Language and Cognition **9** (2017), 332–370. doi:10.1017/langcog.2016.21

When relative clause extraposition is the right choice, it's easier*

ELAINE J. FRANCIS
Purdue University

AND

LAURA A. MICHAELIS
University of Colorado at Boulder

(*Received* 24 *November* 2015 – *Revised* 19 *May* 2016 – *Accepted* 20 *May* 2016 – *First published online* 14 *July* 2016)

ABSTRACT

In one type of Relative Clause Extraposition (RCE) in English, a subject-modifying relative clause occurs in a displaced position following the matrix VP, as in: *Some options were considered that allow for more flexibility*. Although RCE incurs a discontinuous dependency and is relatively infrequent in discourse, previous corpus and acceptability judgment studies have shown that speakers prefer RCE over adjacent ordering when the RC is long in relation to the VP, the subject NP is indefinite, and the main verb is passive/presentative (Francis, 2010; Francis & Michaelis, 2014; Walker, 2013). The current study is the first to relate these conditional preferences to online measures of production. For a spoken production task that required speakers to construct sentences based on visual cues, results showed that the same factors that modulate choice of structure – VP length, RC length, and definiteness of the subject NP – also modulate voice initiation time. That is, when the sentential context warrants a particular structure, that structure becomes easier to produce. Following the approach of

[*] We are grateful to undergraduate research assistants Liz Ancel, Arianna Deldar, Becca Duggan, Corinne Feight, Xuebo Gong, and Morgan Robertson for their assistance with collecting, transcribing, and coding the data, and to Jyotishka Datta from the Purdue Statistical Consulting Service. We would like to thank Jill Duffield, Jack Hawkins, Tom Wasow, and audience members at LSA 2012 for their feedback on preliminary data as summarized in an LSA extended abstract (Francis & Michaelis, 2012). We are especially grateful to an anonymous reviewer for detailed comments which helped us improve the statistical analysis and writing of this paper. We gratefully acknowledge the Department of English and the Linguistics Program at Purdue University and the Department of Linguistics at the University of Colorado for financial support. Address for correspondence: Elaine Francis, Department of English, Purdue University, West Lafayette, IN 47907. e-mail: ejfranci@purdue.edu

MacDonald (2013), we explain these findings in terms of two production biases, one of which favors early placement of shorter, more accessible phrases and the other of which promotes rapid retrieval from memory of the most frequently used subtypes of a construction.

KEYWORDS: relative clause extraposition, grammatical weight, sentence production, syntactic alternation.

1. Introduction

Like the lexicon, grammar offers options to speakers in the form of nearly synonymous competitor forms. How do speakers choose between competing syntactic constructions? The point of departure for this study is the observation that language users make this choice based on both discourse-pragmatic and processing considerations (Arnold & Lao, 2008; Arnold, Losongco, Wasow, & Ginstrom, 2000; Bresnan & Ford, 2010; Bresnan & Nikitina, 2009; de Marneffe, Grimm, Arnon, Kirby, & Bresnan, 2012; Goldberg, 2006; Gregory & Michaelis, 2001; Wiechmann & Kerz, 2013). This study provides evidence, from both sentence production and sentence comprehension, that the discourse-pragmatic considerations underlying constructional choice are closely tied to processing-based considerations. It suggests that the two interact in a way that might initially appear counter-intuitive: the less frequent and structurally more complex option is the easier one to implement when it is the more contextually appropriate option. We will claim that this effect can be attributed to two general biases that language users bring to bear on language production and comprehension, one of which favors early placement of shorter, more accessible phrases and the other of which promotes rapid retrieval from memory of the most frequently used subtypes of a construction.

The grammatical opposition that will be our focus in this study involves the placement of a subject-modifying relative clause in English. In the pattern that we will refer to as RELATIVE CLAUSE EXTRAPOSITION (henceforth, RCE), as in (1a) (from the International Corpus of English Great Britain), a subject-modifying relative clause occurs after the matrix VP, rather than adjacent to the noun it modifies, as in (1b):

(1) a. Evidence has been found that shows one can have the physiological changes without feeling the corresponding emotion one is supposed to.
b. Evidence that shows one can have the physiological changes without feeling the corresponding emotion one is supposed to has been found.

Formal syntactic analyses of RCE, as summarized in Baltin (2006), posit a more complex structure for sentences like (1a) as compared with (1b), to account for

the discontinuous dependency between the subject NP and the relative clause (henceforth, RC) which modifies it. For example, according to Kayne (1994), a sentence such as (1a) involves base-generation of the NP (the noun *evidence* together with its RC modifier) in the post-verbal position, leftward movement of the noun *evidence* into the subject position, and stranding of the RC. In contrast, for a sentence such as (1b), the entire NP is generated in its usual specifier position before the verb and requires no separate movement of the head noun. Other syntactic analyses involve different technical mechanisms but a similar element of complexity. These have included rightward movement of the RC (Baltin, 1981; Ross, 1967), percolation of a non-local EXTRA feature from the head noun to the clause containing it (Kay & Sag, 2012), and syntactic co-indexing of the head noun with a clause-adjoined RC (Culicover & Rochemont, 1990). In addition to syntactic complexity, RCE also incurs a localized processing cost. Levy, Fedorenko, Breen, and Gibson (2012, p. 20) showed that RCE sentences were more difficult to comprehend in word-by-word reading than the corresponding non-extraposed variants.

Given the added complexity incurred by RCE, why should the structure in (1a) occur as an alternative to the canonical word order in (1b)? Previous research has focused on two lines of explanation: discourse function and grammatical weight. With respect to discourse function, RCE is preferred when the subject NP is focal and/or the VP is backgrounded, accounting for the tendency of RCE tokens to contain passive or presentative predicates (Francis, 2010; Kuno & Takami, 2004; Rochemont & Culicover, 1990; Walker, 2013) and indefinite subjects (Huck & Na, 1990; Walker, 2013). In (1a), for example, the writer is asserting the existence of a certain type of evidence, rather than asserting that such evidence has been found. By contrast, (1b) seems rather awkward, in part because it seems to be presupposing the existence of a certain kind of evidence, but doing so using an indefinite NP. Such a sentence can potentially be improved by adding a definite determiner, *the*, before *evidence*. With respect to grammatical weight, RCE provides a means of making the NP shorter while expressing the (typically longer) RC later in the sentence (Wasow, 2002, p. 7). Such an ordering fits with the general observation that shorter, simpler phrases tend to occur earlier in a sentence than longer, more complex phrases. Why should this be? Arnold et al. (2000) and Wasow (2002) propose that postponing heavier constituents affords speakers additional time to plan longer, more complex phrases. Similarly, Hawkins (2004, p. 141) proposes that both speakers and listeners prefer more efficient structures in which integration distances between heads of phrases are minimized. Besides discourse status, then, two additional reasons for the awkwardness of (1b) are: (i) that the speaker must plan and produce a heavy constituent early in the sentence, and (ii) that there is a long integration distance between the head of the subject NP, *evidence*, and the VP, *has been found*.

Our own research on RCE in English has been the first to combine both previous lines of explanation – discourse function and grammatical weight – and the first to provide empirical grounding through a combination of corpus analyses and experimental tasks. Francis (2010) confirmed Hawkins' (2004) efficiency-based predictions for weight effects in comprehension and production. In an experiment that measured whole-sentence reading times, RCE sentences were read faster than non-RCE sentences when the RC was three times longer than the VP. Similarly, in a sample of RCE and non-RCE tokens from the International Corpus of English Great Britain (Nelson, Wallis, & Aarts, 2002), RCE occurred in only about 15% of tokens overall, but frequency of RCE increased to 91% when the RC was five times longer than the VP.[1] Francis and Michaelis (2014) followed up with a multifactorial analysis of the same corpus data, showing independent effects of predicate type (RCE was more common with passive and presentative predicates), definiteness of the subject (RCE occurred more often with an indefinite subject NP), and information status of the predicate (RCE occurred more often when a semantically similar VP had been used in the prior discourse), in addition to grammatical weight (2014, p. 82).

While our previous work on RCE has begun to show how discourse factors and grammatical weight can predict speakers' choices, our reliance on corpus analyses and reading-based tasks has limited our explanatory power. For example, corpus data show that certain factors influence speakers' and writers' choices, but do not allow us to directly manipulate those factors. By contrast, a production task tells us more directly about the reasons for speakers' choices, and further allows us to measure aspects of online planning for production, providing new insights into how RCE and non-RCE structures are processed. In the current study, two experiments – a preference task and an elicited production task – are designed to answer specific questions that have arisen from this earlier work.

1. Does manipulating phrase length and definiteness result in speakers choosing RCE more often under certain conditions? If so, we have direct evidence for the roles of length and definiteness in determining when RCE is chosen.
2. Are the length and definiteness-based conditions under which RCE is chosen most often the same as the conditions under which RCE utterances

[1] The participants in the experiments from Francis (2010) were American English speakers, while the corpus data were from British English. Since the experimental and corpus data were consistent with each other in that study, we believe that dialect does not play a major role in determining the distribution of RCE. However, we acknowledge that a direct comparison of American English and British English corpora could potentially identify differences.

are produced most quickly? If so, this suggests a tight connection between choice of structure and ease of production.

3. Are the length and definiteness-based conditions under which RCE is chosen most often the same for elicited production as for a reading-based preference task? If so, it would appear that production and comprehension mechanisms are drawing on the same knowledge sources.

Findings suggest affirmative answers to questions (1) and (2). RCE was preferred more often when the VP was short, the RC was long, and the subject NP was indefinite. Conversely, non-RCE was preferred more often when the VP was long, the RC was short, and the subject NP was definite. Similarly, RCE was faster to be produced when the subject NP was indefinite and the RC was long, while non-RCE was faster to be produced when the subject NP was definite and the RC was short. (VP length did not affect response times.) Thus, it appears that the structure that the sentential contexts warrants is preferred more often and easier to produce, regardless of dependency type. These findings are perhaps surprising when considering that RCE involves a discontinuous dependency with additional syntactic complexity (Baltin, 2006), is used relatively infrequently in discourse (Francis, 2010), and incurs a localized processing cost in word-by-word reading (Levy et al., 2012). Furthermore, while the effect of RC length on ease of production can be explained in terms of the well-established short-before-long bias in production (Wasow, 2002), the similar effect of definiteness cannot. Invoking the concepts of Easy First and Plan Reuse from MacDonald's (2013) Production–Distribution–Comprehension theory, we argue that these results make sense in the context of general biases that affect sentence production and planning. Specifically, the Easy First bias says that speakers prefer to produce shorter and more accessible phrases earlier in the sentence to allow more time for memory retrieval and production planning of longer and less accessible phrases. This subsumes the established short-before-long bias and helps explain why our participants used RCE most often and were fastest to produce RCE when the VP was short and the RC was long. The Plan Reuse bias says that speakers prefer to reuse frequently occurring structural patterns because they are easier to retrieve from long-term memory. This bias helps explain why our participants used RCE more often and were faster to produce RCE when the subject NP was indefinite.

The answer to question (3) is partially affirmative, showing both expected similarities and interesting differences between tasks. While the definiteness effects were quite consistent, the length effects – in particular the relative influence of VP length and RC length – differed for preference in reading vs. elicited production. We suggest that the differences across task types may reveal different mechanisms at work in comprehension vs. production.

In the remainder of this paper, we elaborate on relevant findings from the literature on RCE and similar constructions (Section 2), and then present the results of the preference task (Section 3) and the elicited production task (Section 4). We then offer a general discussion of the findings and their implications (Section 5) and briefly conclude the paper (Section 6).

2. Grammatical weight, discourse factors, and sentence processing

2.1. GRAMMATICAL WEIGHT: ABSOLUTE OR RELATIVE?

Robust short-before-long effects have been shown in corpus studies of several syntactic alternations, including prepositional/double-object clauses (Bresnan, Cueni, Nikitina, & Baayen, 2007), verb-particle phrases (Gries, 2003; Lohse, Hawkins, & Wasow, 2004), genitive NPs (Rosenbach, 2005), Heavy NP Shift (Arnold et al., 2000), and RCE (Francis, 2010). However, different studies have used different methods of measuring such effects. This section considers which measures of grammatical weight are most useful for predicting phrasal ordering preferences, with an emphasis on RCE in English (Francis, 2010; Francis & Michaelis, 2014) and German (Konieczny, 2000; Strunk, 2014; Uszkoreit et al., 1998).

First, it should be noted that there are a few different ways in which weight can be measured, including syllables, words, and syntactic nodes. Stallings, MacDonald, and O'Seaghdha (1998) argue, based on previous production studies, that length in syllables is not relevant for constituent ordering during sentence production, and instead favor word-based measures. Wasow (1997) considers whether word-based measures are equivalent to measures of structural embedding. Using corpus data from Heavy NP Shift and the prepositional/double-object alternation in English, he found that predictions based on length in words and predictions based on number of syntactic nodes dominated were so highly correlated as to be virtually indistinguishable. Some subsequent studies have shown that structural embedding can have subtle but independent effects (Strunk, 2014; Wasow & Arnold, 2005). Here, though, we follow the majority of current studies and operationalize grammatical weight as phrase length in words.

While some discussions of length effects have focused only on the length of one constituent (e.g., the NP in Heavy NP Shift), Hawkins (1994, 1999) and Wasow (1997) have reported corpus data showing that that the RELATIVE lengths of the two alternating phrases appear to be more predictive of speakers' ordering choices than the length of one phrase alone. Similarly, in their elicited production study of Heavy NP Shift in English, Stallings and MacDonald (2011, p. 184) found that the speakers' choice of ordering was better predicted by the difference in length between the NP and the PP

than by the length of the NP alone. Specifically, they found no difference between the two conditions in which the length of the NP varied, but the length difference between the NP and the PP was the same. In their corpus study of RCE in English, Francis and Michaelis (2014, p. 82) examined four measures of weight: RC length alone, VP length alone, length ratio (VP length divided by RC length), and length difference (RC length minus VP length), finding that length ratio was the most accurate predictor of constituent order according to one statistical measure. It is notable that the same conclusion cannot be drawn for the data from Stallings and MacDonald (2011). If we compare the length ratios of their two conditions where the length difference was five words (2:7 vs. 5:10), the ratio measure would have predicted a higher rate of Heavy NP Shift in the first condition, contrary to fact. By manipulating both RC length and VP length, the current experiments allow us to further examine the predictive accuracy of these various length measurements.

2.2. PHRASE LENGTH AND DISCOURSE FACTORS IN COMBINATION

Current research on constituent order alternations shows that various factors in addition to grammatical weight influence speakers' choice of ordering. Such factors turn out to be somewhat different across languages and constructions, but some examples include information status (Arnold et al., 2000; Bresnan et al., 2007), animacy (Bresnan et al., 2007; Rosenbach, 2005), iconicity of sequence (Diessel, 2008), verb bias (Stallings et al., 1998), and lexical semantic dependency (Lohse et al., 2004; Wiechman & Lohmann, 2013). Most relevant here are previous results for RCE, to which we now turn.

Strunk (2014) used a logistic regression model to investigate thirty-two factors hypothesized to affect writers' choice of RCE or non-RCE structure in the Tübingen Treebank of Written German. Of these, eighteen were significant predictors of RCE. Weight-based factors – RC length and extraposition distance – were shown to be the most important predictors, such that RCE was more frequent as RC length increased and as extraposition distance decreased. Strunk also found independent effects of intervening NPs (Gibson, 1998), such that RCE rarely occurred when one or more NPs (potentially) intervened between antecedent and RC, and of syntactic complexity, such that RCs containing an embedded clause were more likely to be extraposed than other RCs. In addition to these weight-based and complexity-based factors, he found that discourse-related factors, including definiteness of the antecedent NP (RCE was more likely with indefinite NPs) and restrictiveness of the RC (RCE was slightly more likely with restrictive RCs), also affected writers' structural choices.

Francis and Michaelis (2014) used a similar method to Strunk (2014) to examine factors affecting the use of RCE in English. A logistic regression

analysis showed a strong preference for RCE when the RC was at least five times longer than the VP, and a strong preference for non-RCE when the RC was the same length as or shorter than the VP. For those tokens with length ratios falling between these limits, choice of structure depended primarily on definiteness and predicate type. Indefinite subject NPs showed RCE more often than definite subject NPs, while passive and presentative predicates showed RCE more often than other predicate types (where a presentative predicate is an intransitive predicate of appearance or existence, e.g., *come in*, *appear*). Definiteness and predicate type also interacted such that the combination of an indefinite subject NP with a presentative predicate was more likely to occur with RCE than a simple combination of the two factors would predict. Similar effects were found by Walker (2013, p. 161) using acceptability judgment tasks: the biggest boost in acceptability occurred with a presentative predicate in combination with an indefinite NP. (Walker did not manipulate phrase length.) The current study examines the effects of VP length, RC length, and definiteness using two new task types: preference in reading and elicited production.

2.3. ON THE RELATION BETWEEN STRUCTURAL CHOICES, PRODUCTION, AND COMPREHENSION

Hawkins (2004, 2014) proposes that weight effects as shown in corpora are closely related to sentence processing in production and comprehension. For English and other head-initial languages, short-before-long constituent order tends to minimize integration distances for phrasal dependencies, facilitating both comprehension and production (Hawkins, 2014, p. 48). From a more strongly production-based perspective, Arnold et al. (2000) and Wasow (2002) have proposed that postponing longer, more complex phrases allows speakers more time to plan their production of those phrases. In her Production–Distribution–Comprehension (PDC) model, MacDonald (2013) offers similar but more general proposals to account for ordering preferences by means of implicit biases found in utterance planning. According to one such bias, Easy First, speakers produce more accessible words and phrases first, to allow more time for the planning of less accessible words and phrases. This subsumes the short-before-long bias, but also accounts for the tendency to place any kind of information that is highly activated (including animate nouns, frequently occurring words, or recently used words) early in the sentence. For example, one motivation for using passive voice would be to produce an animate theme argument early in the sentence (e.g., *Phil was hit by a baseball*). Another production bias, Plan Reuse, says that speakers can more easily access frequently used syntactic plans (i.e., constructions) from long-term memory. For example, this bias generally favors the more frequent

active voice over the less frequent passive voice, at least in the absence of conflicting pressures (2013, p. 5). MacDonald further argues that the distributional regularities which are shaped (in part) by these production biases aid listeners and readers in comprehension, since listeners are able to predict aspects of the incoming speech signal based on statistical regularities gleaned from past experience, a position similar to that of Levy (2008) and Levy et al. (2012). We follow these authors in claiming that the preferred constituent orderings found in spontaneous speech may benefit both speakers and listeners. Here, we discuss previous experimental studies that bear most directly on the current research.

2.3.1. *Comprehension*

In a study of RCE in German, Konieczny (2000, pp. 638–639) found that, consistent with Hawkins' (2004) locality-based predictions, RCE was judged as most acceptable when the distance between the RC and its antecedent was short (one word) and the RC was long. Similarly, in self-paced reading, Konieczny (2000, p. 641) found that reading times were slower at the relative pronoun when the RC was extraposed, indicating additional processing cost for integrating the displaced RC with its antecedent. Surprisingly, though, there was no reading-time penalty for non-RCE at the clause-final main verb even when the RC was long. Instead, presence of an RC before the main verb actually resulted in faster reading times at the main verb (2000, p. 641). Konieczny interprets this 'antilocality' effect to be a result of expectation-based processing facilitation: the presence of the RC helped readers anticipate the main verb by providing additional information about its object.

Francis (2010) examined RCE in English, using materials similar to those in (1a–b) above. She manipulated RC length and found shorter whole-sentence reading times for RCE sentences as compared with their non-RCE counterparts when the RC was much longer than the VP (15 vs. 5 words). When the RC was shorter than the VP (4 vs. 5 words) or only slightly longer (8 vs. 5 words), there was no significant difference in reading time between RCE and non-RCE variants. Unlike Konieczny's (2000) findings for German, there was no additional processing cost for RCE, but there was in fact a processing cost for non-RCE sentences with a long RC. This can be interpreted as a locality effect, since the long RC increases the distance between the subject NP and main verb in a non-RCE sentence.

In their study of word-by-word reading in English, Levy et al. (2012) found a processing cost for RCE similar to what Konieczny (2000) found for German. Reading times over the first four words of the RC were significantly slower for the RCE sentences (Levy et al., 2012, p. 20). However, there was no difference between RCE and non-RCE variants at the main verb, where

we would expect an advantage for RCE sentences under a locality-based account. Levy et al. interpret their results as favoring an expectation-based model of processing: RCE is difficult when readers are not expecting to encounter an extraposed RC. They further show that the processing cost of RCE can be neutralized when readers are cued to expect an upcoming RC (2012, p. 27). They did not include long RCs among their stimuli, and so it is not clear whether their method would have replicated the facilitation effect for RCE that Francis (2010) found. However, they plausibly interpret Francis' finding not strictly as a locality effect but rather as indicating a comprehension advantage based on the probabilistic expectation that when a reader encounters an extraposed RC, it will be long (Levy et al., 2012, p. 30).

While many questions remain, two main findings emerge from these comprehension studies: (i) in the absence of mitigating factors, RCE incurs a localized processing cost in both English and German; and (ii) long RCs constitute a mitigating factor which appears to give RCE sentences a processing advantage over comparable non-RCE sentences in English, but not in German.

2.3.2. *Production*

Since the current study is the first to use elicited production of RCE, we consider here data from elicited production of two other constituent order alternations in English. Arnold et al. (2000) investigated the role of grammatical weight and discourse status in the structural realization of giving events. Each participant was required to give instructions to a co-participant about giving various objects to various animal characters (*Give the white rabbit the carrot* / *Give the carrot to the white rabbit*). Both the length of the NPs required to describe the objects (theme argument) and the characters (goal argument), and the givenness of the objects and the characters (i.e., whether they were already introduced by the co-participant) were manipulated. The results showed independent effects of weight and givenness: participants tended to produce newer and heavier constituents later in the sentence. Importantly, these experimental results closely mirrored the results from their corpus analysis.

Like Arnold et al. (2000), Stallings et al. (1998) show that speakers' choice of structure in elicited production depends on some of the same factors shown to affect corpus frequencies. Their study, on which the methods for our Experiment 2 are based, examined the effects of grammatical weight and verb bias on speakers' choice of Heavy NP Shift (*Todd delivered to Al the package*) vs. canonical order (*Todd delivered the package to Al*). They used a sentence construction paradigm in which three phrases were presented on a computer screen in varying orders and participants were required to formulate

and speak a sentence based on those phrases. The results showed that speakers used the shifted (NP-last) order more often when the NP was long, and when the verb allowed a clausal complement. They also measured preparation time (time spent reading and preparing the sentence) and voice initiation time (time lag between a visual cue and onset of speech) for speakers' utterances. Preparation time showed a significant effect of NP length, indicating that participants took longer to read and prepare longer sentences. However, initiation time showed no length effect. The authors attribute this finding to the incremental nature of sentence production – in particular, speakers' lack of commitment to a particular production plan beyond the first phrase at the moment of voice initiation (1998, p. 407).[2] Unfortunately, their statistical analyses of voice initiation time and preparation time did not separate responses by choice of constituent order. Thus, it is unclear how NP length and verb bias might have interacted with structural choice. However, they do report the means and standard deviations for voice initiation time broken down by constituent order (1998, p. 404, Table 3). Items with canonical order showed numerically faster initiation times than items with shifted order when the object NP was short, but this difference diminished or disappeared when the NP was long. These mean responses are suggestive that participants may be planning to some extent beyond the first phrase, and that initiation times may be sensitive to the same factors that govern choice of structure.

A stronger case for a connection between structural choice and ease of production is made in Kuperman and Bresnan's (2012) study of word duration in spontaneous speech. In a previous study, Bresnan et al. (2007) had annotated a large set of prepositional (V-NP-PP, *Todd sent the package to Al*) and double-object (V-NP-NP, *Todd sent Al the package*) sentences from the Switchboard corpus of spontaneous spoken US English with respect to variables known to affect structural choice, including the length, definiteness, givenness, person, animacy, pronominality, and number of the recipient, as well as similar properties of the theme. They then used these variables to estimate the probability that a given utterance would occur in the prepositional construction (as opposed to the double-object construction) and assigned each utterance a probability value. Kuperman and Bresnan (2012) collected acoustic data for word duration at different points in each utterance and related it to these probability values. A key finding was that the probability for a particular structural choice affected word duration at the 'choice point'

[2] This interpretation receives support from more recent studies that measured voice initiation times for producing spoken descriptions of moving pictures (Allum & Wheeldon, 2007; Martin et al., 2010; Smith & Wheeldon, 1999; Wheeldon et al., 2013). Such studies have shown that the length of the first phrase affects initiation time, but that the length of subsequent phrases does not.

where speakers definitely committed to a particular structure (i.e., first word of the first object). For example, if someone used the prepositional construction in a context where the double-object construction would more typically be used, word duration was longer at the first word of the theme argument as compared with prepositional tokens used in a context where the prepositional construction was more typical. The authors found similar effects at the verb (i.e., before the choice point), but no effects after the choice point. They interpret these findings to mean that in both the planning of an upcoming syntactic choice (at the verb) and in the initial stages of its production (at the first object), speech requires more effort when the structure chosen is less probable (2012, p. 603).

The current study goes beyond what has been done in previous elicited production studies by directly comparing the two structures of interest with respect to two measures of processing ease – preparation time and voice initiation time – under varying sentential contexts. Similar to Kuperman and Bresnan's (2012) findings for word duration in spontaneous speech, the current results show that speakers can more easily produce a particular structure in contexts where that structure is more frequently used and encountered. The methods, hypotheses, and results of our two experiments are presented in the following sections.

3. Experiment 1: preference in reading

In this forced-choice task, participants were presented with pairs of grammatically acceptable sentences differing only in constituent order and asked to choose which sentence of each pair sounded more natural.

3.1. MATERIALS

The sentence materials incorporated three independent variables in a repeated measures design: (i) definiteness of the subject NP (*the* vs. *some*), (ii) RC length (5 words vs. 12 words), and (iii) VP length (2 words vs. 5 words). All combinations of these factors resulted in eight experimental conditions, which were repeated across eight lexical sets (token sets), for a total of sixty-four experimental items. Each item consisted of a pair of sentences which differed only in constituent order (RCE and non-RCE), with the dependent variable being the participant's choice of order. To ensure the acceptability of both RCE and non-RCE variants, passive predicates were used in all of the experimental stimuli. The following passive verbs were used as the main verb for each set: *conducted*, *raised*, *formed*, *provided*, *considered*, *presented*, *received*, and *made*. See 'Appendix A' for a complete list of sentence materials for the experimental conditions. Ninety-six filler sentences containing several

constituent order alternations other than RCE were used to distract participants from the structure being tested.

3.2. PARTICIPANTS

Forty native speakers of American English (29 female, 11 male), aged 18–53 (mean age 22) were recruited from the Purdue University community to participate in Experiment 1. Each was paid $8 for a session of about 35 minutes.

3.3. PROCEDURES

This experiment measured structural preference in reading, following a procedure similar to that in Rosenbach (2005). Participants were presented with a written questionnaire that contained all of the experimental and filler sentences (160 sentence pairs). For each item on the questionnaire, they were asked to choose which of two versions a sentence sounded more natural (as in (2a–b) and (3a–b) below). Items were arranged in blocks to avoid similar sentences occurring together, and the order of items within each block was randomized. Eight versions of the questionnaire were created using four different orderings of items. Top-bottom ordering of (a–b) options (as in (3a–b) below) was counterbalanced across items and across participants. Responses were later coded as 1 (RCE order) or 0 (non-RCE order) for statistical analysis. Since all of the participants saw all of the items, there is the possibility of a repeated exposure effect, especially for items within the same lexical set. To address this, responses were also coded for the relative order (1–8) in which the members of the same lexical set occurred within the questionnaire.[3] We will refer to this factor as within-set order.

3.4. HYPOTHESES

Based on the corpus results from Francis and Michaelis (2014), specific predictions for the statistical analysis are as follows:

i. Main effect of definiteness: indefinite subject NPs will induce a higher rate of RCE responses than definite subject NPs.
ii. Main effect of VP length: short VPs will induce a higher rate of RCE responses than long VPs.
iii. Main effect of RC length: long RCs will induce a higher rate of RCE responses than short RCs.

[3] We are grateful to an anonymous reviewer for drawing our attention to this factor.

Sample sentences for the two most extreme conditions – those that are predicted to induce the most and the least RCE – are illustrated below in (2) and (3). In (2), sentence (a) should be preferred most often, while in (3), sentence (b) should be preferred most often.[4]

(2) *Indefinite subject NP, short VP, long RC*
   a. Some research was conducted that refutes the existing theories with very clear and convincing new evidence. (RCE)
   b. Some research that refutes the existing theories with very clear and convincing new evidence was conducted. (non-RCE)

(3) *Definite subject NP, long VP, short RC*
   a. The research has been conducted fairly recently that refutes the existing theories. (RCE)
   b. The research that refutes the existing theories has been conducted fairly recently. (non-RCE)

The predictions for relative length depend on how it is measured. The current design, as shown in Table 1, allows us to test which measure of relative length makes more accurate predictions. The two methods – length difference and length ratio – make the same predictions with respect to the shortVP–longRC and longVP–shortRC conditions (top and bottom in Table 1). They differ with respect to the longVP–longRC and shortVP–shortRC conditions. The hypotheses for relative length are as follows:

iv. Ratio method: there should be no difference between longVP–longRC and shortVP–shortRC conditions.
v. Difference method: RCE should be preferred more often in the longVP–longRC condition as compared with the shortVP–shortRC condition.

3.5. RESULTS

Overall, RCE was chosen in 54.3% (1390/2560) of responses. Descriptive statistics are summarized in 'Appendix B', Table 2. Using a mixed logit model with participant specified as a random variable, significant main effects were found for RC length ($F(1,2507) = 40.00$, $p < .001$), definiteness ($F(1,2507) = 31.14$, $p < .001$), and VP length ($F(1,2507) = 5.13$, $p = .02$)

[4] An anonymous reviewer points out that the length manipulation of the VP is confounded with a tense/aspect alternation, since the short VPs were always in the simple past tense, and the long VPs were always in the present perfect. We have no way of resolving this confound in the current study, but note that robust effects of VP length were shown in the corpus data across VPs of varying tense and aspect (Francis & Michaelis, 2014). We know of no previous studies that have identified tense or aspect as a factor in RCE. However, this issue is yet to be investigated systematically.

TABLE 1. *Relative length calculations for experimental conditions*

Length condition	VP length	RC length	Length ratio	Length difference
shortVP–longRC	2	12	0.16	10
longVP–longRC	5	12	0.41	7
shortVP–shortRC	2	5	0.40	3
longVP–shortRC	5	5	1.00	0

(Figures 1–2). All of these effects were in the expected direction: RCE was preferred most often (71.9%, 230/320) with short VP, long RC, and indefinite subject, and least often with a long VP, short RC, and definite subject (34.1%, 109/320), and there were no significant interactions, thus confirming hypotheses i–iii (Figure 3).

Two of the four length conditions were compared directly to address the different predictions of the ratio and difference methods for computing relative length. As shown in Figure 2, RCE was preferred more often in the longVP–longRC condition (59.2%, 379/640) as compared with the shortVP–shortRC condition (52.0%, 333/640), and a chi-squared test showed that this difference was significant (χ^2 (1, N = 1280) = 6.69, $p < .01$). Similar to the result of Stallings and MacDonald (2011), this result is most consistent with the predictions from the difference method, as stated in hypothesis v. The predictions of the ratio method, as stated in hypothesis iv, receive less support.

Top–bottom ordering of the (a–b) options was not significant (F(1,2507) = 0.35, p = .55). However, there was a near-significant effect of within-set order (F(1,2507) = 3.56, p = .06). In the indefinite condition, RCE was chosen more often when the item came earlier in its set, and less often when the item came later in its set. No clear trend was shown in the definite condition (Figure 4). (We will see that Experiment 2 shows the same trend, but with clear statistical significance.) There was also an unexpected significant main effect of lexical set (F(7,2507) = 6.71, $p < .01$). RCE occurred less often for sentences containing the verbs *provided* (41.9%, 134/320) and *considered* (46.9%, 150/320), and more often for sentences containing the verb *conducted* (61.6%, 197/320), with the other five lexical sets falling between 55.3% (177/320) and 59.4% (190/320).[5]

[5] Here, we identify lexical sets by their main verb because we believe that the verb is more likely than other elements to have affected RCE. Previous corpus results have shown that predicate type is an important factor in the distribution of RCE (Francis, 2010; Francis & Michaelis, 2014), while no previous works have identified other types of lexical content words (nouns, adjectives, or adverbs) as a relevant factor.

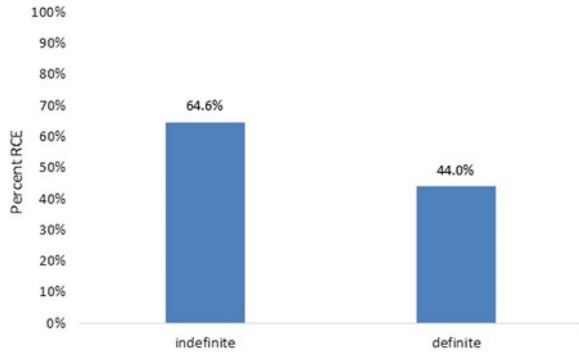

Fig. 1. Choice of structure by definiteness (preference task).

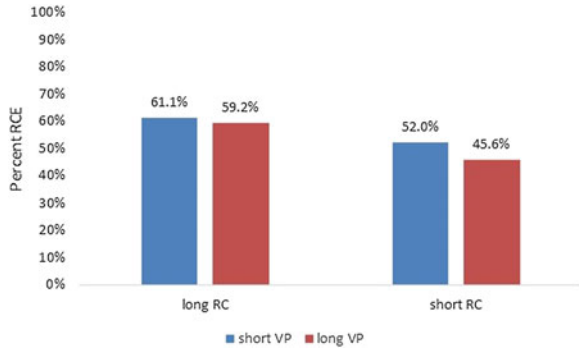

Fig. 2. Choice of structure by VP length and RC length (preference task).

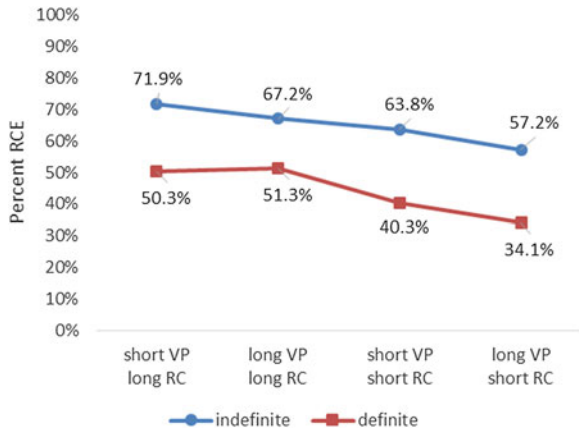

Fig. 3. Choice of structure by VP length, RC length, and definiteness (preference task).

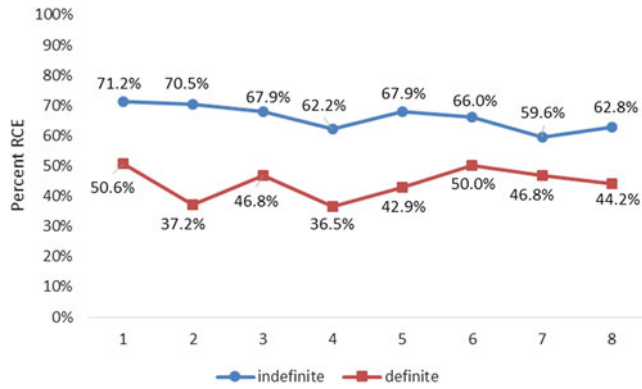

Fig. 4. Choice of structure by within-set order (1–8) and definiteness (preference task).

4. Experiment 2: preference in spoken production

Experiment 2 measured structural preference in spoken production, following the method used in Stallings et al. (1998). The logic is similar to that of Experiment 1: speakers' preferences are reflected in their choice of constituent order. However, unlike the forced-choice task, the elicited production task involved no direct comparison between the two constituent orders. Rather, participants were presented with three phrases on the screen and asked to combine them in whatever way seemed most natural, and then speak the sentence.

4.1. MATERIALS

The sixty-four experimental items were identical to those in Experiment 1. To reduce the length of the experimental sessions, only a subset of the filler items (64 out of 96) was used.

4.2. PARTICIPANTS

Forty native speakers of American English (21 female, 19 male), aged 18–57 (mean age 24) were recruited from the Purdue University community to participate in Experiment 2. Each was paid $10 for a session of about 45 minutes.

4.3. PROCEDURES

Similar to Experiment 1, each participant encountered all of the experimental and filler trials (n = 128). Participants saw sentence constituents arranged on

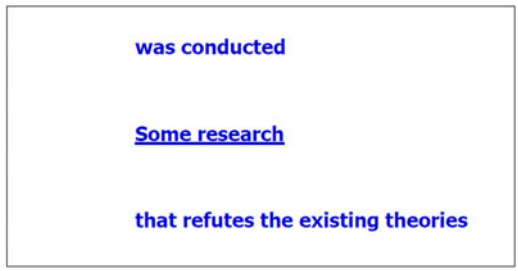

Fig. 5. Elicited production stimulus sample.

the computer screen from top to bottom, and were instructed to formulate and speak a sentence in one of two orders, depending on their personal preference: middle–top–bottom, or middle–bottom–top (Figure 5). When participants had formulated their sentence and were ready to respond, they would press a button, causing the words to disappear from the screen, and wait for a visual cue (the screen changing color) before speaking the sentence. At the same time the visual cue occurred, the words reappeared on the screen. Thus, the participants did not need to completely memorize each sentence before beginning to speak.

Similar to Experiment 1, items were arranged in blocks. The order of items within each block and the order of blocks were randomized separately for each participant by the E-Prime program. The top–bottom ordering of the constituents (i.e., VP top, RC bottom or vice versa) was counterbalanced across items and participants.

Responses were later coded for constituent order (RCE = 1; non-RCE = 0), within-set order, preparation time, and voice initiation time. As in Experiment 1, within-set order was defined as the relative ordering (1–8) of items within the same lexical set. Preparation time was defined as the amount of time participants spent viewing the phrases on the screen prior to pressing a button indicating readiness to respond. Voice initiation time was defined as the amount of time lag between the visual cue (screen changing color) and the onset of speech. Responses were also coded for any deviations from the stimulus sentences. Because the words were available on the screen, the responses contained only a few minor deviations. In 1.4% of trials (35/2496), the participant omitted 1–3 words that were present in the stimulus items, and in 0.8% of trials (21/2496), the participant added an extra 1–2 words. Following Stallings et al. (1998, p. 408), no trials were excluded from the analysis on this basis. However, five trials, or 0.2%, were excluded because one or more phrases was missing from the response. In addition, 2.5% of trials were excluded from the analysis of preparation time because the value was more than three standard deviations above or below the mean for a particular participant's responses.

4.4. HYPOTHESES

The hypotheses related to choice of constituent order are the same as for Experiment 1, as detailed above in hypotheses i–v. Because this experiment included two additional dependent variables, preparation time and voice initiation time, additional hypotheses are needed. In the analysis of these two measures, structure (RCE or non-RCE) was treated as an independent variable rather than as a dependent variable. This is because choice of structure potentially influenced preparation times and initiation times. For these analyses, there were four factors instead of three: (i) definiteness of the subject NP, (ii) RC length, (iii) VP length, and (iv) structure, resulting in a total of sixteen conditions. Although speakers used RCE less often than they used the non-RCE structure (920 tokens as compared with 1572), enough tokens were produced in each condition (at least 52) so that it was possible to run statistical analyses using structure as a factor.

Given that only one previous study of a syntactic alternation (Stallings et al., 1998) measured voice initiation time and preparation time, and did so without including structure as a factor, our specific hypotheses, as detailed below, are necessarily more tentative. As noted above in Section 2.3.2, Stallings et al. found a general length effect on preparation time: longer sentences took longer to prepare. We expect that, in the current study, participants should also take longer to prepare longer sentences. Unlike preparation time, initiation time showed no length effects for Stallings et al., which they interpret as being due to speakers' lack of commitment to a production plan beyond the first phrase (1998, p. 407). Because their study only manipulated the length of the object NP, the first phrase (the subject NP) never varied in length. In the current study, however, the length of the first phrase (the subject NP) varied depending on the participant's choice of structure. The non-RCE variant includes an RC within the subject NP, but the RCE variant does not. Therefore, if there is a first phrase effect, RCE sentences should be initiated faster than non-RCE sentences. Finally, based on previous results for word duration in spontaneous speech (Kuperman & Bresnan 2012), we predict that those conditions that generally facilitate use of RCE (short VP, long RC, indefinite NP) should result in shorter preparation times and initiation times for RCE sentences as compared with the opposite conditions (long VP, short RC, definite NP). Given these considerations, our specific hypotheses are as follows:

Preparation time:

vi. Main effects of VP length and RC length: preparation times will be faster overall for sentences with short VPs as compared with long VPs, and for sentences with short RCs as compared with long RCs, due to the longer time spent reading and preparing longer sentences.

vii. Interaction between structure and RC length: due to the main effect of RC length, preparation times should be longer overall for long RC conditions. However, this effect should be mitigated for RCE sentences, since long RCs are preferred for this structure.
viii. Interaction between structure and VP length: due to the main effect of VP length, preparation times should be longer overall for long VP conditions. However, this effect should be greater for RCE sentences, since short VPs are preferred for this structure.
ix. Interaction between structure and definiteness: for RCE sentences, preparation times should be faster when the subject NP is indefinite than when the subject NP is definite.

Voice initiation time:

x. Main effect of structure: voice initiation times will be faster overall for RCE sentences than for non-RCE sentences, due to the presence of the RC within the subject NP of a non-RCE sentence and the expected effect of first-phrase length on initiation time.
xi. Interaction between structure and RC length: for RCE sentences, voice initiation times will be faster for long RC than for short RC conditions, since long RCs are preferred for this structure.
xii. Interaction between structure and VP length: for RCE sentences, voice initiation times will be faster for short VP than for long VP conditions, since short VPs are preferred for this structure.
xiii. Interaction between structure and definiteness: for RCE sentences, voice initiation times should be faster when the subject NP is indefinite than when the subject NP is definite.

4.5. RESULTS

4.5.1. *Choice of structure*

As shown in Figures 6–8, the results for Experiment 2 were similar but not identical to the results of Experiment 1. Overall, RCE was used in 36.9% (921/2496) of responses, which was a lower rate than for the preference task in Experiment 1 and more in line with the corpus data reported in Francis and Michaelis (2014). As in Experiment 1, RCE was preferred most often (61.9%, 193/312) with short VP, long RC, and indefinite subject, and least often with a long VP, short RC, and definite subject (16.7%, 52/312) (Figure 8). Descriptive statistics are summarized in 'Appendix B', Table 3. There were again significant main effects for RC length ($F(1,2441) = 6.11$, $p = .01$), definiteness ($F(1,2441) = 81.80$, $p < .001$), and VP length ($F(1,2441) = 65.21$, $p < .001$) (Figures 6–7). All of these effects were in the expected direction: RCE occurred more often with indefinite NPs, short VPs, and long RCs, as predicted in hypotheses i–iii.

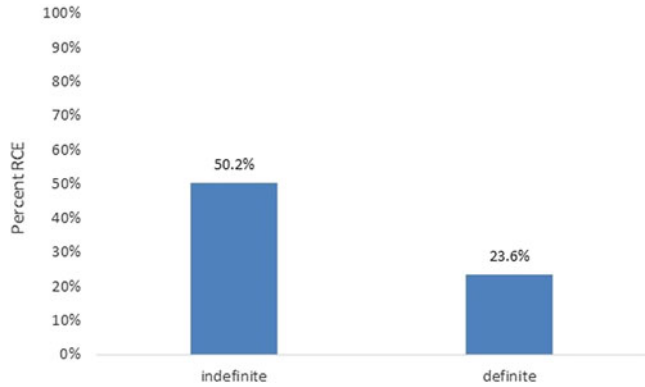

Fig. 6. Choice of structure by definiteness (elicited production).

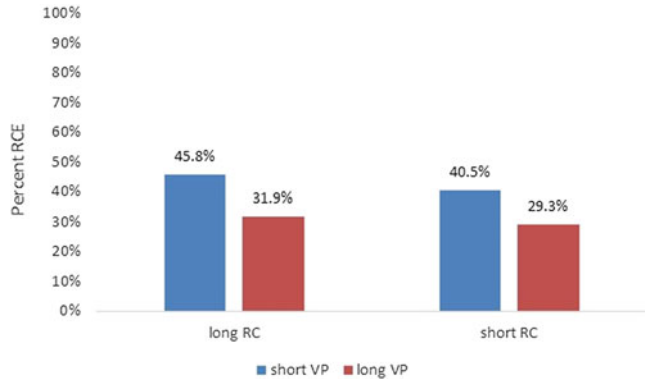

Fig. 7. Choice of structure by VP length and RC length (elicited production).

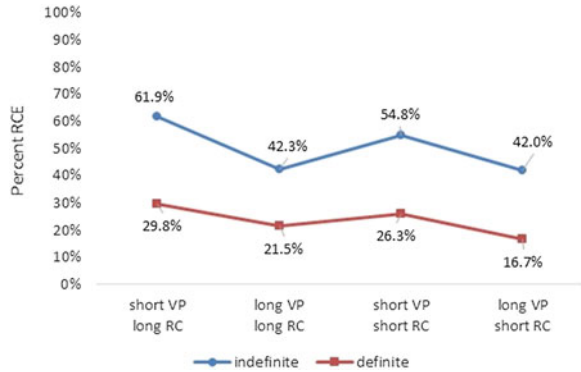

Fig. 8. Choice of structure by VP length, RC length, and definiteness (elicited production).

Relative length effects were again assessed using pairwise comparisons. As shown in Figure 7, the effects were in the opposite direction from Experiment 1. RCE was preferred LESS often in the longVP–longRC condition (31.9%, 199/624) as compared with the shortVP–shortRC condition (40.5%, 253/624), and a chi-squared test showed that this difference was significant (χ^2 (1, N = 1248) = 10.11, $p < .01$). Neither the ratio method nor the difference method predicts a difference in this direction, although the ratio method comes closer by predicting no difference. Thus, there was no clear support for either hypothesis iv or hypothesis v. It appears that VP length exerted a greater influence over choice of constituent order than RC length did, whereas Experiment 1 showed the opposite pattern.

Similar to Experiment 1, top–bottom ordering of the constituents on the screen was not significant (F(1,2441) = 0.06, p = .80), and there were no significant interactions among (various combinations of) VP length, RC length, and definiteness. There was, however, a significant effect of within-set order (F(1,2441) = 4.07, p = .04), and this time also a significant interaction between within-set order and definiteness (F(1,2441) = 3.82, p = .05). Figure 9 shows the same basic trend as in Figure 4 from Experiment 1, but with a clearer interaction between within-set order and definiteness. In the indefinite condition only, RCE was chosen more often when the item came within the first three trials in its set, and less often when the item came later in its set, while no clear trend was shown in the definite condition. Also as in Experiment 1, there was a significant main effect of lexical set (F(7,2441) = 5.79, $p < .001$). RCE occurred less often for sentences containing the verbs *provided* (29.5%, 92/312) and *considered* (29.5%, 92/312), and more often for sentences containing the verbs *conducted* (43.6%, 136/312) and *presented* (41.7%, 130/312), with the other four lexical sets falling between 34.6% (108/312) and 40.4% (126/312).

4.5.2. *Preparation time*

Analyses of preparation time and initiation time were conducted using a linear mixed model, with participant specified as a random factor. In accordance with hypothesis vi, participants took longer to prepare sentences with long RCs and long VPs as compared with short RCs and short VPs, indicating that participants took longer to read and prepare longer sentences (Figure 10).[6] These differences resulted in significant main effects of VP length (F(1,2415) = 35.71, $p < .001$) and RC length (F(1,2415) = 22.00, $p < .001$). Unexpectedly, participants took longer to prepare RCE responses (M = 5073.90, SD = 3838.65) than non-RCE responses (M = 4786.38, SD = 3772.95), and this difference was significant (F(1,2415) = 9.89, p = .002). There were no significant differences in preparation

[6] Error bars in Figure 10 and throughout represent standard error of the mean.

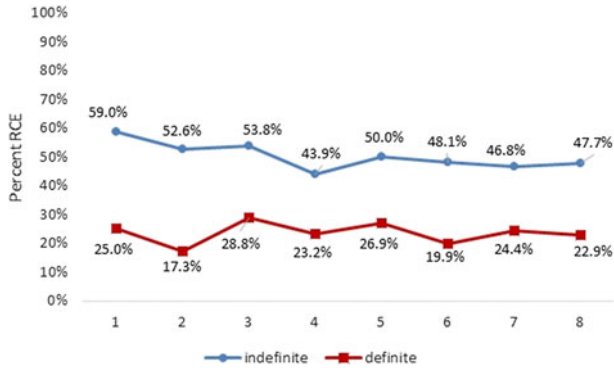

Fig. 9. Choice of structure by within-set order (1–8) and definiteness (elicited production).

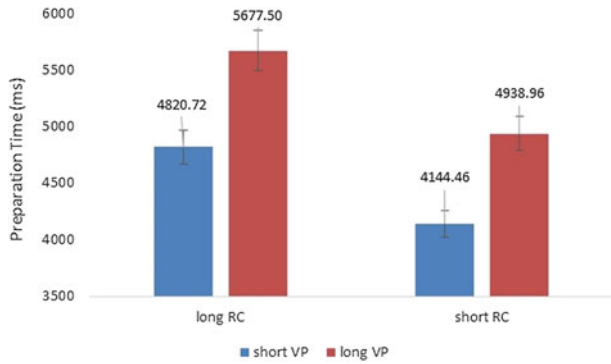

Fig. 10. Preparation time by VP length and RC length.

time due to definiteness ($F(1,2415) = 2.97$, $p = .09$), lexical set ($F(7,2415) = 0.68$, $p = .69$), or top–bottom ordering of phrases ($F(1,2413) = 2.51$, $p = .11$).

Results failed to support the predicted interactions in hypotheses vii–ix. However, trends in the expected direction were shown for the interaction of VP length and structure ($F(1,2415) = 2.03$, $p = .15$), as shown in Figure 11, and for the interaction of definiteness and structure ($F(1,2415) = 3.07$, $p = .08$), as shown in Figure 12. Descriptive statistics for these interactions are summarized in 'Appendix B', Table 4. Although preparation times for long VP conditions were always faster than for short VP conditions, RCE sentences appeared to show a stronger penalty for long VPs than non-RCE sentences did (Figure 11). This trend is confirmed using pairwise t-tests. Preparation times for RCE responses were significantly slower than for non-RCE responses in the long VP condition ($t = 4.08$, $p < .01$), but there was no difference between RCE and non-RCE responses in the short VP condition ($t = 0.52$, $p = .60$).

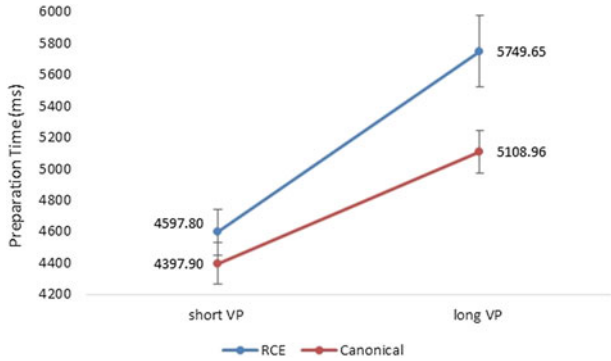

Fig. 11. Preparation time by structure and VP length.

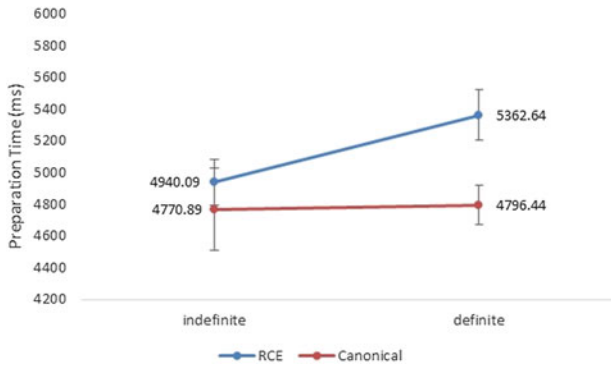

Fig. 12. Preparation time by structure and definiteness.

Similarly, RCE sentences appeared to show a penalty for definite NPs which was not shown by RCE sentences (Figure 12). However, in this case, pairwise comparisons revealed no significant difference between RCE and non-RCE responses, either in the definite condition ($t = 1.55$, $p = .12$) or in the indefinite condition ($t = 0.86$, $p = .39$).

Preparation time showed a significant effect of within-set order ($F(1,2415) = 93.35$, $p < .001$). There was no interaction with definiteness (or any other factor) in this case, and the trend was straightforward: participants took longer to prepare responses that occurred earlier in the lexical set than to prepare responses that occurred later in the same set (Figure 13).

4.5.3. *Voice initiation time*

Initiation times were numerically faster for RCE sentences ($M = 809.56$, $SD = 451.37$) than for non-RCE sentences ($M = 822.70$, $SD = 354.91$).

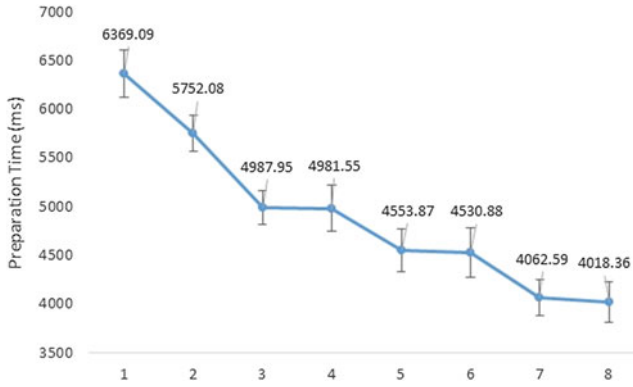

Fig. 13. Preparation time by within set-order (1–8).

However, contrary to our hypothesis x, this difference was not significant ($F(1,2472) = 0.03$, $p = .87$). There were no main effects for lexical set ($F(7,2472) = 1.11$, $p = .35$), or for top–bottom ordering of phrases ($F(1,2472) = 1.02$, $p = .31$). Nor were there any significant differences due to RC length ($F(1,2472) = 1.27$, $p = .26$) or VP length ($F(1,2472) = 0.88$, $p = .35$). Unexpectedly, initiation times were overall faster when the subject NP was indefinite ($M = 795.94$, $SD = 434.48$) than when the subject NP was definite ($M = 839.78$, $SD = 346.03$), and this difference was significant ($F(1,2472) = 9.64$, $p = .002$). However, as shown in Figure 15, this effect was driven primarily by the RCE condition.

Contrary to hypothesis xii, there was no significant interaction between VP length and structure ($F(1,2472) = 0.67$, $p = .41$). However, the other two expected interactions were found. Descriptive statistics for these interactions are summarized in 'Appendix B', Table 5. As predicted in hypothesis xi, there was a significant interaction between RC length and structure ($F(1,2472) = 6.03$, $p = .01$) (Figure 14). Pairwise t-tests show that, in the non-RCE conditions, short RC responses were initiated faster than long RC responses ($t = 3.42$, $p < .01$). Although there appears to be an advantage for RCE over non-RCE in the long RC condition, this difference was not significant ($t = 1.67$, $p = .09$), nor was there a significant advantage for non-RCE sentences in the short RC condition ($t = 0.69$, $p = .40$). A plausible interpretation of this interaction, then, is that for non-RCE sentences, there is a penalty for long RCs, whereas for RCE sentences, there is no such penalty.

As predicted in hypothesis xiii, a significant interaction between definiteness and structure was also found ($F(1,2472) = 3.93$, $p = .047$) (Figure 15). Pairwise comparisons show that, for RCE sentences, responses with indefinite NPs

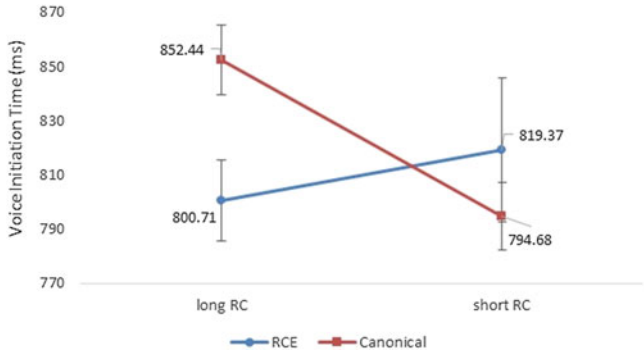

Fig. 14. Initiation time by structure and RC length.

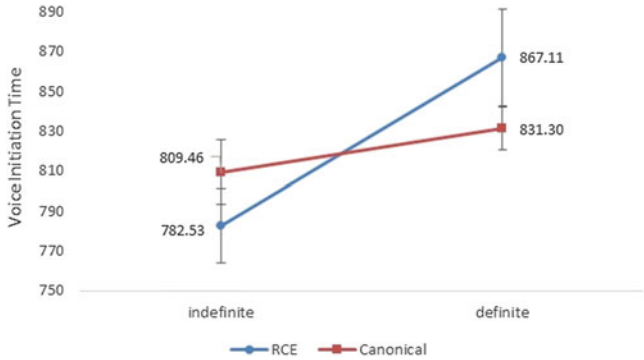

Fig. 15. Initiation time by structure and definiteness.

did not differ significantly from responses with definite NPs, although the difference did approach significance (t = 1.82, p = .07). Within the definite condition, non-RCE sentences were initiated significantly faster than RCE sentences (t = 2.19, p = .03). These results are most consistent with an interpretation in which RCE sentences show a penalty for definiteness which non-RCE sentences do not show.

The results for initiation time showed a significant effect of within-set order (F(1,2472) = 8.51, p = .004). The overall trend was similar to the one shown for preparation time: participants took longer to initiate responses that occurred earlier in the lexical set than to initiate responses that occurred later in the set (Figure 16). However, contrary to the general trend, the initiation time for items in the second position was faster than for items in the third or fourth positions.

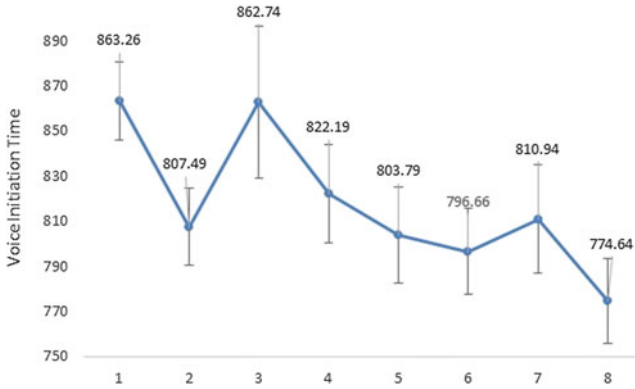

Fig. 16. Initiation time by within set-order (1–8).

5. Discussion

Two experiments established independent effects of constituent length and definiteness on speakers' choice of RCE vs. non-RCE structure. For both experiments, RCE was used most often with an indefinite subject NP, short VP, and long RC, and least often with a definite subject NP, long VP, and short RC. In addition, both experiments showed independent main effects for all three factors. These results provide direct evidence for the importance of phrase length and definiteness in determining when speakers choose RCE.

It is notable that in both tasks, definiteness exerted a stronger effect on RCE use than did phrase length.[7] In Experiment 1, participants chose RCE for 64.6% of indefinite stimuli but only 44.0% of definite stimuli (Figure 1), for a difference of 20.6%, and in Experiment 2, participants used RCE for 50.2% of indefinite stimuli but only 23.6% of definite stimuli, for a difference of 26.6% (Figure 6). By contrast, the difference between the two opposite length conditions (shortVP–longRC and longVP–shortRC) was only 15.5% for Experiment 1 (Figure 2) and 16.5% for Experiment 2 (Figure 6). This is unexpected based on the results of our earlier corpus study, which showed a relatively stronger effect of length ratio as compared with definiteness (Francis & Michaelis, 2014, p. 82). We suspect that the limited number of length values (four combinations of VP and RC length), may have contributed to a somewhat smaller (but still highly significant) length effect. The current stimuli differed from the corpus sample in that RC length was never less than VP length. More puzzling is the question of why definiteness showed such a strong effect even in the absence of any supporting discourse context. One possibility is that the subject NP is specified as indefinite in the representation

[7] We are grateful to an anonymous reviewer for this observation.

of the RCE construction, as in exemplar-based models of syntax, where rules of grammar are statistical catalogs of language experiences (Bod, 2006). This conjecture is further supported by the fact that we also found a definiteness effect on initiation time, as elaborated below in relation to MacDonald's (2013) principle of Plan Reuse.

The consistency found across task types is also notable. Experiment 1 was purely a receptive task, while Experiment 2 also involved production, and both experiments were consistent with previous corpus data showing similar effects of VP length, RC length, and definiteness (Francis & Michaelis, 2014). This consistency across different measurements suggests that comprehension and production are sensitive to the same types of linguistic knowledge. Interestingly, though, the RELATIVE effects of RC length and VP length differed between the two tasks. Our results for Experiment 1 were consistent with the hypothesis that relative length, as defined by length difference in words, can predict choice of constituent order. Specifically, RCE was preferred more often in the longVP–longRC condition (7-word difference) as compared with the shortVP–shortRC condition (3-word difference). The results for Experiment 2 did not, however, show the expected effects of relative length. RCE was used LESS often in the longVP–longRC condition as compared with the shortVP–shortRC condition, contrary to the predictions based on length difference (which predicted the opposite effect) and length ratio (which predicted no difference). This result reflects the stronger effect of VP length in Experiment 2 as compared with Experiment 1.

The different length effects in the two experiments appear to be consistent with differences found in previous results for corpus frequencies as against reading time in Francis (2010). The stronger effect of VP length in Experiment 2 is consistent with corpus data findings showing that while both VP length and RC length correlated with RCE use, VP length was the stronger predictor (2010, p. 62). By contrast, the stronger effect of RC length in Experiment 1 is consistent with the previous result for whole-sentence reading time, which showed a significant reading time advantage for RCE sentences over non-RCE sentences when the RC was long (2010,p. 51).[8] It seems plausible, therefore, that Experiment 1, which required participants to read and compare sentences, might have tapped into comprehension mechanisms to a greater

[8] Francis (2010) did not include VP length as a factor, and so the potential effect of VP length on reading time is not known. It is also interesting that Francis did not find any advantage for RCE sentences in an acceptability judgment task. In the long RC condition, where RCE sentences show an advantage in reading time, there was no difference in acceptability, while in the short RC condition, non-RCE sentences were judged as higher in acceptability than RCE sentences. While both tasks from Francis showed effects of RC length, it appears that the forced-choice task in Experiment 1 aligns more closely with reading time than with acceptability.

extent than Experiment 2. Conversely, Experiment 2, a production task requiring a spoken response, showed a closer alignment with the length effects shown in spontaneous production in the corpus. Further research is needed to understand why VP length may be a more important influence in production as compared with comprehension (cf. Menn & Duffield, 2014).

Both experiments showed significant effects of lexical set, and these effects were consistent across the two tasks. For Experiment 1, the rank order of lexical sets from highest to lowest rate of RCE was: *conduct* (61.6%), *make* (59.4%), *raise* (56.5%), *receive* (57.2%), *form* (56.3%), *present* (55.3%), *consider* (46.9%), *provide* (41.9%). For Experiment 2, the rank order was the same with the exception of one set (*present*): *conduct* (43.6%), *present* (41.7%), *make* (40.4%), *raise* (37.8%), *receive* (38.1%), *form* (34.6%), *consider* (29.5%), *provide* (29.5%). Although we used passive forms of transitive verbs in all of the lexical sets, we did not consider that some transitive verbs are more strongly biased toward passive than others. It is therefore possible that participants used RCE more often when the verb was more strongly biased toward passive use. A search for our eight verbs in the Corpus of Contemporary American English (Davies, 2009) provides some support for this idea. Consistent with both experiments, the verb with the strongest passive bias was *conduct*, with 23.6% of occurrences (9547/40407) in passive voice, and the verb with the second weakest passive bias was *provide*, with only 4.5% of occurrences (7844/172843) in passive voice. However, for the other six verbs, there was no obvious correspondence between the corpus frequencies and the rate of RCE shown in our experiments. For example, although *make* occurred frequently with RCE in our experiments, the corpus shows its rate of passive at only 6.0%. In terms of raw frequencies, however, *make* has six times as many passive tokens as the other verbs (on average), due to its higher overall frequency. These corpus data suggest that verb bias may have had some effect on RCE use. However, an additional study manipulating verb bias in stimulus items matched for overall frequency would be needed to draw any firm conclusions.

The two experiments showed similar trends for within-set order – the order in which stimulus items occurred during the task relative to other items in the same lexical set. (Recall that participants never encountered the exact same sentence twice, and items from the same lexical set were separated by items from other sets and by filler sentences.) When the subject NP was indefinite, participants were more likely to choose RCE on first exposure to an item from a particular set, and less likely to choose RCE following repeated exposure to lexically similar items. It is unclear why the repeated exposure effect only showed up with indefinites, or why incidence of RCE decreased rather than increased with repeated exposure. Possibly, the RCE variant competes most successfully with the non-RCE variant under two conditions: (i) when linguistic features are compatible with the usual discourse function

of RCE (as when the subject NP is indefinite), and (ii) when the non-RCE variant is relatively difficult to process (as when there is no prior exposure to the lexical content or when there is a short VP and a long RC). Further research is needed to better interpret this repeated exposure effect.

In addition to examining choice of structure, Experiment 2 also measured preparation time and voice initiation time. Based on Stallings et al. (1998), we hypothesized that preparation times would be subject to general length effects, and this hypothesis was confirmed. For both RC length and VP length, preparation times were longer for longer phrases. One puzzling result was the main effect of structure: non-RCE sentences were prepared faster than RCE sentences. Possibly, this was due to the general infrequency of the RCE construction as found in corpora (Francis, 2010) and as shown in the current results for choice of structure. In Experiment 2, 37% of trials showed RCE while 63% showed non-RCE. This effect of structure on preparation time is also consistent with the idea of RCE as a marked construction with added syntactic complexity.

We hypothesized that, unlike preparation times, initiation times would be primarily sensitive to the length of the first phrase, and that this first-phrase effect should show up in longer initiation times for non-RCE sentences as compared with RCE sentences. However, results for initiation time showed no overall difference between the two structures. Possibly, there could have been a first-phrase effect on initiation time which was neutralized by a general markedness effect (of the same kind that showed up in slower preparation times for RCE) working in the opposite direction.

Both preparation time and initiation time showed effects of within-set order such that participants were slower to prepare and initiate items which occurred earlier in a lexical set, and faster to prepare and initiate items that occurred later within a lexical set. These results suggest that prior activation of sentences with similar lexical and structural content facilitates sentence production. (Recall that all stimulus items except the fillers were structurally similar in containing a subject NP, a VP with a passive verb, and a relative clause, while only items from the same lexical set were similar in lexical content.) Such findings are consistent with similar findings from structural priming, which showed an increase in structurally matching responses (Cleland & Pickering, 2003; Pickering & Branigan, 1998) as well as faster response times (Corley & Scheepers, 2002) when participants repeated both structural and lexical content (in particular, the verb) from a recently encountered prime sentence.

Our most important findings shed light on the relationship between choice of structure and ease of production. We predicted that the factors which influenced choice of structure – definiteness, VP length, and RC length – should similarly affect preparation times and initiation times, providing evidence

that choice of structure is closely related to processing efficiency (Hawkins, 2004; Kuperman & Bresnan, 2012; MacDonald, 2013; Wasow, 2002). Preparation times for RCE responses were in fact significantly slower than for non-RCE responses when the VP was long, whereas there was no difference between RCE and non-RCE when the VP was short. Consistent with the fact that RCE is used less often with a long VP, RCE sentences with long VPs took longer to prepare. However, the predicted interactions of structure with RC length and definiteness were not found.

The results for voice initiation time showed the expected interactions more clearly. Non-RCE sentences showed faster initiation times for short RCs as compared with long RCs, while RCE sentences did not show this difference. Consistent with the fact that non-RCE constituent order is used more often with a short RC as compared with a long RC, non-RCE sentences were easier to produce when the RC was short. While this result might be due to a first-phrase effect on initiation time, it also fits nicely with the idea put forward by Arnold et al. (2000) and Wasow (2002) that speakers tend to avoid heavy constituents near the beginning of a sentence because such constituents are taxing to the production system. Equivalently, using a non-RCE structure with a long RC violates MacDonald's (2013) Easy First principle, since longer phrases are generally less accessible to the production system than shorter phrases.

Perhaps the most striking result was the interaction between structure and definiteness. Consistent with the fact that RCE is used more often with an indefinite subject NP, participants were faster to initiate an RCE sentence with an indefinite subject NP. Unlike in the case of RC length, however, this result had nothing to do with the short-before-long bias in sentence production. The difference between definite and indefinite NPs in our stimuli was represented only by choice of determiner (*the* vs. *some*). These results therefore show that the close relationship between choice of structure and ease of production is not limited to weight-based effects, but may apply more generally. How can this generalization best be formulated? MacDonald's (2013) principle of Easy First does not account for this particular effect, since there was no prior context given to enhance the accessibility of the first constituent. Rather, the determiner itself (*the* or *some)* seems to be the determining context. Another of MacDonald's proposed production biases – Plan Reuse – appears to be at work here. The idea is that "partially lexically-independent abstract plans" are stored in long-term memory. To the extent that such plans are frequently used, they are easier to retrieve from memory and therefore easier to implement in production (2013, p. 4). If one such plan is a strong (but violable) association between RCE structure and indefiniteness, then the effect we found is predicted.

Such 'abstract plans' are, of course, more familiar to linguists as constructions (Goldberg, 2006; Kay & Michaelis, 2012), and we may therefore

think of Plan Reuse as an effect of the frequency information associated with a particular construction (e.g., passive, ditransitive, or RCE) or, in this case, a particular constructional subtype (RCE with indefinite subject) through implicit learning. This information is stored in long-term memory together with the construction itself. If MacDonald's theory is correct, linguistic regularities in the form of abstract constructions directly aid in production planning, explaining our result that more frequent form–meaning combinations (RCE with indefinite NP) are easier to produce than less frequent ones (RCE with definite NP), even when both are possible. In their study of prepositional and double-object sentences in spontaneous speech, Kuperman and Bresnan (2012) provide further support for the current interpretation. Although they do not distinguish between availability effects in production due to Easy First (or, in their terminology, Principle of Immediate Mention; Ferreira & Dell, 2000) and those due to Plan Reuse, their finding that speakers were sensitive to the probability of a construction WITHIN A PARTICULAR CONTEXT during production planning is compatible with both of these principles. Since the probabilities they used in their analysis were based on multiple factors, including some related to immediate prior context (e.g., givenness), as well as some related to linguistic form (e.g., definiteness, person, and number), it is likely that both immediate accessibility of words/phrases (Easy First) and accessibility of stored constructional subtypes (Plan Reuse) played a role in determining the observed effects on verb duration. By manipulating linguistic form (definiteness) in the absence of discourse context, the current study isolates the effect of Plan Reuse and provides evidence for the role of constructions and constructional subtypes in language production. Thus, we believe that the current findings support a constructionist view of language structure.

6. Conclusion

We began with the observation that RCE is a marked construction in English. In syntax, it involves a discontinuous dependency (Baltin, 2006); in usage, it occurs relatively infrequently (Francis, 2010); and in comprehension, it incurs a heavier processing load at the start of the RC (Levy et al., 2012). Nevertheless, RCE is preferred over non-RCE structure under certain conditions of phrase length, definiteness, and predicate type. While previous studies of RCE in English and German have included corpus analyses (Francis, 2010; Francis & Michaelis, 2014; Strunk, 2014; Uszkoreit et al., 1998), acceptability judgment tasks (Francis, 2010; Konieczny, 2000; Walker, 2013), and reading time measures (Francis, 2010; Konieczny, 2000; Levy et al., 2012), the current study is the first to examine structural choices using preference and elicited production tasks. These measures allow us to show

that the same factors known to affect frequency patterns in corpora also influence participants' choice of structure under controlled experimental conditions. Thus, the current study provides direct evidence for the effects of phrase length and definiteness on speakers' structural choices. Interestingly, the relative length effects differed between the preference task and the elicited production task. Given that a similar difference was shown for corpus frequencies vs. reading time (Francis, 2010), we believe that this difference may relate to the different mechanisms involved in reading vs. speaking. In addition, the elicited production task allowed us to examine measures of processing load in production and relate them to speakers' structural choices. Results for voice initiation time showed that neither RCE nor non-RCE structure was easier in general, and that the same factors that modulate choice of structure also modulate ease of production for each structure. That is, when the sentential context warrants a particular structure, that structure becomes easier to produce. Conversely, when the structure does not fit the sentential context, it becomes harder to produce. These results therefore support theoretical approaches to constituent order alternations which assume a tight connection between speakers' choice of structure and ease of production (Hawkins, 2004; Kuperman & Bresnan, 2012; MacDonald, 2013; Wasow, 2002). Importantly, the effects that we found were not limited to phrase length and so cannot be explained purely in terms of dependency distance (Hawkins, 2004) or in terms of the difficulty associated with producing a heavy phrase early in the sentence (Arnold et al., 2000). Instead, our results suggest a more general connection among sentential context, structural choice, and ease of production that applies even for discourse-related factors such as definiteness. We have argued that MacDonald's (2013) idea of Plan Reuse, in combination with other production biases, can help explain this connection in a way highly amenable to constructionist approaches to language structure.

REFERENCES

Allum, P. H., & Wheeldon, L. R. (2007). Planning scope in spoken sentence production: the role of grammatical units. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **33**(4), 791–810.

Arnold, J. E., & Lao, S. Y. C. (2008). Put in last position something previously unmentioned: word order effects on referential expectancy and reference comprehension. *Language and Cognitive Processes*, **23**(2), 282–295.

Arnold, J. E., Losongco, A., Wasow, T., & Ginstrom, R. (2000). Heaviness vs. newness: the effects of structural complexity and discourse status on constituent ordering. *Language*, **76**(1), 28–55.

Baltin, M. R. (1981). Strict bounding. In C. L. Baker & J. McCarthy (Eds.), *The logical problem of language acquisition* (pp. 257–295). Cambridge, MA: MIT Press.

Baltin, M. R. (2006). Extraposition. In M. Everaert & H. Van Riemsdijk (Eds.), *The Blackwell companion to syntax*, vol. I (pp. 237–271). Malden, MA: Blackwell.

Bod, R. (2006). Exemplar-based syntax: how to get productivity from examples. *Linguistic Review*, **23**(3), 291–320.

Bresnan, J., Cueni, A., Nikitina, T., & Baayen, R. H. (2007). Predicting the dative alternation. In G. Bouma, I. Kraemer, & J. Zwarts (Eds.), *Cognitive foundations of interpretation* (pp. 69–94). Amsterdam: Royal Netherlands Academy of Arts and Sciences.

Bresnan, J., & Ford, M. (2010). Predicting syntax: processing dative constructions in American and Australian varieties of English. *Language*, **86**(1), 168–213.

Bresnan, J., & Nikitina, T. (2009). The gradience of the dative alternation. In L. Uyechi & L. H. Wee (Eds.), *Reality exploration and discovery: pattern interaction in language and life* (pp. 161–184). Stanford, CA: CSLI Publications.

Cleland, A. A., & Pickering, M. J. (2003). The use of lexical and syntactic information in language production: evidence from the priming of noun-phrase structure. *Journal of Memory and Language*, **49**(2), 214–230.

Corley, M., & Scheepers, C. (2002). Syntactic priming in English sentence production: categorical and latency evidence from an internet-based study. *Psychonomic Bulletin & Review*, **9**(1), 126–131.

Culicover, P. W., & Rochemont, M. S. (1990). Extraposition and the complement principle. *Linguistic Inquiry*, **21**, 23–47.

Davies, M. (2009). The 385+ million word Corpus of Contemporary American English (1990–2008+): design, architecture, and linguistic insights. *International Journal of Corpus Linguistics*, **14**(2), 159–190.

de Marneffe, M. C., Grimm, S., Arnon, I., Kirby, S., & Bresnan, J. (2012). A statistical model of the grammatical choices in child production of dative sentences. *Language and Cognitive Processes*, **27**(1), 25–61.

Diessel, H. (2008). Iconicity of sequence: a corpus-based analysis of the positioning of temporal adverbial clauses in English. *Cognitive Linguistics*, **19**(3), 465–490.

Ferreira, V. S., & Dell, G. S. (2000). Effect of ambiguity and lexical availability on syntactic and lexical production. *Cognitive Psychology*, **40**(4), 296–340.

Francis, E. J. (2010). Grammatical weight and relative clause extraposition in English. *Cognitive Linguistics*, **21**, 35–74.

Francis, E. J., & Michaelis, L. A. (2012). Effects of weight and definiteness on speakers' choice of clausal ordering in English. *LSA Annual Meeting Extended Abstracts*, **3**, 5.

Francis, E. J., & Michaelis, L. A. (2014). Why move? How weight and discourse factors combine to predict relative clause extraposition in English. In E. Moravcsik, A. Malchukov, & B. MacWhinney (Eds.), *Competing motivations in grammar and usage* (pp. 70–87). Oxford: Oxford University Press.

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

Goldberg, A. E. (2006). *Constructions at work: the nature of generalization in language*. Oxford: Oxford University Press.

Gregory, M. L., & Michaelis, L. A. (2001). Topicalization and left-dislocation: a functional opposition revisited. *Journal of Pragmatics*, **33**(11), 1665–1706.

Gries, S. T. (2003). *Multifactorial analysis in corpus linguistics: a study of particle placement*. New York: Continuum Press.

Hawkins, J. A. (1994). *A performance theory of order and constituency*. Cambridge: Cambridge University Press.

Hawkins, J. A. (1999). The relative order of prepositional phrases in English: going beyond manner–place–time. *Language Variation and Change*, **11**(3), 231–266.

Hawkins, J. A. (2004). *Efficiency and complexity in grammars*. Oxford: Oxford UniversityPress.

Hawkins, J. A. (2014). *Cross-linguistic variation and efficiency*. Oxford: Oxford University Press.

Huck, G. J., & Na, Y. (1990). Extraposition and focus. *Language*, **66**(1), 51–77.

Kay, P., & Michaelis, L. A. (2012). Constructional meaning and compositionality. In C. Maienborn, K. von Heusinger, & P. Portner (Eds.), *Semantics: an international handbook of natural language meaning* (pp. 2271–2296). Berlin: Mouton de Gruyter.

Kay, P., & Sag, I. A. (2012). Cleaning up the big mess: discontinuous dependencies and complex determiners. In H. C. Boas & I. A. Sag (Eds.), *Sign-based Construction Grammar* (pp. 229–256). Stanford, CA: CSLI Publications.

Kayne, R. S. (1994). *The antisymmetry of syntax*. Cambridge, MA: MIT Press.
Konieczny, L. (2000). Locality and parsing complexity. *Journal of Psycholinguistic Research*, **29**(6), 627–645.
Kuno, S., & Takami, K. (2004). *Functional constraints in grammar: on the unergative–unaccusative distinction*. Amsterdam: John Benjamins.
Kuperman, V., & Bresnan, J. (2012). The effects of construction probability on word durations during spontaneous incremental sentence production. *Journal of Memory and Language*, **66**(4), 588–611.
Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, **106**(3), 1126–1177.
Levy, R., Fedorenko, E., Breen, M., & Gibson, E. (2012). The processing of extraposed structures in English. *Cognition*, **122**(1), 12–36.
Lohse, B., Hawkins, J. A., & Wasow, T. (2004). Domain minimization in English verb-particle constructions. *Language*, **80**(2), 238–261.
MacDonald, M. C. (2013). How language production shapes language form and comprehension. *Frontiers in Psychology*, **4**, 1–16.
Martin, R. C., Crowther, J. E., Knight, M., Tamborello, F. P., & Yang, C. (2010). Planning in sentence production: evidence for the phrase as a default planning scope. *Cognition*, **116**(2), 177–192.
Menn, L., & Duffield, C. J. (2014). Looking for a 'Gold Standard' to measure language complexity: what psycholinguistics and neurolinguistics can (and cannot) offer to formal linguistics. In F. J. Newmeyer & L. Preston (Eds.), *Measuring grammatical complexity* (pp. 281–302). Oxford: Oxford University Press.
Nelson, G., Wallis, S., & Aarts, B. (2002). *Exploring natural language: working with the British component of the International Corpus of English*. Amsterdam: John Benjamins.
Pickering, M. J., & Branigan, H. P. (1998). The representation of verbs: evidence from syntactic priming in language production. *Journal of Memory and Language*, **39**(4), 633–651.
Rochemont, M. S., & Culicover, P. W. (1990). *English focus constructions and the theory of grammar*. Cambridge: Cambridge University Press.
Rosenbach, A. (2005). Animacy versus weight as determinants of grammatical variation in English. *Language*, **81**(3), 613–644.
Ross, J. R. (1967). *Constraints on variables in syntax*. Unpublished doctoral dissertation, Massachusetts Institute of Technology.
Smith, M., & Wheeldon, L. (1999). High level processing scope in spoken sentence production. *Cognition*, **73**(3), 205–246.
Stallings, L. M., & MacDonald, M. C. (2011). It's not just the heavy NP: relative phrase length modulates the production of heavy-NP shift. *Journal of Psycholinguistic Research*, **40**, 177–187.
Stallings, L. M., MacDonald, M. C., & O'Seaghdha, P. G. (1998). Phrasal ordering constraints in sentence production: phrase length and verb disposition in heavy-NP shift. *Journal of Memory and Language*, **39**(3), 392–417.
Strunk, J. (2014). A statistical model of competing motivations affecting relative clause extraposition in German. In E. Moravcsik, A. Malchukov, & B. MacWhinney (Eds.), *Competing motivations in grammar and usage* (pp. 88–106). Oxford: Oxford University Press.
Uszkoreit, H., Brants, T., Duchier, D., Krenn, B., Konieczny, L., Oepen, S., & Skut, W. (1998). Studien zur performanzorientierten Linguistik. *Kognitionswissenschaft*, **7**(3), 129–133.
Walker, H. (2013). Constraints on relative clause extraposition in English: an experimental investigation. In G. Webelhuth, M. Sailer, & H. Walker (Eds.), *Rightward movement in a comparative perspective* (pp. 145–171). Amsterdam: John Benjamins.
Wasow, T. (1997). Remarks on grammatical weight. *Language Variation and Change*, **9**(1), 81–105.
Wasow, T. (2002). *Postverbal behavior*. Stanford: CSLI Publications.
Wasow, T., & Arnold, J. (2005). Intuitions in linguistic argumentation. *Lingua*, **115**(11), 1481–1496.

Wheeldon, L., Ohlson, N., Ashby, A., & Gator, S. (2013). Lexical availability and grammatical encoding scope during spoken sentence production. *Quarterly Journal of Experimental Psychology*, **66**(8), 1653–1673.
Wiechmann, D., & Kerz, E. (2013). The positioning of concessive adverbial clauses in English: assessing the importance of discourse-pragmatic and processing-based constraints. *English Language and Linguistics*, **17**(1), 1–23.
Wiechmann, D., & Lohmann, A. (2013). Domain minimization and beyond: modeling prepositional phrase ordering. *Language Variation and Change*, **25**(1), 65–88.

Appendix A

Sentence materials

Definite and indefinite variants differed with respect to choice of determiner (*the* or *some*). RCE and non-RCE variants differed with respect to the position of the bracketed VP in relation to the RC. Only the RCE version with VP-first order is shown here.

Set 1: *conducted*

The/Some research [was conducted] that refutes the existing theories with very clear and convincing new evidence.
The/Some research [was conducted] that refutes the existing theories.
The/Some research [has been conducted fairly recently] that refutes the existing theories with very clear and convincing new evidence.
The/Some research [has been conducted fairly recently] that refutes the existing theories.

Set 2: *raised*

The/Some points [were raised] that need some more clarification with regard to the expected profit margin.
The/Some points [were raised] that need some more clarification.
The/Some points [have been raised just now] that need some more clarification with regard to the expected profit margin.
The/Some points [have been raised just now] that need some more clarification.

Set 3: *formed*

The/Some committees [were formed] that are investigating the incident thoroughly and with the greatest possible care.
The/Some committees [were formed] that are investigating the incident.
The/Some committees [have been formed just recently] that are investigating the incident thoroughly and with the greatest possible care.
The/Some committees [have been formed just recently] that are investigating the incident.

Set 4: *provided*

The/Some instructions [were provided] that explain the new procedure in a great deal of painstaking detail.
The/Some instructions [were provided] that explain the new procedure.
The/Some instructions [have been provided right here] that explain the new procedure in a great deal of painstaking detail.
The/Some instructions [have been provided right here] that explain the new procedure.

Set 5: *considered*

The/Some options [were considered] that allow for more flexibility in the way changes will be implemented.
The/Some options [were considered] that allow for more flexibility.
The/Some options [have been considered here today] that allow for more flexibility in the way changes will be implemented.
The/Some options [have been considered here today] that allow for more flexibility.

Set 6: *presented*

The/Some evidence [was presented] that makes a strong case for imposing new regulations on big businesses.
The/Some evidence [was presented] that makes a strong case.
The/Some evidence [has been presented of late] that makes a strong case for imposing new regulations on big businesses.
The/Some evidence [has been presented of late] that makes a strong case.

Set 7: *received*

The/Some messages [were received] that indicate a possible threat to the security of the surrounding area.
The/Some messages [were received] that indicate a possible threat.
The/Some messages [have been received just now] that indicate a possible threat to the security of the surrounding area.
The/Some messages [have been received just now] that indicate a possible threat.

Set 8: *made*

The/Some changes [were made] that will improve students' performance on standardized tests of reading and vocabulary.
The/Some changes [were made] that will improve students' performance.
The/Some changes [have been made since then] that will improve students' performance on standardized tests of reading and vocabulary.
The/Some changes [have been made since then] that will improve students' performance.

Appendix B

TABLE 2. *Descriptive statistics for choice of structure, Experiment* 1

Length condition	NP type	Trials	Number RCE	Percent RCE
shortVP–longRC	indef	320	159	71.9
longVP–longRC	indef	320	156	67.2
shortVP–shortRC	indef	320	191	63.8
longVP–shortRC	indef	320	109	57.2
shortVP–longRC	def	320	90	50.3
longVP–longRC	def	320	105	51.3
shortVP–shortRC	def	320	116	40.3
longVP–shortRC	def	320	183	34.1

TABLE 3. *Descriptive statistics for choice of structure, Experiment* 2

Length condition	NP type	Trials	Number RCE	Percent RCE
shortVP–longRC	indef	312	193	61.9
longVP–longRC	indef	312	132	42.3
shortVP–shortRC	indef	312	171	54.8
longVP–shortRC	indef	312	131	42.0
shortVP–longRC	def	312	93	29.8
longVP–longRC	def	312	67	21.5
shortVP–shortRC	def	312	82	26.3
longVP–shortRC	def	312	52	16.7

TABLE 4. *Descriptive statistics for preparation time, Experiment* 2

Condition	Trials	Mean	Standard deviation
shortVP–longRC	609	4820.72	3767.52
longVP–longRC	604	5677.50	4505.38
shortVP–shortRC	614	4144.46	2996.35
longVP–shortRC	605	4938.96	3639.29
RCE–shortVP	528	4597.80	3331.97
RCE–longVP	372	5749.65	4376.69
canonical–shortVP	695	4397.90	3480.76
canonical–longVP	837	5108.96	3972.65

TABLE 5. *Descriptive statistics for voice initiation time, Experiment 2*

Condition	Trials	Mean	Standard deviation
RCE–shortRC	436	819.37	556.78
RCE–longRC	484	800.71	329.09
canonical–shortRC	809	794.68	353.57
canonical–longRC	762	852.44	354.14
RCE–def	294	867.11	415.23
RCE–indef	626	782.53	465.26
canonical–def	952	831.30	321.45
canonical–indef	619	809.46	400.87