysis found no main effect but a highly significant interaction ($p<0.001$).
Discussion. The key patterns observed in this study are (a) non-local RC configurations (ones in which the RC NP occurs preverbally) are consistently disfavored in both structural frequencies and processing difficulty; and (b) in the most difficult conditions—case-syncretized ORCs and case-marked RCs with non-local word orders—maximal processing difficulty is never observed before the RC verb. Our results lend little support to theories of universal structural asymmetry or perspective shift: we find no clear overall RT penalty for ORCs as compared with SRCs (see also (Hsiao and Gibson, 2003, Gibson and Wu, in press) for results inconsistent with such theories in Chinese). Nor do these results give clear support to expectation-based theories such as surprisal, but they do give support to theories including the memory-based DLT, decay-based theories, and SBI, which predict retrieval at the RC verb to be the key factor in determining processing difficulty, with greater difficulty in the non-local conditions. As described in Section 1.1, the ERH is also likely to predict this locus of processing difficulty. However, the ERH does not predict why processing difficulty would be greater at the nonlocal ORC verb in the case-syncretized condition than in the case-marked condition, since in both cases the RC NP’s case marking should already have disambiguated the grammatical function of the head noun. In memory-retrieval theories this result might be accounted for by positing that case marking on both the relative pronoun and the RC NP facilitates retrieval above and beyond effects from case marking on the RC NP alone.

Although the word-by-word predictions of surprisal were not met—it was the local verbs, not the non-local verbs, whose reading times were shortest—at a coarser granularity, in reading times across the RC and the main verb, the predictions of surprisal as determined by the frequencies of the different RC types, limited to full-NP RCs, were met: VO SRCs were read fastest overall, followed by VS ORCs, then SV ORCs, and finally OV SRCs. Additionally, RTs at the RC NP are consistent with the predictions of surprisal: a penalty
is paid at RC-initial accusative NPs, which are highly unlikely among SRCs, but not at
RC-initial nominative NPs, which are not so unlikely for ORCs (Table II).

2.4. Experiment 1b (conducted in Russia)

This experiment follows a design similar to that of Experiment 1a, but with several
differences. The most theoretically crucial difference is that Experiment 1b includes case-
syncretized relative pronouns as well as case-marked relative pronouns, which allows us to
probe the degree to which comprehenders track fine-grained structural statistics in their
input, since the corpus study described earlier revealed substantial differences in how case-
syncretized and case-marked relativepronouns are used. The remaining differences are limitations
in the design of the present experiment, which was designed and conducted before Experiment
1a, but whose results nevertheless corroborate and extend those of Experiment 1a. The first
of these is that, unlike in Experiment 1a, the thematic roles were not counter-balanced
across RC extraction type. This may play a role in some of the differences in the details of
the results that we find between Experiments 1a and 1b. Second, we violated prescriptive
Russian orthography in omitting delimiting commas at the left and right edges of the relative
clauses. Fortunately, there is no evidence in our results that suggests our participants did
not rapidly adapt to this and process our sentences in overall similar ways as in Experiment
1a. Third, we did not include RC-final prepositional phrases in this experiment.

2.4.1. Participants

Forty native Russian speakers participated in Volgograd, Russia for cash compensation.

2.4.2. Materials

Thirty-two items (listed in full in Appendix D) were constructed following the pattern in
(4), with both case-marked and case-syncretized relative pronouns as described previously.
Each item was eight words long in each condition: a one-word sentence-initial subject noun,
followed by a relative pronoun, a one-word past-tense RC verb and one-word RC NP (with
relative order of these two elements determined by the word order condition), a one-word
transitive past-tense main verb, a one-word main-clause direct object, and finally a two-word
adjunct prepositional phrase. As noted earlier, the only punctuation used was a sentence-
final period. The eight conditions of an item would thus look as in (9) below (cf. (4); note
that the English translation depends only on the extraction type):

(9) a. [SRC, DEFAULT, MARKED]

Slesar kotoryj udaril elektrika poterjal terpenie v spore.
    Repairman who.NOM hit electrician.ACC lost patience in argument.

---

5It should be mentioned that, with the comma absent, there is an alternative local interpretation
in Russian of bigrams of the form [Noun.masculine+kotorogo] as meaning “whose Noun”; although this
interpretation is not consistent with sentence-initial context, locally coherent syntactic interpretations of
word sequences have been shown to affect online comprehension [Tabor et al., 2004; Konieczny, 2005; Gibson,
2006]. As will be seen, however, there is no evidence that comprehenders entertained this possibility.

Levy, Fedorenko, Gibson—Draft of 12 March 2012
“The repairman who hit the electrician lost his patience in the argument.”

b. [SRC, SCRAMBLED, MARKED]
Slesar kotoryj elektrika udaril poterjal terpenie v spore.
Repairman who.NOM electrician.ACC hit lost patience in argument.

c. [ORC, DEFAULT, MARKED]
Slesar kotorogo elektrik udaril poterjal terpenie v spore.
Repairman whom.ACC electrician.NOM hit lost patience in argument.

“The repairman whom the electrician hit lost his patience in the argument.”

d. [ORC, SCRAMBLED, MARKED]
Slesar kotorogo udaril elektrik poterjal terpenie v spore.
Repairman whom.ACC electrician.NOM hit lost patience in argument.

e. [SRC, DEFAULT, SYNCRETIZED]
Slesar chto udaril elektrika poterjal terpenie v spore.
Repairman that hit electrician.ACC lost patience in argument.

f. [SRC, SCRAMBLED, SYNCRETIZED]
Slesar chto elektrika udaril poterjal terpenie v spore.
Repairman who hit electrician.ACC hit lost patience in argument.

g. [ORC, DEFAULT, SYNCRETIZED]
Slesar chto elektrik udaril poterjal terpenie v spore.
Repairman who hit electrician.NOM hit lost patience in argument.

h. [ORC, SCRAMBLED, SYNCRETIZED]
Slesar chto udaril elektrik poterjal terpenie v spore.
Repairman who hit electrician.NOM hit lost patience in argument.

Each participant saw only one of the eight conditions of each item according to a Latin square design. These experimental stimuli were interleaved with 68 fillers such that no two experimental sentences were seen consecutively.

2.4.3. Procedure
The procedure was identical to that of Experiment 1a. The study typically took 35—45 minutes to complete.

2.4.4. Results
Statistical analysis procedures. The same analysis procedures were used as in Experiment 1a.

Comprehension accuracy. Question-answering accuracies for each condition are shown in Table 3. Mixed logit model analysis revealed a significant effect of RC type ($p<0.001$) and a three-way significant interaction ($p<0.001$); the primary dynamic here is that sentences containing ORCs with case-syncretized relative pronouns were by far least accurately understood.
Table 3: QA accuracy for Experiment 1b

<table>
<thead>
<tr>
<th></th>
<th>nscri</th>
<th>scri</th>
</tr>
</thead>
<tbody>
<tr>
<td>marked subj</td>
<td>0.86</td>
<td>0.86</td>
</tr>
<tr>
<td>marked obj</td>
<td>0.80</td>
<td>0.76</td>
</tr>
<tr>
<td>syncretized subj</td>
<td>0.88</td>
<td>0.84</td>
</tr>
<tr>
<td>syncretized obj</td>
<td>0.68</td>
<td>0.63</td>
</tr>
</tbody>
</table>

However, a significant main effect of extraction type was found even among the the case-marked conditions alone [todo: redo as mixed analyses.] (F1(1,39) = 9.62, p = 0.004; F2(1,31) = 10.54, p = 0.003). Word order had no reliable effects on question-answering accuracy.

Reading times. As it turned out, reading-time patterns were considerably different for the marked-pronoun and unmarked-pronoun conditions. Hence we plot these conditions separately. We treated each of the first six words of the sentence as its own region; an example is given in (10) below:

(10) Slesar kotoryj udaril ekelktrika poterjal terpenie v spore.
    Subj RelPro RCVerb RCNP MatrixVerb Spillover
    Repairman who hit electrician lost patience in argument.

Figures 4 and 5 show average reading times for each of these six regions of analysis in the marked- and unmarked-pronoun regions respectively (note that the y-axes have substantially different scales in these two figures, since the RT increases in ORCs with case-syncretized relative pronouns were so dramatic). We start with an aggregate 2×2×2 LME analysis. At the RC NP, RC verb, and matrix verb regions these analyses indicated three-way interactions (p<0.001, p<0.025, and p<0.01 respectively), clearly driven by behavior in the two case-syncretized ORC conditions. In these two conditions (see Figure 5), case marking on the RC NP is the first unambiguous indication of the correct grammatical function assignment to the head noun. In the syncretized ORC scrambled condition, where the RC NP appears postverbally, we see highly inflated reading times at the RC NP; these inflated reading times persist into the next region (the matrix verb) and to some extent to the spillover region as well. In the syncretized ORC default-order condition, where the RC NP appears preverbally, we see highly inflated reading times at the RC verb, which persist (though not as much in the ORC scrambled condition, leading to the three-way interaction) onto the next region, the matrix verb. Both these effects can be interpreted as an effective garden-pathing of the comprehender toward an SRC interpretation when the relative pronoun is encountered, and are consistent with the predictions laid out in Section 1.1 of expectation-based theories in which fine-grained syntactic event co-occurrences are tracked: when modifying animate head nouns chto is taken as a strong signal that the upcoming RC is subject-extracted. The highly inflated reading times in both these conditions can thus be interpreted as an expectation-based disambiguation penalty, though the processing penalty not being reflected in RTs until
Figure 3: Reading times in Experiment 1b, unmarked (chto) relative pronoun conditions only. The RCNP region appears twice because it is preverbal in the non-local conditions and postverbal in the local conditions.

Because the case-syncretized ORC conditions pattern so differently from the remaining conditions, we also analyzed the four case-marked conditions on their own. (Inside the RC the case-syncretized SRC conditions patterned indistinguishably from the case-marked SRC conditions.) Figure 4 presents reading times in these conditions. At the RC NP we see a significant main effect of word order, with faster times in the default conditions. Pairwise comparisons indicate that accusative RC NPs are read more slowly RC-initially than postverbally, but that nominative RC NPs are read more quickly RC-initially than postverbally (both \( p < 0.05 \)). At the main verb, we see marginal main effects of extraction type (SRCs faster than ORCs, \( p = 0.057 \)) and scrambling (scrambled order faster than default order, \( p = 0.058 \)). However, these main effects were driven by the overwhelmingly strongest result, a large and highly significant interaction between extraction type and word order such that the local word orders (those in which the verb immediately follows the relative pronoun) are read faster than the non-local word orders. This interaction is seen both on the RC verb and on the main verb (both \( p \ll 0.001 \)).

For the case-marked conditions we also conducted an aggregated analysis of the entire RC region (beginning at the relative pronoun) plus the main verb in which the mean RT across these regions served as the response variable of analysis (analysis procedures otherwise followed those described earlier in this section). This analysis showed no main effects of extraction type or word order, but a highly significant interaction between the two (\( p \ll 0.001 \)).
2.4.5. Discussion

This experiment replicated the key result of Experiment 1a—an interaction between extraction type and word order at the RC verb, with faster reading times in the local conditions than in the non-local conditions. The same pattern was seen at the main verb in Experiment 1a, and is seen here throughout the remainder of the RC and onto the main verb. This experiment also gives additional evidence for comprehenders’ sensitivity to fine-grained syntactic event frequency and processing difficulty within the RC: chto is interpreted as a clear sign that the RC is subject-extracted. This result gives some degree of support for expectation-based theories including word-order theories and surprisal—in particular regarding the magnitude of difficulty observed within the RC—though where processing difficulty is first observed still seems better predicted by other theories, as we concluded in discussion of Experiment 1a.

Although the results of Experiments 1a and 1b are largely consistent with one another, the pattern seen on the RC NP in Experiments 1a and 1b differs and requires careful interpretation. In the case-marked conditions of Experiment 1b, we saw a main effect of scrambling, with higher RTs when the RC NP was in its non-default position (OV for SRCs and VS for ORCs). In Experiment 1a, RTs on the unscrambled object (SRC condition) are just as high as those on the scrambled subject (ORC condition), leading to an interaction rather than a main effect of scrambling. It is not clear which of the two patterns we should consider to more faithfully reflect the contribution of syntactic complexity per se to processing difficulty, but it is possible that this discrepancy may have arisen from plausibility effects in Experiment 1b; if so, the pattern seen in Experiment 1a would be the more faithful
reflection and would thus require interpretation. The challenge in interpreting this pattern can be cast in clearest light by posing the question: what could cause the nominative RC NP for ORCs to be read more slowly in scrambled, postverbal position than in unscrambled, preverbal position (see Figure 2)? We offer two possible explanations. First, the DLT might be taken to predict this pattern on the grounds that only postverbal RC NPs can immediately be integrated with their governing verb. As long as the integration cost for an adjacent dependent is greater than zero, the postverbal RC NP, which is integrated with the immediately preceding verb, will have a higher integration cost than the preverbal RC NP, where no integration occurs at all. However, this alone cannot explain the observed interaction between word order and grammatical function, since the discrepancy between RC NP reading times is seen only in ORCs. This interaction motivates the second possible explanation we propose, namely a role for syntactic expectations: as seen in Table 1, it is quite unlikely that an SRC will begin with an object NP, but not so unlikely that an ORC will begin with a subject NP. Thus the higher reading times for preverbal RC NPs in the SRC than in the ORC condition are predicted by surprisal. An additive combination of surprisal and DLT integration cost (such as that used by Demberg and Keller, 2008, 2009) might thus predict a pattern similar to that seen in Experiment 1a. An alternative possibility is that both the higher RC-verb and higher RC-NP RTs in the non-local conditions might reflect spillover from processing difficulty on the first word of the RC; this alternative is also consistent with the inflated RTs during the RC-final PP and at the main verb in the non-local conditions, whose aggregate difficulty is predicted to be greatest both under memory-based theories and under expectation-based theories using fine-grained syntactic expectations.

2.5. General discussion for Experiment 1

The key result obtaining consistently in both Experiments 1a and 1b is that processing difficulty as measured by reading times at RC verbs is greatest in non-local conditions, where the RC NP intervenes between the relative pronoun and the RC verb (i.e., SOV and OSV word orders in the RC), and least in the local conditions, where the RC verb immediately follows the relative pronoun (i.e., SVO and OVS word orders). This result is directly predicted by memory-based theories and possibly by the ERH. Universal structural asymmetry and perspective-shift theories make no predictions regarding this result. Finally, the result is problematic for word-order frequency theories and for surprisal, although overall whole-RC processing-difficulty pattern matches both these expectation-based theories, if the distinction between full and pronominal NPs is taken into account in determining the relevant structural frequencies. One result obtained in Experiment 1a is predicted only by surprisal—the processing difficulty obtaining at RC-initial accusative NPs. Though a different pattern was seen in Experiment 1b, this may reflect a design limitation of that experiment.

An important limiting factor in interpreting the RT results at the RC verb in Experiment 1, however, is the possibility that spillover may be affecting RTs observed at the RC NP and RC verb; it could well be the case, for example, that the inflated RTs observed on the RC verb in the non-local conditions reflect spillover from processing difficulty initiated at the RC-initial NP. Furthermore, the design of Experiment 1 does not permit us to discriminate
clearly between entropy reduction and memory-based theories, or among different memory-based theories. Experiment 2 goes some way toward addressing these issues.

3. Experiment 2

Although Experiment 1 demonstrated that self-paced reading can be used to find large and reliable differences in syntactic comprehension difficulty in Russian relative clauses of different extraction types and word orders, it did not give us full confidence in determining the precise origin sites of processing difficulty, or in distinguishing sharply between expectation-based and memory-based theories. In Experiment 2 we thus endeavor to achieve these latter goals, testing more precisely the word-by-word predictions of different theories and minimizing the possibility that spillover may obscure the origin sites of processing difficulty. We do so by parametrically varying the number of preverbal dependents in a syntactically constrained context to yield contrasts as clear as possible between the predictions of expectation-based and memory-based theories (Figure [1]). We focus our attention on subject-extracted relative clauses and consider cases where zero, one, or two constituents intervene between the relative pronoun and the RC verb. We further allow these constituents to be either arguments or adjuncts of the RC verb; we use ditransitive RC verbs so that up to two arguments are available to intervene. To maximize the possibility of distinguishing the processing difficulty associated with a given clausal constituent from spillover processing difficulty due to the onset of the preceding constituent, we make each of the constituents at least two words long. For the intervening constituents we do this by using postmodifiers; for the RC verb itself we achieve this by using a verb complex consisting of a finite verb and an infinitival verb-form (e.g., “forgot to bring”). Finally, we note that some authors have suggested that reading times may tend to decrease as the position of a word within the sentence increases [Ferreira and Henderson 1993]; although we did not see such an effect in Experiment 1, any such effect would confound a result here favoring expectation-based theories. To prevent such a confound we use two-clause sentences with the RC in the second clause and, in the adjunct manipulation, place any adjuncts that are not within the RC in the first clause of the sentence, so that the linear position of the critical RC verb complex is identical across adjunct-manipulation conditions. The maximally local variant of one of our items is given in (11) below; underscores indicate words presented together in a single region in self-paced reading:

(11) a. Shef-povara xvalili za ego_masterstvo rannim_vecherom okolo_shesti, no
chef.ACC praised for his_mastery early_evening around_six, but
ofitsiant, kotoryj zabyl prinesti bljudo iz teljatiny posetitelju
waiter.NOM who.NOM forgot to bring dish of veal customer.DAT
v_chernom_kostjume vovremja, ne poluchil chaevyx posle uzhina.
in_black_suit on_time, not received tip after dinner.
“The chef was praised for his mastery early in the evening around six o’clock, but the waiter, who forgot to bring the dish of veal to the customer in the black suit on time, didn’t receive a tip after dinner.”
Here, the RC verb complex is *zabyl prinesiti* (“forgot to bring”); its direct object (DO) is *bljudo iz teljatiny* (“dish of veal”) and its indirect object (IO) is *posetitelju v chernom kostume* (“customer in black suit”). In the argument manipulation, either the accusative or both accusative and dative arguments are fronted before RC verb, leading to the three possibilities for RC-internal word order given in (12) below:

(12) a. [zero interveners; RC-internal word order V DO IO, identical to (11)]

\[ \ldots ofitsiant, kotorijj zabyl prinesiti BLJUDO IZ TELJATINY \]
\[ \ldots waiter.NOM who.NOM forgot to bring DISH OF VEAL posetitelju v_chernom_kostjume vovremja\ldots \]
\[ \text{customer.DAT in_black_suit on_time}\ldots \]

b. [one argument intervener; word order DO V IO]

\[ \ldots ofitsiant, kotorijj BLJUDO IZ TELJATINY zabyl prinesiti \]
\[ \ldots waiter.NOM who.NOM DISH OF VEAL forgot to bring posetitelju v_chernom_kostjume vovremja\ldots \]
\[ \text{customer.DAT in_black_suit on_time}\ldots \]

c. [two argument interveners; word order DO IO V]

\[ \ldots ofitsiant, kotorijj BLJUDO IZ TELJATINY posetitelju \]
\[ \ldots waiter.NOM who.NOM DISH OF VEAL customer.DAT v_chernom_kostjume zabyl prinesiti vovremja\ldots \]
\[ \text{in_black_suit forgot to bring on_time}\ldots \]

In the argument manipulation, the rest of the sentence is left the same, and the meaning of the sentence (at least in terms of predicate-argument structure) is the same in all three variants.

In the first clause of the sentence, two temporal phrases appear as well—*rannim vecherom* (“early in the evening”) and *okolo shesti* (“around six o’clock”). In the adjunct manipulation, one or both of these temporal phrases are shifted into the RC, between the relative pronoun and the RC verb complex, leading to the following three possibilities for RC-internal word order:

(13) a. [zero interveners; identical to (11)]

\[ \ldots ofitsiant, kotorijj zabyl prinesiti bljudo iz teljatiny posetitelju \]
\[ \ldots waiter.NOM who.NOM forgot to bring dish of veal customer.DAT v_chernom_kostjume vovremja, \ldots \]
\[ \text{in_black_suit on_time}\ldots \]

b. [one temporal-NP intervener]

\[ \ldots ofitsiant, kotorijj OKOLO_SHESTI zabyl prinesiti bljudo iz \]
\[ \ldots waiter.NOM who.NOM AROUND_SIX forgot to bring dish of teljatiny posetitelju v_chernom_kostjume vovremja, \ldots \]
\[ \text{veal customer.DAT in_black_suit on_time, \ldots} \]
\[ \ldots the waiter, who around six o’clock forgot to bring the dish of veal to the}
customer in the black suit on time, . . . ”

c. [two temporal-NP interveners]
   . . . ofitsiant, kotoryj rannym_vecherom OKOLO_SHESTI zabyl
   . . . waiter.NOM who.NOM early_evening AROUND_SIX forgot
   prinesti bljudo iz teljatiny posetitelju v_chernom_kostjume vovremja,
   to_bring dish of veal customer.DAT in_black_suit on_time,
   . . .
   . . .
   . . .
   “. . . the waiter, who early in the evening around six o’clock forgot to bring the
dish of veal to the customer in the black suit on time, . . . ”

Each temporal phrase that is shifted into the RC is removed from the initial clause; thus
the same set of words appears before the RC verb complex in all versions of the adjunct
manipulation. Note that [12a] and [13a] are the same sentence, hence there are five conditions
in this experiment: two $1 \times 3$ manipulations, with the maximally local variant shared across
the two.

We now go over the predictions of each class of theory for this experimental design.
Perspective shift and universal structural theories of the SRC/ORC asymmetry make no
predictions regarding difficulty, since all our conditions are SRCs. Memory-based theories
make the simple prediction that greater numbers of interveners should lead to greater
processing difficulty at the RC verb. There may be a further prediction that more difficulty
per intervener may ensue for argument interveners than for adjunct interveners, either
because of greater interference due to the greater similarity of the arguments than the
adjuncts to the NP head (e.g., only the NP head and the dative argument are animate)
or to the fact that each argument intervener contains two NPs.

Intuitively, expectation-based theories predict that the presence of additional preverbal
dependents intervening between the relative pronoun and the RC verb should generally
facilitate comprehension of verb, because these dependents will help sharpen the comprehender’s
expectations regarding where the verb will appear and which verb will be encountered,
on principles similar to those described in Section 1.1 for English RCs. The sharpening
of expectations regarding verb identity should be especially pronounced for the argument
manipulation: simply put with respect to the example in [12], there are far fewer things that
a waiter can do to a dish of veal with a customer in a black suit fulfilling the benefactive role
than a waiter can do in general, and to “forget to bring” the dish is one of those things.
As for the sharpening of expectations regarding verb location, it is plausible that once
an accusative argument is seen the comprehender knows that the next constituent is less
likely to be an accusative NP, hence expectations for other constituent types, including the
RC verb, should strengthen; the same should happen for both dative NPs and temporal
adjuncts (though the effect for temporal modifiers might be expected to be less pronounced,
since multiple temporal phrases can sometimes be found in a single clause). To determine
whether this reasoning is corroborated in the empirical distribution of Russian word order
Table 4: Syntactic conditional probabilities of verb given preceding RC-internal context for argument and adjunct interveners in Experiment 2.

<table>
<thead>
<tr>
<th>Event &amp; Conditioning Structure</th>
<th>Support</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>p(V</td>
<td>kotoryj)</td>
<td>1574</td>
</tr>
<tr>
<td>p(V</td>
<td>kotoryj, NP)</td>
<td>51</td>
</tr>
<tr>
<td>p(V</td>
<td>kotoryj, Adjunct)</td>
<td>325</td>
</tr>
<tr>
<td>p(V</td>
<td>kotoryj, Adjunct, Adjunct)</td>
<td>27</td>
</tr>
</tbody>
</table>

frequencies, we conducted a corpus study using the Russian Dependency Treebank similar to those conducted for Experiment 1. Tree-search patterns are given in Appendix A and results of these searches are given in Table 4. Consistent with our original reasoning, adding either an intervening NP or adjunct immediately after the relative pronoun increases the conditional probability that the next clausal constituent encountered will be the RC verb. The data were insufficient for us to estimate the effect of adding a second intervening constituent except in the case of the adjunct-intervener search in the Dependency Treebank, for which the second intervening constituent raises the conditional probability of seeing a verb next even further. Thus we conclude that expectations regarding verb location are indeed likely to sharpen as the number of interveners increases. Ideally we would also estimate the effects of our intervening arguments and adjuncts on expectations regarding verb identity, but corpus data are currently far too sparse to give hope of obtaining reliable estimates.

3.1. Experiment 2a (conducted in Russia)

3.1.1. Participants

Forty native Russian speakers participated in Kazan, Moscow, and St. Petersburg, Russia, for cash compensation.

3.1.2. Materials

Twenty items (listed in full in Appendix E) were constructed following the pattern seen in (12)–(13). For each item, the RC of interest consisted of an unambiguously nominative relative pronoun, an inanimate unambiguously accusative-marked direct object NP, an animate unambiguously dative-marked indirect object NP, a finite verb followed by a non-finite verb, an RC-final phrase, and (in the One and Two adjunct-intervener conditions) one or two temporal adjuncts. The accusative and dative NPs each had a postmodifier. Each sentence consisted of a coordination of two conjunct sentences with the end of the first sentence conjunct delimited by a comma; the RC of interest always modified the initial subject of the second sentence. The main verb region of the second sentence conjunct always immediately followed the end of the first RC. Any temporal adjunct not appearing inside the RC appeared inside the first sentence conjunct, so that the set of words appearing before the critical RC verb complex was identical across the three adjunct conditions. Each participant saw only one of the five conditions of each item according to a Latin square design. These experimental stimuli were interleaved with 60 fillers such that no two experimental sentences were seen consecutively.
3.1.3. Procedure

Sentences were presented to participants in a non-cumulative region-by-region moving-window self-paced procedure on a Dell laptop PC running DMDX software [Forster and Forster, 2003]. Each trial began with a series of dashes displayed on the computer screen in place of the words in the sentence. Due to the length of these sentences, it was impossible to present them on a single line of the screen. Therefore we broke text across lines such that the critical RC verb complex was always preceded by at least one region of presentation and followed by at least one region of presentation on the same line. Participants controlled sentence presentation with a Logitech USB gamepad; the first press of a button on the gamepad revealed the first region in the sentence, and each subsequent press of the gamepad revealed the next region in the sentence and masked the previous region. The adjunct interveners, the critical RC finite and non-finite verbs, and the main verb each always appeared as individual regions; the accusative and dative NPs each appeared as either two or three regions depending on the item. Times between button presses were recorded to the nearest millisecond. Each sentence was followed by a yes-or-no comprehension question probing the participant’s understanding of the content of the sentence. The study typically took 30–40 minutes to complete.

3.1.4. Results

Statistical analysis procedures. Due to a programming error, the finite and non-finite verb in the RC were presented as a single region in Item 8. We thus excluded this item from all data analysis. Question-answering accuracies and region-by-region reading times were each analyzed in two sets of LME analyses corresponding to the argument and adjunct manipulations respectively. These analyses involved fitting one model with and one model without fixed effects of the manipulation, and using the likelihood-ratio test (with two degrees of freedom, since number of interveners is a three-level factor in each case) to assess whether the fixed effect significantly improves model fit. Both models with and without the fixed effect included condition-specific random effects (i.e., “maximal” random-effects structure) for both subjects and items. These analyses can be thought of as crossed random-effects analogues of traditional $1 \times 3$ “by-subjects” and “by-items” analyses, but which yield a single p-value rather than two separate p-values; and indeed traditional $1 \times 3$ ANOVA analyses (not reported here) yielded qualitatively similar results in all cases. Procedures within each of these analyses were otherwise the same as in Experiment 1a unless otherwise specified. As can be seen in Appendix E, there was some variability in the number of regions of presentation of the AccMod and DatMod regions of analysis, which were usually 2 and 1 regions of presentation respectively. No qualitative changes in results obtain when items with different numbers of regions of presentation are excluded from analyses.

Comprehension accuracy. Comprehenders answered questions with accuracy of 72% ($\pm 3\%$) in the zero-intervener condition, 74% ($\pm 3\%$) in the one-argument-intervener condition, 69% ($\pm 4\%$) in the two-argument-intervener condition, 66% ($\pm 4\%$) in the one-temporal-intervener condition, and 68% ($\pm 4\%$) in the two-temporal-intervener condition. Neither argument nor adjunct manipulation had a statistically significant effect in this measure, though there
is a hint of question-answering accuracy being lower in the non-local (1 or 2 interveners) conditions than in the most local condition.

*Reading times.* For purposes of reading-time analysis we broke the RC-internal region of the sentence down as illustrated below for the two-intervener adjunct condition (13c):

```
kotoryj rannym_vecherom okolo_shesti zabyl prinesti bljudo iz teljatiny
RelPro Tmp1 Tmp2 Vfin Vinf Acc AccMod Dat
who early_evening around_six forgot to bring dish of veal
posetitelju v_chernom_kostjume vovremja, ...  
DatMod RCFinal MatrixVerb
to_customer in_black_suit on_time, ...
```

Figures 5 and 6 show region-by-region reading times for the argument and adjunct manipulations respectively. LME analyses of the argument condition recovered significant effects of number of interveners only at the ACC_MOD, finite-verb, non-finite verb, and RC-final region (all $p<0.05$). Analyses of the adjunct condition recovered significant effects ($p<0.05$) at all regions except for the RC-final and main-verb regions. At both the finite and non-finite verbs in the RC verb complex, reading times increase monotonically with the number of verbal dependents intervening between the relative pronoun and the RC verb, regardless of whether these dependents are arguments or adjuncts. The magnitude of the increase in verb RTs was similar for both arguments and adjuncts. This overall pattern of verb-complex RT increasing with number of interveners suggests a simple summarization in which average RT is linear in the number of interveners between the relative pronoun and the RC verb complex. To test this summarization, we fit LME models in which mean predicted RT is a linear function of the number of interveners irrespective of adjunct/argument intervener status, with random condition-specific by-participant and by-item effects, for each of the finite-verb and non-finite-verb regions. We then compared these models against other models with the same random-effects structure but different fixed-effects structures, using likelihood-ratio tests. At both the finite and non-finite verbs, this model was significantly better than a baseline model of no condition-specific fixed effects (both $p<0.01$), but not significantly worse than models in which number of interveners is treated as a categorical predictor and/or interacts with intervener type (all $p>0.4$). The slopes for the linear-in-number-of-interveners models are 50.13ms for the finite verb and 57.16ms at the non-finite verb. Since each of these models uses only two parameters (an intercept and a slope) and is not a significantly worse fit to the data than more complex models (up to 5-parameter) models, parsimony suggests that the linear-in-number-of-intervener models are a reasonable summary of our data. Thus we can say that each additional preceding interveners increases the amount of time spent reading the RC verb by an average of 100—110ms more time reading the RC verb complex per intervening constituent.

We also ran planned pairwise comparisons on reading times within the accusative and dative argument NPs, treating preverbal versus postverbal realization as a dichotomous variable. For these comparisons, since number of regions of presentation differed across items...
in the ACC_MOD and DAT_MOD regions, to normalize quantity of visually presented material we computed residual reading times by computing, for each region of analysis, a linear regression of reading time against (i) number of regions of presentation and (ii) total number of characters in the region of analysis. LME analyses revealed a significant effect of pre/post-verbal positioning at the ACC_MOD region (effect size 64.21ms, \( p < 0.01 \)), but not at any other position.

Finally, at the final phrase of the RC ("vovremya “on time" in Example (12)), we saw a main effect of number of interveners in the argument manipulation—driven by inflated reading times in the two-intervener condition—but not in the adjunct manipulation; this effect seems most likely a result of spillover from the RC verb complex, which immediately precedes the RC-final phrase only in the two-intervener argument condition.

Since a number of theoretically relevant effects emerge at locations where there were differences in the material being read in the previous few regions, we also conducted a spillover analysis (cf. Mitchell, 1984; Vasisht and Lewis, 2006; Jaeger et al., 2008) accounting for reading times at previous regions. This analysis is reported in Appendix B. Crucially, no qualitative differences in RT patterns emerged from this analysis; the effect seen at the nonfinite verb became insignificant, but the effect at the finite verb remained significant, and the effect at the ACC_MOD region became marginal.
3.1.5. Discussion

Reading-time results at the RC verb complex support memory-based theories, which predicted that increasing the number of interveners between the relative pronoun and the RC verb complex would increase processing load at the verb complex. However, the results do not clearly discriminate between the DLT and interference-based theories of retrieval difficulty. In favor of the DLT one might point out that there is no clear evidence that the NP argument interveners, which are presumably more featurally similar to the head noun than the temporal adjunct interveners, generated greater retrieval difficulty. Against the DLT one might point out that the argument interveners uncontroversially contain discourse referents, whereas the adjunct interveners may not (depending on what counts as a discourse referent).

These results at the verb complex are unsupportive of surprisal, since the most natural prediction under surprisal would be that the additional interveners would sharpen the comprehender’s expectations regarding RC verb identity and location, and that these benefits would be realized in easier verbal processing. The results are similarly unsupportive of the Entropy-Reduction Hypothesis: the ERH’s account of the English SRC/ORC processing difficulty asymmetry relied on the fact that an immediately post-nominal position is a high-entropy position, and that discovering a verb at that point reduces entropy sharply; but this account does not predict that adding a second noun-final intervener (either a dative NP argument or a PP adjunct) would increase the processing load at the RC verb beyond the load already arising from a single noun-final intervener. However, the effect observed at the accusative NP object (i.e., slower RTs on the RC-initial accusative NPs compared to post-verbal NPs) is clearly consistent with the predictions of surprisal, and possibly with those of the ERH (on the assumption that the accusative NP rules out more possibilities about how the RC may unfold when it appears at the RC onset than when it appears postverbally). The RTs on the accusative NP are unsupportive of memory-based models. If such models were to predict any difference in difficulty on the accusative NP it would be the post-verbal position that would be more difficult, when the NP can actually be integrated with the RC verb.

Because of the effect observed at the accusative NP, the interpretation of the RT patterns at the RC verb complex must be tempered with a note of methodological caution: we cannot completely rule out the possibility that the differences in reading times observed on the verb complex are not the product of spillover from preceding material, especially because the material immediately preceding the verb complex differs across conditions. In the design of the experiment, these NPs were made longer (all at least three words) so as to minimize this danger of spillover. Nevertheless, since our understanding as a field of the detailed nature of spillover, including how long spillover effects can persist under a given set of experimental conditions, remains poor we cannot be sure that the length of these NPs was sufficient for the difficulty induced by their onset to be completely over by the time the RC verb complex is reached. (Additionally, we cannot be sure that the length of these phrases themselves is not itself surprising and difficulty-inducing.) The best evidence against a spillover explanation comes from two sources. First, as shown in Appendix B, even when spillover effects are partialed out of reading times the same qualitative effects of experimental
manipulation are present and significant in the theoretically critical parts of our sentences. Second, reading times at and immediately following the dative NP argument are instructive: there were no significant differences on reading times within this argument, and reading times on the postmodifier of the dative argument were nearly identical across all three argument-manipulation conditions. Since this postmodifier is immediately preverbal in the two-intervener condition, it suggests that at least some part of the inflated reading times at the RC complex in the two-intervener condition is not due to any particular difficulty arising from syntactic processing of the immediately preceding constituent.

Given the importance of reading time data from regions immediately preceding and following the RC verb complex, one final methodological limitation must be mentioned: while the physical presentation of sentences was designed to ensure that the critical RC verb complex did not fall near the beginning or end of a line of text, the same cannot be said for the other regions of the sentence under discussion. This concern is remedied in Experiment 2b.

3.2. Experiment 2b (conducted in USA)

In order to address the above concern regarding physical positioning of non-critical regions with respect to line breaks, and also to test the replicability of the theoretically critical patterns observed in Experiment 2a, we ran a slightly modified version of this experiment on native Russian speakers in the United States.

3.2.1. Participants

Twenty-five native Russian speakers living in or visiting the United States participated in this experiment in Boston and San Diego for cash compensation. None had arrived in the United States before age 13, and all reported that they continue to use Russian on a regular basis and consider it the language they are most comfortable with.

3.2.2. Materials

We used the same materials as in Experiment 2a.

3.2.3. Procedure

Sentences were presented to participants in a non-cumulative word-by-word moving-window self-paced procedure on a Mac or a PC computer running Linger. Each experimental trial was presented as three lines of text. The first line break always occurred immediately before the relative pronoun. In the argument conditions, the second line break always occurred immediately after the end of the RC; in the adjunct conditions with one or two interveners, the second line break always appeared after the accusative RC-internal NP object. Unlike Experiment 2a, the postmodifiers of the accusative and dative RC-internal NPs (AccMod and DatMod) were always a single region of presentation. Procedures were otherwise identical to the preceding three experiments. The study typically took 30–40 minutes to complete.

3.2.4. Results

Statistical analysis procedures. The same procedures were used as in the previous experiments.
Comprehension accuracy. Comprehenders answered questions with accuracy of 62% (±5%) in the zero-intervener condition, 70% (±5%) in the one-argument-intervener condition, 71% (±5%) in the two-argument-intervener condition, 73% (±5%) in the one-temporal-intervener condition, and 68% (±4%) in the two-temporal-intervener condition. As with Experiment 2a, neither effect of adjunct nor of argument manipulation was statistically significant. The numeric trend of Experiment 2a for questions to be answered most accurately the maximally local condition has reversed here, suggesting that there may be no true between-condition differences in accuracy.

Reading times. Figures 7 and 8 show average reading times in each region for the argument and adjunct manipulations respectively. LME analyses indicated significant effects in the argument manipulation only in the ACC_MOD and finite-verb regions (both $p<0.01$), and in the adjunct manipulation at the TMP1, TMP2, and finite-verb regions (all $p<0.05$), with a marginal effect at the RC-final region ($p=0.097$). In the argument manipulation, results look generally similar to those found in Experiment 2a: we see numerical patterns in which reading times at the RC verb complex and the final region of the RC increase monotonically with number of verbal dependents intervening between the relative pronoun and the RC verb. Reading times are significantly different by condition only at the non-finite verb.

To assess the evidence for the generalization from Experiment 2a that mean RTs at the
verb complex increased linearly with number of intervening constituents, we again fit and compared multi-level linear models for each of the finite-verb and nonfinite-verb regions as described in Section 3.1.4.3. The parametric finite-verb model is significantly better than baseline \( (p<0.01) \) and no model is significantly better than it, but for the nonfinite-verb region it is not significantly better than baseline \( (p=0.47) \). The estimated slope in the finite-verb model is 60.66ms; if we include the (non-significant) slope in the nonfinite-verb model of 13.05ms, we obtain an estimate of about 74ms additional reading time at the RC verb complex per intervening constituent, a qualitatively similar but smaller effect than in Experiment 2a.

As with Experiment 2a, we also conducted analyses of the accusative and dative NP regions with pre/post-verbal realization as a dichotomous predictor. Consistent with the results of Experiment 2a, these analyses revealed a robust effect for the AccMod region (189.91ms effect size, \( p<0.001 \)); the effect was marginal at the ACC region (90ms effect size, \( p=0.070 \)), but not for the DAT or DAT_MOD regions (both \( p>0.6 \)).

Also as with Experiment 2a, we conducted a spillover analysis, reported in Appendix B. Crucially, none of the qualitative RT patterns changed in this analysis; and both the effects at the finite verb and the ACC_MOD region remained significant.

3.2.5. Discussion

Although with only twenty-five participants this experiment had less statistical power than Experiment 2a, the results largely corroborate those of the previous experiment. The main difference is that there is less evidence for differences in reading times at the finite verb of the RC in the present experiment than in Experiment 2a, though the numerical patterns at this region matched those of Experiment 2a with the exception of the two-intervener condition in the adjunct manipulation. At accusative object NPs the reading-time penalty in preverbally realized conditions is qualitatively the same as in Experiment 2a, but is significant only in the AccMod region, whereas in Experiment 2a it was significant in both the Acc and AccMod regions. The overall pattern of results in this experiment can thus be summarized as qualitatively similar to that found in Experiment 2a, except that effects here tend to emerge as reliable one region further downstream.

The key results of Experiment 2 can be summarized as follows: in ditransitive subject-extracted RCs in Russian where both RC-internal NPs are full, processing difficulty increases monotonically with the number of interveners between the relative pronoun and the RC verb, and the RC verb complex is a locus of the inflated reading times corresponding to that processing difficulty. These inflated reading times were significant at the finite RC verb in Experiment 2b, and in both the non-finite verb and the immediately preceding finite verb in Experiment 2a. These results are predicted by retrieval-based accounts but not by expectation-based accounts. Although disentangling this effect from possible spillover effects is rather delicate (see also Appendix B), the results in the two-intervener argument manipulation seem to indicate that the RC verb complex is itself a genuine locus of processing difficulty.

One effect reliably observed in Experiment 2 is, however, more consistent with the predictions of expectation-based theories than with those of memory-based theories: the
reading-time penalty paid at an RC-initial accusative object NP. We saw similar evidence for such an effect in Experiment 1. This effect is reminiscent of the result reported by Staub (2010), who compared eye movement behavior in reading of English ORCs with superficially similar complement clauses:

(14) a. The employees that the fireman noticed hurried across the open field.
    b. The employees hoped that the fireman noticed the people who were still in the building.

Staub found an increase in first-pass regression rate and go-past reading times on the RC-initial determiner the in (14a) in comparison with (14b). As noted by Staub, this result specifically confirmed a long-outstanding prediction of surprisal, according to which there should be some processing cost associated with ruling out the possibility that the subject is extracted from the embedded clause—a possibility absent in (14b). Notably, Staub’s data also confirmed previous findings that the ORC verb (noticed in (14)) is itself an independent locus of processing difficulty (e.g., Grodner and Gibson, 2005). In Staub’s data, however, signs of differential difficulty also emerged before the determiner—at the relative pronoun that, where first-fixation and go-past durations were inflated in (14a) compared with (14b) raising the possibility that the effect at the determiner is spillover from the lower expectation for the word that after a noun than after a complement clause-selecting verb.

No such confound exists for our results, strengthening the case that, at least in some situations, encountering an NP at the beginning of an RC where it is unexpected can induce immediate processing difficulty.

4. General Discussion

Across four experiments with two different designs, we find two consistent patterns in reading time within Russian relative clauses varying in extraction type and word order. First, reading times at RC verbs increase monotonically with the number of dependents of the RC verb intervening between it and the relative pronoun. In Experiment 1, this pattern was seen as an interaction between the extraction type of the RC and the “defaultness” of transitive RC-internal word order: for subject-extracted RCs, the “default” verb-object order yielded faster reading; for object-extracted RCs, the “scrambled” verb-subject order yielded faster reading. In experiment 2, the pattern was seen as a monotonic increase in reading times at RC verb complexes as the number of preverbal dependents intervening between the relative pronoun and the verb complex is increased, regardless (for the most part) of what type of preverbal dependent intervened. Although it was not clear from Experiment 1 whether this effect originates at the RC verb independently of processing difficulty associated with the onset of the RC, the results of Experiment 2 provide some degree of evidence that the RC verb is indeed an independent locus of comprehension difficulty. These results are broadly

6 Though this pattern was not significant in Staub’s analysis, the first author of the present paper has replicated it at $p < 0.05$ in an unpublished self-paced reading study.
Experimental result

<table>
<thead>
<tr>
<th></th>
<th>No overall SRC&gt;ORC preference</th>
<th>RC-global locality pref. with marked rel. pronouns in Expt 1</th>
<th>SRC bias for <em>chto</em> in Experiment 1b</th>
<th>RC verb in Expt 1 &amp; RC verb complex in Expt 2</th>
<th>Fronted accusative NP penalty in Expts 1 &amp; 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>DLT</td>
<td>✓</td>
<td>✓</td>
<td>–</td>
<td>✓</td>
<td>?</td>
</tr>
<tr>
<td>SBI</td>
<td>✓</td>
<td>✓</td>
<td>–</td>
<td>✓</td>
<td>?</td>
</tr>
<tr>
<td>Word-order theories</td>
<td>(√)</td>
<td>✓</td>
<td>✓(? )</td>
<td>×</td>
<td>?</td>
</tr>
<tr>
<td>Surprisal</td>
<td>(√)</td>
<td>✓</td>
<td>✓</td>
<td>×</td>
<td>✓</td>
</tr>
<tr>
<td>Perspective Shift</td>
<td>×</td>
<td>–</td>
<td>✓(? )</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Structural</td>
<td>×</td>
<td>–</td>
<td>✓(? )</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Subject Preference</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5: Support for/inconsistency with predictions of different theories. ✓: theory makes a clear prediction of an effect matched by data; (√): theory makes prediction that an effect (SRC>ORC asymmetry) will be weaker than seen in other studied languages, and lack of any observed asymmetry is weakly consistent with predictions; ×: theory makes a clear prediction which is disconfirmed by the experimental result; –: theory makes no prediction; ✓(?): theory is vague in some respect but seems to make prediction consistent with empirical data; ?: theory needs further formal clarification to determine whether predictions are consistent or inconsistent with observed data.

supportive of memory-based theories.

Second, in three out of four experiments there is evidence that an RC-initial accusative NP induces immediate comprehension difficulty. In Experiment 1a this effect is seen among RCs with case-marked relative pronouns as an interaction of RC extraction type and locality at the RC-internal NP; in Experiments 2a and 2b it is seen as an effect of RC-internal argument NP ordering within the accusative object NP. In all these cases, RTs are greater when the accusative NP appears RC-initially than when it appears postverbally. The exception is Experiment 1b, where the accusative NP was read numerically (though not significantly) faster RC-initially than postverbally; but even here, the interaction between word order and RC extraction type indicates that there is a processing penalty associated with RC-initial placement for accusative NPs relative to the cost of RC-initial placement of nominative NPs. This processing cost associated with RC-initial accusative object NPs is directly predicted by expectation-based theories. The predictions on this front are unclear for memory-based theories: under these theories a postverbal NP is integrated with its governing verb but a preverbal NP cannot be, so the integration cost would be greater for preverbal NPs. However, the preverbal NP could be more taxing in terms of the overall representation of syntactic structure in memory: for DLT, this would be manifested in storage cost, and for SBI theories such as that of [Lewis and Vasishth 2005], in which incremental syntactic structure is always fully connected, it would be manifested in greater time spent structure-building to
accommodate the preverbal NP than to accommodate the postverbal NP (Vasishth, p.c.). The direct support from this result for expectation-based theories is not confounded by spillover, since it seems clear under all theories that the difficulty of immediately preceding word is consistently greater in the RC-initial conditions (a relative pronoun) than in the postverbal conditions, and since Experiment 1 demonstrates that this fronting penalty occurs only for accusative NPs in SRCs, not for nominative NPs in ORCs.

How the results of these experiments bear on a wide variety of prominent theories is summarized in Table 5. In Experiment 1, the lack of clearly greater processing difficulty for ORCs than for SRCs overall is damaging to perspective-shift and structural-subject-preference theories, but not to memory-based theories in which it is word order rather than grammatical function that predicts processing difficulty, or to expectation-based theories in which relative frequency of a structure determines its difficulty (word-order theories and surprisal), since the frequency ratio between ORCs and SRCs is far less skewed in Russian than in English (see Section 2.2). The overall processing preference for verb-local configurations observed was predicted by DLT and SBI (on the basis of locality) and by word-order and surprisal theories (on the basis of frequency). The strong bias for interpreting the case-syncretized relative pronoun chto as a cue indicating subject-extraction of the upcoming RC is predicted clearly by surprisal, on the hypothesis that comprehenders track fine-grained co-occurrence frequencies; this pattern is what was predicted by perspective-shift and structural-subject-preference theories, but these theories do not account for why the pattern disappears when the relative pronoun is case-marked; memory-based theories are silent on the matter; and it is not clear what prediction would be made by the ERH. The effect of number of interveners on RC verb processing times specifically is clearly predicted by memory-based theories, but contravenes the predictions of expectation-based theories. Finally, only surprisal clearly predicts the consistent processing-time penalty observed at accusative NPs placed at the beginning of SRCs (the low-frequency, non-default position).

Overall our results thus provide support for both memory-based and expectation-based theories, and by the same token are thus damaging to extant unitary accounts of processing difficulty. These results support ideas explored by researchers such as Boston et al. (2011) that “a complete theory of syntactic complexity must integrate insights from both expectation-based and memory-based theories.” The existing data might be consistent with Boston et al.’s model and with the Psycholinguistically Motivated Tree-Adjoining Grammar model of Demberg and Keller (2009), both of which include both surprisal and working-memory (retrieval for Boston et al., verification for Demberg & Keller) components of processing difficulty; it may turn out in these models that the surprisal component naturally dominates at the RC-initial accusative NP, whereas the memory component dominates at the RC verb. It is worth recalling, however, that verbs in many other language/construction combinations show expectation-based patterns that are not unambiguously predicted by memory retrieval-based theories; these cases include German main and subordinate clauses (Konieczny, 2000, Konieczny and Döring, 2003; Levy and Keller, 2012), Hindi relative clauses (Vasishth and Lewis, 2006), and Japanese main clauses (Nakatani and Gibson, 2008). This observation returns us to the original question we posed in Section 1 of what properties of language/construction...
combinations might determine when the processing-difficulty patterns we see will most closely reflect the predictions of expectation-based theories or memory-based theories. We had raised several possibilities: the key factors might include the dominant word order of the language, the morphological richness of the language, and/or the construction type investigated (most notably whether relativization is involved). Based on our present results in Russian, we can reasonably conclude that morphological richness on its own is not a key factor. Russian is much closer to German, Hindi, and Japanese than to English in its morphological complexity, and if anything should probably be said to be more morphologically complex than German insofar as it has six cases as compared to German’s four (the comparison with Hindi and Japanese is more difficult as the distribution and functional role of case marking in these languages overlaps significantly with that of prepositions in European languages). Given that Vasishth and Lewis (2006) demonstrated expectation-based processing patterns in relative clauses for verb-final Hindi, it seems reasonable to suspect that the dominant word order of a language plays a key role in determining the syntactic complexity of relative clauses in that language. One generalization that might profitably be pursued is that the verb-medial languages tend to exhibit the general patterns predicted by memory-based theories, whereas verb-final languages tend to exhibit the general patterns predicted by expectation-based theories. Vasishth et al. (2010) provide collateral evidence for this generalization, finding that native German speakers maintain more accurate expectations for upcoming sentence structure through multiple center-embeddings in German than native English speakers do in English. This generalization is not exceptionless—Vasishth and Drenhaus (2011) have recently found evidence for integration cost effects in German when memory load is made extremely high, and Jaeger et al. (2008) presented evidence for anti-locality effects in English at matrix-clause verbs—but it may serve as a useful rubric to guide further research on memory and expectations in syntactic comprehension. Another possibility for generalizing the present results would be to hypothesize that, as suggested by Vasishth and Drenhaus, “expectation plays a dominant role only when working memory load is relatively low”, though the question would remain as to why the working-memory threshold seems to be higher for verb-final languages like German and Japanese than for verb-medial languages like English and Russian. (A corollary of this discussion is that empirical research on RC comprehension in verb-initial languages is sorely lacking and could be of considerable theoretical value.)

Overall, these experiments underscore the value of cross-linguistic empirical breadth in advancing our understanding of both the syntactic complexity of relative clauses—a topic of theoretical interest in its own right—and more generally the interplay between memory and expectations in online sentence comprehension. The patterns observed in the study of speakers’ comprehension of sentences in their native languages are clearly emergent from a combination of the universal cognitive capacities of our species with contingent facts about the language in question. When study is restricted to a single language, however, it is impossible to discern which of these patterns are universal and which are language-contingent. Although no single theory yet explains why we see precisely the memory-based and expectation-based patterns in the circumstances we do, expanding the scope of inquiry across languages raises prospects for clarifying this picture and thereby advancing our fundamental
understanding of online language comprehension. In cases of ambiguity resolution, our understanding has already benefited considerably from a broader cross-linguistic view (Brysbaert and Mitchell, 1996; Cuetos and Mitchell, 1988; Cuetos et al., 1996; Desmet et al., 2002; Gibson et al., 1996, 1999; Mitchell and Brysbaert, 1998). With the present studies we hope to contribute to similar advances in our understanding of syntactic complexity.

5. Acknowledgments

This work has benefited from presentations at Dundee University, Edinburgh University, Glasgow University, the University of Rochester, the 2007 CUNY Sentence Processing Conference, and Harvard University. We are grateful to our native Russian speaker participants in Volgograd, Kazan, Moscow, St. Petersburg, Boston, and San Diego for their interest in this study and willingness to participate. We are grateful to Igor Boguslavsky for making the Russian Dependency Treebank available to us. Frank Keller gave us valuable advice regarding experiment design. Natalia Roudakova gave invaluable assistance in the construction of materials for Experiment 2. Svetlana & Georgiy Fedorenko, Nina Isakova and Kristina Fedorenko were tremendously helpful in conducting Experiment 1b. The data for Experiment 2a would never have been collected without the facilitation of Natalia Kondryateva and Irina Ryabkova. Margo Schwartz collected some of the Experiment 1a data, and Maria Sokolov collected the Experiment 2b data in San Diego. Maria Polinsky has given us valuable feedback and suggestions related to this work on numerous occasions. Finally, we are grateful for the thoughtful comments of Shravan Vasishth and two anonymous reviewers. All mistakes are, of course, our own.

Appendix A. Tregex patterns for corpus searches

Appendix A.1. Experiment 1

<table>
<thead>
<tr>
<th>Condition</th>
<th>All RCs</th>
<th>Only full-NP RCs</th>
</tr>
</thead>
<tbody>
<tr>
<td>SRC, VO</td>
<td>^S-(ЧТО</td>
<td>КОТОРЫЙ</td>
</tr>
<tr>
<td></td>
<td>^S-(Я</td>
<td>ТЫ</td>
</tr>
<tr>
<td>SRC, OV</td>
<td>^S-(ЧТО</td>
<td>КОТОРЫЙ</td>
</tr>
<tr>
<td></td>
<td>^S-(Я</td>
<td>ТЫ</td>
</tr>
<tr>
<td>ORC, SV</td>
<td>^S-(ЧТО</td>
<td>КОТОРЫЙ</td>
</tr>
<tr>
<td></td>
<td>^S-(Я</td>
<td>ТЫ</td>
</tr>
<tr>
<td>ORC, VS</td>
<td>^S-(ЧТО</td>
<td>КОТОРЫЙ</td>
</tr>
<tr>
<td></td>
<td>^S-(Я</td>
<td>ТЫ</td>
</tr>
</tbody>
</table>
Appendix A.2. Experiment 2

All SRCs
(_ _ < /^S-(ЧТО|КОТОРЫЙ|КТО)-caseИМ$/) >> /^.*-релят/ 
Verb-initial SRC
/^S-(ЧТО|КОТОРЫЙ|КТО)-caseИМ$/ >> /^.*-релят/ . @V 
Noun-initial SRC
/^S-(ЧТО|КОТОРЫЙ|КТО)-caseИМ$/ >> /^.*-релят/ .

"followed by verb" 

Adjunct-initial SRC
(_ _ < /^S-(ЧТО|КОТОРЫЙ|КТО)-caseИМ$/) >> /^.*-релят/ 

2-adjunct-initial SRC
(_ _ < /^S-(ЧТО|КОТОРЫЙ|КТО)-caseИМ$/) >> /^.*-релят/ 

Appendix B. Spillover analysis of Experiment 2

The materials of Experiment 2 were designed to minimize the possibility that any possible surprisal-based processing difficulty upon encountering preverbal accusative or dative NP arguments would spill over to the RC verb complex, by introducing additional regions postmodifying each of these NPs to “absorb” the spillover (see Experiment 2 of Grodner and Gibson, 2005, for a similar design). Nevertheless, there is no guarantee that these regions are sufficient to capture all such spillover, nor is the material immediately preceding either the RC verb complex or the accusative NP—both regions showing results of critical theoretical interest—identical across condition. We therefore linear mixed models to conduct analyses of reading-time effects at these conditions attempting to account for these possible effects of spillover. There are a number of ways one might attempt to model spillover effects (e.g., Mitchell, 1984; Vasishth and Lewis, 2006; Jaeger et al., 2008), and conducting an exhaustive comparison of possible approaches is beyond the scope of the present paper. The approach we took here can be thought of a relatively straightforward extension of standard practices for computing length-adjusted residual reading times (Trueswell et al., 1994). For each of the two experiments (2a and 2b), we fit mixed linear regression models to each non-sentence-final word of the filler materials where the length of the current word and the RTs of each of up to five preceding words were included as predictors, both fixed and by-subjects random effects. For each experiment we attempted to fit a set of thirty models comprised of these predictors subject to the following constraints: (i) Word length must be included as both a fixed and a random effect; (ii) If $RT_{n-k}$ was to be included a fixed (respectively random) effect in the regression regression, then $RT_i$ for all $n-k+1 < i < n$ had to be included as a fixed (respectively random) effect as well; (iii) An “item” random effect was included indexing either specific filler-sentence/word-number combination (fine-grained item effect) or filler sentence (coarse-grained item effect). (iv) The model must converge and be non-singular. Of the converging, non-singular models we obtained, we picked the one with the best Bayesian Information Criterion for each experiment. This gave us 4-back fixed, 1-back random effects for Experiment 2a, and 2-back fixed, 1-back random effects for Experiment 2b.
2b; in both experiments fine-grained item effects were preferred. The final resulting word-length/spillover estimated model for Experiment 2a was as follows (random-effects tables are standard-deviation/correlation matrices):

<table>
<thead>
<tr>
<th>Estimate</th>
<th>Std. Error</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>246.288</td>
<td>21.960</td>
</tr>
<tr>
<td>wordlen</td>
<td>31.808</td>
<td>3.656</td>
</tr>
<tr>
<td>rt.1back</td>
<td>0.150</td>
<td>0.018</td>
</tr>
<tr>
<td>rt.2back</td>
<td>0.077</td>
<td>0.010</td>
</tr>
<tr>
<td>rt.3back</td>
<td>0.044</td>
<td>0.009</td>
</tr>
<tr>
<td>rt.4back</td>
<td>0.040</td>
<td>0.009</td>
</tr>
</tbody>
</table>

and the final model for Experiment 2b was as follows:

<table>
<thead>
<tr>
<th>Estimate</th>
<th>Std. Error</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>329.102</td>
<td>12.145</td>
</tr>
<tr>
<td>wordlen</td>
<td>47.573</td>
<td>16.241</td>
</tr>
<tr>
<td>rt.1back</td>
<td>0.043</td>
<td>0.013</td>
</tr>
<tr>
<td>rt.2back</td>
<td>0.003</td>
<td>0.007</td>
</tr>
</tbody>
</table>

Note that although we see spillover effects for both experiments, there is considerably more spillover in Experiment 2a—e.g., a 15% effect from the previous word—than for 2b (a 4.3% effect).

For each of these models, we extracted the fixed-effect parameter estimates plus the conditional modes of the subject-specific random effects (the “best linear unbiased predictors”) to derive subject-specific models of word-length and spillover effects. We used these models to derive RT predictions for each word in our experimental materials (using actual, not predicted, previous-word RTs in each prediction). For each word in our experimental materials, the residual between this prediction and the true recorded RT then became the response measure for LME analyses based on our experimental manipulation, using maximal random effects structures and likelihood-ratio tests (treating the likelihood ratio as approximately normally distributed; Baayen et al., 2008; Barr et al., under revision). The qualitative patterns at the theoretically critical finite-verb, nonfinite-verb, ACC, and ACC_MOD regions are quite similar to those shown in the main body of the paper; region-by-region reading times are shown in Figures 9–12. For the ACC and ACC_MOD conditions we merged the one- and two-intervener conditions, as they had not diverged from one another at this point within the sentence. For Experiment 2a, these analyses in the argument manipulation yielded a significant effect at the finite verb ($p<0.025$) and a marginal effect at the ACC_MOD region ($p=0.052$) but no significant effects at the non-finite verb or the ACC region; in the adjunct manipulation they yielded a significant effect at the non-finite verb ($p<0.05$) but not at the non-finite verb. For Experiment 2b, these analyses in the argument manipulation yielded a significant effect at the finite verb ($p<0.01$) and a significant effect at the ACC_MOD region ($p<0.025$), but not at the non-finite verb or ACC region; in the adjunct manipulation they yielded a significant effect at the finite verb ($p<0.025$) but not at the non-finite verb. Finally, we conducted a meta-analysis in which we combined the residual-RT data from both
experiments and included a fixed effect of experiment in the mixed models. In the argument manipulation this meta-analysis yielded significant effects at the finite verb ($p<0.001$) and the ACC_MOD region ($p<0.01$), but not at the ACC or nonfinite-verb regions. Although we do not claim that this analysis gives a definitive treatment of the problem of spillover in our experiments, it provides at least some evidence that the theoretically critical reading-time effects recovered in Experiment 2 are not obviously due to spillover effects.

Appendix C. Materials for Experiment 1a

Only the subject-extracted, default (SVO) word order, variant “A” version of each item is given; the other versions can be constructed straightforwardly on the pattern of Example (4) in Section 2.3, which is constructed from Item (29).

1. Диктатор, который возненавидел диссидента после недавних событий, Dictator, who.nom came_to_hate dissident.ACC after recent events, pronounced speech at meeting. “The dictator, who came to hate the dissident after the recent events, made a speech at the meeting.”

2. Терапевт, который обвинил кардиолога после неудачной операции, Therapist, who.acc accused cardiologist.ACC after unsuccessful operation, checked files in office. “The therapist, who accused the cardiologist after the unsuccessful operation, checked the files in the office.”

3. Мотоциклист, который игнорировал таксиста на центральной дороге, Motorcyclist, who.ignored taxi_driver.ACC on central road, made turn at intersection. “The motorcyclist, who ignored the taxi driver on the central road, turned right at the intersection.”

4. Инженер, который раздражал аналитика уже который год, написал Engineer, who.irtitated analyst.ACC already many year, wrote report about project. “The engineer, who had irritated the analyst for many years, wrote a report about the project.”

5. Скрипач, который разгневал флейтиста во время репетиции, отменил Violinist, who.angered flautist.ACC at time_of practice, cancelled концерт в пятницу. concert on Friday.
“The violinist, who angered the flautist during the practice, cancelled the concert on Friday.”

(6) Сыщик, который заметил подозреваемого около китайского ресторана, Detective, who noticed suspect.ACC near Chinese restaurant, pulled cap over eyes.
“The detective, who noticed the suspect near the Chinese restaurant, pulled his cap over his eyes.”

(7) Художник, который утомил коллекционера во время беседы, свернул холст с портретом. Artist, who tired collector.ACC at time_of visit, rolled_up canvas with portrait.
“The artist, who tired the collector during the visit, rolled up the canvas with the portrait.”

(8) Бухгалтер, который предупредил статиста перед важной встречей, Accountant, who warned statistician.ACC before important meeting, summarized results in a_hurry.
“The accountant, who warned the statistician before the important meeting, hurriedly summarized the results.”

(9) Солдат, который недолюбливал офицера с первого знакомства, получил медаль за отвагу. Soldier, who disliked officer.ACC from first meeting, received medal for bravery.
“The soldier, who had disliked the officer from their first meeting, received a medal for his bravery.”

(10) Врач, который посетил пациента до обеденного перерыва, записал рекомендации в блокноте. Doctor, who visited patient.ACC until lunch break, wrote recommendations on notepad.
“The doctor, who visited the patient until the lunch break, wrote his recommendations on a notepad.”

“The parman, who pushed the brunette near the bar, dropped a glass full of wine.”
(12) Entrepreneur, who interested investor at time_of discussion, considered offer of collaboration.
“The entrepreneur, who had interested the investor during the discussion, considered the offer of collaboration.”

(13) Grandfather, who hugged grandson after walk in courtyard, composed story about elephant.
“The grandfather, who hugged his grandson after the walk in the courtyard, made up a story about an elephant.”

(14) Mayor, who pleased governor after important press-conference, expected success in elections.
“The mayor, who pleased the governor after the important press-conference, expected success in the elections.”

(15) Swede, who defeated Frenchman in tennis match, dropped racquet on court.
“The Swede, who defeated the Frenchman in tennis match, dropped his racquet on court.”

(16) Chemist, who congratulated biologist with publication article, received prize for research.
“The chemist, who congratulated the biologist on the article’s publication, received a prize for his research.”

(17) Bully, who provoked classmate after second class, explained situation after fight.
“The bully, who provoked his classmate after the second class, explained the situation after the fight.”

(18) Expert, who disappointed manager in end year, cut_off contact with firm.
“The expert, who disappointed the manager at the year’s end, cut off contact with the firm.”

(19) Administrator, who reproached librarian for small blunder, wrote complaint in anger.
“The administrator, who reproached the librarian for the small blunder, wrote a complaint in anger.”

(20) Firefighter, who noticed medic in burning house, summoned help by radio.
“The firefighter, who noticed the medic in the burning house, summoned help by radio.”

(21) Director, who puzzled participant during quiz, answered question with another question.
“The director, who puzzled the participant during the quiz, answered the question with another question.”

(22) Lawyer, who recommended notary before judicial hearing, arranged meeting on Tuesday.
“The lawyer, who recommended the notary before the judicial hearing, arranged the meeting on Tuesday.”

(23) Senator, who disturbed president after important meeting, sent letter to Congress.
“The senator, who disturbed the president after the important meeting, sent a letter to Congress.”

(24) Reporter, who notified investigator after horrible killing, described crime in details.
“The reporter, who notified the investigator after the horrible killing, described the crime in detail.”
(25) Генерал, который унижил лейтенанта у всех на глазах, допустил ошибку в битва.
"The general, who humiliated the lieutenant in front of everyone, committed an error in battle."

(26) Философ, который процитировал лингвиста на международной конференции, преподавал курс в Кембридже.
"The philosopher, who cited the linguist in an international conference, taught a course at Cambridge."

(27) Соло-гитарист, который любил ударника с ранней юности, основал группу в 1988 году.
"The solo guitarist, who loved the drummer from early youth, established the group in 1988."

(28) Ныряльщик, который испугал пловца в большом бассейне, сделал движение в сторону.
"The diver, who frightened the swimmer in the big pool, moved to one side."

(29) Слесарь, который ударил электрика со всего размаха, ушел домой с синяком под глазом.
"The mechanic, who hit the electrician with all his strength, went home with a bruise under his eye."

(30) Полицейский, который ранил преступника во время перестрелки, обронил револьвер во время погони.
"The policeman, who injured the criminal during the shootout, dropped his revolver during the chase."

(31) Аптекарь, который запутал посетителя во время визита, прочитал рецепт еще раз.
"The druggist, who confused the visitor during the visit, reread the prescription."
(32) Вор, который подставил гангстера во время ограбления, спрятал выручку в сейфе.
“The thief, who framed the gangster during the robbery, hid the loot in a safe.”

Appendix D. Materials for Experiment 1b

These materials are highly similar to those for Experiment 1a; the key differences are that in (i) in 1a, commas were used to delimit RC onsets and offsets, but in 1b these commas were omitted; (ii) in 1a an adjunct phrase was included at the end of the RC, but in 1b there was no such adjunct phrase; and (iii) in 1a every RC NP had to serve equally as a potential head NP (see experiment design in Section 2.3) but this constraint was not present in 1b, so different NPs and main-clause predicates were used across the two experiments in a few cases.

An example item is given below in all eight conditions; further items are given only in the subject-extracted RC, default (SVO) word order, case-marked relative pronoun condition, from which the other conditions can be derived.

(33) a. Диктатор который ненавидел диссидента произнес речь на собрании.
“The dictator who hated the dissident gave a speech at the meeting.” [CASE-MARKED relative pronoun, SRC, DEFAULT]

b. Диктатор который диссидента ненавидел произнес речь на собрании.
“The dictator who hated the dissident gave a speech at the meeting.” [CASE-MARKED relative pronoun, SRC, SCRAMBLED]

c. Диктатор которого диссидент ненавидел произнес речь на собрании.
“The dictator who the dissident hated gave a speech at the meeting.” [CASE-MARKED relative pronoun, ORC, DEFAULT]

d. Диктатор которого ненавидел диссидент произнес речь на собрании.
“The dictator who the dissident hated gave a speech at the meeting.” [CASE-MARKED relative pronoun, ORC, SCRAMBLED]

e. Диктатор что ненавидел диссidenta произнес речь на собрании.
“The dictator who hated the dissident gave a speech at the meeting.” [CASE-SYNCRETIZED relative pronoun, SRC, DEFAULT]

f. Диктатор что диссидента ненавидел произнес речь на собрании.
“The dictator who hated the dissident gave a speech at the meeting.” [CASE-SYNCRETIZED relative pronoun, SRC, SCRAMBLED]

g. Диктатор что диссидент ненавидел произнес речь на собрании.
“The dictator who the dissident hated gave a speech at the meeting.” [CASE-SYNCRETIZED relative pronoun, ORC, DEFAULT]

h. Диктатор что ненавидел диссident произнес речь на собрании.
“The dictator who hated the dissident gave a speech at the meeting.” [CASE-SYNCRETIZED relative pronoun, ORC, SCRAMBLED]
“The dictator who the dissident hated gave a speech at the meeting.” [CASE-SYNCRETIZED relative pronoun, ORC, SCRAMBLED]

(34) Терапевт который обвинил кардиолога проверил файлы в кабинете.
physician who.NOM cardiologist.ACC accused checked files at office.
“The physician who accused the cardiologist checked the files in the office.”

(35) Врач который посетил пациента выписал сироп от кашля.
doctor who.NOM visited patient.ACC prescribed syrup from cough.
“The doctor who visited the patient prescribed syrup for the cough.”

(36) Бармен который толкнул брюнета уронил стакан с вином.
barman who.NOM pushed dark_haired_man.ACC dropped glass with wine.
“The barman who pushed the dark-haired man dropped the glass with the wine.”

(37) Предприниматель который заинтересовал инвестора рассмотрел предложение о сотрудничестве.
entrepreneur who.NOM interested investor.ACC evaluated offer of cooperation.
“The entrepreneur who interested the investor evaluated the offer of cooperation.”

(38) Дедушка который обнял внука сочинил историю про слона.
grandfather who.NOM hugged grandson.ACC composed story about elephant.
“The grandfather who hugged the grandson made up a story about an elephant.”

(39) Мэр который обрадовал губернатора ожидал успеха на выборах.
mayor who.NOM pleased governor.ACC expected success in elections.
“The mayor who pleased the governor expected to succeed in the election.”

(40) Теннисист который обыграл брата уронил ракетку на корт.
tennis_player who.NOM defeated brother.ACC dropped racket on court.
“The tennis player who beat his brother dropped the racket on the court.”

(41) Химик который поздравил коллегу разделил приз за исследования.
chemist who.NOM congratulated colleague.ACC shared prize for research.
“The chemist who congratulated the colleague shared the prize for the research.”

(42) Парень который спровоцировал однокурсника объяснил ситуацию после драки.
young_man who.NOM fellow_student fellow_student.ACC explained situation after fight.
“The young man who provoked a fellow student explained the situation after the fight.”

(43) Эксперт который разочаровал представителя прервал контракт с фирмой.
expert who.NOM disappointed representative.ACC terminated contract with firm.
“The expert who disappointed the representative terminated the contract with the firm.”

(44) Администратор который упрекнул библиотекаря написал жалобу в гневе.
administrator who.NOM reproached librarian.ACC wrote complaint in wrath.
“The administrator who reproached the librarian filed a complaint in anger.”

(45) Мотоциклист который игнорировал таксиста сделал поворот на перекрестке.
motorcyclist who.NOM ignored taxi_driver.ACC made turn at intersection.
“The motorcyclist who ignored the taxi-driver made a turn at the crossing.”
(46) Пожарник который спас медика пострадал ожоги первой степени.

"The firefighter who saved the medic suffered burns first level."

(47) Ведущий который озадачил участника ответил вопросом на вопрос.

"The program host who puzzled the contestant answered the question with a question."

(48) Адвокат который рекомендовал нотариуса назначил встречу на вторник.

"The lawyer who recommended the notary arranged a meeting on Tuesday."

(49) Сенатор который обеспокоил президента послал письмо в Конгресс.

"The senator who worried the president sent a letter to the Congress."

(50) Репортер который пригласил следователя описал убийство в деталях.

"The reporter who invited the investigator described the murder in great detail."

(51) Генерал который унизил лейтенанта допустил ошибку в битве.

"The general who put down the lieutenant made a mistake in the battle."

(52) Философ который процитировал лингвиста преподавал класс в Кембриджке.

"The philosopher who cited the linguist taught a class at Cambridge."

(53) Соло-гитарист который любил ударника основал группу в 1988.

"The solo-guitarist who loved the drummer founded the band in 1988."

(54) Ныряльщик который испугал пловца сделал движение в сторону.

"The diver who scared the swimmer moved over."

(55) Слесарь который ударил электрика потерял терпение в споре.

"The mechanic who hit the electrician lost patience during the argument."

(56) Инженер который раздражал аналиста написал доклад о проекте.

"The engineer who irritated the analyst wrote a report about the project."

(57) Полицейский который ранил преступника обронил револьвер в погоне.

"The policeman who injured the dissident dropped his revolver during the chase."

(58) Аптекарь который запутал посетителя прочитал рецепт еще раз.

"The druggist who confused the visitor reread the prescription."

(59) Вор который подставил напарника спрятал выручку в сейфе.

"The thief who framed his accomplice hid his loot in the safe."
“The thief who framed his accomplice hid the loot in a safe.”

(60) **Musician** who.NOM angered organizer.ACC cancelled concert on Friday.
“The musician who angered the organizer cancelled Friday’s concert.”

(61) **Dissident** who.NOM hated dissident.ACC pronounced speech at meeting.

(62) **Organizer** who angered the organizer cancelled Friday’s concert.

(63) **Dissident** who.tired collector.ACC rolled up canvas with portrait.
“The artist who noticed the suspect pulled his cap over his eyes.”

(64) **Collecter** who.tired the collector rolled up the canvas with the portrait.”

(65) **Detective** who noticed the suspect pulled his cap over his eyes.

**Appendix E. Materials for Experiment 2**

We only present the maximally local condition (11a & 12a) of each item. Underscores denote word sequences presented as a single region. The two temporal phrases are the last regions in the first clause, the dative argument starts at the dative noun and all following regions up to but not including the last region in the RC, and the accusative argument starts at the accusative noun and includes everything before the dative noun.

(1) **Mayor_**city_**visited** new factory after press-conference at 3 o’clock, and director factory, who promised tell plans.ACC for future official.DAT city_administration in_details, was very happy visit.
“The mayor of the city visited the new factory after the press conference at 3 o’clock, and the director of the factory, who promised the official of the city administration to explain the plans for the future in detail, was very happy with the visit.”

(2) **Cardiologist** not_excluded possibility repeat_infarction after visit at 11 o’clock, and attending_doctor, who wanted_to reduce dose.ACC strong medicine patient.DAT with_high_blood_pressure in_half, decided order additional tests.
“The cardiologist did not exclude the possibility of another heart attack after the visit at eleven...
o’clock, and the attending physician, who had wanted to cut the patient with high blood pressure’s
dose of the strong medication in half, decided to order additional tests.”

(3) Marina was walking her dog in the park before lunch at two o’clock, and her neighbor, who was
planning to send a New Year’s card to his relative in Kiev, went to the post office to buy an envelope
and a stamp.”

(4) Embassy urgently needed documents from the legal service, early in the morning,
and the postman, who managed to deliver the refugee’s passport to the visa officer on time, was
rewarded for his punctuality.

(5) The chef was praised for his mastery early in the evening around six o’clock, but the waiter, who
forgot to bring the dish of veal to the customer in the black suit on time, didn’t receive a tip after
dinner.”

(6) The dean sat in his office during the post-class break, but the professor, who had enthusiastically
decided to demonstrate the chemistry experiment to the student in the older department, forgot
about the appointment.”

(7) There were big sales at the market on December 31 before New Year, but the merchant, who was
trying to sell Moroccan oranges to the buyer with the dog, didn’t want to drop his price.”
(8) Business in the company had gone fine this year since August, and the director of development, who suggested that profit from transactions be invested entirely in the Saint Petersburg branch, finally obtained agreement from the head of the company.

(9) The director prepared a laudatory speech on Sunday at 8 in the evening, but the host, who was going to give the award to the Italian actress with great pomp and circumstance, was so nervous he forgot to even introduce the director.

(10) The hostess bought fruit for dessert on Saturday early in the morning, but the host, who was going to bake the guest from the capital a cherry pie for the holiday, couldn’t find the needed recipe.

(11) The designer signed a big contract with the USA this Wednesday, and his agent, who was supposed to present the spring-season collection to the Greek entrepreneur at the exhibition, decided to ask the designer for a raise.

(12) The students organized a concert in honor of May 9 [Victory Day] after the parade in the evening, and the girl who had promised to sing a war song for the veterans from Volgograd a cappella expended the most energy.
(13) The biologist disappointed his colleagues with his lack of success at the conference in March, but his colleague, who tried to explain the results of her research to the Australian scientists in great detail, was rewarded for her contribution to science. "The biologist disappointed his colleagues with his lack of success at the conference in March, but his colleague [fem], who tried to explain the results of her research to the Australian scientists in great detail, was rewarded for her contribution to science."

(14) The parents came to pick up their children after work around five in the evening, and the day-care worker, who started to read story about animals children.DAT with_expression, decided to finish the story the next day. "The parents came to pick up their children after work around five in the evening, and the day-care worker, who had started to read the children in the day-care center the story about animals with great expression, decided to finish the story the next day."

(15) The policeman looked for fingerprints in the room in the morning before the questioning, and the investigator, who wanted to show photographs of two suspects to the witness to the crime as soon as possible, hoped to catch him before lunch at work. "The policeman looked for fingerprints in the room in the morning before the questioning, and the investigator, who wanted to show photos of two suspects to the witness to the crime as soon as possible, hoped to catch him before lunch at work."

(16) The secretary received an important fax on Friday after work, and dealer, who planned to send package with goods to the client in Boston by express mail, added some more samples to the package. "The secretary received an important fax on Friday after work, and the dealer, who planned to send the package with goods to the client in Boston by express mail, added some more samples to the package."

(17) The assistant received an order for outfits this month after the holidays, and the dressmaker, who was preparing to sew a silk dress with special patterns for the client from London, ordered fabric from a specialized shop. "The assistant received an order for outfits this month after the holidays, and the dressmaker, who was preparing to sew a silk dress with special patterns for the client from London, ordered fabric from a specialized shop."
18) Секретарша закончила делать ксерокопии к началу собрания в 2 часа, и administrator, который должен был раздать описание нового проекта, was her very grateful.

“The secretary finished making photocopies for the start of the meeting at 2 o’clock, and the administrator, who was supposed to immediately hand out descriptions of the new projects to the department’s professors, was very grateful to her.”

19) Слесарь сидел на больничном на выходных в субботу днем, и его коллега, who could fix faucet.АCC in bathtub woman.DAT in_house_number_5 очень быстро, спешил по вызову в несколько квартир.

“The repairman was on sick leave during the day on Saturday over the weekend, but his co-worker, who was able to quickly fix the bathtub faucet for the woman in house number 5, hurried on calls to several apartments.”

20) Офицер рассказывал о героических подвигах во время церемонии в День Победы, и the general, who wanted to give medal.АCC for bravery lieutenant.DAT from_Krasnoyarsk торжественно, решил наградить еще нескольких военных.

“The officer spoke of the heroic deeds during the ceremony on Victory Day, and the general, who was going to award the lieutenant from Krasnoyarsk a medal for his bravery with great pomp and circumstance, decided to honor several other veterans as well.”

References


