Available online at www.sciencedirect.com
ScienceDirect
Journal of Memory and Language 57 (2007) 1–23
Journal of Memory and Language
www.elsevier.com/locate/jml

Processing of relative clauses is made easier by frequency of occurrence

Florencia Reali *, Morten H. Christiansen

Department of Psychology, Cornell University, Ithaca, NY 14853, USA

Received 17 April 2006; revision received 28 August 2006
Available online 27 October 2006

Abstract

We conducted a large-scale corpus analysis indicating that pronominal object relative clauses are significantly more frequent than pronominal subject relative clauses when the embedded pronoun is personal. This difference was reversed when impersonal pronouns constituted the embedded noun phrase. This pattern of distribution provides a suitable framework for testing the role of experience in sentence processing: if frequency of exposure influences processing difficulty, highly frequent pronominal object relatives should be easier to process but only when a personal pronoun is in the embedded position. We tested this hypothesis experimentally: We conducted four self-paced reading tasks, which indicated that differences in pronominal object/subject relative processing mirrored the pattern of distribution revealed by the corpus analysis. We discuss the results in the light of current theories of sentence comprehension. We conclude that object relative processing is facilitated by frequency of the embedded clause, and, more generally, that statistical information should be taken into account by theories of relative clause processing.

Keywords: Sentence processing; Relative clauses; Distributional information; Corpus analysis; Constraint-based approaches

Introduction

Over the past couple of decades a tremendous amount of effort has been put into elucidating the types of information used during incremental sentence comprehension. Recent research in psycholinguistics has shed much light on this issue and many theories have been proposed to account for differences in processing difficulties. A wide range of information sources has been shown to influence language processing, including lexical, contextual, syntactic and probabilistic information. However, the intricate ways in which different constraints interact with each other during sentence processing has been a matter of intense debate (for a review, see MacDonald, Pearlmutter, & Seidenberg, 1994; Tanenhaus & Trueswell, 1995). One of the recent topics of research has been the study of the information influencing the comprehension of nested structures, in particular sentences containing relative clauses that modify head noun phrases.

When the head noun phrase is the object of the verb in the relative clause, it is called an object relative clause. Conversely, sentences containing subject relative clauses are those in which the head noun phrase is the subject of the embedded verb. Examples 1(a) and (b) are subject relative and object relative sentences that have been

* Corresponding author. Fax: +1 607 255 8433.
E-mail address: fr34@cornell.edu (F. Reali).

0749-596X/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.jml.2006.08.014

previously used in the psycholinguistic literature (e.g., Holmes & O'Regan, 1981; King & Just, 1991):

(1) a. The reporter that the senator attacked admitted the error [Object Relative]
b. The reporter that attacked the senator admitted the error [Subject Relative]

It is a well-established finding that subject relative sentences such as (1b) are easier to process than object relative sentences like (1a). Such a difference in processing difficulty has been shown using different measurement procedures including online lexical decision, reading times, and response accuracy to probe questions (e.g., Ford, 1983; Holmes & O'Regan, 1981; King & Just, 1991; for a review see, Gibson, 1998).

Different theories have been proposed to explain the difference in processing difficulty between object relative and subject relative clauses. For example, structure-based accounts (e.g., Miyamoto & Nakamura, 2003) explain the subject-relative preference in terms of syntactic factors rather than functional factors such as cognitive resources. Following a generative approach, structure-based accounts emphasize a universal preference for syntactic gaps in the subject position. This approach predicts a universal preference for subject relative clauses, independently of cognitive and discourse constraints.

Working-memory-based approaches differ from syntactic-based approaches in that they rely on functional factors such as cognitive resources and integration constraints. These theories propose that the storage of incomplete head-dependencies in phrase structure causes the increase in complexity in object relative sentences compared to subject relatives (Chomsky & Miller, 1963; Gibson, 1998; Lewis, 1996). Thus, object relative sentences are harder because there is a larger number of temporally incomplete dependencies in the processing of object extractions. Along these lines, the dependency locality theory (DLT) (Gibson, 1998; Gibson, 2000; Grodner & Gibson, 2005; Hsiao & Gibson, 2003; Warren & Gibson, 2002) is based on the principle that dependencies between lexical items are constrained by both storage and integration resources. The integration component in DLT accounts for the cost associated with performing structural integrations. The object relative clauses require more resources because the integrations at the embedded verb involve connecting the object position to the wh-filler, an integration that crosses the subject noun phrase. Integration cost is increased, among other factors, by the discourse complexity of the intervening material between the elements being integrated. In particular, building new discourse structure (such as a discourse referent) is more expensive than accessingpreviously constructed discourse elements. Thus, according to DLT, the processing cost of integrating structures to their head constituents increases with the number of new discourse referents introduced between the phrasal heads that must be integrated. For example, in object relative clauses, the integration across a subject definite noun phrase (e.g., *the senator* in (1a)) is more costly than the integration across a subject noun phrase that is part of the discourse (e.g., first-/second-person pronoun).

Some working-memory-based theories include the additional component of interference by syntactic similarity between subject noun phrases that need to be simultaneously held in memory (Bever, 1970; Gordon, Hendrick, & Johnson, 2001; Gordon, Hendrick, & Johnson, 2004; Gordon, Hendrick, & Levine, 2002; Lewis & Vasishth, 2005; Van Dyke & Lewis, 2003). In object relatives, representations for both the matrix and embedded nouns are accessed before either noun phrase is integrated with the verb of the modifying clause. Thus, according to the similarity-based interference approach, the processing difficulty in object relatives is explained because unintegrated nouns in the sentence interfere with each other in working memory. Similar to DLT, this is a memory-retrieval-based theory: integrations are made difficult by the syntactic interference of the intervening material.

Finally, according to experience-based accounts, the observed difference between processing of object and subject relative clauses may be explained, at least in part, by differences in exposure to statistical regularities of the language (MacDonald & Christiansen, 2002; Mitchell, Cuetos, Corley, & Brysbaert, 1995; Tabor, Juliano, & Tanenhaus, 1997). For example, according to constraint-based models (e.g., MacDonald et al., 1994) syntactic processing is constrained by a wide variety of probabilistic factors at the syntactic, lexical, contextual and semantic levels. Under this view, statistical regularities may influence sentence comprehension, more particularly, the processing of object relative and subject relative sentences.

Recent work has explored the influence of the embedded noun phrase type on sentence complexity (Gordon et al., 2001; Gordon et al., 2004; Mak, Vonk, & Schriefers, 2002; Warren & Gibson, 2002). For example, Warren and Gibson (2002) examined the extent to which referential properties of the second noun phrase affect the complexity of center-embedded sentences. Using both complexity rating and self-paced reading tasks, they found that the processing difficulty in nested sentences depends on the degree to which the embedded subject was old or new in the discourse according to the *Giveness Hierarchy* (Gundel, Hedberg, & Zacharski, 1993). As an example, consider the doubly nested sentences (2) used in Warren and Gibson (2002):

(2) a. The student who the professor who I collaborated with had advised copied the article.
b. The student who the professor who the scientist collaborated with had advised copied the article.

DLT states that the integration cost increases with the number of new discourse referents that are introduced between the phrasal heads that must be integrated. In sentence (2b) the most deeply embedded noun phrase introduces new discourse referents, while the first personal pronoun *I* in (2a) is considered part of the discourse. Thus, DLT predicts that sentence (2a) should be easier to process than (2b). Warren and Gibson (2002) showed that processing difficulty increased as a function of the rank of the embedded subject according to the *Giveness Hierarchy*.

In a different series of studies, Gordon et al. (2001) showed that the well-established difference in processing difficulty between subject relatives and object relatives could be eliminated when the embedded noun phrase was the indexical pronoun *you* and reduced when it was a proper name. The authors interpreted the results from a similarity-based interference perspective: memory interference during encoding and retrieval may not occur because the matrix and the embedded noun phrases produce non-interfering representations.

Both DLT and similarity-based interference approaches account for the reduction of complexity in pronominal object relative sentences, suggesting that the data could be explained by a combination of factors. Other constraints may also be involved in explaining these results. For example, in pronominal object relative clauses, the embedded noun phrase is a prototypical subject (a pronoun), suggesting that discourse and distributional information may play a role in the reduction of processing difficulty. Despite the striking pattern of results recently observed in pronominal relative clauses (e.g., Gordon et al., 2001; Warren & Gibson, 2002), the distributional properties of pronominal object/subject relatives in English remained mostly unexplored. What is the relative frequency of subject relative and object relative clauses containing personal pronouns naturally occurring in language? Does the relative distribution of pronominal object/subject relative clauses influence processing difficulty? Here, we take the first steps toward answering these questions. First, we conduct a corpus analysis to explore the relative frequency of subject relative and object relative clauses with embedded pronouns, finding an overwhelming majority of pronominal object relative clauses compared to pronominal subject relative clauses. We suggest that the observed regularities are expected under discourse-based explanations of the type previously proposed by Fox and Thompson (1990). Second, we conduct a series of self-paced reading experiments to explore the extent to which the distributional patterns revealed by the corpus analysis mirror the differences in processing difficulty between pronominal object/subject relative clauses. Our results provide strong support to experience-based approaches.

The role of statistical information during online sentence processing

Recently, there has been a reappraisal of statistical approaches to language processing, partly motivated by research indicating that probabilistic information influences language acquisition and comprehension (e.g., Crocker & Corley, 2002; Jurafsky, 1996; MacDonald et al., 1994; Spivey-Knowlton & Sedivy, 1995; Trueswell, 1996). The role of statistical information has been studied mostly in the context of ambiguity resolution (e.g., Crocker & Corley, 2002; Jurafsky, 1996; MacDonald et al., 1994; Spivey-Knowlton & Sedivy, 1995; Trueswell, 1996). Some studies, such as those conducted by Mitchell et al. (1995), provide evidence that distributional information tabulated at the structural level influences initial parsing strategies in English and Spanish (but see Fodor, 1998). Gibson and Schütze (1999) conducted a study of English in which disambiguation preferences were not found to mirror corpus frequencies, seemingly disconfirming the predictions of experience-based theories. Using similar materials, Desmet and Gibson (2003) provided a reevaluation of the discrepancies between disambiguation preferences and corpus frequencies reported by Gibson and Schütze (1999). In the latest study, specific features of the test sentences were analyzed and corpus frequencies were tabulated at a finer grain. Interestingly, the results in Desmet and Gibson (2003) revealed that online disambiguation preferences matched corpus frequencies when lexical variables were taken into account. The authors nevertheless acknowledge the difficulty in understanding the cause-effect relations underlying this correlation.

Other studies provide support for constraint-based *lexicalist* approaches in that they have shown that the interpretation of ambiguities is also constrained by combinatorial distributional information associated with specific lexical items (Desmet, De Baecke, Drieghe, Brysbaert, & Vonk, 2005; MacDonald, 1994; McRae, Spivey-Knowlton, & Tanenhaus, 1998; Pearlmutter & MacDonald, 1992; Tabossi, Spivey-Knowlton, McRae, & Tanenhaus, 1994; Trueswell, Tanenhaus, & Garnsey, 1994). Despite the growing number of studies designed to explore whether statistical information affects the resolution of syntactic ambiguities, much less is known about its potential role in the processing of unambiguous utterances. Some recent studies have explored the influence of fine-grained statistics during online processing of simple sentences. For example, using a self-paced reading task, McDonald and

Shillcock (2003) demonstrated that reading times of individual words are affected by the transitional probabilities of the lexical components (but see Frisson, Rayner, & Pickering, 2005). However, very little research has been conducted to explore the role of distributional information during comprehension of sentences containing nested grammatical structure.

In a recent paper, MacDonald and Christiansen (2002) proposed that distributional constraints might play a role in explaining the differences in processing difficulties found in subject relative and object relative clauses. They argued in favor of experience-based accounts according to which comprehension difficulties that have been observed during the processing of nested structure may be explained, at least in part, by differences in statistical regularities of the language (see also Christiansen, 1994; Reali & Christiansen, 2006). This view is consistent with probabilistic-constraint approaches that emphasize the need for an essential continuity between language acquisition and processing (e.g., Bates & MacWhinney, 1987; Farmer, Christiansen, & Monaghan, 2006; Seidenberg, 1997; Seidenberg & MacDonald, 1999; Snedeker & Trueswell, 2004). Along these lines, we advocate a model of structure representation that is affected by language use.

Recently, Bybee (2002) proposed that the representation of constituent structure is highly influenced by frequent sequential co-occurrence of linguistic elements. According to this view, when words repeatedly co-occur together in a specific order, such multi-word sequences may fuse together into a single processing unit. As a consequence of this '*chunking*' process, the repeated exposure to sequential stretches of words within a linguistic constituent would create a supra-lexical representation of this construction, making it easier to access. Recent studies suggest that the adult human parser might adopt a chunk-by-chunk strategy (e.g., Abney, 1991; Konieczny, 2005; Tabor, Galantucci, & Richardson, 2004; Tabor & Hutchins, 2003; Wray, 2002). In a series of studies, Tabor et al. (2004) provided experimental evidence suggesting that the human processor constructs partial parses that are syntactically compatible with only a subpart of the sentence being read. For example, using syntactically unambiguous materials like *The coach smiled at the player tossed a Frisbee*, they showed interference from locally coherent structures (such as *the player tossed*) as reflected by distractive effects of irrelevant Subject-Predicate interpretations. They argued in favor of bottom-up dynamical models in which locally coherent structures are constructed during parsing, at least temporarily. From a computational perspective, Abney (1991, 1996) proposed that the notion of *chunk* corresponds to one or more content words surrounded by function words, matching a fixed template. According to this view, co-occurrence of chunks is determined not only by their syntactic categories but also by the precise words that constitute them, and crucially, the order in which the chunks occur is much more flexible than the order of words within chunks.

In line with the view that the human parser follows a chunk-by-chunk strategy, our goal is to explore whether the frequency of the chunks affects processing difficulty when they constitute pronominal relative clauses. In the spirit of the constructivist approach outlined in Bybee (2002; Bybee & Scheibman, 1999), our theoretical proposal is grounded in the view that language use, and in particular frequency of *chunk* use, plays a crucial role in the representation of constituent structure. Bybee (2002) argues that repetition of word sequences triggers a chunking mechanism that binds them together to form constituent representations. Importantly, elements that are frequently used together would bind tighter into constituents. Therefore, constructions may have different degrees of cohesion due to the differences in their co-occurrence patterns (Bybee & Scheibman, 1999). Frequent word-sequences (chunks) would fuse into amalgamated processing units that can be accessed and produced more easily.

Along these lines, we hypothesize that frequent word sequences forming relative clauses may lead to more cohesive representations that are easier to access than less frequent ones. We focus on the case of pronominal relative clauses to explore this hypothesis. Importantly, our thesis is not that frequency is the only constraint affecting the comprehension of embedded structure. On the contrary, we believe that discourse and referential information, as well as cognitive limitations, play a crucial role. However, our goal is to provide evidence indicating that the role of statistical information may have been underestimated in most current models of relative clause processing. We combine corpus analysis and self-paced reading experiments to determine the extent to which the difficulties encountered during online processing of pronominal relative clauses mirror distributional patterns occurring naturally in language. We contrast the results with the predictions of other theories of sentence processing. To do this, we take advantage of the fact that working-memory-based models in their current form do not predict object relative clauses to be *easier* to process than their subject relative counterparts, while experience-based approaches do, but only under some circumstances.

The corpus analysis presented in the next section revealed that pronominal object relative clauses are significantly more frequent than pronominal subject relative clauses when the embedded pronoun is personal. This difference was reversed when impersonal pronouns constituted the embedded noun phrase. In light of these intriguing statistical differences, the following predictions were made: first, if clause frequency affects relative clause processing we should find some measurable

facilitation of pronominal object relative clauses compared to pronominal subject relative clauses when a personal pronoun constitutes the second noun phrase. However, pronominal subject relative clauses should be harder when an impersonal pronoun (e.g., *it*) is in the second noun phrase position. In Experiment 1, we conducted a self-paced reading task to compare the processing difficulty of object relative and subject relative clauses in which a second-person pronoun was the embedded noun phrase. Although a similar experiment has been previously conducted by Gordon et al. (2001), we argue that a critical analysis is missing to rule out object relative facilitation across the embedded region. Crucially, Experiment 1 reproduces Gordon et al.'s (2001) main results, and, in addition, reading-time comparisons across the embedded two-word region revealed facilitation of the object relative condition compared to the subject relative condition. In Experiments 2 and 3 we conducted a self-paced reading task to explore the processing of object/subject relative constructions in which the second noun phrase was a first-person pronoun (*I*) and a third-person pronoun (*they*/*them*), respectively. Similar to Experiment 1, we found an effect of relative-clause-type condition in the region comprising the two words after the relativizer, indicating that object relative clauses were read faster in Experiments 2 and 3. In Experiment 4 we compared processing difficulties in object/subject relative constructions in which an impersonal pronoun (*it*) was in the second noun phrase position. Because the corpus analysis revealed a larger proportion of pronominal subject relative clauses compared to pronominal object relative clauses of this type, we predicted that the latter should be harder to process. The experiment results confirmed this prediction.

All experiments showed a robust difference between high and low frequency conditions. The results indicate that the processing of relative clauses is facilitated by the frequency of the embedded clause and, more generally, that statistical information must be taken into account by theories of relative clause processing.

Corpus analysis

Previous corpus analyses have started to shed light on the distributional regularities underlying the use of relative clause constructions. For example, Fox and Thompson (1990) examined transcripts of naturally occurring conversations, exploring distributional characteristics of a sample of 414 relative clauses. They found that the distribution of object relative and subject relative clauses varied according to the properties of the head noun phrase of the main clause. For example, if the head noun phrase was an inanimate subject, object relatives were more frequent than subject relatives, while if the head noun phrase was an inanimate object, then subject relatives were more frequent than object relatives. They argued that the tendency of nonhuman subject heads to occur with object relatives was due to fact that nonhuman head noun phrases tend to be anchored by a referent in the object relative clause. Fox and Thompson provide an explanation for this phenomenon consisting of two parts: first, nonhuman full-noun phrases tend to occur initially in the sentence and are typically ungrounded. Second, nonhuman head noun phrases are typically inanimate and therefore good *objects*. Thus, the most typical grounding for a nonhuman head noun phrase is one in which a relative-clause-internal good *agent* (e.g., a pronoun) is the subject of the embedded verb. Consider the following example taken from Fox and Thompson (1990): *Well you see that the problem I have is my skin is oily and that lint just flies into my face* (p.303). The authors observed that this type of anchoring is usually done by subject pronouns. Fox and Thompson conclude that " ... there are clear cognitive and interactional pressures at work to favor constructions in which nonhuman Subject Heads have relative clauses with pronominal subjects." (p. 304) Fox and Thompson explored the characteristics of the head noun phrase in the main clause position associated with each type of relative clause. However, they did not investigate the relative frequency of second-noun-phrase types in object relative and subject relative clauses; that is, they did not distinguish between pronominal and non-pronominal relative clauses in their frequency counts.

The goal of our corpus analysis is to explore the relative frequencies of object vs. subject relative clauses in which the embedded subject is a pronoun and to compare them with the relative frequencies of non-pronominal object and subject relative clauses. Converging evidence from psycholinguistic studies indicates that subject relative clauses containing definite and indefinite noun phrases are easier to process than their object relative counterparts. Thus, a higher frequency of non-pronominal subject relative clauses would indicate the existence of a correlation between statistical biases and processing difficulty predicted by working-memory-based accounts and structural-based theories. However, such a correlation is difficult to anticipate in the case of pronominal subject/object relative clauses.

Methods

Materials

The corpus analysis was conducted using the first released version of the American National Corpus (ANC) (Ide & Suderman, 2004). The corpus contains over 11 million words from both spoken and written language sources. It is compiled from seven different sources: CallHome (50,494 words), Switchboard

(3,056,062 words), Charlotte narratives (117,832 words), New York Times (3,207,272 words), Berlitz Travel Guides (514,021 words), Slate Magazine (4,338,498 words), and Oxford University Press (OUP) (224,037 words). The CallHome corpus includes transcripts and documentation files for 24 unscripted telephone conversations between native speakers of English. The transcripts cover a contiguous 10-min segment of each call. The Switchboard corpus includes the transcriptions of the LDC Switchboard corpus. It consists of 2320 spontaneous conversations averaging 6 min in length and comprising about 3 million words of text, spoken by over 500 speakers of both sexes from every major dialect of American English. The Charlotte Narrative and Conversation Collection (CNCC) corpora contains 95 narratives, conversations and interviews representative of the residents of Mecklenburg County, North Carolina, and surrounding communities. The New York Times component of the ANC First Release consists of over 4000 articles from the *New York Times* newswire for each of the odd-numbered days in July 2002. The Berlitz Travel Guide corpus contains travel guides written by and for Americans that were contributed by Langensheidt Publishers. The Slate Magazine is an on-line publication with articles on various topics. The ANC Slate Magazine corpus contains 4694 short articles from the Slate archives published between 1996 and 2000, including articles on topics of current interest, including news and politics, arts, business, sports, technology, travel, food, etc. Finally, the various non-fiction OUP corpora contains about a quarter million words of non-fiction stories drawn from five Oxford University Press publications authored by Americans.

We used the tagged version of the first release of the ANC corpus, which uses the morpho-syntactic tags from the tagset developed by Biber (1988, 1995).

Procedure

All the corpus analyses were done using software developed in our lab in a Linux environment. A combined tagged version of the corpora was used to perform the analyses. Sentences containing relative clauses were selected from the corpora by pulling out phrases containing relative pronouns from one of the following categories:

1- '*That'* as dependent clause head of an object relative clause (Biber tag description: *tht* + *rel* + *obj* ++)
2- '*That'* as dependent clause head of a subject relative clause (Biber tag description: *tht* + *rel* + *subj* ++)
3- 'Wh' pronoun as head of an object relative clause (Biber tag description: *whp* + *rel* + *obj* ++)
4- 'Wh' pronoun as head of a subject relative clause (Biber tag description: *whp* + *rel* + *subj* ++)

Within the subject relative clauses, those phrases containing a pronoun in the embedded position (*relativizer* + *VP* + *pronoun*) were counted. Similarly, object relative clauses with pronominal noun phrases (*relativizer* + *pronoun* + *VP*) were counted. Five types of pronouns were considered in the analyses: first-person pronouns (*I, we, me, us*), second-person pronoun (*you*), third-person personal pronouns (*she, he, they, her, him, them*), third-person impersonal pronoun (*it*) and nominal pronouns

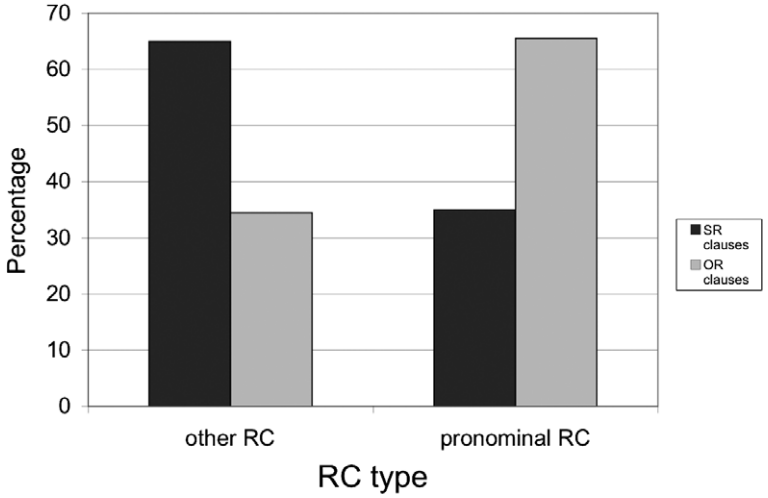

Fig. 1. Results from the corpus analysis. Bars represent the percentage of object relative clauses (OR, light bars) and subject relative clauses (SR, dark bars) in pronominal (right) and non-pronominal relative clauses (left).

(e.g., *someone*). Different types of pronouns were identified using their Biber tag descriptions.

Results and discussion

We found a total of 69,503 phrases tagged as relative clauses. Of these, 44,492 were tagged as subject relative clauses (65%) while 25,011 were tagged as object clauses (35%). For practical reasons, only relative clauses with relative pronouns were analyzed, that is, we did not consider reduced relative clauses (e.g., *the man I know*) in the analysis. When pronominal clauses of the form '*relativizer+VP+pronoun*' and '*relativizer + pronoun + VP*' were excluded, subject-relative phrases (41,458) significantly outnumbered the object-relative phrases (19,251) ($\chi^2 > 100$; $p < .0001$). As shown in Fig. 1, the tendency was dramatically reversed when the embedded noun phrase was a pronoun: subject relative constructions (3034) comprised 34.5 % of pronominal relative clauses while object relative constructions (5760) accounted for the remaining 65.5% of them ($\chi^2 > 100$; $p < .0001$).

Fig. 2 shows the distribution of object relative and subject relative clauses for each type of embedded pronoun. Object relatives were more frequent than subject relatives when the second noun phrase was a personal pronoun (first-person pronouns: 82% were object relatives; second-person pronouns: 74% were object relatives; third-person pronouns: 68% were object relatives). However, this tendency was reversed when the pronoun was impersonal (*it*) (34% were object relatives) or nominal (22% were object relatives). The number of pronominal subject/object relative clauses across individual corpora is provided in Table 1. Although the proportion of pronominal object relatives was greater in the spoken corpora than in written corpora, qualitative trends are the same across all sources.

Nominal pronouns could be animate (*everyone, everybody, anybody*) or inanimate (*anything, something*). We therefore investigated the relative frequencies of nominal object/subject relative clauses when the subject was animate. To do that, we repeated the analysis, but considered only the following eight quantifying pronouns: *everyone, everybody, anybody, anyone, no one, nobody, someone* and *somebody*. The results revealed that object relative clauses were more frequent than subject relative clauses of this type (see Table 1). This tendency suggests that pronominal object relative clauses tend to be more frequent than their subject relative counterpart when the pronoun in the embedded noun phrase position is animate.

Much recent research has shown that non-pronominal object relative sentences are more difficult to process than subject relative sentences. Thus, the higher frequency of non-pronominal subject relatives indicates a correlation between distribution and complexity that might reflect choices during production. However, the larger proportion of pronominal object relatives compared to pronominal subject relatives cannot be explained as a result of choices in production associated with difficulties derived from working-memory-related factors. One possibility is that the distributional pattern of pronominal relative clauses derives from discourse constraints. Fox and Thompson (1990) suggested that object relative clauses are frequently found

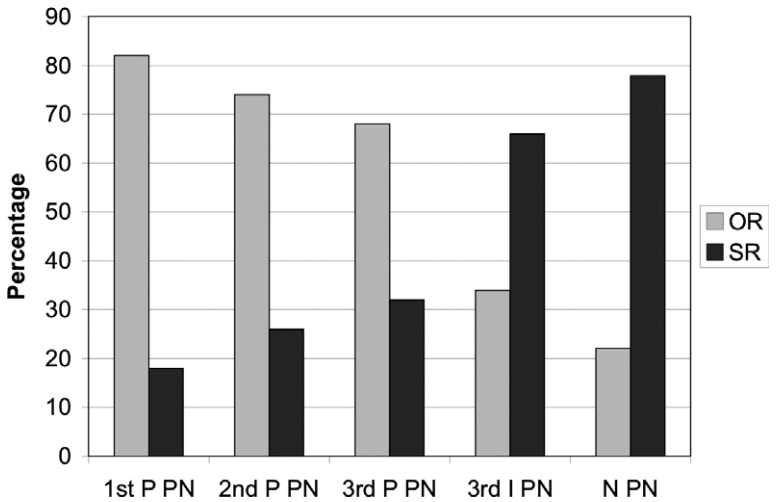

Fig. 2. Bars represent the percentage of object relative (light bars) and subject relative (dark bars) clauses across different types of pronominal relative clauses (1st P PN = first-person pronoun; 2nd P PN = second-person pronoun; 3rd P PN = third-person personal pronoun; 3rd I PN = third-person impersonal pronoun; N PN = nominal pronoun; SR = subject relative; OR = object relative).

Table 1
American National Corpus

Spoken corpus	RC-internal-PN	OR	SR	χ^2
Switchboard				
	Personal	3317	478	>100[d]
	Impersonal	128	302	>100[d]
	Animate nominal	46	29	3.8
	Inanimate nominal	8	113	91.1[d]
Callhome				
	Personal	24	5	12.4[c]
	Impersonal	7	6	<1
	Animate nominal	0	0	—
	Inanimate nominal	0	1	<1
Charlotte				
	Personal	269	45	>100[d]
	Impersonal	6	7	<1
	Animate nominal	5	1	<1
	Inanimate nominal	0	12	12[c]
Written corpus	**RC-internal-PN**	**OR**	**SR**	χ^2
New York Times				
	Personal	537	392	22.6[d]
	Impersonal	85	173	30[d]
	Animate nominal	13	14	<1
	Inanimate nominal	4	80	68[d]
Slate Magazine				
	Personal	920	742	19[d]
	Impersonal	199	354	43[d]
	Animate nominal	49	26	7[b]
	Inanimate nominal	5	139	>100[d]
Berlitz Travel				
	Personal	66	37	8.16[b]
	Impersonal	12	28	6.4[a]
	Animate nominal	0	0	—
	Inanimate nominal	0	6	6[a]
OUP				
	Personal	47	29	4.26[a]
	Impersonal	10	10	<1
	Animate nominal	0	1	<1
	Inanimate nominal	3	4	<1
Total		5760	3034	

Note. RC-internal-PN = Relative-Clause-internal-Pronoun; OR = Object Relative; SR = Subject Relative.
[a] $p < .05$.
[b] $p < .01$.
[c] $p < .001$.
[d] $p < .0001$.

modifying nonhuman head noun phrases in the sentential subject position because they provide a way to anchor the head noun phrase to the ongoing discourse context. In addition, it has been previously found that anchoring to discourse is nearly always done by a pronoun (Fox, 1987). This led Fox and Thompson (1990) to suggest that constructions in which subject head noun phrases have relative clauses with pronominal subjects should be high in frequency.

Importantly, the observed bias suggests that distributional information might be an additional factor in the facilitation of pronominal object relative constructions reported in recent studies (Gordon et al., 2001; Warren & Gibson, 2002). The challenge of studying the information influencing sentence processing complexity is made difficult by the fact that similar processing difficulties may be expected under experience-based and working-memory-based accounts. Fortunately, the distributional pattern of pronominal relative clauses provides a suitable framework to investigate the relative influence of statistical regularities on relative clause processing. This is because working-memory-based approaches do not predict pronominal object relatives to be easier than pronominal subject relatives, whereas experience-based approaches do. Thus, if such trend were to be found, it would reveal the influence of statistical information. We conducted three experiments to investigate object/subject relative processing difficulty when the second noun phrase is a second-person pronoun (Experiment 1), a first-person pronoun (Experiment 2), and a third-person pronoun (Experiment 3). In Experiment 4 we explored object/subject relative differences in processing difficulty when the second noun phrase is an impersonal pronoun. The experimental results indicate a correlation between differences in object/subject relative processing difficulty and the relative frequency of each type of pronominal relative clause.

Experiment 1

Experiment 1 was a self-paced moving-window reading task conducted to explore whether object relative clauses were read faster than subject relative clauses when the embedded noun phrase was an indexical pronoun. Working-memory-based theories predict a reduction or elimination of the traditional object/subject relative clause difference. However, neither DLT nor similarity-based interference theories predict object relatives to be easier than their subject relative counterparts.

Previously, Gordon et al. (2001) conducted a similar reading task experiment comparing the processing of object and subject relative clauses in which the indexical pronoun *you* was the embedded noun phrase. They found an elimination of the well-established difference in processing difficulty across relative-clause type. The stimuli in Gordon et al. (2001, Experiment 2) included both sentences with the indexical pronoun as the second noun phrase and sentences with a definite noun phrase (e.g., the lawyer) as the second noun phrase. The following sentences are examples of their stimuli:

(3) a. The barber that the lawyer/you admired climbed the mountain.
b. The barber that admired the lawyer/you climbed the mountain.

Reading times in the pronoun condition were analyzed separately for two critical words. They found no difference across relative-clause type at the second critical word, namely the main verb of the sentence (e.g., *climbed* in sentences (3)). In addition, they found no effect of relative-clause type at the first critical word consisting of the indexical pronoun (*you*) in the subject relative condition and the embedded verb in the object relative condition (e.g., *admired* in example (3a)). The lack of differentiation in reading times on the first critical word indicated that the word *you*—a short and frequent lexical item—was read at the same speed in the subject relative condition as the embedded verb in the object relative condition, which included infrequent and long words (e.g., *questioned* or *complimented*). Thus, a more reasonable comparison would involve the analysis of reading times averaged across the two-word region that follows the relativizer (e.g., *you admired* in the object relative condition vs. *admired you* in the subject relative condition in example (3)). According to an experience-based account, the processing at the chunk '*you admired*' occurring in the pronominal object relative condition should be facilitated by frequency of occurrence relative to the chunk '*admired you*' occurring in the pronominal subject relative condition. Unfortunately, numerical values of reading times averaged across this two-word region were not provided in Gordon et al. (2001). However, a close look at Fig. 2 in Gordon et al. (2001, p. 1415) indicates that the first word after the relative pronoun (the word *you* in the object relative condition and the verb in the subject relative condition) was read numerically faster in the object relative condition, while the second word (the verb in the object relative condition and the word *you* in the subject relative condition) was read equally fast in both conditions. Thus, numerical values displayed graphically suggest that reading times averaged across this two-word region are faster in the object relative condition.

Gordon et al. (2001) conducted statistical comparisons across the region that included the words after the relative pronoun (*that*) and before the matrix verb. However, their analysis of variance was collapsed across both types of embedded noun-phrase-type (definite common noun phrase and indexical pronoun), revealing no significant reading-time difference across relative clause type condition and a significant interaction between relative-clause type and noun-phrase type. Gordon et al. (2001) did not report statistical comparisons across this two-word critical region for the pronoun condition only.

In Experiment 1 we therefore employ a self-paced reading task designed to compare processing difficulty between pronominal object relative and subject relative sentences at the level of the two-word region in the relative clause. The stimuli used here are similar to those used in Gordon et al. (2001).

Methods

Participants

Twenty-eight native English speakers from Cornell undergraduate classes participated in this study.

Materials

Fourteen experimental items were tested with two conditions per item. The stimuli consisted of sentences with a relative clause that modified the subject noun phrase of the main clause. The two conditions varied in the type of embedded clause (subject vs. object relative). All sentences had a second-person pronoun as the noun phrase in the relative clause. The corpus analysis revealed a higher frequency of object relative clauses than subject relative clauses in which the pronoun *you* was the second noun phrase. Thus, experience-based accounts predict object relatives to be easier than subject relatives.

Sentences provided in (4) are examples of the stimuli used in the object relative condition (4a) and subject relative condition (4b):

(4) a. The consultant that you called emphasized the need for additional funding.
b. The consultant that called you emphasized the need for additional funding.

Two lists were created, each comprising fourteen experimental items and fifty-two fillers. In this and subsequent experiments, lists were randomized across participants, and the two conditions were counterbalanced across lists so that each participant only saw one version of each item. A complete list of materials for all the experiments described herein is included in the Appendix A.

In order to ensure that our stimuli were not biased in terms of plausibility, we conducted a norming study in which an additional 20 participants rated the plausibility of the experimental sentences on a 1–7 scale where 1 was "not plausible" and 7 was "very plausible". Each questionnaire comprised fourteen experimental items and fifty fillers. In this and subsequent experiments, the two conditions were counterbalanced across lists so that each participant only saw one version of each item. The lists were pseudo-randomized so that no two experimental items occurred back to back and the order of the questionnaire pages was varied. Analyses of variance revealed that participants found no difference in plausibility between object relative (mean = 5.75; $SD = .76$)

and subject relative (mean = 5.81; $SD = .64$) sentences ($F_1(1,19) < 1$; $F_2(1,13) < 1$).

Procedure

The experimental task involved self-paced reading in a word-by-word moving window display (Just, Carpenter, & Woolley, 1982) using the Psyscope experimental software package (Cohen, MacWhinney, Flatt, & Provost, 1993) on a Macintosh computer. At the start of each trial, a sentence appeared on the screen with all characters replaced by dashes. Participants pressed a key to change a string of dashes into a word. Each time the key was pressed, the next word appeared and the previous word reverted back into dashes. The time between key-presses was recorded. After each sentence, participants answered a yes/no comprehension question about its content. No feedback was provided for responses. Participants were asked to read at a natural pace and were given a small set of practice items and questions before the experimental items were presented in order to familiarize them with the task.

Results and discussion

Comprehension accuracy in the object relative and subject relative conditions was 96.3 and 97.2%, respectively, and did not differ significantly across conditions. In this and subsequent experiments, reading times were removed if they exceeded 3000 ms.

Differences across conditions were analyzed using pairwise contrasts. We provide 95% confidence intervals for the differences between condition means, which were calculated using mean square error terms taken from the analysis by participants (Masson & Loftus, 2003). A halfwidth-size confidence interval that does not exceed the difference across condition means indicates that this difference is significant at a .05 level.

Fig. 3 shows mean reading times per word. First, we analyzed the region consisting of the matrix verb of the sentence. Similarly to Gordon et al. (2001) we found no effect of relative-clause type in this region (mean = 473 ms, $SD = 199$ ms in object relatives, and mean = 444 ms, $SD = 203$ ms in subject relatives), $F_1(1,27) = 1.52$, MSE = 7929, $p = .23$; $F_2(1,13) = 0.75$, MSE = 6201, $p = .4$. Reading times were 29 ms slower in the object relative clauses; however, the difference was not significant, with a confidence interval of ±34 ms.

The second critical region of study consisted of the two words following the relativizer *that* (*you called* in the object relative condition vs. *called you* in the subject relative condition), a region that was crucial to test our experimental hypothesis. A 2 (Subject Relative vs. Object Relative) × 2 (word1 vs. word2) ANOVA revealed an effect of relative-clause-type, $F_1(1,27) = 8.01$, MSE = 11,048, $p = .008$; $F_2(1,13) = 7.51$, MSE = 6375, $p = .017$; minF'(1,34) = 3.9. In the object relative condition, the mean reading time *averaged* across the two-word region was 370 ms (mean = 353 ms, $SD = 98$ ms in word1, and mean = 388 ms, $SD = 161$ in word2). In the subject relative condition, the mean in the same region was 427 ms (mean = 431 ms in word1, $SD = 220$ ms, and mean = 423 ms in word2, $SD = 140$ ms). The 95% confidence interval for this 57 ms difference between condition means (427 − 370 ms) was ±47 ms, indicating that the object relative condition was read significantly faster. Fig. 4

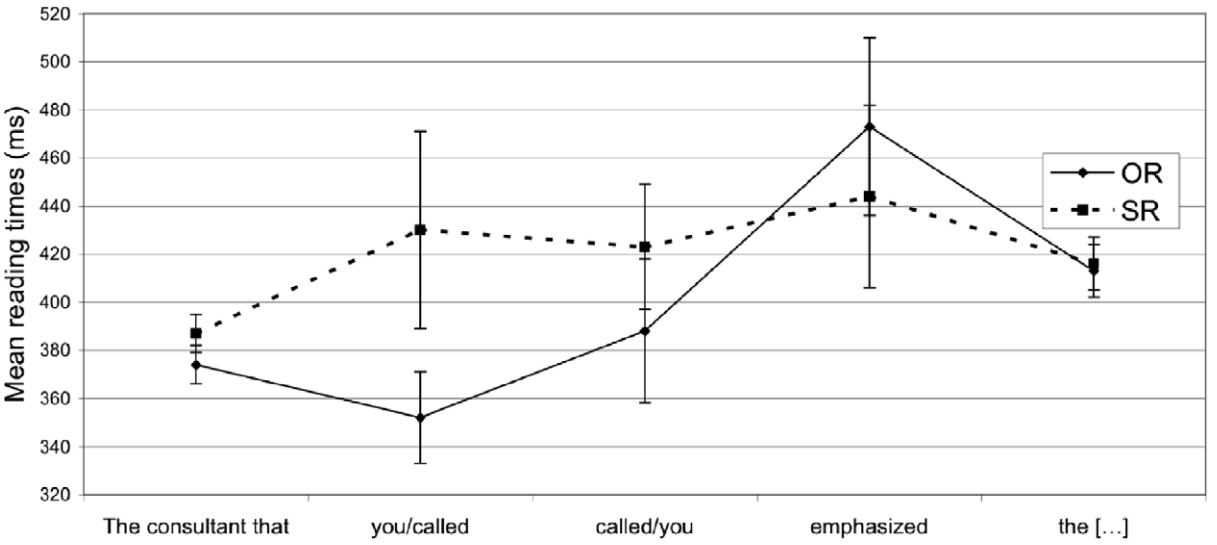

Fig. 3. Results from Experiment 1: mean reading times across regions for subject relative (dashed line) and object relative (solid line) conditions. Error bars correspond to the standard error for each reading time mean (SR = subject relative; OR = object relative).

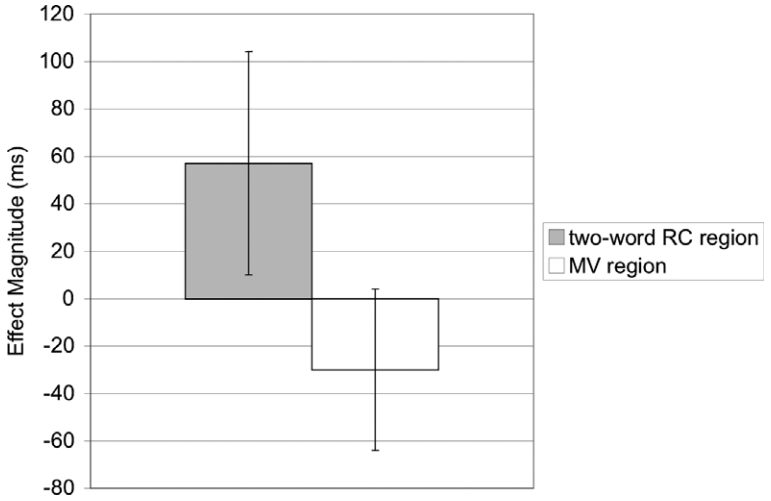

Fig. 4. Results of Experiment 1: differences between reading time means (subject relative condition minus object relative condition) in the relative-clause-internal two-word region (dark bar) and main-verb region (light bar). The error bars correspond to the 95% confidence interval for each difference (MV = main verb; RC = relative clause).

shows the difference between condition means (subject-relative condition minus object relative condition) for the main verb region and the two-word critical region. The error bars in the figure represent the 95% confidence interval for each region.

The results indicate a clear difference in reading times across object relative and subject relative clauses in that longer reading times were observed in the subject relative condition across the two-word region constituting the embedded clause. These results reproduced those obtained by Gordon et al. (2001) at the matrix verb region. However, our analyses differ from theirs in that we directly compared reading times across the broader two-word region, revealing an overall facilitation of the object relative condition.

Experiment 2

Experiment 2 was a self-paced reading time task designed to compare processing difficulty in object/subject relative-clause sentences in which a first-person pronoun was the second noun phrase. Following a similar line of reasoning, it provides a natural extension to Experiment 1 in order to further substantiate its results.

Methods

Participants

Thirty-two native English speakers from Cornell undergraduate classes participated in this study.

Materials

Fourteen experimental items were tested with two conditions per item. The stimuli consisted of object/subject relative-clause sentences in which a singular first-person pronoun (*Il/me*) was the second noun phrase. Sentences 5(a) and (b) are examples of the stimuli:

(5) a. The lady that I visited enjoyed the meal.
b. The lady that visited me enjoyed the meal.

Using identical methods to Experiment 1, two experimental lists were created, each with fourteen experimental items and forty-two fillers.

As in the previous experiments, we conducted a norming study in which an additional 20 participants rated the plausibility of the experimental sentences. Analyses of variance revealed that participants found no difference in plausibility between object relative (mean = 6; $SD = 0.21$) and subject relative (mean = 5.9; $SD = 0.26$) sentences ($F_1(1,19) < 1$; $F_2(1,13) < 1$).

Procedure

Same as in Experiment 1.

Results and discussion

Comprehension accuracy in the object relative and subject relative conditions was 95.9 and 96.8%, respectively, and did not differ significantly across conditions.

Reading times per word are plotted in Fig. 5. We found no significant effect of relative-clause type at the matrix verb region (mean = 382 ms, $SD = 176$ ms in subject relatives, and mean = 403 ms, $SD = 158$ ms in object relatives), $F_1(1,31) = 1.6$, $p = .21$; $F_2(1,13) = 1.52$, $p = .24$. This 21 ms difference was not significant, with a 95% confidence interval of ± 24 ms.

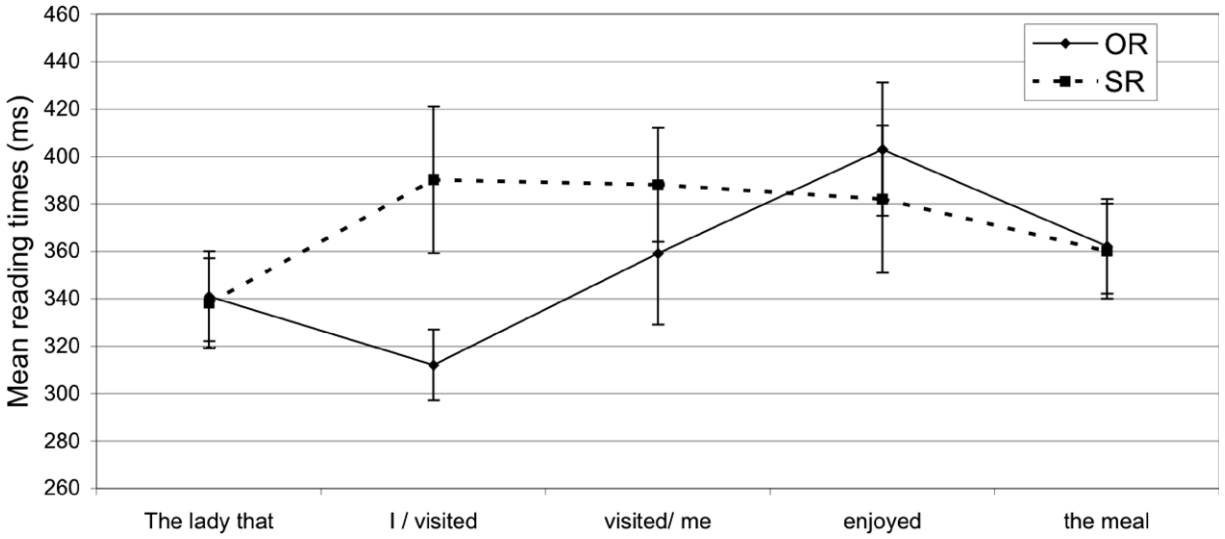

Fig. 5. Results from Experiment 2: mean reading times across regions for subject relative (dashed line) and object relative condition (solid line). The error bars correspond to the standard error for each reading time mean (SR = subject relative; OR = object relative).

Next, we analyzed the two-word critical region following the relativizer (*I visited* in the object relative condition vs. *visited me* in the subject relative condition). A 2 (Subject Relative vs. Object Relative) × 2 (word1 vs. word2) ANOVA revealed a significant effect of relative-clause type, $F_1(1,31) = 30.58$, MSE = 3004, $p < .0005$; $F_2(1,13) = 26.56$, MSE = 1513, $p < .0005$; minF'(1,34)= 14.21. In the object relative condition, the mean reading time *averaged* across the two-word region was 335 ms (mean = 312 ms, $SD = 87$ ms in word1, and mean = 359 ms, $SD = 174$ in word2). In the subject relative condition, the mean reading time in the same region was 389 ms (mean = 390 ms, $SD = 175$ ms in word1, and mean = 388 ms, $SD =$ 137 ms in word2). The 95% confidence interval for this 53.5 ms difference between condition means (389 ms minus 335 ms) was ±25.5 ms, indicating that the object relative condition was read significantly faster. Fig. 6 shows the differences between reading time means in the two analyzed regions.

Similar to Experiment 1, DLT and similarity-based interference accounts predict a significant reduction of the well-established subject-relative preference in English. However, only the experience-based accounts predict the observed preference for object relative clauses.

Experiment 3

Experiment 3 provides a further extension of Experiments 1 and 2. We replicated the design, but used materials in which a referring third-person pronoun constituted the second noun phrase. Because object relative clauses of this type are significantly more frequent than their subject relative counterparts, experience-based accounts predict a facilitation of the former.

The nature of the referential constraints associated with the stimuli of Experiment 3 differs from that of Experiments 1 and 2. The noun phrase occurring in the embedded position was a referring third person pronoun, which needed to be grounded to its referent during online processing. DLT postulates that integration costs during object relative processing are a function of referential demands (e.g., Warren & Gibson, 2002). Thus, it is not clear to what extent DLT predicts an elimination of the subject relative preference across the conditions of Experiment 3. To be conservative in testing our experimental hypothesis, we assumed that both DLT and similarity-based interference approaches could explain an elimination of the subject relative preference in Experiment 3. However, similar to Experiments 1 and 2, neither DLT nor similarity-based interference theories predict object relatives to be easier than subject relatives. On the other hand, experience-based accounts predict a facilitation of object relative clauses.

Methods

Participants

Thirty native English speakers from Cornell undergraduate classes participated in this study.

Materials

Fourteen experimental items were tested with two conditions per item. The stimuli consisted of subject/object relatives in which a referring plural third

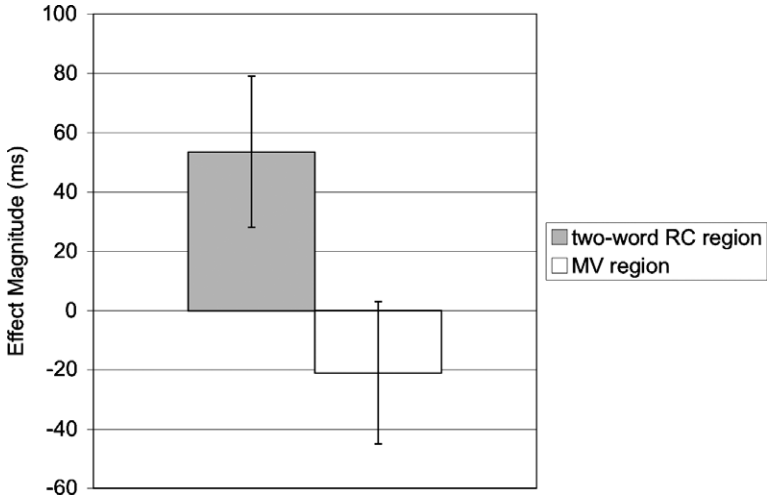

Fig. 6. Results of Experiment 2: differences between reading time means (subject relative condition minus object relative condition) in the relative-clause-internal two-word region (dark bar) and main-verb region (light bar). The error bars correspond to the 95% confidence interval for each contrast (MV = main verb; RC = relative clause).

person pronoun was the second noun phrase. Plural third-person pronouns were chosen over singular third-person pronouns because the pronoun *her* is ambiguous, in that the same word is used as a possessive marker. Sentences 6(a) and (b) are examples of the stimuli:

(6) a. According to the Taylors, the landlord that they telephoned offered a nice apartment.
b. According to the Taylors, the landlord that telephoned them offered a nice apartment.

Two experimental lists were created, each comprising fourteen experimental items and 52 filler sentences. Additionally, twenty participants rated the plausibility of the experimental sentences, finding no difference between object relative (mean = 5.8; $SD = 0.15$) and subject relative (mean = 5.8; $SD = 0.16$) sentences ($F_1(1,19) < 1$; $F_2(1,13) < 1$).

Procedure

Same as in Experiments 1 and 2.

Results and discussion

One participant answered less than 80% of the comprehension questions across the experiment correctly and was excluded from the analyses. Comprehension accuracy in the object relative and subject relative conditions was 91 and 93%, respectively, and did not differ significantly across conditions.

Reading times per word are plotted in Fig. 7. The analysis of reading times at the matrix verb region revealed no significant effect of relative-clause-type in this region (mean = 447 ms, $SD = 133$ ms in subject relatives, and mean = 436 ms, $SD = 162$ ms in object relatives), $F_1(1,28) = .196$, $p = .661$; $F_2(1,13) = 0.553$, $p = .47$. The difference between condition means was 11 ms, with a 95% confidence interval of ±36 ms.

Second, we analyzed the two-word critical region following the relativizer *that* (*they telephoned* in the object relative condition vs. *telephoned them* in the subject relative condition). A 2 (Subject Relative vs. Object Relative) × 2 (word1 vs. word2) ANOVA revealed a significant effect of relative-clause-type, $F_1(1,28) = 15.605$, MSE = 1909, $p < .001$; $F_2(1,13) = 11.44$, MSE = 1497, $p = .005$; minF'$(1,31) = 6.6$.

The reading time mean averaged across the two-word region was 343 ms in the object relative condition (mean = 339 ms in word1, $SD = 90$ ms; and mean = 347 ms, $SD = 95$ in word2). In the subject relative condition, the reading time mean in the same region was 375 ms (mean = 387 ms in word1, $SD = 104$ ms, and mean = 363 ms, $SD = 96$ ms in word2). The 95% confidence interval for the 32 ms difference between condition means in the two-word region was ±16 ms. Differences between reading-time means in the two analyzed regions are plotted in Fig. 8.

Similar to the other two experiments, this study indicates a preference for the object relative condition. The results in Experiments 1, 2 and 3 cannot be accommodated by working-memory-based theories in their current form because they do not predict object relatives to be read faster. Such a bias, nonetheless, correlates with the distributional patterns observed in the linguistic corpora. However, before concluding in favor of experience-based accounts, we should evaluate the contribution of an additional factor that might have exerted an influence on these results.

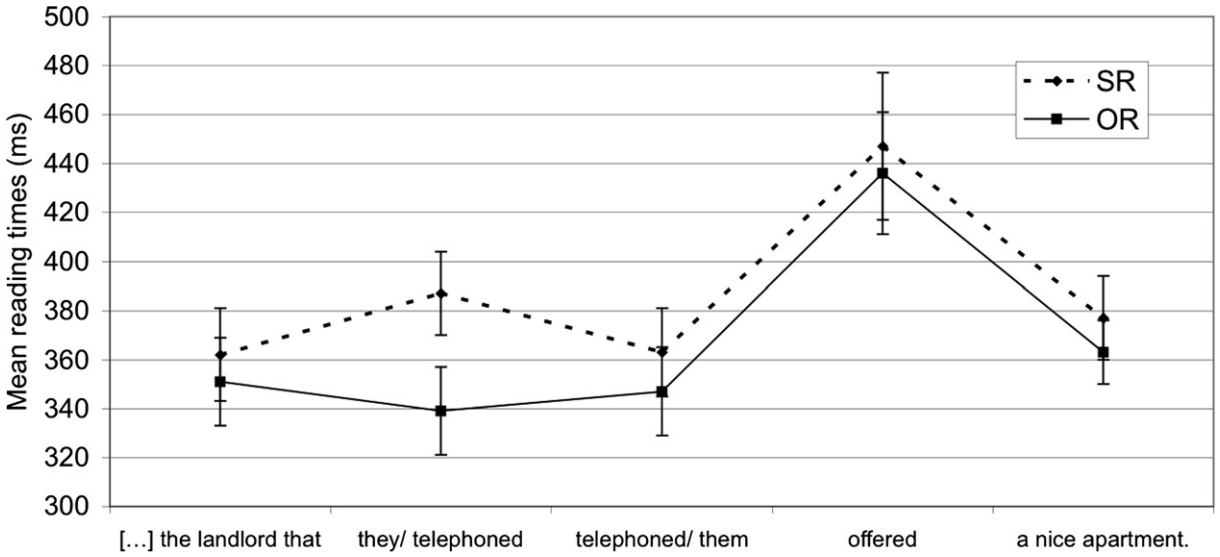

Fig. 7. Results from Experiment 3: mean reading times across regions for subject relative condition (dashed line) and object relative condition (solid line). The error bars correspond to the standard error for each reading time mean (SR = subject relative; OR = object relative).

The facilitation observed in the object relative condition could be partly due to processing spillover effects. It has been noted that one of the methodological problems related to the self-paced reading task is that the latency between key-presses often reflects processing associated with earlier displays (Mitchell, 1984; Danks, 1986). Mitchell (1984) conducted a systematic investigation of this problem, showing that reading times for a three-word display were positively correlated with the number of characters in the preceding three-word display.

In the materials used here, the verbs in the relative clause have a greater number of characters than the pronouns. Thus, the number of characters occurring in the *first word* of the target region is greater in the subject relative conditions (e.g., *called you*) than in the object relative condition (e.g., *you called*), and therefore, the processing spillover costs associated with longer verbs would remain within the two-word region in subject relatives but not in object relatives. The subordinate verbs are, on average, 3.35 characters longer than the pronoun (*you*) in Experiment 1, 5.14 characters longer than the pronoun (*I*) in Experiment 2, and 3.5 characters longer than the pronoun (*they*) in Experiment 3. Previous investigations of the processing spillover effect (Mitchell, 1984) revealed that reading times for a display showed a reliable increase of 7.68 ms per character in the immediate preceding display. Extrapolating these figures to the present studies, a comparable effect would result in an average reading-time difference between condition means of 25.8 ms at the second word of the two-word region in Experiment 1, 39.4 ms in Experiment 2 and 26.9 ms in Experiment 3. Note that spillover effects would increase subject relative reading times in *the second* word of the target region. However, in our experiments, we considered the mean reading times *averaged* across the two words comprising the target region ([reading times for word1 + reading times for word2] divided by 2). Therefore, an effect comparable to the one found by Mitchell (1984) across the two-word region would produce a difference between condition means of 12.9 ms in Experiment 1, 19.7 ms in Experiment 2 and 13.5 ms in Experiment 3. These numbers are notably smaller than the size of the observed difference between condition means in this region, and more importantly, they fall out of the 95% confidence interval in Experiment 2 (mean = 53.5 ms ± 25.5 ms) and Experiment 3 (mean = 32 ms ± 16 ms) (but not in Experiment 1 (mean = 57 ± 47 ms)). Thus, the likelihood that our results are due to processing spillover is small.

However, these considerations are not entirely conclusive. To provide a more thorough investigation of this issue, we conducted a series of regression analyses designed to explore whether the number of characters in the subordinate verb predicted the size of clause-type effect in the two-word region. Because the cost of processing spillover should be proportional to the number of characters in the first-occurring word, we should find a correlation between the length of the subordinate verb and the difference between condition means. That is, we should find a greater effect of clause type in items with longer verbs (e.g., *supervised you*) compared to shorter ones (e.g., *met you*).

We conducted 3 regression analyses (one for each Experiment) in which the dependent variable was the difference between condition means per item (e.g., the

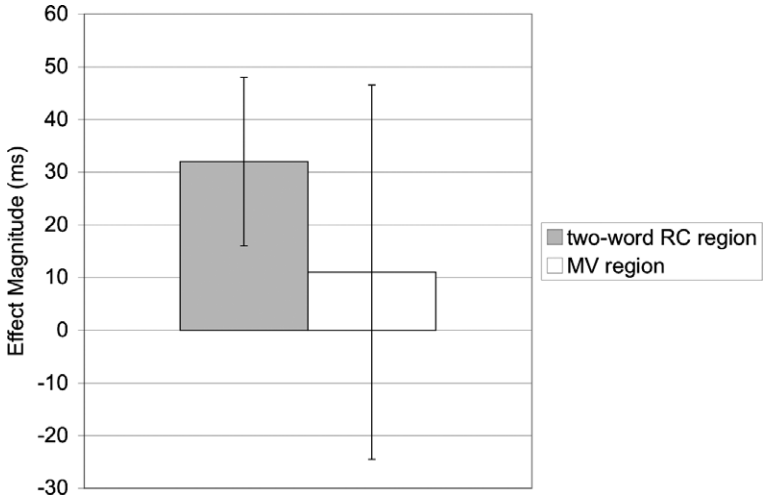

Fig. 8. Results of Experiment 3: differences between reading time means (subject relative condition minus object relative condition) in the relative-clause-internal two-word region (dark bar) and main-verb region (light bar). The error bars correspond to the 95% confidence interval for each contrast (MV = main verb; RC = relative clause).

reading time for [*supervised you*] minus the reading time for [*you supervised*]), while the independent variable was the number of characters of the subordinate verb (e.g., 10 in *supervised*). The independent variable ranged from 3 to 10 characters in Experiment 1, 5 to 11 characters in Experiment 2, and 5 to 10 characters in Experiment 3. All three regression analyses revealed no significant correlation between size of effect and the number of characters in the subordinate verb (Experiment 1: $R^2 = .12$, $F(1,12)=1.6$, $p = 0.22$; Experiment 2: $R^2 = .13$, $F(1,12) = 1.7$, $p = 0.21$; Experiment 3: $R^2 < .005$, $F(1,12)=.04$, $p = 0.83$). Although the presence of spillover effects cannot be entirely ruled out on these grounds, these results taken together with the above considerations suggest that, if present, processing spillover would not be sufficient to provide a full account of the data.

In sum, the results of Experiments 1, 2 and 3 are consistent with the predictions of the experience-based models. Although processing spillover effects may have favored the object relative condition, they fail to provide a thorough account of the results. Thus, the most parsimonious interpretation is one in which statistical information is the primary factor accounting for the reversal of the standard preference in subject/object relative clauses.

Experiment 4

Experiment 4 involved a self-paced reading task designed to continue testing the degree to which processing difficulties in pronominal object/subject relative clauses mirror the distributional patterns revealed in the corpus analysis. The stimuli used in Experiment 4 comprised object/subject relative-clause sentences in which the second noun phrase was a referring impersonal pronoun (*it*). Crucially, experience-based theories predict that, contrarily to the case of personal pronouns, object relatives should be harder to process than subject relatives. A positive integration cost may be associated to the process of referring the impersonal pronoun to the ongoing discourse. Thus, DLT also predicts subject relative clauses to be easier, in line with experience-based accounts. However, according to similarity-based interference accounts, the pronoun *it* should produce syntactic interference that is comparable to the interference produced by other pronouns during the processing of object relative constructions. Thus, similarity-based interference theories would predict a comparable amount of object relative facilitation across all Experiments including the present one.

Methods

Participants

Thirty-two native English speakers from Cornell undergraduate classes participated in this study.

Materials

Fourteen experimental items were tested with two conditions per item. The stimuli consisted of object relatives and subject relatives in which the second noun phrase was the impersonal pronoun *it* (7(a) and (b)):

(7) a. The research was very illuminating. The studies that it motivated converged to similar results.
b. The research was very illuminating. The studies that motivated it converged to similar results.

Two experimental lists comprised fourteen experimental items and forty-two fillers. Twenty participants rated the plausibility of the experimental sentences. They found no difference between object relative (mean = 5.5; $SD = 0.2$) and subject relative (mean = 5.6; $SD = 0.2$) sentences ($F_1(1,19) < 1$; $F_2(1,13) < 1$).

Procedure

Same as in Experiments 1, 2 and 3.

Results and discussion

Comprehension accuracy in the object relative and subject relative condition was 88.8 and 88.7%, respectively, and did not differ significantly across conditions.

The matrix verb region revealed an effect of relative-clause type indicating that the subject relative condition was read faster (mean = 498 ms, $SD = 149$ ms in subject relatives, and mean = 567 ms, $SD = 194$ ms in object relatives), $F_1(1,31) = 6.87$, MSE = 11,137, $p = .013$; $F_2(1,13) = 6.202$, MSE = 5606, $p = .027$; minF'$(1,35) = 3.25$. The 95% confidence interval for this 69 ms difference was ±38 ms.

Fig. 9 shows the mean reading times per word. We analyzed the two-word critical region (*it motivated* in the object relative condition vs. *motivated it* in the subject relative condition). A 2 (Subject Relative vs. Object Relative) × 2 (word1 vs. word2) ANOVA revealed an effect of relative clause condition, $F_1(1,31) = 7.42$, MSE = 4024, p=.01.; $F_2(1,13) = 6.87$, MSE = 4116, $p = .021$; minF'$(1,35) = 3.36$, indicating that subject relative clauses were read significantly faster than object relatives. Also, the analysis revealed a significant interaction between word type (word1 vs. word2) and condition $F_1(1,31) = 12.102$, MSE = 16,077, $p = .002$; $F_2(1,13) = 19.36$, MSE = 4485, 4842, $p = .001$; minF'$(1,42) = 7.44$. The reading time mean averaged across the two-word region was 487.5 ms in the object relative condition (mean = 409 ms, $SD = 108$ ms in word1, and mean = 566 ms, $SD = 264$ in word2), while in the subject relative condition the mean was 444.5 ms (mean = 444 ms, $SD = 119$ ms in word1, and mean = 445 ms, $SD = 121$ ms in word2). The 95% confidence interval for the 43 ms difference between condition means (487.5 ms minus 444.5 ms) in the two-word region was ±41 ms. Fig. 10 shows the differences between reading time means in the two analyzed regions.

As in the first three experiments, differences in object/subject relative processing difficulties mirrored the frequencies observed in the corpus analysis, providing further support for experience-based models. However, it should be noted that there may be other interpretations for the reading-time effects observed in Experiment 4. The presence of other factors is partly suggested by the fact that reading times in Experiment 4 are much slower (range, 409–567 ms per word) than in the first three experiments (range, 312–473 ms per word). It is therefore possible that the results of this study are not exclusively due to frequency matching, but rather, to a combination of processing constraints simultaneously favoring the subject relative condition.

For example, the results could be accommodated by DLT-based explanations: according to the locality hypothesis, the cost associated with syntactic integration in object relatives is a function of the complexity of the intervening discourse structure between the elements

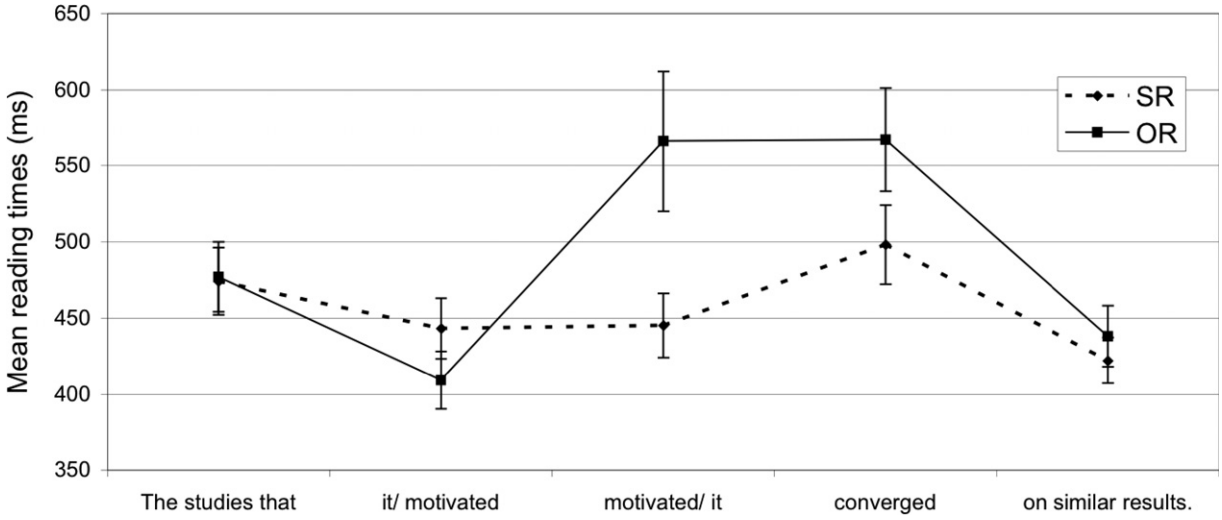

Fig. 9. Results from Experiment 4: mean reading times across regions for subject relative condition (dashed line) and object relative condition (solid line). The error bars correspond to the standard error for each reading time mean (SR = subject relative; OR = object relative).

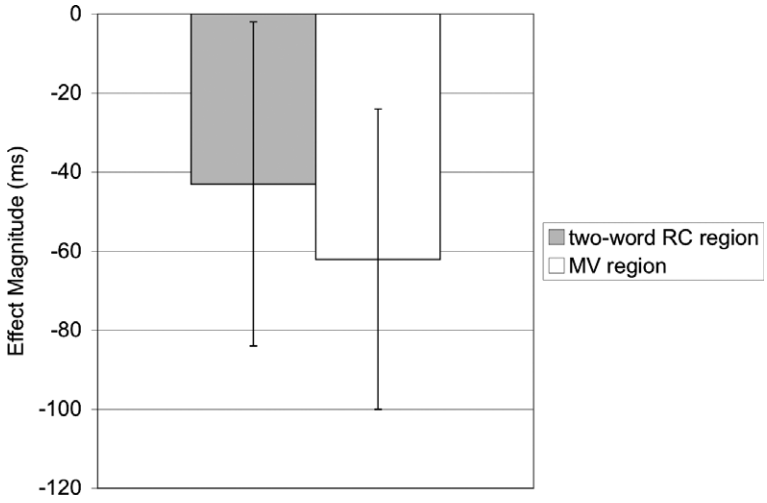

Fig. 10. Results of Experiment 4: differences between reading time means (subject relative condition minus object relative condition) in the relative-clause-internal two-word region (dark bar) and main-verb region (light bar). The error bars correspond to the 95% confidence interval for each contrast (MV = main verb; RC = relative clause).

being integrated (e.g., Warren & Gibson, 2002). In Experiment 4, discourse complexity may have increased as a result of the difficulties in storing and retrieving the antecedent of the inanimate pronoun *it*, causing reading times to slow down at the level of the embedded and matrix verb.

Additional retrieving difficulties could come from the inanimate nature of the embedded pronoun (Mak et al., 2002, Mak, Vonk, & Schriefers, 2006; Traxler, Morris, & Seely, 2002). For example, Mak et al. (2002, 2006) explored the effect of animacy on the processing of object/subject relative clauses in Dutch. They found no reading-time differences between subject and object relative sentences with inanimate sentential subjects and animate relative-clause-internal noun phrases. In contrast, they found that subject relatives were read faster in sentences with inanimate sentential subjects and inanimate relative-clause-internal noun phrases, such as in *According to the brochure, the leakages that the gel remedies, should disappear at once*, and *According to the brochure, the gel that remedies the leakages, should work at once*. Their findings suggest that the presence of an inanimate noun phrase in the relative clause increases the processing difficulty in object relatives (for an extended discussion, see Mak et al., 2006; Traxler et al., 2002).

The results of Experiments 3 and 4 provide an interesting contrast in sentences having referring pronouns in the second noun phrase position. However, a source of concern in establishing direct comparisons is the difference in the nature of discourse context used in the two studies. In Experiment 3 the antecedent of the pronoun was given within the sentence (e.g., "According to *X*, etc."), while in Experiment 4 the antecedent of the inanimate pronoun *it* was introduced in a separate sentence preceding the test one. Thus, the way in which the antecedent was introduced could have affected the results in Experiment 4. This factor may have interacted with the difficulties that arise when establishing anaphoric reference, leading to longer reading times in the object relative condition. On the one hand, the results of Experiment 3 contradict the hypothesis that anaphoric reference costs would favor the subject relative condition. However, this effect could have been more marked in Experiment 4 because the antecedent is located outside the target sentence. Thus, the reading-time effects in this last experiment may be partly due to the difficulties associated with anaphoric reference and its interaction with the discourse properties.

In sum, the results of Experiment 4 can be accommodated by a variety of different explanations. The presence of overall slower reading times suggests that non-statistical factors may be associated with the subject relative preference. When all experiments are taken together, the results favor multiple-constraint approaches according to which sentence processing is simultaneously affected by a variety of factors, including statistical information, animacy, discourse constraints and retrieval difficulties. However, it should be noted that, while experience-based accounts are not alone in explaining the results of Experiment 4, they provide the most plausible explanation for the reading-time effects observed in Experiments 1, 2 and 3.

General discussion

The large-scale corpus analysis conducted here revealed an overwhelming majority of pronominal object relative clauses compared to pronominal subject relative clauses. Recently, it has been argued that

distributional information may be important to both comprehension and production (MacDonald, 1999; Race & MacDonald, 2003). It is clear that prior comprehension experiences have an effect on the choices of syntactic structure during production (for an extended discussion, see MacDonald, 1999). For example, Race and MacDonald (2003) explored the use of the relativizer *that* in the production and comprehension of object relative clauses. They found that producers less frequently insert *that* in object relative clauses when the embedded subject is a pronoun. Other factors such as the length of the clause increased the inclusion of *that* during production, suggesting that the word *that* may be inserted to alleviate production difficulties. An additional experiment showed that comprehenders are sensitive to the observed production biases. Race and MacDonald (2003) argued for constraint-based interactions between the production and comprehension systems.

In the case of relative clause constructions, theories of complexity largely agree in that object relative sentences require more memory resources than subject relative ones. Thus, it seems implausible that the pattern of distribution observed in pronominal relative clauses emerges from pressures on the production system to reduce processing difficulties. Rather, discourse-related explanations may explain the results more naturally. In line with Fox and Thompson (1990), we suggest that the high frequency of pronominal object relative constructions may be a consequence of discourse demands. A speaker's choice for relative clauses may be guided by the need to anchor a nonhuman head noun phrase to the discourse. Two combined factors may explain our results: first, Fox and Thompson (1990) showed that nonhuman noun phrases in the sentential subject position tend to be modified by object relative rather than subject relative clauses. Second, anchoring is nearly always done by a pronoun (Fox, 1987). Thus, the majority of pronominal object relatives may result from the need to anchor a nonhuman head noun phrase to the ongoing discourse. The observed bias would therefore arise from discourse constraints that, in turn, create distributional patterns that are likely to be extremely helpful to the comprehender.

The corpus analysis suggests that exposure to language might provide an additional factor involved in the facilitation of sentences containing pronominal object relative clauses. Experiments 1, 2 and 3 showed that object relatives with embedded personal pronouns were read faster than subject relatives across the relative clause region. Finally, Experiment 4 indicated an opposite pattern of results when the impersonal pronoun *it* was the second noun phrase. Importantly, the experimental data showed that the object/subject relative differences in processing difficulty consistently mirrored the distributional patterns found in the linguistic corpora.

Our results strongly disconfirm structure-based accounts that predict a universal preference for subject relative clauses (e.g., Miyamoto & Nakamura, 2003). Instead, they support theories that rely on a combination of functional factors and cognitive constraints. However, the data cannot be accommodated by working-memory-based theories in their current form. The DLT proposed by Gibson and colleagues (Gibson, 1998; Hsiao & Gibson, 2003; Warren & Gibson, 2002) specifies the nature of the memory difficulties involved in syntactic integration, which, in turn, depends on the referential properties of the embedded noun phrase. Because first and second person pronouns have little or negligible integration cost, DLT predicts a reduction of processing difficulty in pronominal object relative clauses. However, it does not predict object relative clauses to be easier than subject relative ones, therefore failing to provide a complete explanation of the findings. It should be noted that DLT could be revised to accommodate these results, provided that it incorporates *chunk*-frequency as a factor capable of affecting memory demands during comprehension.

Similarity-based interference theories (e.g., Bever, 1974; Gordon et al., 2001; Gordon et al., 2004) predict a reduction in the object-relative/subject-relative difference in processing difficulty in Experiments 1, 2 and 3. This is because the head noun phrase is syntactically dissimilar to personal pronouns, and therefore, they would produce little interference in memory (for an extended discussion see Gordon et al., 2001, 2004). However, similarity-based interference theories do not predict object relative sentences to be *easier* than subject relatives. Finally, similarity-based interference approaches predict that the amount of syntactic interference produced by the impersonal pronoun (*it*) in the second noun phrase position in object relative constructions should be comparable to the amount of syntactic interference produced by other types of pronouns. The results in Experiment 4 are at odds with this prediction, thus disfavoring similarity-based interference interpretations.

In sum, although memory and referential constraints are likely to play a crucial role during sentence comprehension, they are not sufficient to account for the observed data. Taken together, our results point towards the need for a model that includes statistical information as a factor that may interact with other factors derived from cognitive and discourse constraints. The most parsimonious account is one in which a wide variety of constraints, including memory, referential, syntactic and statistical information simultaneously affect sentence comprehension— a view to which some working-memory-based theories also subscribe (e.g., Grodner & Gibson, 2005).

The role of multiple functional constraints is supported by recent studies suggesting that frequency factors may not be sufficient to explain some reading-time effects. For example, Gordon et al. (2004) showed that embedded

definite vs. indefinite noun phrases did not influence the magnitude of the processing difficulty even though their corpus analysis indicated a significant association between these noun phrase classes and the type of relative clause. In a different series of studies, Gordon et al. (2002) showed reliable differences in processing difficulties between subject- and object-extracted cleft constructions depending on whether the head noun and the embedded noun phrase was matched in type (e.g., occupation/proper name) to a word that participants had to remember while reading the test sentences. Frequency alone is unlikely to provide an explanation for these results, pointing toward a memory-based interpretation (see also Van Dyke & Lewis, 2003). A second kind of phenomenon that is unlikely to be explained by frequency alone is the *location* of difficulty observed during the online processing of non-pronominal relative clauses reported by Grodner and Gibson (2005). They found that the reading-time effects favoring the subject relative condition were observed only at the level of the embedded and matrix verbs. The authors argued that frequency-based accounts would predict difficulty as soon as low frequency structure starts to be processed, but the difficulty observed in this study occurred later in the sentence.

In light of these findings, we subscribe to the view that both functional and statistical factors actively interact during online comprehension. However, our approach differs from those working-memory-based theories that tend to conceive language-processing mechanisms as governed by explicit rules (e.g., Gordon et al., 2001; Lewis, 1996). We prefer to look toward language processing models in which linguistic structures and parsing strategies might arise in a self-organized fashion through learning (e.g., Elman, 1991; MacDonald & Christiansen, 2002; Tabor et al., 1997), and in which processing constraints derive from domain-general cognitive mechanisms. Along these lines, our results could be interpreted from the constraint-based perspective (MacDonald et al., 1994; Spivey, Fitneva, Tabor, & Ajmani, 2002; Stevenson, 1994; Tabor et al., 2004; Tabor & Hutchins, 2003) that conceives parsing as a dynamical process in which the construction of linguistic representations is continuously valued as a function of how much the input supports them. Other factors being equal, the strength of linguistic representations would be influenced by the reader's experience but it is hard to determine the exact nature of the interaction between competing factors, such as, for example, contextual constraints defined at the discourse level vs. statistical biases. However, a thorough investigation of the circumstances under which other factors would override statistical biases goes beyond the purpose of the present work.

To some extent, our account of the data is consistent with models of sentence processing that propose that structural frequencies influence online processing (e.g., Mitchell et al., 1995). In our view, however, frequency tabulation should not be restricted to a single level of granularity. Rather, in the spirit of connectionist approaches, we look toward a model of sentence processing in which the system is influenced by statistical information defined at multiple levels of abstraction. According to this view, the parsing system is continuously making graded expectations based on contextual information, thus defining a probability landscape for possible sentence continuations (Elman, 1990, 1995). Expectations should then be a function of a wide variety of factors including syntax, semantics, discourse and, importantly, exposure to language. Thus, the probability landscape for possible continuations may be defined at the level of lexical items or classes of lexical items ranging from individual words to part-of-speech categories as a function of contextual constraints. In turn, exposure to sequential material at the constituent level may play a crucial role during the processing of relative clause structure. This is because, before encountering the relative clause, readers have processed sentence fragments such as *The N that*...that lead to expectations toward a set of possible continuations. When the system is then exposed to a high-frequency sentence continuation (such as *that I VERB*), integration with the ongoing sentence should be facilitated. This is, in part, because readers' expectations are influenced by exposure to sequences of words (or classes of words) that have been repeatedly used in similar contexts.

In addition, access to constituent representation may be facilitated by exposure to frequent word sequences defined at the constituent level. In the spirit of the constructivist view outlined in Bybee (2002; Bybee & Scheibman, 1999), we hypothesize that the representation of constituent structure may be shaped by language use and frequency of occurrence. Representation of constituents may therefore have different degrees of cohesion due to the differences in the co-occurrence patterns of specific word sequences in the input. Thus, while the system constructs a parse exposure to relative clauses formed by frequent chunks (such as *that I VERB*) would lead to stronger representations that have become fluent through language use and repetition, and therefore are quite relatively easy to access. Importantly, sequential material may be defined at different levels of abstraction (Bybee, 2002). For example, repetition of sentence fragments of the type '*The N that I like*' etc., '*The N I know*' or '*The N, which I saw under the table*' etc., would lead to schematized relative clause representations formed by sequential material with shared parts, such as *(Relativizer) I VERB*, in which the elements differ in their level of abstraction. Because pronominal object relative clauses of this type are extremely frequent, their representation is likely to have been consolidated through experience and repetition.

To conclude, our results provide strong evidence indicating that, consistent with experience-based theories,

object relative clauses with personal pronouns in the second noun phrase position are easier to process than their subject relative counterparts. More generally, these findings suggest that statistical information must be taken into account by theories of relative clause processing.

Acknowledgments

We are grateful to Michael Spivey, Thomas Farmer and Jennifer Misyak for helpful discussions on this work. We also thank Edward Gibson, Don Mitchell and an anonymous reviewer for providing insightful comments and suggestions regarding an earlier version of this manuscript.

Appendix A. Experiment 1

1. The lady that [visited you]/[you visited] enjoyed the pool in the back of the house.
2. The professor that [met you]/[you met] was extremely impressed with your verbal skills.
3. The person that [loved you]/[you loved] attended your graduation.
4. The teacher that [disliked you]/[you disliked] distributed an extremely hard assignment.
5. The detective that [distrusted you]/[you distrusted] resigned before the case was solved.
6. The consultant that [called you]/[you called] emphasized the need for additional funding.
7. The students that [liked you]/[you liked] invited all your friends to the party.
8. The guy that [supervised you]/[you supervised] worked for the company for twenty years.
9. The manager that [promoted you]/[you promoted] acted in the best interest of the company.
10. The landlord that [phoned you]/[you phoned] offered a nice apartment on the west side of the town.
11. The woman that [hosted you]/[you hosted] planned a delicious dinner to celebrate the occasion.
12. The girl that [dated you]/[you dated] decided to throw a party to celebrate the beginning of the semester.
13. The executive that [fired you]/[you fired] knew the truth about the missing letters.
14. The agency that [hired you]/[you hired] doubled the benefits last year.

Appendix B. Experiment 2

1. The lady that [visited me]/[I visited] enjoyed the meal.
2. The teacher that [praised me]/[I praised] wrote excellent recommendation letters.
3. The neighbor that [despised me]/[I despised] lived in the same building for twenty years.
4. The clerk that [trusted me]/[I trusted] had been working for the company since 1980.
5. The landlord that [called me]/[I called] offered a nice apartment on the west side of the town.
6. The woman that [hosted me]/[I hosted] planned a delicious dinner to celebrate the occasion.
7. The businessman that [contacted me]/[I contacted] set up a meeting to close the deal.
8. The girl that [hated me]/[I hated] criticized the article in the Ithaca journal.
9. The professor that [disliked me]/[I disliked] distributed an extremely hard assignment.
10. The boy that [liked me]/[I liked] sent flowers for my birthday.
11. The dancer that [phoned me]/[I phoned] decided to invite her friends to the opening.
12. The director that [interviewed me]/[I interviewed] had a small but cozy office.
13. The guy that [teased me]/[I teased] left the room after everyone else.
14. The salesman that [denounced me]/[I denounced] was aware of the fraud since last year.

Appendix C. Experiment 3

1. According to the Smiths, the lady that [visited them]/[they visited] enjoyed the pool in the back of the house.
2. According to the plumbers, the person that [supervised them]/[they supervised] fixed the problem very easily.
3. According to the students, the teacher that [praised them]/[they praised] wrote excellent recommendation letters.
4. According to the girls, the professor that [disliked them]/[they disliked] distributed an extremely hard assignment.
5. According to the witnesses, the lawyer that [distrusted them]/[they distrusted] resigned before the final trial.
6. According to the authors, the scientist that [cited them]/[they cited] failed to understand the underlying message of the book.
7. According to the businessmen, the consultant that [called them]/[they called] emphasized the need for additional funding.
8. According to the costumers, the waitress that [liked them]/[they liked] lost her job the very next day.
9. According to the owners, the employee that [trusted them]/[they trusted] worked for the same company since 1980.
10. According to the Taylors, the landlord that [telephoned them]/[they telephoned] offered a nice apartment on the west side of the town.
11. According to the Moores, the woman that [hosted them]/[they hosted] planned a delicious dinner to celebrate the occasion.
12. According to the workmen, the salesman that [denounced them]/[they denounced] was aware of the fraud since last year.
13. According to the policemen, the agent that [questioned them]/[they questioned] overlooked an important detail.
14. According to the guys, the agency that [hired them]/[they hired] doubled the benefits last year.

Appendix D. Experiment 4

1. The research was very illuminating. The studies that [motivated it]/[it motivated] converged on similar results.
2. The attack was completely unexpected. The situation that [triggered it]/[it triggered] was shown on TV.
3. The chemical reaction was unique. The mechanisms that [activated it]/[it activated] had been studied for many years.
4. The blue jet ski accelerated. The speedboat that [followed it]/[it followed] had been repaired recently.
5. The finding was very controversial. The articles that [refuted it]/[it refuted] stimulated the debate.
6. The minivan was really fast. The car that [chased it]/[it chased] lost control suddenly.
7. The decision had been made. The events that [caused it]/[it caused] were covered by the media.
8. The argument was long lasting. The fight that [created it]/[it created] centered on hot political issues.
9. The major-league club was training. The team that [defeated it]/[it defeated] hired a new coach.
10. The show was about country-life. The sitcom that [supplanted it]/[it supplanted] had been developed from a play.
11. The hand-drawn picture was peculiar. The watercolor that [emulated it]/[it emulated] portrayed a battle.
12. The nightingale was sitting quietly. The female that [saw it]/[it saw] flew from one tree to another.
13. The wild creature was amazing. The animal that [hunted it]/[it hunted] lived in the African savanna.
14. The gossip was very harmful. The quarrels that [started it]/[it started] grew out of proportions.

References

Abney, S. P. (1991). Parsing by chunks. In R. C. Berwick, S. P. Abney, & C. Tenny (Eds.), *Principle-based parsing: Computation and psycholinguistics* (pp. 257–278). Dordrecht: Kluwer Academic Publishers.

Abney, S. (1996). Part-or-speech tagging and partial parsing. In K. Church, S. Young, & G. Bloothooft (Eds.), *Corpus-based methods in language and speech* (pp. 1–23). Dordrecht: Kluwer Academic Publishers.

Bates, E., & MacWhinney, B. (1987). Competition, variation, and language learning. In B. MacWhinney (Ed.), *Mechanisms of Language Acquisition* (pp. 157–193). Hillsdale, NJ: Erlbaum.

Bever, T. G. (1970). The cognitive basis for linguistic structures. In J. R. Hayes (Ed.), *Cognition and the development of language* (pp. 279–332). New York: Wiley.

Bever, T. G. (1974). The ascent of the specious, or there's a lot we don't know about mirrors. In D. Cohen (Ed.), *Explaining linguistic phenomena* (pp. 173–200). Washington: Hemisphere.

Biber, D. (1988). *Variation across speech and writing*. New York: Cambridge University Press.

Biber, D. (1995). *Dimension of register variation*. New York: Cambridge University Press.

Bybee, J. (2002). Sequentiality as the basis of constituent structure. In T. Givón & B. Malle (Eds.), *The evolution of language out of pre-language* (pp. 107–132). Philadelphia, PA: John Benjamins.

Bybee, J., & Scheibman, J. (1999). The effect of usage on degrees of constituency the reduction of *don't* in English. *Linguistics, 37*, 575–596.

Chomsky, N., & Miller, G. A. (1963). Introduction to the formal analysis of natural languages. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.). *Handbook of mathematical psychology* (vol. 2, pp. 269–321). New York: Wiley.

Christiansen, M. H. (1994). *Infinite languages, finite minds: Connectionism, learning and linguistic structure*. Unpublished Ph.D. dissertation. University of Edinburgh, Scotland.

Cohen, J. D., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: an interactive graphic system for designing and controlling experiments in the psychology laboratory using Macintosh computers. *Behavioral Research Methods, Instruments and Computers, 25*(2), 257–271.

Crocker, M. W., & Corley, S. (2002). Modular architectures and statistical mechanisms. In P. Merlo & S. Stevenson (Eds.), *The lexical basis of sentence processing* (pp. 157–180). Amsterdam, Philadelphia: John Benjamins Publishing Company.

Danks, J. (1986). Identifying component processes in test comprehension: comment on Haberlandt and Graesser. *Journal of Experimental Psychology: General, 115*, 193–197.

Desmet, T., De Baecke, C., Drieghe, D., Brysbaert, M., & Vonk, W. (2005). Relative clause attachment in Dutch: on-line comprehension corresponds to corpus frequencies when lexical variables are taken into account. *Language and Cognitive Processes, 21*, 453–485.

Desmet, T., & Gibson, E. (2003). Disambiguation preferences and corpus frequencies in noun phrase conjunction. *Journal of Memory and Language, 49*, 353–374.

Elman, J. L. (1991). Distributed representations, simple recurrent networks, and grammatical structure. *Machine Learning, 7*, 195–225.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science, 14*, 179–211.

Elman, J. L. (1995). Language as a dynamical system. In R. F. Port & T. van Gelder (Eds.), *Mind as motion: explorations in the dynamics of cognition* (pp. 195–223). Cambridge, MA: MIT Press.

Farmer, T. A., Christiansen, M. H., & Monaghan, P. (2006). Phonological typicality influences on-line sentence comprehension. *Proceedings of the National Academy of Sciences, 103*, 12203–12208.

Fodor, J. D. (1998). Learning to parse? *Journal of Psycholinguistic Research, 27*, 285–319.

Ford, M. (1983). A method for obtaining measures of local parsing complexity throughout sentences. *Journal of Verbal Learning and Verbal Behavior, 22*, 203–218.

Fox, B. A. (1987). The noun phrase accessibility hierarchy: subject primacy or the absolutive hypothesis? *Language, 63*, 856–870.

Fox, B. A., & Thompson, S. A. (1990). A discourse explanation of the grammar of relative clauses in English conversation. *Language, 66*(2), 297–316.

Frisson, S., Rayner, K., & Pickering, M. J. (2005). Effects of contextual predictability and transitional probability on eye movements during reading. *Journal of Experimental Psychology: Learning, Memory and Cognition, 31*, 862–877.

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

Gibson, E. (2000). The dependency locality theory: a distance-based theory of linguistic complexity. In Y. Miyashita, A. Marantz, & W. O'Neil (Eds.), *Image, language, brain* (pp. 95–126). Cambridge, MA: MIT Press.

Gibson, E., & Schütze, C. T. (1999). Disambiguation preferences in noun phrase conjunction do not mirror corpus frequency. *Journal of Memory and Language, 40*, 263–279.

Gordon, P. C., Hendrick, R., & Johnson, M. (2001). Memory interference during language processing. *Journal of Experimental Psychology: Learning, Memory and Cognition, 27*, 1411–1423.

Gordon, P. C., Hendrick, R., & Johnson, M. (2004). Effects of noun phrase type on sentence complexity. *Journal of Memory and Language, 51*, 97–114.

Gordon, P. C., Hendrick, R., & Levine, W. H. (2002). Memory load interference in syntactic processing. *Psychological Science, 13*, 425–430.

Grodner, D., & Gibson, E. (2005). Consequences of the serial nature of linguistic input. *Cognitive Science, 29*, 261–291.

Gundel, J., Hedberg, H., & Zacharski, R. (1993). Referring expressions in discourse. *Language, 69*, 274–307.

Holmes, V. M., & O'Regan, J. K. (1981). Eye fixation patterns during the reading of relative clause sentences. *Journal of Verbal Learning and Verbal Behavior, 20*, 417–430.

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

Ide, N., & Suderman, K. (2004). In *Proceedings of the Fourth Language Resources and Evaluation Conference* (pp. 1681–1684) Lisbon, Portugal.

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

Just, M. A., Carpenter, P. A., & Woolley, J. D. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General, 3*, 228–238.

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

Konieczny, L. (2005). The psychological reality of local coherences in sentence processing. *Proceedings of the 27th Annual Conference of the Cognitive Science Society*.

Lewis, R. (1996). A theory of grammatical but unacceptable embeddings. *Journal of Psycholinguistic Research, 25*, 93–116.

Lewis, R. L., & Vasishth, D. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science, 29*, 375–421.

MacDonald, M. C. (1994). Probabilistic constraints and syntactic ambiguity resolution. *Language and Cognitive Processes, 12*, 121–136.

MacDonald, M. C. (1999). Distributional information in language comprehension, production, and acquisition: three puzzles and a moral. In B. MacWhinney (Ed.), *The emergence of language* (pp. 177–196). Mahwah, NJ: Erlbaum.

MacDonald, M. C., & Christiansen, M. H. (2002). Reassessing working memory: a comment on Just & Carpenter (1992) and Waters & Caplan (1996). *Psychological Review, 109*, 35–54.

MacDonald, M., Pearlmutter, N., & Seidenberg, M. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review, 101*, 676–703.

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

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

Masson, M. E. J., & Loftus, G. R. (2003). Using confidence intervals for graphically based data interpretation. *Canadian Journal of Experimental Psychology, 57*, 203–220.

McDonald, S. A., & Shillcock, R. C. (2003). Eye-movements reveal the on-line computation of lexical probabilities during reading. *Psychological Science, 14*, 648–652.

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

Mitchell, D. C. (1984). An evaluation of subject-paced reading tasks and other methods for investigating immediate processes in reading. In D. Kieras & M. A. Just (Eds.), *New Methods in Reading Comprehension Research* (pp. 69–89). Hillsdale, NJ: Earlbaum.

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

Miyamoto, E., & Nakamura, M. (2003). Subject/Object asymmetries in the processing of relative clauses in Japanese. In G. Garding, & M. Tsujimura (Eds.), *Proceedings of WCCFL*, 22, 342–355.

Pearlmutter, N. J., & MacDonald, M. C. (1992). Plausibility and syntactic ambiguity resolution. In *Proceedings of the 14th annual conference of the cognitive science society* (pp. 498–503). Hillsdale, NJ: Lawrence Erlbaum Associates.

Race, D. S., & MacDonald, M. C. (2003). The use of "that" in the production and comprehension of object relative clauses. In *Proceedings of the 25th Annual Meeting of the Cognitive Science Society* (pp. 946–951). Mahwah, NJ: Lawrence Erlbaum.

Reali, F., & Christiansen, M. H. (2006). Word-chunk frequencies affect the processing of pronominal object-relative clauses. *Quarterly Journal of Experimental Psychology*, in press.

Seidenberg, M. S. (1997). Language acquisition and use: learning and applying probabilistic constraints. *Science, 275*, 1599–1603.

Seidenberg, M. S., & MacDonald, M. C. (1999). A probabilistic constraints approach to language acquisition and processing. *Cognitive Science, 23*, 569–588.

Snedeker, J., & Trueswell, J. C. (2004). The developing constraints on parsing decisions: the role of lexical-biases and referential scenes in child and adult sentence processing. *Cognitive Psychology, 49*, 238–299.

Spivey, M., Fitneva, S., Tabor, W., & Ajmani, S. (2002). The timecourse of information integration in sentence process-

ing. In P. Merlo & S. Stevenson (Eds.), *The lexical basis of sentence processing: formal, computational, and experimental perspectives* (pp. 207–232). Amsterdam, Philadelphia: John Benjamins Publishing.

Spivey-Knowlton, M., & Sedivy, J. (1995). Resolving attachment ambiguities with multiple constraints. *Cognition, 55*, 227–267.

Stevenson, S. (1994). Competition and recency in a hybrid network model of syntactic disambiguation. *Journal of Psycholinguistic Research, 23*, 295–322.

Tabossi, P., Spivey-Knowlton, M. J., McRae, K., & Tanenhaus, M. K. (1994). Semantic effects on syntactic ambiguity resolution: evidence for a constraint-based resolution process. In C. Umilta & M. Moscovitch (Eds.), *Attention and Performance XV* (pp. 589–616). Cambridge, MA: MIT Press.

Trueswell, J. C. (1996). The role of lexical frequency in syntactic ambiguity resolution. *Journal of Memory and Language, 35*, 566–585.

Tabor, W., Galantucci, B., & Richardson, D. (2004). Effects of merely local syntactic coherence on sentence processing. *Journal of Memory and Language, 50*, 355–370.

Tabor, W., & Hutchins, S. (2003). Evidence for self-organized sentence processing: digging-in effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*, 431–450.

Tabor, W., Juliano, C., & Tanenhaus, M. K. (1997). Parsing in a dynamical system: an attractor-based account of the interaction of lexical and structural constraints in sentence processing. *Language and Cognitive Processes, 12*, 211–272.

Tanenhaus, M. K., & Trueswell, J. C. (1995). Sentence Comprehension. In J. L. Miller & P. D. Eimas (Eds.), *Speech, Language, and Communication* (pp. 217–262). San Diego: Academic Press.

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

Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. M. (1994). Semantic influences on parsing: use of thematic role information in syntactic disambiguation. *Journal of Memory and Language, 33*, 285–318.

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

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

Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge: Cambridge University press.