

The nature of the working memory system underlying language processing and its relationship to the long-term memory system

by

Evelina G. Fedorenko

A.B. Psychology/Linguistics
Harvard University, 2002

SUBMITTED TO THE DEPARTMENT OF BRAIN AND COGNITIVE SCIENCES IN
PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY IN COGNITIVE SCIENCE
AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

SEPTEMBER 2007

© 2007 Evelina G. Fedorenko. All rights reserved.

The author hereby grants to MIT permission to reproduce
and to distribute publicly paper and electronic
copies of this thesis document in whole or in part
in any medium now known or hereafter created.

Signature of Author: _____

Department of Brain and Cognitive Sciences

June 6, 2007

Certified by: _____

Nancy Kanwisher

Department of Brain and Cognitive Sciences

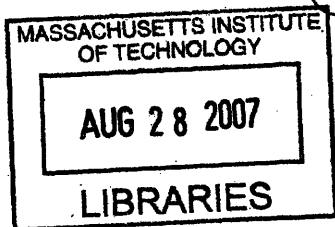
Thesis Supervisor

Accepted by: _____

Matthew Wilson

Department of Brain and Cognitive Sciences

Chairman, Committee for Graduate Students



ARCHIVES

The nature of the working memory system underlying language processing and its relationship to the long-term memory system

by

Evelina G. Fedorenko

Submitted to the Department of Brain and Cognitive Sciences
on June 6, 2007 in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy
in Cognitive Science

ABSTRACT

This thesis examines two questions concerning the working memory system underlying language processing: (1) To what extent is the working memory system underlying language processing domain-specific? and (2) What is the relationship between the working memory system and the long-term memory system in language processing? In **Chapter 1**, I describe ten experiments investigating the extent to which the working memory system underlying linguistic integrations is domain-specific. I argue that the results of these experiments demonstrate that at least some aspects of the working memory system used for linguistic integrations are not domain-specific, being involved in arithmetic, and possibly, musical processing. In **Chapter 2**, I describe six experiments investigating the relationship between the two retrieval operations that are required when an incoming word is integrated into an evolving structure: the retrieval of the lexical properties of the word from long-term memory and the retrieval of its structural dependents from working memory. I demonstrate that the relative ease or difficulty of retrieving the lexical properties of an incoming word affect the difficulty of retrieving its structural dependents. I therefore argue that the two retrieval operations rely on overlapping pools of resources.

Thesis supervisor: Nancy Kanwisher

Title: Ellen Swallow Richards Professor of Cognitive Neuroscience

Acknowledgments

First and foremost, I would like to thank my advisor Nancy Kanwisher for continuing support and inspiration in both academic and personal aspects of my life. I would also like to thank the members of my committee: David Caplan, John Gabrieli, and Ken Wexler, for their insightful comments about different aspects of this work throughout my graduate career.

In addition, I would like to say big thanks to all the faculty members at BCS – past and present – who have made my experience at MIT the best years of my life. In particular, I'd like to thank Lera Boroditsky, Steve Pinker, Molly Potter, Rebecca Saxe, and Josh Tenenbaum.

I am also more grateful than I could ever express in words to my former advisor (now husband) Ted Gibson. I am incredibly lucky to have met Ted on my life path. Thanks, Ted, for making me so happy, for making me laugh with your silly jokes and stories without a point (which you tell me over and over again), for holding me whenever I need to be held, for accepting me with all my crazy quirks and my inability to cook, for deciding to spend your life with me, and for letting me be the boss in the family.

I would also like to thank all the members of TedLab – past and present – for helpful discussions and never-ending fun over the last five years: especially Mara Breen, Timmy Desmet, Mike Frank, Dan Grodner, John Kraemer, Steve Piantadosi, Michael Wagner, and Florian Wolf. I would like to express particular thanks to Doug Rohde who was a collaborator on a lot of the work described in the first chapter.

In addition to my friends at TedLab, I would like to thank all my other friends and collaborators who have always been so supportive and caring and many of whom have provided great feedback on my research over the years: Asaf Bachrach, Bill Badecker, Daniel Casasanto, Laura Dilley, Tali Ditman, Jill Hardee, Barbara Hemforth, Phil Holcomb and Kathy Midgley, Franny Hsiao, Florian Jaeger, Roger Levy, Brad Mahon, Josh McDermott, Meryane and Mark Murphy, Jared Novick, Ani Patel, Neal Pearlmuter, Amy Perfors, Ivan Sag, Lauren Schmidt, Becca Schwarzlose, Paul Smolensky, Laura Staum, Adam Szczegieliak, Sashank Varma, Gloria Waters, Jon Winawer.

I would also like to express gratitude to the many UROPs who have worked in the lab and helped me with various projects. In particular, I'd like to thank Nicole Bieber, Meredith Brown, Stephanie Chan, Charlene Chuang, Tess Diduch, Sarah Gibson, Denise Ichinco, Daniel Jaramillo, Joyce Kim, Jenny Lee, Anna Lo, Diana Lo, Nikki Longe, Rina Patel, Anna Schapiro, and Lisa Shank.

Finally, I would like to thank my family without whose love and support I would never have gotten to where I am right now. Thanks for believing in me.

I would like to dedicate this thesis to my parents Svetlana and Georgiy.

Table of contents

Introduction	6
Layout of the thesis	8
Chapter 1:	
The extent of domain-specificity of the working memory system underlying linguistic integrations	10
General background	10
1A. The heterogeneity of the working memory resources underlying language processing: storage costs vs. integration costs	12
Background	12
Experiment 1A	15
Conclusions	20
1B. Investigating working memory tasks with similar integration processes	21
Background	21
Background for Experiments 1B-1 through 1B-6	22
Experiment 1B-1: Linguistic vs. Arithmetic integrations	24
Experiments 1B-2 and 1B-3: Linguistic vs. Arithmetic integrations (controlling for morpho-phonological complexity in the arithmetic task)	31
Experiment 1B-2	31
Experiment 1B-3	35
Experiment 1B-4: Linguistic vs. Spatial integrations (the “Pie” spatial-rotation task)	40
Experiment 1B-5: Linguistic vs. Spatial integrations (the “Grid” spatial-integration task)	50
Experiment 1B-6: The spatial-rotation task with verbal instructions	54
Conclusions for Experiments 1B-1 through 1B-6	58
Background for Experiments 1B-7a and 1B-7b	62
Experiments 1B-7a and 1B-7b: Linguistic vs. Musical integrations	63
Experiment 1B-7a	64
Experiment 1B-7b	73
Conclusions for Experiments 1B-7a and 1B-7b	78
General conclusions	78
1C. Investigating verbal working memory tasks with similar materials	79
Background	79
Experiment 1C-1: Linguistic integrations and a verbal working memory task with similar materials	80
Conclusions	90
Chapter 2:	
The relationship between the working memory system and the long-term memory system in language processing	91
General background	91
2A. The difficulty of retrieving the lexical properties of an incoming word increases or delays the difficulty of retrieving the word’s structural dependents.	96
Background for Experiments 2A-1 through 2A-3	96

Experiments 2A-1 – 2A-3: The effects of the difficulty of retrieving the word’s thematic properties on the difficulty of retrieving its structural dependents	100
Experiment 2A-1	100
Experiment 2A-2	104
Experiment 2A-3	109
Conclusions for Experiments 2A-1 through 2A-3	114
Background for Experiment 2A-4	115
Experiment 2A-4: The effects of the difficulty of retrieving the word’s meaning on the difficulty of retrieving its structural dependents	115
Conclusions for Experiment 2A-4	124
General conclusions	124
2B. The ease of retrieving the lexical properties of an incoming word alleviates the difficulty of retrieving the word’s structural dependents.	125
Experiments 2B-1 and 2B-2: Effects of contextual predictability on the difficulty of retrieving the word’s structural dependents	125
Experiment 2B-1	126
Experiment 2B-2	130
Conclusions for Experiments 2B-1 and 2B-2	134
Summary and Conclusions	136
References	139
Appendices	
Appendix A	146
Appendix B	150
Appendix C	151
Appendix D	152
Appendix E	154
Appendix F	155
Appendix G	156
Appendix H	157
Appendix I	159
Appendix J	161
Appendix K	164
Appendix L	165
Appendix M	166
Appendix N	168

Introduction

In every cognitive domain, in order to process information we rely on the *long-term memory* system where the domain-relevant knowledge is stored and on the *working memory* system which is used for retrieving the relevant information from the long-term store and processing and manipulating the incoming information.

The long-term memory system used in language processing contains information about language-specific information units, such as phonemes, morphemes, words, sentences and discourse units, as well as more domain-general information units, such as, for example, events (i.e. world knowledge about the relative likelihoods of different kinds of events). What we learn about different information units over the course of our lifetimes affects our processing of subsequent occurrences of these units and may guide our subsequent learning processes. All these experience-based constraints can be referred to as *informational constraints*. Informational constraints are likely to differ across individuals due to the uniqueness of people's experiences with the environment. However, they are likely to be similar across groups of individuals that share some aspects of their environment, such as people speaking the same language.

The working memory system used in language processing – similar to other domains – has a limited capacity, such that there are constraints on how much information we can process and manipulate at any given point in time (e.g., Miller, 1956; Cowan, 2001). Unlike the informational constraints, these constraints on the cognitive architecture of our system (the *resource constraints*) are much less shaped by experience: we are born with these resource limitations and they remain relatively constant throughout our life spans. Resource constraints do differ across individuals but arguably to a much lesser extent.

This thesis will investigate the nature of the working memory system underlying language processing and its relationship to the long-term memory system.

Two important properties of human language are: (1) sequentiality, and (2) the presence of non-local structural dependencies. First, language is sequential in nature with the words coming in one at a time. Second, linguistic utterances are characterized by a dependency structure, which involves dependencies between the immediately adjacent elements (local dependencies), as well as dependencies between elements separated by other elements (non-local dependencies). For example, in (1a) the dependency between the noun phrase *the student* and the verb *filed* is local, but in (1b) and (1c), the noun phrase *the student* is separated from its verb by a prepositional phrase (PP) *in the office by the window* and by a relative clause (RC) *who was obsessive-compulsive*, respectively:

(1a) The student filed the papers.

(1b) The student in the office by the window filed the papers.

(1c) The student who was obsessive-compulsive filed the papers.

Local dependencies – where the two elements of the dependency are immediately adjacent – do

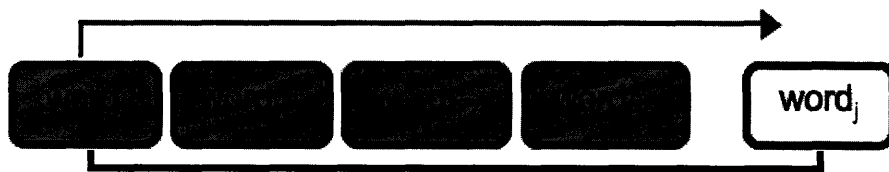
not pose difficulty for the working memory system because at the point of encountering the second element of the dependency, the representation of the first element of the dependency is still highly active in working memory. In non-local dependencies, however, the issue of the limited capacity of the working memory system becomes highly relevant because upon encountering the second element of the dependency, some aspects of the first element of the dependency have to be retrieved from memory.

Gibson (1998, 2000) has argued that in studying the working memory system underlying language processing, it is important to distinguish between two types of working memory costs associated with the processing of non-local structural dependencies. In particular, Gibson argued that working memory is required for (1) keeping track of incomplete syntactic dependencies (**storage costs**; e.g., Chomsky & Miller, 1963; Wanner & Maratsos, 1978; Lewis, 1996), and (2) for integrating incoming words to earlier words/positions in the sentence (**integration costs**; e.g., Ford, 1983). The *storage cost component* was quantified in terms of the number of syntactic categories minimally required to complete the input string as a grammatical sentence and was postulated to arise at the point of predicting a syntactic category and to last until the target syntactic category is encountered. The *integration cost component* was quantified in terms of the distance between the two ends of a syntactic dependency and was postulated to arise at the second element of the dependency where the information about the first element needs to be retrieved from memory. Since Gibson's proposal, other proposals have been advanced arguing for alternative ways to quantify these two types of working memory costs (e.g., *integration costs*: Gordon et al., 2001, 2004; Van Dyke & Lewis, 2003, Vasishth & Lewis, 2005).

Consider the graphic illustrations of the storage and integration costs in (2a) and (2b), respectively. In (2a), a prediction for an upcoming element ($Word_j$) is generated at $Word_i$ and so the cost (associated with maintaining this prediction) is incurred during the processing of $Word_i - Word_{i+3}$ and it is alleviated at $Word_j$ when the prediction is met. In (2b), a cost is incurred during the processing of $Word_j$, because in order to integrate this word to $Word_i$, it is necessary to retrieve aspects of $Word_i$ from memory. Unlike in (2a), no cost is incurred during the processing of $Word_{i+1} - Word_{i+3}$.

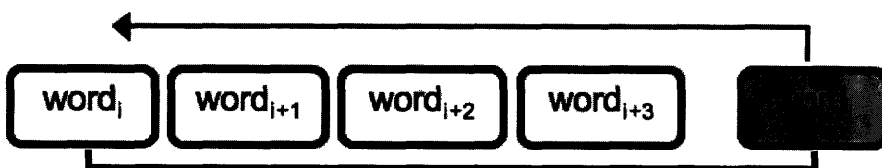
(2a)

Storage costs:



(2b)

Integration costs:



These two types of linguistic processes – keeping track of incomplete syntactic dependencies and integrating incoming words to earlier words/positions in the sentence – can be conceived of as *eager*, or early (forward-looking), processing (*storage costs*) and *lazy*, or late¹ (backward-looking) processing (*integration costs*). Depending on the assumptions about what the *storage* process involves, it may seem that these two types of linguistic processes are redundant. In particular: if the storage process – keeping track of the number of syntactic categories minimally required to complete the input string as a grammatical sentence – involves keeping the first element of a dependency highly active in working memory (e.g., a noun which requires a verb), then upon encountering the second element of the dependency (the verb), there should be no cost associated with retrieving aspects of the first element from memory (since it is still highly active in working memory), i.e. no *integration* cost. If, on the other hand, the storage process involves predicting aspects of the upcoming second element of a dependency (such as, for example, lexico-semantic properties of an upcoming verb) without maintaining the first element of the dependency highly active in working memory, then upon encountering the second element of the dependency, there may still be a cost associated with retrieving aspects of the first element from memory, i.e. an *integration* cost.

If the former conceptualization of the storage process is correct, then – in terms of computational efficiency – it is plausible that performing one or the other type of processing (*eager* or *lazy*) is a choice that the human parser can make: resources can be allocated to keeping the first elements of structural dependencies active in working memory, thereby reducing the average processing speed but avoiding the costs associated with completing the dependencies; or, alternatively, instead of allocating resources to keeping the first elements of structural dependencies active in working memory, the comprehenders can use these resources to increase the average processing speed, but this will result in costs associated with completing the dependencies. The choice of the parser between *eager* and *lazy* processing may depend on factors like construction type, language, general cognitive load, or even differences among individual cognitive architectures. However, the nature of the storage process remains an open question, and addressing this question is beyond the scope of this thesis².

The thesis will be primarily concerned with the nature of the working memory system underlying the process of linguistic integrations and with the relationship between this working memory system and the long-term memory system.

Layout of the thesis

While research in the past several decades has substantially advanced our understanding of the language system, as well as both the long-term and the working memory systems across different domains, several key questions with regard to the memory systems underlying language

¹ This terminology emerged in the course of my discussions with Sashank Varma (Stanford University).

² Some recent proposals have attempted to account for a number of linguistic complexity effects in terms of entropy reduction / predictability (e.g., Hale, 2001; Levy, 2005, 2007). These proposals, however, have focused on measuring the difficulty of processing an incoming word depending on how expected it is, rather than on the costs associated with maintaining structural predictions over time.

processing remain unanswered. This thesis will examine two such questions:

1. To what extent is the working memory system underlying language processing domain-specific?
2. What is the relationship between the working memory system and the long-term memory system in language processing?

In **Chapter 1**, I will describe ten experiments investigating the extent to which the working memory system underlying linguistic integrations is domain-specific. I will argue that the results of these experiments demonstrate that at least some aspects of the working memory system used for linguistic integrations are not domain-specific, being involved in arithmetic, and possibly, musical processing.

When a word is integrated into an evolving structure, the processing mechanism must retrieve (1) the lexical properties of the incoming word, which include both syntactic information (e.g., the syntactic category of the word, the phrase structure rules associated with this category, etc.) and semantic information (e.g., the meaning of the word, the thematic properties of the word), and (2) the word(s) to which the incoming word is connected in the current dependency structure. Both of these retrieval operations have been shown to affect the relative ease or difficulty of the integration process. In **Chapter 2**, I will describe six experiments investigating how the two retrieval operations – the retrieval of lexical properties of a word and the retrieval of its structural dependents – relate to each other. I will demonstrate that the relative ease or difficulty of retrieving the lexical properties of an incoming word affect the relative difficulty of retrieving its structural dependents. I will therefore argue that the two retrieval operations rely on overlapping pools of working memory resources.

Chapter 1: The extent of domain-specificity of the working memory system underlying linguistic integrations

General background

The degree of domain-specificity, or modularity, present in the human mind and brain has puzzled scientists and philosophers alike for centuries: is our cognitive system comprised of modules – subserved by highly specialized neural structures – dedicated to specific cognitive functions, or is it more domain-general in nature, such that the same neural/cognitive resources are used for multiple cognitive functions? Here I examine the extent to which the working memory (WM) system underlying linguistic integrations is domain-specific.

The nature of WM resources used for language processing has long been debated. Whereas there is ample evidence in favor of the independence of working memory resource pools used for processing verbal vs. visuo-spatial material (e.g., Baddeley, 1986; Logie, 1986; Vallar & Shallice, 1990; Hanley et al., 1991; Jonides et al., 1993; Paulesu et al., 1993; Logie, 1995; Shah & Miyake, 1996), it is less clear whether on-line language processing relies on the general verbal working memory resource pool (King & Just, 1991; Just & Carpenter, 1992; Gordon et al., 2002), or whether it relies on a domain-specific resource sub-pool – within the verbal working memory resource pool – dedicated solely to language processing (Caplan & Waters, 1999).

Two approaches have traditionally been used to address the question of working memory resources used in on-line language processing: (1) a dual-task approach; and (2) an individual-differences approach. In the *dual-task approach*, participants perform two tasks simultaneously: (1) on-line sentence processing, and (2) a non-linguistic verbally-mediated task (usually a digit-/word-span task). In the *individual-differences approach*, on the other hand, participants are divided into two or more groups on the basis of their performance on some form of a verbal working memory task (usually a reading span task (Daneman & Carpenter, 1980)) and tested on linguistic structures of varying syntactic complexity³. The underlying assumption of the two approaches is that syntactic complexity – determined by the amount of working memory resources necessary for processing a particular structure – will interact with the difficulty of the secondary task or with group-type, respectively, only if the non-linguistic verbally-mediated memory task and on-line linguistic processing rely on overlapping pools of verbal working memory resources.

Most of the previous research on the extent of domain-specificity of the working memory resources underlying language processing have assumed that a homogeneous pool of WM resources is used for processing linguistic structures. However, as briefly discussed above, evidence in the last couple of decades has begun to suggest that two different types of working memory costs may be involved in language processing (storage costs and integration costs). Given the non-homogeneous nature of the WM resource pool underlying language processing, it

³ The classic individual-differences approach – where the relationship between an individual's performance on task(s) in Domain 1 and task(s) in Domain 2 is examined – can also be used, but in the sentence processing literature, the group-based approach has traditionally been used.

is possible that the relationship between the WM system underlying language processing and working memory systems underlying other cognitive processes is more complex than previously thought. Specifically, it is possible that the two pools of working memory resources used in online language processing differ in the extent of their domain-specificity and in the extent of their overlap with other working memory systems. This would imply that in investigating the nature of working memory resources underlying language processing, it might be important to consider the two different resource pools used in language processing – the resource pool underlying the storage process and the resource pool underlying linguistic integrations – separately. In this thesis I will be concerned with the WM system underlying *linguistic integrations*.

Before proceeding to the discussion of the experiments investigating the extent to which the WM system underlying linguistic integrations is domain-specific, I will summarize the evidence for and describe a behavioral experiment demonstrating the independence of the two types of working memory costs in language processing: storage costs and integration costs.

1A. The heterogeneity of the working memory resources underlying language processing: storage costs vs. integration costs

Background

This section will consist of two parts. In the first part, I will summarize some of the evidence for the existence of storage costs and integration costs in on-line language processing. In the second part, I will summarize the evidence for the independence of the resource pools underlying the storage process and the process of linguistic integrations.

According to resource-based theories of linguistic complexity, the relative processing difficulty of a structure is determined by the amount of working memory resources it requires. The standard linguistic contrast that resource-based theories have tried to explain is that between subject- and object-extracted relative clauses. Object-extracted relative clauses (3a) have been shown to be more difficult to process than subject-extracted relative clauses (3b) (e.g., Holmes, 1973, Wanner & Maratsos, 1978, Ford, 1983).

(3a) The reporter that the senator attacked admitted the error.

(3b) The reporter that attacked the senator admitted the error.

There have been a number of proposals in the literature. For example, Wanner & Maratsos (1978) argued that the underlying source of difficulty in object-extracted relative clauses is that the distance over which the head noun (“the reporter”) has to be maintained in memory is longer in (3a) than in (3b). Ford (1983) suggested that the reason that object-extractions are more difficult is because at the point of processing the embedded verb (attacked) in (2) the noun “the reporter” has to be retrieved from memory, and this reactivation is costly. Gibson (1991) proposed that the difficulty in object-extractions results from maintaining multiple nouns in memory, which have not been assigned a thematic role.

As discussed above, in his later work Gibson (1998, 2000) has argued that in studying the working memory system underlying language processing, it is important to distinguish between two types of working memory costs associated with the processing of non-local structural dependencies. In particular, Gibson argued that working memory is required for (1) keeping track of incomplete syntactic dependencies (**storage costs**; e.g., Chomsky & Miller, 1963; Wanner & Maratsos, 1978; Lewis, 1996), and (2) for integrating incoming words to earlier words/positions in the sentence (**integration costs**; e.g., Ford, 1983). The storage cost component was quantified in terms of the number of syntactic categories minimally required to complete the input string as a grammatical sentence and was postulated to arise at the point of predicting a syntactic category and to last until the target syntactic category is encountered. The integration cost component was quantified in terms of the distance between the two ends of a syntactic dependency and was postulated to arise at the second element of the dependency. Specifically, it was hypothesized that the first element of the dependency decays passively over time, and at the point of encountering the second element of the dependency there is a cost associated with retrieving the first element from memory. The original distance metric proposed in Gibson (1998) was in terms of new discourse referents (nouns and verbs). However, it is also possible that retrieving the first element of the dependency from memory is difficult due to the

interference of the intervening elements with the representation of the first element (e.g., Gordon et al., 2001, 2004; VanDyke & Lewis, 2003; Lewis et al., 2006). Thus, another possible distance metric in long-distance dependencies is in terms of similar intervening elements.

Gibson's resource-based theory (the Dependency Locality Theory) accounted for a large number of linguistic complexity effects in English and other languages, including the subject-/object-extracted relative clause contrast above. Below, I review some behavioral and neurophysiological evidence for working memory costs associated with keeping track of incomplete syntactic dependencies and for working memory costs associated with integrating incoming words to earlier words/positions in the sentence.

Chen et al. (2005) tested the predictions of the Dependency Locality Theory with regard to the storage cost component. Specifically, they evaluated the idea that there is a cost associated with keeping track of incomplete syntactic dependencies. The number of minimally required syntactic categories was parametrically varied in structures, like the ones shown in (4a)-(4d).

- (4a) The detective suspected that the thief knew that the guard protected the jewels and so he reported immediately to the museum curator.
- (4b) The detective suspected that the knowledge that the guard protected the jewels came from an insider.
- (4c) The suspicion that the thief knew that the guard protected the jewels **worried** the museum curator.
- (4d) The suspicion that the knowledge that the guard protected the jewels **came** from an insider **worried** the museum curator.

During the critical region (the clause *that the guard protected the jewels*), the number of required syntactic categories (verbs) in (4a) is zero. The number of required verbs is increased by nominalizing one (in (4b) and (4c)) or both (in (4d)) of the verbs in the main clause (*suspected* and *knew*).

Chen et al. found that the reading times over the critical region increase as a function of the number of required verbs, such that the reading times are slower with more pending required verbs. This result provided evidence for the cost associated with keeping track of incomplete syntactic dependencies. Furthermore, investigations of ambiguous structures have also yielded evidence for storage costs (e.g., Grodner et al., 2002).

In addition to the behavioral evidence, there exists neurophysiological evidence from studies using event-related potentials (ERPs) in support of the costs associated with keeping track of incomplete syntactic dependencies. For example, Kluender & Kutas (1993) examined subject and object wh-questions using ERPs and demonstrated that a sustained left-lateralized sustained anterior negativity (LAN) is observed in object wh-questions, where a wh-element has to be held in memory. The LAN was observed to begin at the wh-element and to last until the second element of the dependency (the verb) is encountered. Similarly, King & Kutas (1995) recorded ERPs from participants reading subject- and object-extracted relative clauses and observed a left-lateralized sustained anterior negativity starting after the head noun and lasting throughout the RC (until the embedded verb in object-extractions). Kluender & Kutas (1993) and King & Kutas

(1995) interpreted the observed ERP component as indexing the working memory load associated with holding onto incomplete syntactic dependencies.

Grodner & Gibson (2005) tested the predictions of the Dependency Locality Theory with regard to the integration cost component. Specifically, they evaluated the idea that there is a cost associated with integrating incoming words to earlier words/positions in the sentence. First, subject- and object-extracted relative clauses (5a)-(5b) were compared; and second, a set of structures where the distance between a noun and a verb was parametrically varied (6a)-(6c) was examined.

(5a) The **reporter** who the photographer sent to the editor hoped for a story.

(5b) The **reporter** who sent the photographer to the editor hoped for a story.

(6a) The **nurse** supervised the administrator while...

(6b) The **nurse** from the clinic supervised the administrator while...

(6c) The **nurse** who was from the clinic supervised the administrator while...

In object-extracted relative clauses (5a), at the point of processing verb *sent* the relative pronoun *who* co-indexed with the noun phrase *the reporter* needs to be retrieved from memory, while in subject-extracted relative clauses (5b), the verb *sent* can be locally integrated with the relative pronoun *who*. Grodner & Gibson found that the reading times at the embedded verb (*sent* in (5)) were longer in the object-extracted conditions, compared to the subject-extracted conditions.

In structures in (6), where the distance between the noun phrase *the nurse* and the verb *supervised* in parametrically varied, assuming the discourse-referent distance metric (local dependency, intervening prepositional phrase (one new discourse referent – the noun phrase *the clinic*), intervening relative clause (two new discourse referents – the verb *was* and the noun phrase *the clinic*)), Grodner & Gibson observed that the reading times at the verb (*supervised* in (6)) increased with an increasing distance between the noun phrase and the verb.

This result is more easily accommodated within a decay-based, as opposed to an interference-based, framework, because the intervening material is not similar to the first element of the dependency. Specifically, the noun phrase *the clinic* is inanimate and thus quite different from the noun phrase *the nurse*. Furthermore, the verb in the relative clause (*was*) is a different syntactic category than the noun phrase *the nurse*. However, it is difficult to talk about relative similarity of various elements without a formal account of what constitutes similarity⁴.

The results of Grodner & Gibson (2005), as well as a number of other studies in the literature (e.g., Gordon et al., 2001, 2004; Warren & Gibson, 2002; McElree et al., 2003; VanDyke & Lewis, 2003; c.f. Konieczny, 2000; Vasishth, 2003; Vasishth & Lewis, 2005) provide support for the cost associated with integrating incoming words to earlier words/positions in the sentence.

In addition to the behavioral evidence, there exists neurophysiological evidence from studies using event-related potentials (ERPs) in support of the costs associated with retrieving the

⁴ There are good reasons to think that animacy plays an important role in the processing of long-distance dependencies (e.g., Traxler et al., 2002), as will be discussed in Section 2A.

incoming word's structural dependent from memory. Kaan et al. (2000) observed a late posterior component (the P600) at the end of a wh-dependency (at the verb), compared to control cases where no such dependency was completed at the verb. Kaan et al. hypothesized that the amplitude of the P600 component reflects the difficulty of syntactic integrations.

Although behavioral and neurophysiological evidence exists in support of the two types of working memory costs in sentence comprehension, the relationship between the pools of resources underlying keeping track of incomplete syntactic dependencies, on the one hand, and integrating incoming words to earlier words/positions in the sentence, on the other hand, is not clear. Specifically, it is not well understood whether these two types of processes rely on the same pool of resources, or whether they rely on two independent pools of resources. This question is relevant for understanding the functional architecture of the working memory system underlying language processing and more generally, for determining the extent to which our cognitive system is modular in nature (e.g., Fodor, 1983; Cosmides & Tooby, 1994; Sperber, 1994).

Several ERP studies have directly compared storage costs and integration costs (e.g., in German: Fiebach et al., 2002; Felser et al., 2003; and in English: Phillips et al., 2005) and observed similar patterns of results, consistent with the earlier ERP studies investigating these two types of working memory costs independently. In particular, these studies found that maintaining incomplete syntactic dependencies in memory is associated with a left-lateralized anterior negativity (LAN), and completing non-local syntactic dependencies is associated with a centro-parietally distributed positivity (the P600).

The fact that – using event-related potentials – researchers have found two distinct signatures for the storage costs and for the integration costs is suggestive of two independent pools of working memory resources, especially given that the two components (the LAN and the P600) appear to have different scalp distributions. Additional evidence comes from an fMRI study demonstrating that syntactic storage costs and syntactic integration costs appear to rely on different neural substrates: specifically, Fiebach et al. (2005) demonstrated that syntactic storage costs, but not syntactic integration costs, resulted in higher activation in Brodmann area 44 (an area traditionally included in the Broca's area), as well as in the inferior frontal and superior temporal areas bilaterally. The activation in the frontal areas is consistent with the scalp distribution of the LAN component in the ERP studies.

The goal of the experiment reported here is to further investigate the relationship between the resource pools underlying storage and integration cost, using a behavioral method (self-paced reading).

Experiment 1A

[Fedorenko & Gibson, submitted]

The logic of this experiment relies on the assumption that if two cognitive processes rely on the same pool of resources or on overlapping pools of resources, then with an increase in complexity of each of these two processes the resource pool (or the shared component, in the case of overlapping resource pools) will become overtaxed, resulting in super-additive processing

difficulty. Two factors – storage costs and integration costs – were manipulated in a 2 x 2 design. If we observe a super-additive processing difficulty when both types of costs are high, then we may infer that the working memory resources that are required for keeping track of incomplete syntactic dependencies overlap – partially or completely – with those that are required for integrating words to earlier words/positions in the sentence. In contrast, if no such interaction is observed, this would be consistent with the two types of costs relying on independent resource pools.

Methods

Participants Forty-four participants from MIT and the surrounding community were paid for their participation. All were native speakers of English and were naive as to the purposes of the study.

Design and materials The experiment had a 2 x 2 design, crossing storage costs (low, high) with integration costs (low, high).

The materials consisted of 40 sets of sentences, having four different versions as in (7):

(7a) *Low Storage / Low Integration:*

After the appraiser notified the dealer / *from a small town in Greece*, / the thief knew / that the painting / which the millionaire / **inherited** / could sell for millions of dollars.

(7b) *Low Storage / High Integration:*

After the appraiser notified the dealer, / the thief knew / that the painting / which the millionaire / *from a small town in Greece* / **inherited** / could sell for millions of dollars.

(7c) *High Storage / Low Integration:*

After the appraiser notified the dealer / *from a small town in Greece*, / the thief's knowledge / that the painting / which the millionaire / **inherited** / could sell for millions of dollars / proved to be useless.

(7d) *High Storage / High Integration:*

After the appraiser notified the dealer, / the thief's knowledge / that the painting / which the millionaire / *from a small town in Greece* / **inherited** / could sell for millions of dollars / proved to be useless.

The storage costs were manipulated in a similar way to the manipulation used in Chen et al. (2005). Specifically, the verb *knew* in the low-storage-cost conditions ((7a) and (7b)) is nominalized to *knowledge* in the high-storage-cost conditions ((7c) and (7d)), such that in these conditions, an expectation for a verb is created. This expectation continues through the critical region (*inherited*) and is satisfied with the verb phrase *proved to be useless*.

The integration costs were manipulated in a similar way to the manipulation used in Grodner & Gibson (2005). Specifically, at the critical region there is a local integration of the verb *inherited* to its subject *the millionaire* in the low-integration-cost conditions ((7a) and (7c)), but in the high-integration-cost conditions ((7b) and (7d)) this integration is crossing a prepositional phrase *from a small town in Greece*. Note that in order to align the critical region (*inherited*) across the

four conditions⁵, in the low-integration-cost conditions we attached the prepositional phrase *from a small town in Greece* to the noun phrase in the first clause.

In addition to the target sentences, 60 sentences from two unrelated experiments were included. The three experiments served as fillers for one another. The length and syntactic complexity of the sentences in the two unrelated experiments was similar to that of the target sentences. The stimuli were pseudo-randomized separately for each participant with at least one filler separating the target sentences. Each participant saw only one version of each sentence, following a Latin-Square design (see Appendix A for a complete list of linguistic materials).

Procedure The task was self-paced phrase-by-phrase reading with a center-screen presentation (Just, Carpenter & Woolley, 1982). The reason that we used center-screen presentation, as opposed to the more standard moving-window presentation, is that given the length of the sentences (more than 30 words) it was impossible to fit the whole sentence on the screen, and thus it would have been difficult to interpret the results due to line breaks. The experiment was run using the Linger 2.85 software by Doug Rohde. Each experimental sentence had seven (low-storage-cost conditions) or eight (high-storage-cost conditions) regions (as shown in (7a)-(7d)): (1) a subordinate temporal clause, initiated by “before”, “after” or “when”, and including a noun phrase, a transitive verb, and another noun phrase, (2) a prepositional phrase modifying the object of the subordinate clause (low-integration-cost conditions only), (3) the main clause subject and verb (low-storage-cost conditions), or the nominalized version of the verb (high-storage-cost conditions), (4) complementizer *that* initiating a sentential complement clause and an inanimate noun phrase, (5) relative pronoun *which* initiating an object-extracted relative clause modifying the inanimate noun phrase from the preceding region, and an animate subject, (6) a prepositional phrase modifying the subject of the relative clause (high-integration-cost conditions only), (7) the verb phrase for the subject of the complement clause (the critical region), and (8) the verb for the main clause (high-storage-cost conditions only).

Each trial began with a fixation cross. Participants pressed the spacebar to reveal each region of the sentence. The amount of time the participant spent reading each region was recorded as the time between key-presses.

To assure that the participants read the sentences for meaning, a comprehension question was presented at the end of each trial, asking about the propositional content of the sentence. Participants pressed one of two keys to respond “yes” or “no”. After a correct response, the word “CORRECT” flashed briefly on the screen, and after an incorrect response, the word “INCORRECT” flashed briefly.

Before the experiment started, a short list of practice items and questions was presented in order to familiarize the participants with the task. Participants took approximately 35 minutes to complete the experiment.

⁵ There is often a general tendency to speed up towards the end of the sentence observed in reading studies, so it is important that the critical region is in a similar position in the sentence across conditions.

Results

Comprehension question performance Across the conditions, participants answered the comprehension question correctly 75% of the time. Table 1 presents the mean accuracies across the four conditions of Experiment 1A. A two-factor ANOVA crossing storage costs (low, high) and integration costs (low, high) on the responses to the comprehension questions revealed two main effects and an interaction. The two main effects appear to be driven by the interaction. First, we observed a main effect of integration costs ($F(1,43)=11.6$; $MSe=1536$; $p < .002$; $F(1,39)=4.36$; $MSe=1397$; $p<.05$). Second, we observed a main effect of storage costs in the items analysis only ($F(1,43)=2.67$; $MSe=657$; $p=.11$; $F(1,39)=4.35$; $MSe=597$; $p<.05$). Finally, we observed an interaction – which did not quite reach significance in the participants analysis – such that the accuracies were lower in the high-storage-cost/high-integration-cost condition than would be expected if the two effects were purely additive ($F(1,43)=3.91$; $MSe=820$; $p= .055$; $F(1,39)=7.19$; $MSe=746$; $p<.02$).

<u>Storage Cost</u>	<u>Integration Cost</u>	
	Low integration cost	High integration cost
Low storage cost	77.7 (1.8)	76.1 (2.1)
High storage cost	78.2 (2.0)	68.0 (3.1)

Table 1. Comprehension accuracies in percent correct, as a function of storage costs and integration costs in Experiment 1A (standard errors in parentheses).

Reaction times Because the comprehension question accuracies were not very high (75%) – which is not surprising given the complexity of the materials – we analyzed all the trials, regardless of whether the comprehension question was answered correctly. The statistical data patterns were very similar in the analysis of only the trials where the comprehension question was answered correctly.

To adjust for overall differences in participants' reading rates, a regression equation predicting reading times from word length was derived for each participant, using all filler and target items (Ferreira & Clifton, 1986; see Trueswell, Tanenhaus & Garnsey, 1994, for discussion). For each word, the reading time predicted by the participant's regression equation was subtracted from the actual measured reading time to obtain a residual reading time. The analyses of the raw reading time data produced the same statistical patterns. Finally, residual reading times more than three standard deviations away from the mean for a position within condition were removed from the analyses, excluding 1.6% of the data.

We present the analysis of the critical region. It is difficult to compare the earlier regions in the sentence, because of the different materials in the earlier regions across the conditions (see Appendix B for a complete reaction time table for all the regions across the four conditions). The critical region included the verb of the main clause (*inherited* in (7)). A 2 x 2 ANOVA (low storage cost / high storage cost, low integration cost / high integration cost) in this region revealed two significant main effects. First, the high-storage-cost conditions were read slower than the low-storage-cost conditions ($F(1,43)=8.79$; $MSe=32778$; $p < .005$; $F(1,39)=8.80$; $MSe=32664$; $p < .01$). Second, the high-integration-cost conditions were read slower than the low-integration-cost conditions ($F(1,43)=18.7$; $MSe=32778$; $p < .001$; $F(1,39)=20.33$; $MSe=143069$; $p < .001$). Critically, however, there was no suggestion of an interaction between the two factors ($F_s < .01$, $p_s > .9$). Figure 1 presents the mean residual RTs for the critical region across the four conditions of Experiment 1A.

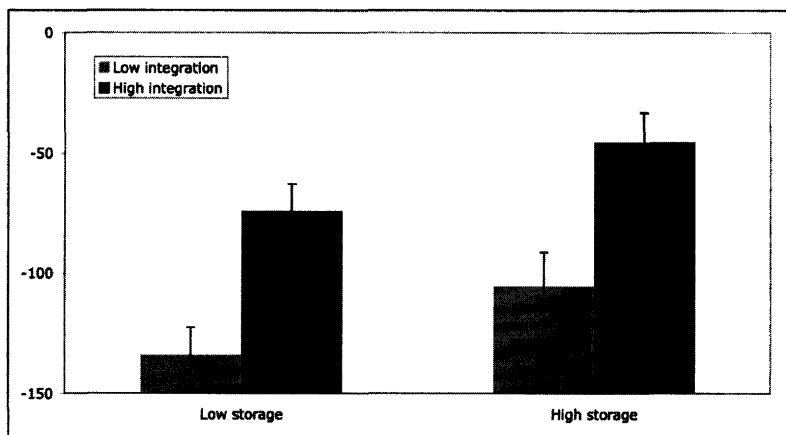


Figure 1: Residual reading times for the critical region in Experiment 1A.

Discussion

We observed different patterns of results in our two dependent measures. Whereas the online measure (reading times at the critical region, where both storage costs and integration costs were manipulated) revealed two main effects with no suggestion of an interaction, the off-line measure (comprehension question accuracies) revealed a significant interaction in the participants analysis (borderline significant in the items analysis) whereby the accuracies in the condition where both storage and integration costs were high were lower than would be expected if the effects of storage costs and integration costs were purely additive.

As will be discussed in more detail below, in discussing the nature of the working memory resources underlying language processing Caplan & Waters (1999) made a distinction between the resources used for processing linguistic representations in real time (interpretive processing) and the resources used for processing linguistic representations off-line (post-interpretive processing). While interpretive processing was argued to involve extracting “meaning from a linguistic signal” (Caplan & Waters, 1999, p. 79), post-interpretive processing was argued to use this extracted meaning to accomplish tasks, like reasoning, planning actions, and storing

information in long-term memory. By these definitions, reading times reflect interpretive processing and comprehension accuracies reflect post-interpretive processing. Caplan & Waters argued that in studying the working memory system underlying language processing we should be concerned with *interpretive* processing.

With this distinction in mind, the results of Experiment 1A – suggesting that keeping track of incomplete syntactic dependencies and integrating incoming words to earlier words/positions in the sentence rely on independent resource pools – are therefore consistent with the previous ERP findings of two distinct components corresponding to these two types of processes.

With regard to the interaction observed in the response accuracy data, I can speculate that the cost of accessing the sentence representation (the propositional content) may rise non-linearly as a function of the amount of working memory resources spent in constructing the representation in on-line processing.

Conclusions

The results of Experiment 1A revealed two significant main effects, with no trace of an interaction, at the critical region in the reading time data, providing further evidence for the independence of the two resource pools underlying keeping track of incomplete syntactic dependencies and integrating incoming words to earlier words/positions in the sentence. I will now proceed to the discussion of the experiments investigating the extent to which the working memory system underlying linguistic integrations is domain-specific.

1B. Investigating working memory tasks with similar *integration processes*

Background

King and Just (1991) and Just and Carpenter (1992) claimed to have provided some evidence in support of the hypothesis whereby language processing relies on a general pool of verbal working memory resources. This evidence consisted of differential behavior of low- and high-span readers, classified using Daneman and Carpenter's (1980) reading span task, in the processing of syntactic structures of low and high complexity (subject- vs. object-extracted relative clauses). However, Caplan and Waters (1999) could not replicate these findings in a series of studies.

Moreover, Waters et al. (1995) used the dual-task approach crossing syntactic complexity and the complexity of a digit-span task. Specifically, participants were asked to perform a sentence-picture matching task with and without concurrent verbal load which involved maintaining a string of digits (equal to the subject's span, or equal to subject's span minus one). The sentences varied in syntactic complexity, involving subject- vs. object-extracted relative clauses. Waters et al. observed a main effect of concurrent task in the accuracy scores for the sentence-picture matching task, but no effect of syntactic complexity, and crucially, no interaction between the difficulty of concurrent task and syntactic complexity. Furthermore, Waters & Caplan (1999) replicated these findings using an enactment task instead of the sentence-picture matching task⁶. On the basis of these results and on the basis of the data from the individual differences studies, Caplan and Waters (1999) argued for a specialized pool of verbal working memory resources used for on-line language processing.

I would like to argue that there are two possible reasons for why the previous attempts to find an interaction between linguistic complexity and non-linguistic verbally-mediated tasks have failed. *First, the cognitive processes* involved in the language processing task and in the digit-/word-span task are qualitatively different. Specifically, the digit-/word-span task involves storing a string of digits or unrelated words. In contrast, the language processing task involves integrating each incoming word into the evolving structural representation, updating this representation, then integrating the next word, and so on. It is therefore plausible that the digit-/word-span task and linguistic integrations rely on independent pools of working memory resources. *Second, the materials* involved in the language processing task and in the digit-/word-span task are qualitatively different. Specifically, the digit-/word-span task involves storing a string of digits or nouns that are very different from the materials used in the language processing task, which typically involves occupation-like nouns (e.g., the fireman, the doctor) and verbs. In this section (1B) I will discuss research addressing the first problem by examining pairs of cognitive processes that are similar in the types of mental operations they involve. In section IC I will

⁶ Caplan & Waters (1999) discuss the results from Waters et al. (1987) where participants were asked to make timed plausibility judgments about sentences containing subject- and object-extracted relative clauses, with and without a concurrent digit-span task, and the reaction time and the accuracy data revealed a main effect of the digit span task, a main effect of syntactic complexity and an interaction between the two factors in the participants, but not in the items analysis. Caplan & Waters do not discuss the implications of this specific set of results for their claim that on-line language processing relies on a specialized pool of verbal working memory resources.

discuss research addressing the second problem by examining pairs of cognitive tasks that involve similar materials.

It is worth noting that there have been several reports of off-line interactions between memory load and syntactic complexity in the literature. For example, Waters et al. (1987) and Waters & Caplan (1996) found that syntactic complexity had an effect on the number of sentence-final words recalled in a sentence-acceptability-judgment task. Similarly, Wanner & Maratsos (1978) used a task where sentence presentation was interrupted by a list of words, which had to be recalled at the end of the sentence. They reported poorer word recall performance in more complex object-extracted RCs. Caplan & Waters (1999) used two different lines of argumentation to show that the off-line interactions observed in some of the previous experiments are still consistent with the idea of an independent pool of verbal WM resources dedicated to on-line sentence comprehension. First, as briefly discussed in the previous section, they made a distinction between interpretive (on-line) and post-interpretive (off-line) processes, which are involved in sentence comprehension. Interpretive processing, according to Caplan & Waters, involves the “extraction of meaning from a linguistic signal” (Caplan & Waters, 1999, p. 79), whereas post-interpretive processing involves using this extracted meaning to accomplish tasks, like reasoning, planning actions, and storing information in long-term memory. Caplan & Waters then argued that the off-line interactions observed between linguistic processing and non-linguistic verbally-mediated tasks do not directly address the question of an overlap in verbal WM resources, because post-interpretive processing (used in off-line tasks) is likely to involve a variety of cognitive processes beyond linguistic processing. Second, Caplan & Waters argued that because the only experiments where interactions between syntactic complexity and memory load have been observed involved one task interrupting the other task, it is likely that such interactions resulted from the necessity to shift attention back and forth between the two tasks, rather than from an overlap in verbal WM resource pools between the two tasks.

Section 1B will be structured as follows: **Experiments 1B-1 – 1B-6** will investigate the relationship between the working memory system underlying linguistic integrations and those underlying arithmetic and spatial integrations; **Experiments 1B-7a and 1B-7b** will investigate the relationship between the working memory system underlying linguistic integrations and that underlying structural integrations in music.

Background for Experiments 1B-1 through 1B-6

In most of the previous dual-task experiments, the standard complexity manipulation in the digit-/word-span task has involved varying the number of elements that have to be remembered (cf. Gordon et al., 2002; Fedorenko et al., 2006; to be discussed in Section 1C). In contrast, the standard complexity manipulation in the language processing task in the dual-task experiments has usually involved the contrast between subject- and object-extracted relative clauses. As discussed above, the difference between subject- and object-extracted relative clauses is plausibly related to the difference in integration lengths in the relative clause region (Gibson, 1998, 2000; Gordon et al., 2001; McElree et al., 2003; Gordon et al., 2004; Grodner & Gibson, 2005; c.f. Wanner & Maratsos, 1978; Ford, 1983; King & Just, 1991). Specifically, in subject-extracted relative clauses (8a), the embedded verb *frustrated* is integrated locally to the

immediately preceding pronoun *who* co-indexed with the noun phrase *the janitor*. In contrast, in object-extracted relative clauses (8b), at the point of processing the embedded verb *frustrated* it is necessary to retrieve the noun phrase *the janitor* from memory to interpret it as the object of *frustrated*, since it occurs earlier in the input.

(8a) Subject-extracted: The **janitor** who frustrated the plumber lost the key on the street.

(8b) Object-extracted: The **janitor** who the plumber frustrated lost the key on the street.

As alluded to above, the difficulty of retrieving the noun phrase *the janitor* at the point of processing the verb *frustrated* in the object-extracted relative clause in (8b) compared to a local integration between the relative pronoun co-indexed with *the janitor* and *frustrated* in the subject-extracted relative clause in (8a) might be due to either (1) the passive decay of the representation of the noun phrase *the janitor* over time, or (2) interference of the intervening noun phrase *the plumber*. There is evidence for both of these factors contributing to the difficulty of processing object-extracted relative clauses and other long-distance dependencies (e.g., Lewis, 1996; Gordon et al., 2001; McElree et al., 2003; VanDyke & Lewis, 2003; Gordon et al. 2004; Grodner & Gibson, 2005; Lewis et al., 2006).

Six dual-task experiments were conducted in order to investigate the relationship between the working memory resources involved in the integration processes in language and those involved in the integration processes in (1) non-linguistic verbally-mediated tasks, and (2) spatial tasks. Specifically, we wanted to test whether non-linguistic verbally-mediated tasks involving integration processes might interact with linguistic integrations due to some overlap in the nature of the integration processes involved.

We first conducted three dual-task experiments which investigated the relationship between the resources involved in linguistic processing and those involved in arithmetic processing (Experiments 1B-1, 1B-2, and 1B-3). Arithmetic additions are similar to linguistic integrations, such that in a series of consecutive additions, an incoming element – a number – is integrated into the current representation, resulting in an intermediate sum. The intermediate sum is then updated with the integration of each incoming number. If it is the case that all cognitive processes involving integrations of verbal material are relying on the same / overlapping pools of working memory resources, then we should observe an interaction between the complexity of linguistic and arithmetic integrations.

To control for a possible confound in terms of domain-general attention-switching costs, which might contribute to the interactions observed in dual-task paradigms (as noted by Caplan & Waters, 1999), we conducted two additional dual-task experiments (Experiments 1B-4 and 1B-5) where we substituted the arithmetic task with a spatial integration task, which involves similar integration processes, but critically does not require the use of verbal working memory resources. The attention-switching account predicts an interaction, regardless of the nature of the secondary task, as long as the secondary tasks are matched for complexity across the experiments. In contrast, the shared working memory resource pool hypothesis predicts an interaction only in the experiments with arithmetic tasks, but not in the experiments with the spatial integration tasks.

Before presenting our experimental results, it is important to acknowledge a possible limitation in interpreting the presence of super-additive interactions in dual-task experiments. Previous dual-task experiments in different areas of cognitive psychology, as well as the experiments reported in this paper, rely on the additive factors logic (Sternberg, 1969, following Donders, 1868-1869), as summarized, for example, by McClelland (1979): “the assumption that one experimental manipulation influences the duration of one stage and another manipulation influences the duration of another stage leads to the conclusion that the two factors will have additive effects on reaction time. On the other hand, factors that influence the duration of the same stage will generally interact with one another” (McClelland, 1979, p. 311). In the experiments presented in this paper the additive factors logic applies as follows. If one experimental manipulation (the difficulty of the language comprehension task) draws on one resource pool, and another experimental manipulation (the difficulty of the arithmetic or spatial integration task) draws on another resource pool, then reaction times should reveal strictly additive effects. If, however, the two experimental manipulations draw on the same / overlapping resource pools then reaction times should reveal super-additivity. A potential problem with this logic arises if reaction times increase super-linearly, which could result in a super-additive interaction even when the two experimental manipulations draw on different resource pools (c.f. Loftus (1978) who discusses this issue with respect to cases where probability (e.g., accuracy) is the dependent measure). However, this issue is mitigated in the current experimental design because reaction time curves, unlike probability curves, do not tend to show super-linear trends (e.g., Sternberg, 1969).

Experiments 1B-1 through 1B-6

[The results of Experiments 1B-1, 1B-3, 1B-4 and 1B-5 appeared in Fedorenko, Gibson, and Rohde (2007).]

Experiment 1B-1: Linguistic vs. Arithmetic integrations

This experiments had a dual-task design, in which participants read sentences phrase-by-phrase, and at the same time were required to perform a series of arithmetic additions. The on-line addition task is similar to on-line sentence comprehension in that an incoming element – a number – must be integrated into (i.e., added to) the representation constructed thus far: the working sum. Both tasks had two levels of complexity, resulting in a 2 x 2 design. Critically, there was no difference in linguistic complexity between the easy and hard arithmetic conditions: the complexity of the arithmetic task was manipulated in terms of the difficulty of the arithmetic integrations (by making the addends larger), while keeping the linguistic form of the two conditions identical (*number plus number plus number*, etc.). Therefore, if we observe a super-additive interaction between the two tasks when the complexity of both tasks is high, then we may infer that the working memory resources that are involved in performing the arithmetic task overlap with those that are involved in syntactic integration processes. In contrast, if language processing relies on an independent working memory resource pool, there should be no such interaction.

Methods

Participants Forty-eight participants from MIT and the surrounding community were paid for their participation. All were native speakers of English and were naive as to the purposes of the study.

Design and materials The experiment had a 2 x 2 design, crossing syntactic complexity (subject-extracted RCs, object-extracted RCs) with arithmetic complexity (simple additions (low initial addend, subsequent addends between 1 and 3) vs. complex additions (higher initial addend, subsequent addends between 4 and 6)).

The language materials consisted of 32 sets of sentences, having four different versions as in (9):

(9a) *Subject-extracted, version 1:*

The janitor / who frustrated the plumber / lost the key / on the street.

(9b) *Subject-extracted, version 2:*

The plumber / who frustrated the janitor / lost the key / on the street.

(9c) *Object-extracted, version 1:*

The janitor / who the plumber frustrated / lost the key / on the street.

(9d) *Object-extracted, version 2:*

The plumber / who the janitor frustrated / lost the key / on the street.

There were two levels of syntactic complexity – subject- and object-extractions – with four versions of each sentence in order to control for potential plausibility differences between the subject- and object-extracted versions of each sentence. As a result, no independent plausibility control is needed in this design. Each participant saw only one version of each sentence, following a Latin-Square design (see Appendix C for a complete list of linguistic materials).

The numbers for the addition task were randomly generated on-line for each participant with the following constraints: (1) the value of the initial addend in the easy-arithmetic condition varied from 1 to 10, whereas the value of the initial addend in the hard-arithmetic condition varied from 11 to 20, and (2) the addends varied from 1 to 3 in the easy-arithmetic condition and from 4 to 6 in the hard-arithmetic condition. There is evidence from the mathematical cognition literature (e.g., Ashcraft, 1992, 1995) showing that reaction times as well as error rates for performing addition operations increase as a function of the size of the addends. That was the motivation for our complexity manipulations.

In addition to the target sentences, 40 filler sentences with various syntactic structures other than relative clauses were included. The length and syntactic complexity of the filler sentences was similar to that of the target sentences. The stimuli were pseudo-randomized separately for each participant, with at least one filler separating the target sentences.

Procedure The language task was self-paced phrase-by-phrase reading with a moving-window display (Just, Carpenter & Woolley, 1982). The experiment was run using the Linger 2.85

software by Doug Rohde. Each experimental sentence had four regions (as shown in (9a)-(9d)): (1) a noun phrase, (2) an RC (subject-/object-extracted), (3) a main verb with a direct object (an inanimate noun phrase) and (4) an adjunct prepositional phrase. The addends for the addition task were presented simultaneously with the sentence fragments, above and aligned with the second character of each fragment. The first sentence region had a number above it (e.g., “12”) and all the subsequent regions had a plus sign followed by a number (e.g., “+4”), as shown in Figure 2.

Time 1:	12 The janitor
Time 2:	+4 who frustrated the plumber
Time 3:	+5 lost the key
Time 4:	+4 on the street.

Figure 2: Sample frame-by-frame presentation of an item in Experiment 1B-1.

Each trial began with a series of dashes marking the length and position of the words in the sentence. Participants pressed the spacebar to reveal each region of the sentence. As each new region appeared, the preceding region disappeared along with the number above it. The amount of time the participant spent reading each region and performing the accompanying arithmetic task, was recorded as the time between key-presses.

To make sure the participants performed the arithmetic task, a window appeared at the center of the screen at the end of each sentence and the participants were asked to type in the sum of their calculations. If the answer was correct, the word “CORRECT” flashed briefly on the screen, if the answer differed by up to 2 from the correct sum, the word “CLOSE” flashed briefly, and if the answer was off by more than 2, the word “INCORRECT” flashed briefly on the screen. To assure that the participants read the sentences for meaning, two true-or-false statements were presented sequentially after the sum question, asking about the propositional content of the sentence they just read. Participants pressed one of two keys to respond “true” or “false” to the statements. After a correct answer, the word “CORRECT” flashed briefly on the screen, and after an incorrect answer, the word “INCORRECT” flashed briefly.

Participants were instructed not to concentrate on one task (reading or additions) more than the other. They were asked to read sentences silently at a natural pace and to be sure that they understood what they read. They were also told to answer the arithmetic and sentence questions as quickly and accurately as they could, and to take wrong answers as an indication to be more careful.

Before the experiment started, a short list of practice items and questions was presented in order to familiarize the participants with the task. Participants took approximately 35 minutes to complete the experiment.

Results

Arithmetic accuracy Participants answered the arithmetic sum correctly 89.5% of the time. Table 2 presents the mean arithmetic accuracies across the four conditions of Experiment 1B-1. A two-factor ANOVA crossing arithmetic complexity (easy, hard) and syntactic complexity (easy, hard) on these data revealed a main effect of arithmetic complexity ($F(1,47)=7.87$; $MSe=0.0941$; $p < .01$; $F_2(1,31)=8.12$; $MSe=0.0627$; $p < .01$; $\min F'(1,75)=3.99$; $p<.05$), but no other reliable effects.

<u>Arithmetic complexity</u>	<u>Syntactic complexity</u>	
	Subject-extraction (Easy)	Object-extraction (Hard)
Easy arithmetic	93.5 (1.7)	89.8 (1.8)
Hard arithmetic	86.7 (2.4)	87.8 (2.0)

Table 2. Arithmetic accuracies in percent correct, as a function of arithmetic complexity and syntactic complexity in Experiment 1B-1 (standard errors in parentheses).

Comprehension question performance There were two comprehension questions following each experimental trial. Participants answered the first question correctly 81.5% of the time, and the second question 79.4% of the time. The percentages of correct answers by condition were very similar for the two questions, so we collapsed the results in our analyses. Table 3 presents the mean accuracies across the four conditions of Experiment 1B-1. A two-factor ANOVA crossing arithmetic complexity (easy, hard) and syntactic complexity (easy, hard) on the responses to the two comprehension questions revealed a main effect of syntactic complexity ($F(1,47)=13.37$; $MS=.1270$; $p < .001$; $F_2(1,31)=6.41$; $MS=.0846$; $p<.02$; $\min F'(1,59)=4.33$; $p<.05$) and a main effect of arithmetic complexity in the participants analysis ($F(1,47)=6.08$; $MS=.0661$; $p < .02$; $F_2(1,31)=3.52$; $MS=.0441$; $p = .07$; $\min F'(1,63)=2.22$; $p=.14$), but no significant interaction ($F_s < 1.5$).

Arithmetic complexitySyntactic complexity

	Subject-extraction (Easy)	Object-extraction (Hard)
Easy arithmetic	85.8 (2.1)	78.8 (2.7)
Hard arithmetic	80.2 (2.4)	77.0 (2.1)

Table 3. Comprehension accuracies in percent correct, as a function of arithmetic complexity and syntactic complexity in Experiment 1B-1 (standard errors in parentheses).

Reaction times Because participants had to answer three questions (one arithmetic, two language) for each sentence, the odds of getting all three correct were not very high overall (57.9%). As a result, we analyzed all trials, regardless of how the arithmetic and the comprehension questions were answered. The data patterns were very similar in analyses of smaller amounts of data, in which we analyzed (1) trials in which one or both of the comprehension questions were answered correctly, or (2) trials in which the arithmetic question was answered correctly. To adjust for differences in region lengths as well as overall differences in participants' reading rates, a regression equation predicting reaction times from region length was derived for each participant, using all filler and target items (Ferreira & Clifton, 1986; see Trueswell, Tanenhaus & Garnsey, 1994, for discussion). For each region, the reaction time predicted by the participant's regression equation was subtracted from the actual measured reaction time to obtain a residual reaction time. Reaction time data points that were more than three standard deviations away from the mean residual RT for a position within a condition were excluded from the analysis, affecting 2.3% of the data. Figure 3 presents the mean residual RTs per region across the four conditions of Experiment 1B-1.

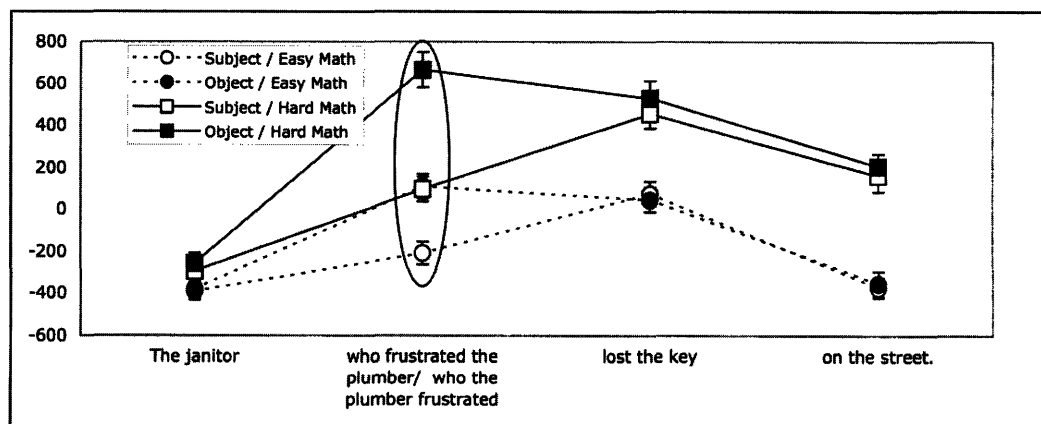


Figure 3: Reaction times per region in the four conditions of Experiment 1B-1. Error bars indicate standard errors. The critical region is circled.

We conducted a 2 x 2 ANOVA crossing syntactic complexity and the complexity of the arithmetic task for each of the four regions. The results are presented in Table 4. For comparisons between means of conditions, we report 95% confidence intervals (CIs) based on the mean squared errors of the relevant effects from the participants analyses (see Masson & Loftus, 2003). We first present the analysis of the critical region, Region 2, which included the relative clause (“who frustrated the plumber” / “who the plumber frustrated”), followed by the analyses of the other regions. At the critical region, the hard-arithmetic conditions were read significantly slower than the easy-arithmetic conditions (380.8 ms vs. -49.5 ms; 95% CI=120.6 ms), and the syntactically more complex object-extracted RC conditions were read significantly slower than the subject-extracted conditions (387.8 ms vs. -56.5 ms; 95% CI=147 ms). Most interestingly, there was a significant interaction, such that in the hard arithmetic conditions, the difference between subject- and object-extracted RCs (569.7 ms) was larger than in the easy arithmetic conditions (318.8 ms). The statistical analyses of the raw reaction time data produced the same numerical patterns: specifically, the two main effects were significant in the participants and in the items analyses; and the interaction was significant in the participants analysis, but did not reach significance in the items analysis ($p=.12$). The interaction between syntactic and arithmetic complexity is predicted by the hypothesis whereby linguistic processing and arithmetic processing rely on overlapping pools of working memory resources, but not by the hypothesis whereby the pools of resources are independent.

In the other three regions (Region 1, Region 3, and Region 4) there was a main effect of arithmetic complexity, but no other significant effects.

Table 4. Analysis of Variance results for Experiment 1B-1.

Source of variance	By participants			By items		min F'	
	<i>df</i>	F1 value	MSe	<i>df</i>	F2 value	<i>df</i>	min F'
Region 1							
Syntactic Complexity	1,47	<1	39434	1,31	<1	1,64	<1
Arithmetic Complexity	1,47	10.74*	52455	1,31	7.07*	1,66	4.26*
Interaction	1,47	<1	58064	1,31	<1	1,63	<1
Region 2 (critical region containing the relative clause)							
Syntactic Complexity	1,47	36.96*	256350	1,31	43.46*	1,77	19.9*
Arithmetic Complexity	1,47	51.53*	172459	1,31	41.67*	1,71	23.0*
Interaction	1,47	4.40*	171584	1,31	5.78*	1,78	2.49
Region 3							
Syntactic Complexity	1,47	<1	149644	1,31	<1	1,77	<1
Arithmetic Complexity	1,47	40.78*	220174	1,31	73.36*	1,77	26.2*
Interaction	1,47	<1	141423	1,31	<1	1,76	<1

Region 4

Syntactic Complexity	1,47	<1	102210	1,31	<1	1,78	<1
Arithmetic Complexity	1,47	86.75*	164414	1,31	152.81*	1,78	55.3*
Interaction	1,47	<1	85723	1,31	<1	1,39	<1

Note: Significant effects are marked by asterisk.

Discussion

The results of Experiment 1B-1 provide support for a working memory framework where linguistic integrations and arithmetic integrations rely on overlapping resource pools. Most importantly, there was an interaction between syntactic complexity and arithmetic complexity in the critical region of the linguistic materials, where syntactic complexity was manipulated between subject-extracted RCs (easy integrations) and object-extracted RCs (more difficult integrations). There was no evidence of any interaction of this kind in any of the other three regions. Critically, linguistic complexity did not vary across the two conditions of the arithmetic task (both conditions used expressions like *number plus number plus number*, etc.), so the observed interaction is not due to an overlap in the linguistic processes that are involved in the two tasks. In other words, the fact that the arithmetic task uses verbal material cannot, by itself, account for the observed interaction.

It should be noted, however, that there are two possible confounding factors present in Experiment 1B-1. The first confounding factor involves a difference between the easy and the hard conditions of the arithmetic task in terms of low-level verbal complexity, and the second confounding factor involves a possible explanation of the interaction in terms of a domain-general attention-switching mechanism.

First, even though the easy and the hard conditions in the arithmetic task are the same in terms of syntactic complexity, there might be a difference between the two conditions in terms of low-level morpho-/phonological complexity, which might result in the hard conditions having higher rehearsal demands. Specifically, because the hard condition involves adding larger numbers, both the length and the morphological complexity of the numbers in the hard condition are on average higher. In the working memory framework proposed by Baddeley and colleagues (Baddeley & Hitch, 1974; Baddeley, 1986), this would involve the phonological loop system used for storing and rehearsing verbal material. On the assumption that both the linguistic and the arithmetic task make use of the phonological loop, one could argue that the observed interaction might be due to the fact that the hard language condition (object-extracted relative clauses) and the hard arithmetic condition (more difficult additions) have higher rehearsal demands and are thus overtaxing the rehearsal system. To rule out the possibility that an overlap in the rehearsal system use between the linguistic and the arithmetic task contributed to the interaction observed in Experiment 1B-1, we conducted Experiments 1B-2 and 1B-3 – where the morpho-/phonological complexity was kept constant across the easy and the hard conditions of the arithmetic task and only the difficulty of the arithmetic operations was manipulated. (It is worth noting, however, that the explanation in terms of overtaxing the shared rehearsal system is not very plausible, given the patterns of data in the previous experiments in the literature. Specifically, as discussed above, in the earlier dual-task experiments where a digit-span or a

word-span task was used as a secondary task, the complexity manipulation (more vs. fewer items to remember) inevitably varied the amount of required rehearsal, and yet no reliable interactions between digit-/word-span task complexity and syntactic complexity have been observed (Waters et al., 1995; Caplan & Waters, 1999).)

Another alternative explanation for the observed pattern of results in Experiment 1B-1 is in terms of attentional resources required for the simultaneous performance of the two tasks, as discussed in Caplan and Waters (1999). In dual-task paradigms, resources are needed in order to direct attention to one task or another. It is possible that in the difficult conditions, more attention switches are required, or the switches between tasks are more costly. The observed interaction could therefore be a result of additional task-switching costs in the high syntactic complexity / high arithmetic complexity condition. Experiments 1B-4 and 1B-5 – where an arithmetic task was substituted by a spatial integration task – were conducted to address this issue.

Experiments 1B-2 and 1B-3: Linguistic vs. Arithmetic integrations (controlling for morpho-phonological complexity in the arithmetic task)

These experiment had a very similar design to that in Experiment 1B-1. The main difference was in that a similar range of numbers was used in the easy and hard versions of the arithmetic task in both experiments, which ensured that the easy and hard arithmetic conditions did not differ in terms of morpho-phonological complexity. Experiment 1B-2 used numbers in a similar range (between 30 and 60) and the complexity was manipulated by using additions in the easy version of the task and subtractions in the hard version of the task. Experiment 1B-3 used large numbers (in the range from 1 to 399) and forced the participants to perform approximate calculations by rounding off the numbers to the nearest hundred. The complexity of the rounding off process was manipulated. I will first present the Methods and the Results for each of these experiments and then proceed to discuss the results of the two experiments together.

Experiment 1B-2

This experiment had a similar dual-task design, in which participants read sentences phrase-by-phrase, and at the same time were required to perform arithmetic calculations. In contrast to Experiment 1B-1, the difficulty of the arithmetic task was manipulated by using different operations in the easy and hard conditions: additions were used in the easy condition and subtractions were used in the hard condition. There is evidence that children learn additions before subtractions (e.g., Siegler, 1987), and also that subtractions take longer and are more error prone than additions (Campbell & Xue, 2001), suggesting that there is something more difficult about the process of subtraction, compared to addition. The range of numbers used in both conditions was very similar. Therefore, if we observe a super-additive interaction between syntactic integrations and the new arithmetic task, then we may infer that the interaction observed in Experiment 1B-2 was not due to the difference in rehearsal demands between the easy and hard conditions of the arithmetic task, and that the working memory resources that are involved in performing the arithmetic task overlap with those that are involved in syntactic integration processes. In contrast, if language processing relies on an independent working memory resource pool, there should be no such interaction.

Methods

Participants Forty participants from MIT and the surrounding community were paid for their participation. All were native speakers of English and were naive as to the purposes of the study. None participated in Experiment 1B-1.

Design and materials The experiment had a 2 x 2 design, crossing syntactic complexity (subject-extracted RCs, object-extracted RCs) with arithmetic complexity (simple arithmetic operations (initial addend between 30 and 50, subsequent addends between 3 and 6, additions) vs. complex arithmetic operations (initial addend between 40 and 60, subsequent addends between 3 and 6, subtractions)). The size of the initial addend differed slightly between the easy and hard conditions, so that across the regions participants have to work with the numbers in approximately the same range (and hence of the same morpho-phonological complexity).

The language materials, including 40 fillers, were the same as those used in Experiment 1B-1. The numbers for the arithmetic task were randomly generated on-line for each participant with the constraints described above. For the filler sentences, the arithmetic task had the following constraints: (1) the initial addend was between 30 and 60, and (2) the subsequent addends (with the values between 3 and 6) could be either added or subtracted.

Procedure The procedure was exactly the same as in Experiment 1B-1. Participants took approximately 35 minutes to complete the experiment.

Results

Arithmetic accuracy Participants answered the arithmetic sum correctly 85% of the time. Table 5 presents the mean arithmetic accuracies across the four conditions of Experiment 1B-2. A two-factor ANOVA crossing arithmetic complexity (easy, hard) and syntactic complexity (easy, hard) on these data revealed no significant effects and no interaction (all $F_s < 1.5$). Notice that this pattern of results differs from that in Experiment 1B-1 where a main effect of arithmetic complexity was observed.

<u>Arithmetic complexity</u>	<u>Syntactic complexity</u>	
	Subject-extraction (Easy)	Object-extraction (Hard)
Easy arithmetic	85.6 (2.4)	85.3 (2.3)
Hard arithmetic	82.5 (2.4)	85.6 (2.2)

Table 5. Arithmetic accuracies in percent correct, as a function of arithmetic complexity and syntactic complexity in Experiment 1B-2 (standard errors in parentheses).

Comprehension question performance There were two comprehension questions following each experimental trial. Participants answered the first question correctly 78.5% of the time, and the second question 74.5% of the time. As in Experiment 1B-1, we collapsed the results in our analyses. Table 6 presents the mean accuracies across the four conditions of Experiment 1B-2. A two-factor ANOVA crossing arithmetic complexity (easy, hard) and syntactic complexity (easy, hard) on the responses to the two comprehension questions revealed no main effects and no interaction (all $F_s < 2$). Notice that this pattern of results is different from that in Experiment 1B-1, where an effect of syntactic complexity and an effect of arithmetic complexity (significant in the participants analysis) were observed.

<u>Arithmetic complexity</u>	<u>Syntactic complexity</u>	
	Subject-extraction (Easy)	Object-extraction (Hard)
Easy arithmetic	78.6 (2.5)	74.4 (2.4)
Hard arithmetic	76.1 (2.3)	76.9 (2.6)

Table 6. Comprehension accuracies in percent correct, as a function of arithmetic complexity and syntactic complexity in Experiment 1B-2 (standard errors in parentheses).

Reaction times As in Experiment 1B-1, we analyzed all trials, regardless of how the arithmetic and the comprehension questions were answered. Also, as in Experiment 1B-1, reaction time data points that were more than three standard deviations away from the mean residual RT for a position within a condition were excluded from the analysis, affecting 1.9% of the data. Figure 4 presents the mean residual RTs per region across the four conditions of Experiment 1B-2.

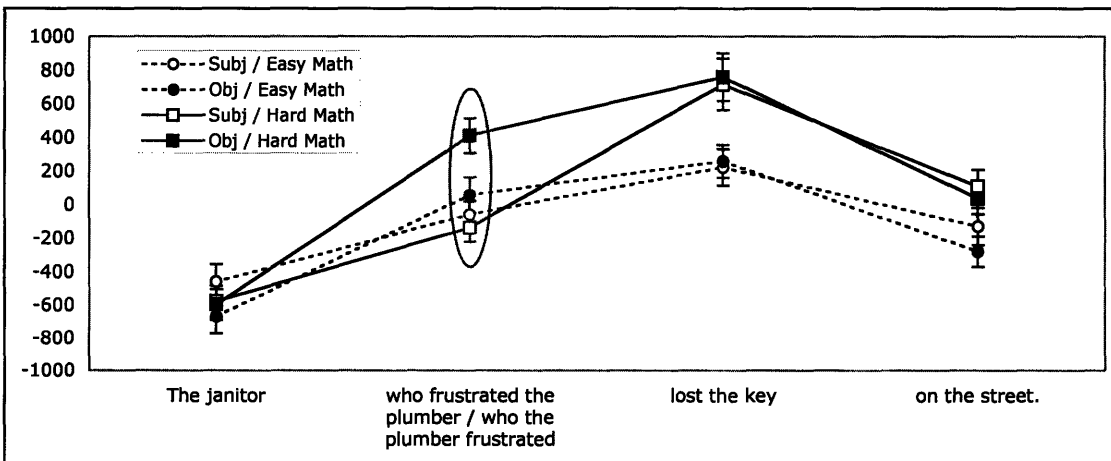


Figure 4: Reaction times per region in the four conditions of Experiment 1B-2. Error bars indicate standard errors. The critical region is circled.

We conducted a 2 x 2 ANOVA crossing syntactic complexity and the complexity of the arithmetic task for each of the four regions. The results are presented in Table 6. We first present the analysis of the critical region, Region 2, which included the relative clause (“who frustrated the plumber” / “who the plumber frustrated”), followed by the analyses of the other three regions. At the critical region, the hard-arithmetic conditions were read slower than the easy-arithmetic conditions (137.8 ms vs. -0.43 ms; 95% CI=139.4 ms), and the syntactically more complex object-extracted RC conditions were read slower than the subject-extracted conditions (234 ms vs. -96.7 ms; 95% CI=159.5 ms). Most importantly, there was a significant interaction, such that in the hard arithmetic conditions, the difference between subject- and object-extracted RCs was larger (546.4 ms) than in the easy arithmetic conditions (114.9 ms). This interaction is predicted by the hypothesis whereby sentence processing and arithmetic processing rely on overlapping pools of WM resources, but not by the hypothesis whereby the pools of resources are independent.

In the other three regions, the patterns of reaction times were as follows. In Region 1, there was an unpredicted effect of syntactic complexity, such that the object-extracted RCs (-633.3 ms) were read faster than the subject-extracted RC (-519.2 ms). There is no reason to expect a difference between subject- and object-extracted RCs in this region, because the linguistic materials were exactly the same. Similarly, there was an unpredicted interaction, such that the difference between the easy and the hard arithmetic conditions was larger in the subject-extracted conditions than in the object-extracted conditions. Again, there is no reason to expect any differences among the four conditions in this region, because the linguistic materials were exactly the same. In Regions 3 and 4, there was a main effect of arithmetic complexity, but no other effects and no interaction.

Table 7. Analysis of Variance results for Experiment 1B-2.

Source of variance	By participants			By items		min F'	
	<i>df</i>	F1 value	MSe	<i>df</i>	F2 value	<i>df</i>	min F'
Region 1							
Syntactic Complexity	1,39	5.29*	98456	1,31	4.88*	1,68	2.53
Arithmetic Complexity	1,39	<1	99559	1,31	<1	1,60	<1
Interaction	1,39	2.93	120448	1,31	5.60*	1,67	1.92
Region 2 (critical region containing the relative clause)							
Syntactic Complexity	1,39	17.58*	248728	1,31	15.76*	1,68	8.31*
Arithmetic Complexity	1,39	4.02	189919	1,31	2.70	1,64	1.61
Interaction	1,39	12.03*	154865	1,31	5.49*	1,56	3.76
Region 3							
Syntactic Complexity	1,39	<1	414711	1,31	<1	1,45	<1
Arithmetic Complexity	1,39	13.86*	714308	1,31	31.09*	1,65	9.58*
Interaction	1,39	<1	275918	1,31	<1	1,41	<1
Region 4							
Syntactic Complexity	1,39	1.79	271729	1,31	2.39	1,70	1.02
Arithmetic Complexity	1,39	15.11*	202847	1,31	12.32*	1,67	6.78*
Interaction	1,39	<1	137942	1,31	<1	1,44	<1

Note: Significant effects are marked by asterisk.

Experiment 1B-3

This experiment had a similar dual-task design, in which participants read sentences phrase-by-phrase, and at the same time were required to perform arithmetic calculations. In contrast to Experiments 1B-1 and 1B-2, this experiment used large numbers and the task did not require the participants to perform exact calculations; instead, the participants were instructed to round off the numbers to the nearest hundred. The difficulty of the arithmetic task was manipulated by changing the difficulty of the rounding off process which was necessary, or at least very advantageous, in performing the task. The range of numbers used in both conditions was the same. Therefore, if we observe a super-additive interaction between syntactic integrations and the new arithmetic task, this would provide further evidence that the interaction observed in Experiment 1B-1 was not due to the difference in rehearsal demands between the easy and hard conditions of the arithmetic task, and that the working memory resources that are involved in performing the arithmetic task overlap with those that are involved in syntactic integration processes.

Methods

Participants Forty participants from MIT and the surrounding community were paid for their participation. All were native speakers of English and were naive as to the purposes of the study. None participated in Experiments 1B-1 or 1B-2.

Design and materials The experiment had a 2 x 2 design, crossing syntactic complexity (subject-extracted RCs, object-extracted RCs) with arithmetic complexity. The addends in both the easy and the hard versions of the arithmetic task were between 1 and 399. Participants were explicitly instructed to approximate in their calculations. At the end of the sentence – unlike in Experiments 1B-1 and 1B-2 where participants were instructed to type in the result of their calculations – participants were instructed to make a forced-choice decision about which of the two numbers was closer to the result of their calculations. Neither of the two numbers in this forced-choice task corresponded exactly to the sum of the calculations, which was done to discourage the participants from performing the exact calculations. The correct answer was within 5% from the correct answer, and the incorrect answer was within 10-20% off.

The complexity manipulation had to do with how easy/hard it was to round off the exact numbers. We reasoned that it should be easier to round off the exact numbers if the last two digits are either close to the previous hundred (if they range from 1 to 20) or close to the next hundred (if they range from 81 to 99). For example, it intuitively seems easier to round off, for example, 210 to “approximately 200”, compared to 260.

The language materials, including 40 fillers, were the same as those used in Experiments 1B-1 and 1B-2. The numbers for the arithmetic task were randomly generated on-line for each participant with the constraints described above. For the filler sentences, the arithmetic task had the following constraints: (1) the numbers were in the same range as the target items (1-399), and (2) the last two digits covered the whole range (1 to 99).

Procedure The procedure was almost the same as in Experiments 1B-1 and 1B-2, except for the forced-choice task at the end. Participants took approximately 35 minutes to complete the experiment.

Results

Arithmetic accuracy Participants answered the forced-choice question correctly 80% of the time. Table 8 presents the mean arithmetic accuracies across the four conditions of Experiment 1B-3. A two-factor ANOVA crossing arithmetic complexity (easy, hard) and syntactic complexity (easy, hard) on these data revealed an unpredicted cross-over interaction ($F(1,39)=5.62$, $MSe=.1196$, $p<.05$; $F2(1,31)=8.81$, $MSe=.0957$, $p<.01$), such that while in the subject-extracted conditions the hard-arithmetic condition was more accurate than the easy-arithmetic condition, in the object-extracted conditions, the easy-arithmetic condition was more accurate than the hard-arithmetic condition.

<u>Arithmetic complexity</u>	<u>Syntactic complexity</u>	
	Subject-extraction (Easy)	Object-extraction (Hard)
Easy arithmetic	78.1 (3.4)	83.4 (2.4)
Hard arithmetic	81.3 (2.8)	75.6 (3.0)

Table 8. Arithmetic accuracies in percent correct, as a function of arithmetic complexity and syntactic complexity in Experiment 1B-3 (standard errors in parentheses).

Comprehension question performance There were two comprehension questions following each experimental trial. Participants answered the first question correctly 74.5% of the time, and the second question 73.1% of the time. As in Experiments 1B-1 and 1B-2, we collapsed the results in our analyses. Table 9 presents the mean accuracies across the four conditions of Experiment 1B-3. A two-factor ANOVA crossing arithmetic complexity (easy, hard) and syntactic complexity (easy, hard) on the responses to the two comprehension questions revealed no main effects and no interaction ($F_s < 2.8$).

<u>Arithmetic complexity</u>	<u>Syntactic complexity</u>	
	Subject-extraction (Easy)	Object-extraction (Hard)
Easy arithmetic	75.2 (2.2)	75.2 (2.7)
Hard arithmetic	71.7 (2.8)	73.0 (2.5)

Table 9. Comprehension accuracies in percent correct, as a function of arithmetic complexity and syntactic complexity in Experiment 1B-3 (standard errors in parentheses).

Reaction times As in Experiments 1B-1 and 1B-2, we analyzed all trials, regardless of how the arithmetic and the comprehension questions were answered. Also, as in Experiments 1-B1 and 1-B2, reaction time data points that were more than three standard deviations away from the mean residual RT for a position within a condition were excluded from the analysis, affecting 1.9% of the data. Figure 5 presents the mean residual RTs per region across the four conditions of Experiment 1B-2.

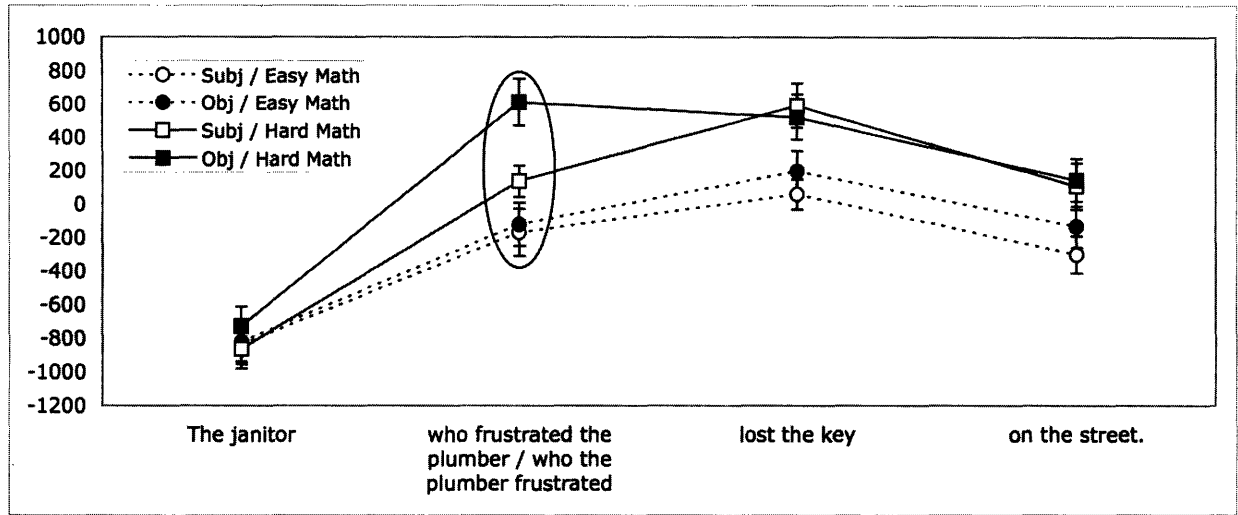


Figure 5: Reaction times per region in the four conditions of Experiment 1B-3. Error bars indicate standard errors. The critical region is circled.

We conducted a 2 x 2 ANOVA crossing syntactic complexity and the complexity of the arithmetic task for each of the four regions. The results are presented in Table 10. We first present the analysis of the critical region, Region 2, which included the relative clause (“who frustrated the plumber” / “who the plumber frustrated”), followed by the analyses of the other three regions. At the critical region, the hard-arithmetic conditions were read slower than the easy-arithmetic conditions (374.9 ms vs. -143.7 ms), and the syntactically more complex object-extracted RC conditions were read significantly slower than the subject-extracted conditions (246.2 ms vs. -15.04 ms). Most importantly, there was a significant interaction, such that in the hard arithmetic conditions, the difference between subject- and object-extracted RCs was larger (473.8 ms) than in the easy arithmetic conditions (48.6 ms). This interaction is predicted by the hypothesis whereby sentence processing and arithmetic processing rely on overlapping pools of WM resources, but not by the hypothesis whereby the pools of resources are independent.

In the other three regions, the patterns of reaction times were as follows. In Region 1, there were no significant effects ($F_s < 2.4$). In Regions 3 and 4, there was a main effect of arithmetic complexity, but no other effects or interactions.

Table 10. Analysis of Variance results for Experiment 1B-3.

Source of variance	By participants			By items	
	<i>df</i>	F1 value	MSe	<i>df</i>	F2 value
Region 1					
Syntactic Complexity	1,39	1.17	232455	1,31	2.302
Arithmetic Complexity	1,39	<1	42481	1,31	<1
Interaction	1,39	<1	136368	1,31	<1

Region 2 (critical region containing the relative clause)					
Syntactic Complexity	1,39	6.02*	2729244	1,31	11.3*
Arithmetic Complexity	1,39	10.3*	10758025	1,31	34.8*
Interaction	1,39	4.09*	1807954	1,31	4.51*
Region 3					
Syntactic Complexity	1,39	<1	44844	1,31	<1
Arithmetic Complexity	1,39	12.9*	7372391	1,31	18.2*
Interaction	1,39	<1	436851	1,31	<1
Region 4					
Syntactic Complexity	1,39	<1	418004	1,31	<1
Arithmetic Complexity	1,39	12.5*	4752458	1,31	19.2*
Interaction	1,39	<1	170501	1,31	<1

Note: Significant effects are marked by asterisk.

Discussion of Experiments 1B-2 and 1B-3

In Experiments 1B-2 and 1B-3 potential differences in rehearsal demands were controlled between the easy and hard versions of the arithmetic task. Despite this, we observed interactions similar to that in Experiment 1B-1. Specifically, the difference between the object-extracted and subject-extracted conditions was larger when the arithmetic integrations were hard, compared to when they were easy.

The results of Experiments 1B-2 and 1B-3 allow us to rule out the explanation of the interaction observed in Experiment 1B-1 in terms of the difference in rehearsal demands between the easy and the hard arithmetic conditions. Specifically, unlike in Experiment 1B-1, in Experiments 1B-2 and 1B-3 the numbers used in the easy and the hard arithmetic conditions did not differ in terms of length and/or morphological complexity; only the difficulty of the operations themselves was manipulated. Despite this fact, we still observed an interaction between syntactic and arithmetic complexity in the critical region, such that when both tasks were difficult, participants experienced more difficulty than would be expected if the two effects were purely additive. Thus, the results of Experiments 1B-2 and 1B-3 provide further support for a working memory framework where language processing and arithmetic processing rely on overlapping WM resource pools.

However, as discussed above, there is another confound present in Experiments 1B-1, 1B-2 and 1B-3. Specifically, it is possible to account for the observed interactions in terms of attention-switching costs: it is possible that in the hard conditions, more switches between the tasks are required, or the switches are more costly. To address this issue, Experiments 1B-4 and 1B-5 were conducted. As discussed above, ample evidence exists showing that different pools of working memory resources are used for verbal vs. visuo-spatial processing (e.g., Baddeley & Hitch, 1974; Baddeley, 1986; Vallar & Shallice, 1990; Hanley et al., 1991; Jonides et al., 1993;

Shah & Miyake, 1996). The attention-switching account predicts that an interaction similar to those observed in Experiments 1B-1 – 1B3 should be observed regardless of the nature of the two tasks involved, as long as they are matched for difficulty with the tasks used in Experiments 1B-1 – 1B-3. In Experiments 1B-4 and 1B-5 we used the same linguistic materials as in Experiments 1B-1 – 1B-3, but we substituted the arithmetic task with a spatial integration task. We used two different versions of a spatial integration task in the two new experiments.

Experiment 1B-4: Linguistic vs. Spatial integrations (the “Pie” spatial-rotation task)

This experiment used a similar dual-task paradigm as the first three experiments. In contrast to Experiments 1B-1 – 1B-3, however, the secondary task was a spatial-rotation task. In this task, participants were instructed to visually imagine adding different-size sectors of a circle and to keep track of the angle subtended by the combined segments. The most natural way to solve this task is to mentally rotate each incoming sector until it abuts the estimated sum of the previous sectors, as shown in Figure 6.

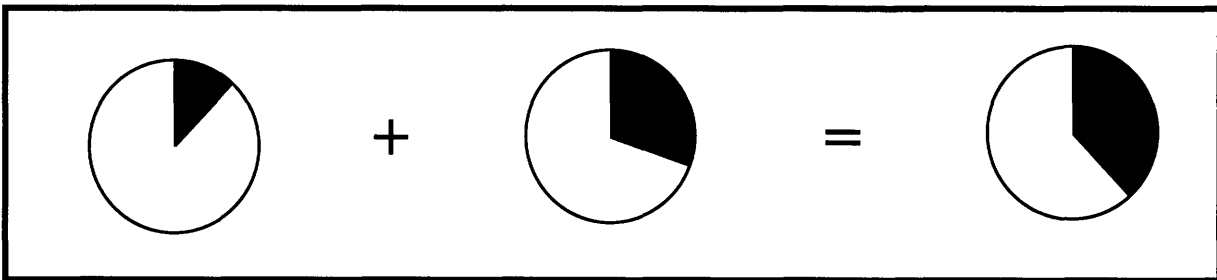


Figure 6: The “Pie” spatial-rotation task in Experiment 1B-4.

The on-line spatial-rotation task is similar to the addition task in that an incoming element – a sector – must be integrated into, or added to, the representation constructed thus far. Critically though, the spatial-rotation task does not rely on verbal working memory resources, and should not therefore interact with the sentence-processing task if the cause for the observed interaction in Experiments 1B-1 – 1B-3 is an overlap in the use of verbal working memory resources. However, if the attentional costs are responsible for the interaction, we should observe a similar interaction, regardless of the nature of the secondary task.

In order to draw conclusions of this sort, however, it is necessary to assure that the spatial integration task is of approximately the same difficulty as the arithmetic tasks used in the first three experiments. Specifically, one reason for why an interaction might be observed when a given task (let’s call it the primary task, for ease of discussion) is paired with one secondary task, but not when it’s paired with another secondary task could be that one secondary task is easier than the other secondary task. In this case, in the experiment where the secondary task is easier, resources might be abundant, and thus the results would not speak to the relationship between the pools of resources used to perform the two tasks. In other words, even if the two tasks rely on the same pool / overlapping pools of resources, it is possible that no super-additive interaction would be observed due to the fact that the pools do not get overtaxed even in the condition where

both tasks are complex. To make sure that our arithmetic tasks and our spatial integration tasks were comparable in difficulty, we had an independent group of participants perform the arithmetic task from Experiment 1B-1 and the spatial integration task from Experiment 1B-4 in isolation and we analyzed the reaction times and accuracies in these two tasks (see the section called *Norming for Task Difficulty* below).

Methods

Participants Sixty-four participants from MIT and the surrounding community were paid for their participation. All were native speakers of English and were naive as to the purposes of the study. None participated in Experiments 1B-1 – 1B-3.

Design and materials The experiment had a 2 x 2 design, crossing syntactic complexity (subject- / object-extracted RCs) with the complexity of the spatial-rotation task (simple rotations with small-angle sectors / complex rotations with larger-angle sectors). The language materials were the same as those used in Experiments 1B-1 – 1B-3.

The sectors for the spatial-rotation task were randomly generated on-line for each participant in the following way: the size of the sectors for the easy condition varied from 5 to 90 degrees, whereas the size of the sectors for the hard condition varied from 30 to 180 degrees. As a result, it was possible in the hard condition – but not in the easy condition – for the sum of the sectors to be more than 360 degrees, thus to “wrap around” the circle. Previous research (e.g., Shepard, 1971) has shown that the time it takes subjects to rotate a two- or three-dimensional figure is related to the angle of the rotation, such that larger angles take longer. Furthermore, pilot testing of the pie task by itself suggested that the task is easier to perform with smaller sectors.

Procedure The procedure was identical to that of Experiments 1B-1 – 1B-3, except for substituting the spatial-rotation task for the arithmetic tasks. Above each sentence fragment, participants saw a small circle. They were instructed to think of it as a plate for a pie. On each “plate”, there was a “pie-slice” shown in blue. The size of the “pie-slices” varied (as described in Materials and Design above), but they all started at the vertically-pointing radius position, as shown in Figure 7.





Time 1:	 The janitor -----
Time 2:	 ----- who frustrated the plumber -----
Time 3:	 ----- lost the key -----
Time 4:	 ----- on the street.

Figure 7: Sample frame-by-frame presentation of an item in Experiment 1B-4.

Participants were instructed to visually imagine adding each new “pie-slice” to the previous one(s) by mentally “putting” them next to each other. To assure that the participants performed the task, at the end of each trial a large blank circle appeared at the center of the screen with a vertically-pointing radius. Participants were instructed to drag this radius (using the mouse) to the end-point where all the “pie-slices” they just saw would come to when placed next to each other. If the answer was within 10 degrees of the correct answer, the words “Very Close!” flashed briefly on the screen; if the answer was within 35 degrees, the words “Pretty Good” flashed briefly; if the answer was within 90 degrees, the words “In The Ballpark” flashed briefly; finally, if the answer was not within 90 degrees, the words “Not Very Good” flashed briefly on the screen. The participants were warned that sometimes the “pie-slices”, when added together, would form more than a complete pie. In such cases, they were told to assume that the slices “wrapped around” and to ignore the complete portion of the pie when keeping track of the end-point. As in Experiments 1B-1 – 1B-3, this task was followed by two comprehension questions about the content of the sentences.

Thus far, we have been referring to the task in Experiment 1B-4 as a spatial integration task. In order to establish that participants were, in fact, performing this task using spatial working memory resources, and not verbal working memory resources, we administered a post-experimental questionnaire to try to understand the strategies the subjects might use in performing the pie task. The question about the strategies was open-ended, giving the subjects a chance to give any feedback they felt was relevant as to how they were performing the task. The open-ended nature of the question resulted in about half of the answers being impossible to code in terms of the strategy – spatial or verbal – used by the subject. The rest of the answers were coded as either “spatial-strategy” or “verbal-strategy”: we marked the strategy as being “spatial” if the answer explicitly mentioned a spatial process, and we marked the strategy as being “verbal” if the answer explicitly mentioned a verbally-mediated process, i.e. a process where

verbal labels could be mapped onto the spatial chunks. Of the answers where the type of strategy could be identified (56% of the responses), 73% of the responses were coded as “spatial-strategy”, and 27% - were coded as “verbal-strategy”. Examples of spatial-strategy answers included things like “tried to visualize it”, “I imagined the line rotating along with each piece”, “would try to visualize as I went along”, etc. Examples of verbal-strategy answers included things like “clock face patterns”, “usually rounding the pie slices to easy chunks was helpful, i.e. if the slice looked almost like a quarter, I rounded it to a quarter”, etc. Note that 27% is a conservative estimate, because even some of the answers, which were coded as “verbal-strategy” could, in principle, be performed in a spatial way: for example, quarters and halves are meaningful spatial chunks, and thus might be easier to operate on, compared to less meaningful sector-sizes. In other words, the fact that a participant would mention “quarters” does not necessarily imply the use of a verbal strategy. Thus, 27% represents an upper bound on the subjects that used a verbal strategy of those whose answers were codable. In addition to the questionnaire responses, several of the participants verbally reported that they initially tried a verbally-based strategy (e.g. encoding the pie-slices in terms of the number of hours), but had to quickly switch to the spatial-rotation strategy, because they found it too difficult to perform the pie-task using a verbal strategy.

Given the questionnaire feedback, it seems safe to conclude that most of the participants performed the task via spatial rotation, as instructed, and thus relied on spatial, and not verbal, working memory resources. Furthermore, foreshadowing Experiment 1B-5, it is worth noting that the spatial integration task there is much less subject to the criticism of potential reliance on verbal strategies, because unlike the spatial integration task in Experiment 1B-4 – where some verbal strategies seem possible – in Experiment 1B-5 it seems difficult to devise a verbal strategy for solving the spatial integration task.

Norming for Task Difficulty

In the Norming Study, an independent group of 37 participants, none of whom participated in any of the experiments described in section 1B, were asked to perform two tasks in turn: (1) the arithmetic addition task from Experiment 1B-1, and (2) the spatial integration task from Experiment 1B-4. The order of tasks was counterbalanced across the participants.

In theory, it should be possible to compare two tasks in terms of their relative difficulty using two dependent measures – reaction times and accuracies. However, it is very difficult to use accuracies as a dependent measure of performance on these two tasks, because the answers are qualitatively very different, and it is difficult to compare them. Specifically, in the arithmetic task, participants provide an answer (a sum), which is either correct or incorrect, thus the accuracies are calculated as percent correct. In contrast, in the spatial integration task, participants are asked to drag the radius to the position subtended by all the sectors added together, as described above, and the accuracies are calculated as degrees off from the correct answer. There is no obvious way to map these two measures onto each other. Therefore, the primary dependent measure we use is reaction time. Moreover, because we are interested in how working memory resources are used in *on-line* processing, a reaction time measure is more informative. An anonymous reviewer has observed that it is difficult in general to meaningfully

compare reaction times in two tasks that don't have comparable accuracy measures. Specifically, in order to argue that one task is more difficult than another, reaction times in the first task should be equal or longer, and the accuracies should be equal or lower. If the accuracies cannot be compared, then the possibility of speed-accuracy trade-offs arises, such that even though one task may take longer, it may be the case that participants are expending more effort to perform the task (and are therefore more accurate).

Whereas this issue can be a problem in comparing reaction times in some pairs of tasks, it is less relevant to the current comparison because the performance on the task which takes less time to perform (the arithmetic task) is at ceiling (97-99%). That is, (1) reaction times in the arithmetic task are faster than in the spatial-rotation task, and (2) the accuracies in the arithmetic task are at least as high as in the spatial-rotation task, because they are at ceiling. Thus, it is plausible that the spatial-rotation task is more difficult than the arithmetic task.

We first present a summary of the reaction time data and the accuracy data from the arithmetic task. Then, we present a similar summary for the spatial integration task. Finally, we present a direct comparison analysis for the two tasks in terms of reaction times.

In the arithmetic task, reaction times in every region revealed a significant effect of task complexity, such that more difficult additions took longer ($F_s > 5$, $p_s < .05$). Furthermore, the arithmetic task accuracies also revealed a significant effect of task complexity, such that more difficult additions were less accurate – 97% versus 99% ($F(1,36)=10.2$; $p < .05$; $F_2(1,31)=4.57$; $p < .05$).

Similarly, in the spatial integration task, reaction times in every region revealed a significant effect of task complexity, such that more difficult rotations took longer ($F_s > 5$, $p_s < .05$). The spatial integration task accuracies also revealed a significant effect of task complexity, such that more difficult rotations were less accurate – 29 degrees off from the correct answer versus 23 degrees ($F(1,36)=17.9$; $p < .001$; $F_2(1,31)=15.5$; $p < .001$). Notice that the accuracies for both the arithmetic task and the spatial integration task were somewhat higher (although very comparable) in the Norming Study, compared to Experiments 1B-1 and 1B-4, respectively. This is expected given that the task demands are higher in the dual-task experiments, compared to the Norming Study where each task is performed in isolation.

We used paired-samples two-tailed t-tests to compare raw reaction times for the two tasks during (1) Region 2 (the critical region in the four experiments described in this paper), and (2) across all four regions. Both t-tests revealed that participants took longer to perform the spatial integration task. The average reaction times during the critical region were 1099 ms for the arithmetic task, and 1867 ms for the spatial integration task ($t(1,36)=-7.85$, $p < .001$). The average reaction times during all the regions were 1125 ms for the arithmetic task, and 1709 ms for the spatial integration task ($t(1,36)=-6.67$, $p < .001$). These results suggest that the spatial integration task was more difficult for participants to perform than the arithmetic task, when performed in isolation.

Results

Spatial-rotation task accuracy On average, participants' estimates were 38.4 degrees off from the correct answer. Table 11 presents the mean accuracies (in degrees off from the correct answer) across the four conditions of Experiment 1B-4. A two-factor ANOVA crossing spatial-rotation task complexity (easy, hard) and syntactic complexity (easy, hard) revealed a main effect of complexity of the spatial-rotation task ($F(1,63)=19.31$; $MSe=4621$; $p < .001$; $F(1,31)=25.63$; $MSe=2295$; $p < .001$; $\min F'(1,90)=11.0$, $p<.002$), but no other significant effects ($F_s < 1$).

<u>Spatial task complexity</u>	<u>Syntactic complexity</u>	
	Subject-extraction (Easy)	Object-extraction (Hard)
Easy rotations	35.1 (2.9)	33.3 (2.4)
Hard rotations	42.9 (2.6)	42.4 (2.5)

Table 11. Spatial-task accuracies in degrees off from the correct answer, as a function of spatial task complexity and syntactic complexity in Experiment 1B-4 (standard errors in parentheses).

Comprehension question performance There were two comprehension questions following each experimental trial. Participants answered the first question correctly 78.7% of the time, and the second question 77.8% of the time. As in the other experiments, we collapsed the results in our analyses. Table 12 presents the mean accuracies across the four conditions of Experiment 1B-4. A two-factor ANOVA crossing spatial-rotation task complexity (easy, hard) and syntactic complexity (easy, hard) on the responses to the comprehension questions revealed a marginal effect of the spatial-rotation task complexity in the participants analysis ($F(1,63)=3.13$; $MSe=.0325$; $p = .082$; $F_2 < 1$; $\min F' < 1$), but no other effects or interactions ($F_s < 1$).

<u>Spatial task complexity</u>	<u>Syntactic complexity</u>	
	Subject-extraction (Easy)	Object-extraction (Hard)
Easy rotations	79.7 (2.0)	79.0 (2.1)
Hard rotations	76.9 (2.0)	77.3 (1.9)

Table 12. Comprehension accuracies in percent correct, as a function of spatial task complexity and syntactic complexity in Experiment 1B-4 (standard errors in parentheses).

Reaction times As in Experiments 1B-1 – 1B-3, we analyzed all trials, regardless of how the comprehension questions were answered and how the spatial-rotation task was performed. Also, as in Experiments 1B-1 – 1B-3, reaction time data points that were more than three standard deviations away from the mean residual RT for a position within a condition were excluded from the analyses, affecting 1.8% of the data. Figure 8 presents the mean residual reaction times per region across the four conditions in Experiment 1B-4.

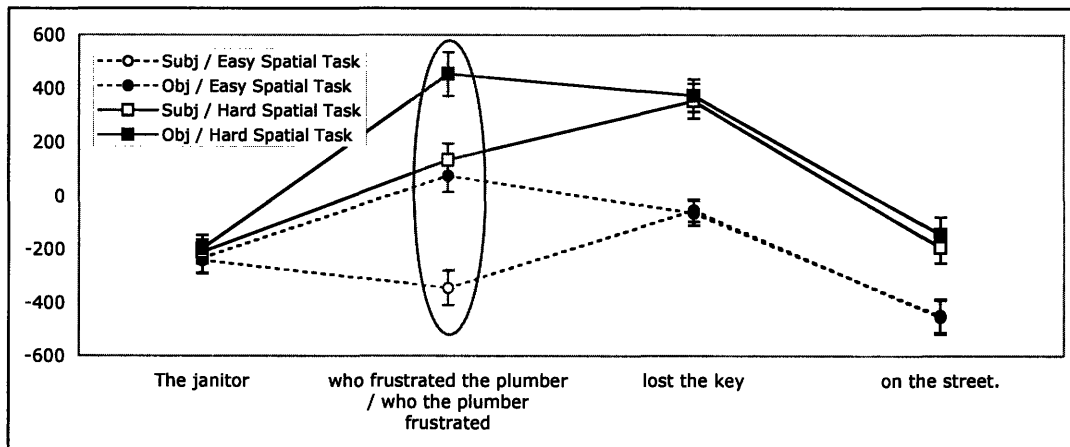


Figure 8: Reaction times per region in the four conditions of Experiment 1B-4. Error bars indicate standard errors. The critical region is circled.

We conducted a 2 x 2 ANOVA crossing syntactic complexity and the complexity of the spatial integration task for each of the four regions. The results are presented in Table 13. We first present the analysis of the critical region, Region 2, which included the relative clause (“who frustrated the plumber” / “who the plumber frustrated”), followed by the analyses of the other three regions. At the critical region, the hard-spatial-task conditions were read slower than the easy-spatial-task conditions (293.8 ms vs. -133.9 ms; 95% CI=158.2 ms), and the syntactically more complex object-extracted RC conditions were read slower than the subject-extracted RC conditions (264.9 ms vs. -105.0 ms; 95% CI=118.6 ms). Critically, there was no interaction between syntactic complexity and the complexity of the spatial task. Moreover, the effect of syntactic complexity in the hard-spatial-task conditions was numerically *smaller* (319.5 ms) than that in the easy-spatial-task conditions (420.3 ms). This result argues against the attentional explanation of the interaction that was observed in Experiments 1B-1 – 1B-3.

In Region 1, there were no significant effects. In Regions 3 and 4, there was a main effect of spatial task complexity, but no other significant effects.

Table 13. Analysis of Variance results for Experiment 1B-4.

Source of variance	By participants			By items		min F'	
	<i>df</i>	F1 value	MSe	<i>df</i>	F2 value	<i>df</i>	min F'
Region 1							
Syntactic Complexity	1,63	<1	76995	1,31	<1	1,94	<1
Spatial-Task Complexity	1,63	<1	83902	1,31	<1	1,69	<1
Interaction	1,63	<1	77597	1,31	<1	1,52	<1
Region 2 (critical region containing the relative clause)							
Syntactic Complexity	1,63	38.82*	225555	1,31	47.77*	1,88	21.4*
Spatial-Task Complexity	1,63	29.19*	401031	1,31	67.44*	1,94	20.3*
Interaction	1,63	1.10	147629	1,31	1.38	1,89	<1
Region 3							
Syntactic Complexity	1,63	<1	113081	1,31	<1	1,93	<1
Spatial-Task Complexity	1,63	52.45*	217636	1,31	88.86*	1,93	32.9*
Interaction	1,63	<1	138094	1,31	<1	1,39	<1
Region 4							
Syntactic Complexity	1,63	<1	69797	1,31	<1	1,61	<1
Spatial-Task Complexity	1,63	30.78*	166628	1,31	107.41*	1,89	23.9*
Interaction	1,63	<1	126894	1,31	<1	1,91	<1

Note: Significant effects are marked by asterisk.

Analysis of Experiment 1B-1 and Experiment 1B-4 with experiment as a factor

To further strengthen the conclusions we draw from the different patterns of results we observed in Experiment 1B-1 (an interaction between linguistic and arithmetic complexity) and in Experiment 1B-4 (a lack of an interaction between linguistic and spatial integration task complexity), we analyzed the two datasets – from Experiments 1B-1 and 1B-4 – using a 2 x 2 x 2 ANOVA, with the following factors: (1) syntactic complexity (subject-extractions / object-extractions), (2) non-linguistic task (arithmetic/spatial-rotation) complexity (easy / hard), and (3) experiment (Experiment 1B-1 / Experiment 1B-4). At the critical relative clause region we observed a significant three-way interaction, such that the interaction between syntactic and non-linguistic task complexity was observed only in Experiment 1B-1, and not in Experiment 1B-4 ($F(1,110)=4.84$; $MSe=779539$; $p < .05$; $F(1,31)=12.2$; $MSe=466202$; $p < .002$; $\min F'(1,139)=3.27$; $p=.07$). There was no such interaction in any of the other regions ($F_s < 1$). We further examined the raw reaction times both at the critical region and across all the regions in Experiment 1B-1 and Experiment 1B-4, and we found that the time ranges were very similar. Specifically, across all the regions, the mean raw reaction time in Experiment 1B-1 was 2187 ms

(SE=43 ms), and the mean raw reaction time in Experiment 1B-4 was 2066 ms (SE=59 ms); at the critical region, the mean raw RT in Experiment 1B-1 was 2780 ms (SE=66 ms), and the mean raw reaction time in Experiment 1B-4 was 2631 ms (SE=79 ms).

Discussion

The attention-switching account of the interaction between syntactic and arithmetic complexity that was observed in Experiments 1B-1 – 1B-3 predicted a similar interaction between syntactic and spatial integration task complexity in Experiment 1B-4. No comparable interaction was observed.

There are at least four possible reasons for why one might not observe an interaction between the language processing task and the spatial integration task in Experiment 1B-4. First, the spatial integration task might have been too easy, with the consequence that participants were not overly taxed in the condition where the complexity of both tasks was high. A prediction of this hypothesis is that the spatial integration task should be easier to process than the arithmetic processing task, because the arithmetic processing task in Experiments 1B-1 – 1B-3 interacted with the language processing task (either due to a shared pool of working memory resources, or due to the attention-switching costs). Contrary to this prediction, the Norming Study established that the spatial integration task was not easier than the arithmetic task. As discussed above, it took participants longer to perform the spatial integration task than the arithmetic task when the tasks were presented in isolation, and the accuracy on the arithmetic task was at ceiling, with the consequence that the accuracy on the spatial integration task could not be higher. Thus, the lack of an interaction in Experiment 1B-4 was not due to the low complexity of the spatial integration task.

Second, it is possible that the spatial integration task was too difficult, with the consequence that the difficulty of the spatial integration task would swamp the syntactic complexity effect. If this were the case, the following patterns of data would be predicted: either (1) no syntactic complexity effect in the hard spatial integration conditions, or (2) no syntactic complexity effect in both the easy and the hard spatial integration conditions. Contrary to this prediction, our data revealed a main effect of syntactic complexity which was present in both the easy spatial integration task conditions ($F_{s} > 34$, $p < .001$) and the hard spatial integration task conditions ($F_{s} > 13$, $p < .001$). Thus, the lack of an interaction in Experiment 1B-4 was not due to the high complexity of the spatial integration task.

Third, it is possible that Experiment 1B-4 did not have enough power to detect the interaction between syntactic complexity and spatial integration task complexity. The standard practice in performing post-hoc power analyses is to estimate the expected effect size (f) in the experiments at question based on the effect sizes observed in similar experiments in previously conducted research (e.g., Rosenthal & Rosnow, 1984). However, because of the novelty of the experimental paradigm used in the experiments reported here, there were no prior similar studies from which we could estimate the expected effect size for our experiments. We therefore estimated the effect size based on Experiments 1B-1 and 1B-2, where we observed the critical interaction. The f -values for these two experiments – calculated using the partial eta squared values for the interaction effect in each of the experiments – were .307 and .556. To perform the

power analysis for the spatial-task experiments (Experiment 1B-4, and Experiment 1B-5 to be presented below), we calculated the mean f -value for Experiments 1B-1 and 1B-2 (.438). The resulting power levels (calculated using G*Power, available at <http://www.psych.uni-duesseldorf.de/aap/projects/gpower/>) were as follows: the power in Experiment 1B-4 was .932, and the power in Experiment 1B-5 was .809. These power levels are higher than the power threshold of .80 accepted as a standard in the field (e.g., Rosenthal & Rosnow, 1984; Cohen, 1988). We therefore conclude that our spatial-task experiments (Experiments 1B-4 and 1B-5) had sufficient power to detect an interaction of the size observed in Experiments 1B-1 and 1B-2.

Finally, the lack of an interaction could result from the fact that linguistic and spatial integrations rely on independent pools of working memory resources. Because the first three reasons are not likely to be able to account for the lack of an interaction in Experiment 1B-4, as discussed above, the independence of resource pools for linguistic vs. spatial integration processes is plausibly responsible for the lack of an interaction.

An anonymous reviewer pointed out an additional concern in comparing the arithmetic task and the spatial integration task. Specifically, it is possible that in the arithmetic task participants might integrate the addends online as they go along in the sentence, but in the spatial integration task they might store the pie-slices and add them up at the end of the sentence, thus not performing the integrations online, as instructed. If that were the case, the online reaction time data from the two tasks (additions vs. the spatial integration task) would not be very meaningful, as the underlying processes involved in the performance of the two tasks would be drastically different. Two sources of evidence suggest that it is unlikely that participants were following this proposed strategy. First, the storing strategy predicts that reaction times should increase from region 1 to region 4, peaking at region 4 where participants would be holding on to three pie-slices from the previous regions and adding the fourth one to the stack. However, this is not the pattern of reaction times we observed in Experiment 3: reaction times are slowest for regions 2 and 3, but at region 4 they come down to the reaction time level of region 1 (see Figure 8 above). Second, in the post-experimental questionnaire mentioned above, we coded the subjects' responses for whether they contained any mention of the type of strategy with regard to the time-course of performing the task. This constituted 43% of the subjects. Out of these responses, 93% strongly suggested that the task was performed incrementally. Examples of such responses included "made mental hash marks on the circle to keep track of how much space was covered by the blue at each step", "adding past slices as the new slices were added", and "added the pieces together as I went", etc. The remaining 7% of the responses were unclear with regard to the time-course issue. None of the subjects explicitly mentioned performing the task by storing individual pie slices while moving along the sentence and adding up the slices at the end. However, it is difficult to conclusively rule out the possibility that on some trials some participants may have used the storage-based strategy, which could have contributed to the lack of an interaction in the critical region. In summary, based on the pattern of results in reaction times and on the questionnaire responses, we conclude that participants usually performed the spatial integration task incrementally on-line.

Experiment 1B-5: Linguistic vs. Spatial integrations (the “Grid” spatial-integration task)

In order to evaluate the generality of the results from Experiment 1B-4, we investigated the relationship between the working memory resources used for online language processing and online spatial integration processing using a different version of a spatial integration task in a similar dual-task paradigm. In this task, participants were presented with a series of three-by-four grids with some squares filled in in blue. Participants were instructed to imagine combining the squares into a geometrical shape, as shown in Figure 9.

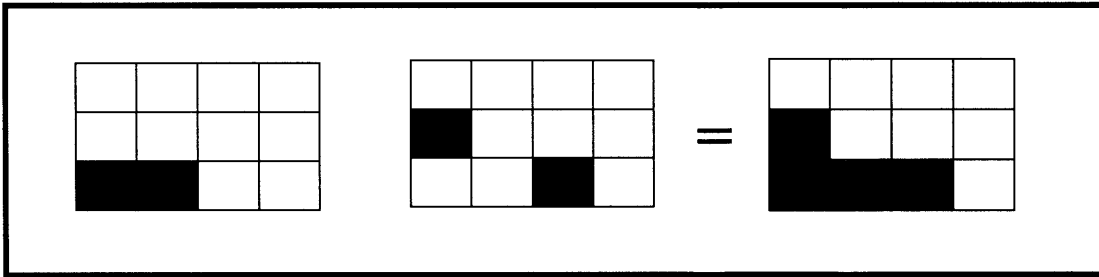


Figure 9: The “Grid” spatial-integration task in Experiment 1B-5.

This spatial integration task is similar to the arithmetic tasks in Experiments 1B-1 – 1B-3 in that an incoming element / incoming elements – a square / squares – must be integrated into, or added to, the evolving representation. Similar to the spatial rotation task in Experiment 1B-4, the spatial integration task does not rely on verbal working memory resources, and should not therefore interact with the sentence-processing task if the cause for the observed interactions in Experiments 1B-1 – 1B-3 is an overlap in the use of verbal working memory resources.

Methods

Participants Forty-four participants from MIT and the surrounding community were paid for their participation. All were native speakers of English and were naive as to the purposes of the study. None participated in Experiments 1B-1 – 1B-4.

Design and materials The experiment had a 2 x 2 design, crossing syntactic complexity (subject- / object-extracted RCs) with the complexity of the spatial integration task (simple integrations with one square per grid / complex integrations with two squares per grid). The language materials were the same as those used in the other experiments.

The squares for the spatial integration task were randomly generated on-line for each participant in the following way: one square or two adjacent (sharing sides) squares were shown in the first grid; in each subsequent grid, the square(s) that were shown in the previous grid were hidden, and one or two squares were added, such that it/they shared sides with the square(s) in the previous grid.

Procedure The procedure was identical to that of the other experiments, except for the new spatial integration task. Above each sentence fragment, participants saw a three-by-four grid.

The first grid had one or two squares filled in in blue. If two squares were filled in, they were always adjacent. On the next grid, the square(s) that were filled in in the first grid were hidden and one or two other squares were filled in, such that if the square(s) from the first grid were shown, the new square(s) would share sides with them. The participants were instructed to imagine constructing a geometrical shape out of the squares, as shown in Figure 10.

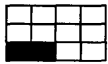



Time 1:		The janitor
Time 2:		who frustrated the plumber
Time 3:		lost the key
Time 4:		on the street.

Figure 10: Sample frame-by-frame presentation of an item in Experiment 1B-5.

To assure that the participants performed the task, at the end of each trial a blank grid appeared at the center of the screen. Participants were instructed to click on the squares (using the mouse) that have been highlighted across the four grids. If all the squares were filled in correctly, the word “Right!” flashed briefly on the screen; if all but one square were filled in correctly (including either a false positive or a missing square), the word “Almost” flashed briefly; finally, if the answer was two or more squares off, the words “Not quite” flashed briefly on the screen. As in the other experiments, this task was followed by two comprehension questions about the content of the sentences.

Results

Spatial integration task accuracy The performance on the spatial integration task was measured in the following way: the number of errors was divided by the total number of squares that should have been selected, where the number of errors was calculated as the maximum of the number of misses and false alarms (i.e. leaving a square out, adding an extra square, or swapping a correct square for an incorrect square would all count as a single error). On average, participants made 5.7% errors. Table 14 presents the mean accuracies (in percent of errors) across the four conditions of Experiment 1B-5. A two-factor ANOVA crossing spatial integration task complexity (easy, hard) and syntactic complexity (easy, hard) revealed a main effect of the spatial integration task complexity ($F(1,43)=44.4$; $MSe=0.0931$; $p < .001$;

$F(1,31)=97.8$; $MSe=0.0677$; $p < .001$; $\min F'(1,71)=30.5$; $p<.001$), but no other significant effects ($F_s<2.5$).

<u>Spatial task complexity</u>	<u>Syntactic complexity</u>	
	Subject-extraction (Easy)	Object-extraction (Hard)
Easy integrations	2.89 (.71)	3.48 (.86)
Hard integrations	7.78 (1.1)	8.24 (1.2)

Table 14. Spatial-task accuracies in percent of errors, as a function of spatial task complexity and syntactic complexity in Experiment 1B-5 (standard errors in parentheses).

Comprehension question performance There were two comprehension questions following each experimental trial. Participants answered the first question correctly 81.9% of the time, and the second question 78.3% of the time. As in the other experiments, we collapsed the results in our analyses. Table 15 presents the mean accuracies across the four conditions of Experiment 1B-5. A two-factor ANOVA crossing spatial integration task complexity (easy, hard) and syntactic complexity (easy, hard) on the responses to the comprehension questions revealed a main effect of the spatial integration task complexity ($F(1,43)=25.39$; $MSe=0.2784$; $p < .001$; $F(1,31)=9.71$; $MSe=0.2025$; $p < .005$; $\min F'(1,54)=7.02$; $p<.02$), but no other significant effects and no interaction ($F_s<1$).

<u>Spatial task complexity</u>	<u>Syntactic complexity</u>	
	Subject-extraction (Easy)	Object-extraction (Hard)
Easy integrations	84.2 (2.5)	83.9 (2.1)
Hard integrations	77.1 (2.6)	75.1 (2.3)

Table 15. Comprehension accuracies in percent correct, as a function of spatial task complexity and syntactic complexity in Experiment 1B-5 (standard errors in parentheses).

Reaction times As in the other experiments, we analyzed all trials, regardless of how the comprehension questions were answered and how the spatial integration task was performed. Also, as in the previous experiments, reaction time data points that were more than three

standard deviations away from the mean residual RT for a position within a condition were excluded from the analyses, affecting 1.5% of the data. Figure 11 presents the mean residual reaction times per region across the four conditions in Experiment 1B-5.

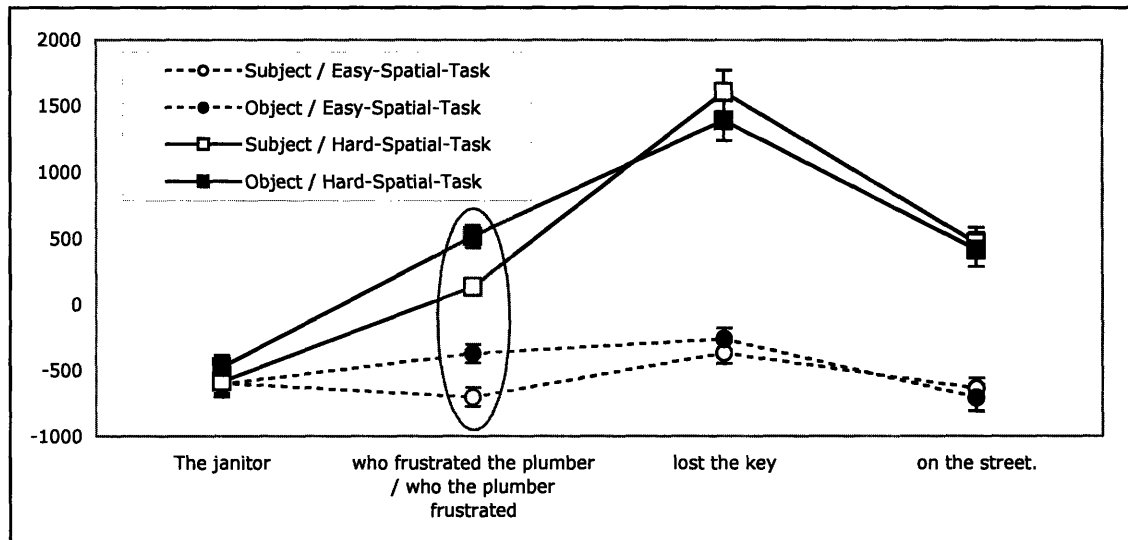


Figure 11: Reaction times per region in the four conditions of Experiment 1B-5. Error bars indicate standard errors. The critical region is circled.

We conducted a 2 x 2 ANOVA crossing syntactic complexity and the complexity of the spatial integration task for each of the four regions. The results are presented in Table 16. We first present the analysis of the critical region, Region 2, which included the relative clause (“who frustrated the plumber” / “who the plumber frustrated”), followed by the analyses of the other three regions. At the critical region, the hard-spatial-task conditions were read slower than the easy-spatial-task conditions (326.7 ms vs. -536.8 ms; 95% CI=160.4 ms), and the syntactically more complex object-extracted RC conditions were read slower than the subject-extracted RC conditions (73.2 ms vs. -283.3 ms; 95% CI=121.8 ms). Critically, there was no trace of an interaction between syntactic complexity and the complexity of the spatial task. This pattern of results is similar to that in Experiment 1B-4, and it provides additional evidence against the attention-switching explanation of the interaction that was observed in Experiments 1B-1 – 1B-3.

In Region 1, there was an effect of the spatial task complexity, which wasn’t significant in the items analysis. There were no other effects. In Region 3, there was a main effect of spatial task complexity. There was also an unpredicted interaction, such that in the easy-spatial-task conditions, the object-extracted condition is numerically slower, and in the hard-spatial-task conditions, the object-extracted condition is numerically faster than the subject-extracted condition. There is no reason to expect an interaction of this sort here, as the linguistic materials were exactly the same in the subject- and object-extracted conditions. Finally, in Region 4, there was again a main effect of spatial task complexity, but no other significant effects.

Table 16. Analysis of Variance results for Experiment 1B-5.

Source of variance	By participants			By items		min F'	
	<i>df</i>	F1 value	MSe	<i>df</i>	F2 value	<i>df</i>	min F'
Region 1							
Syntactic Complexity	1,43	1.70	69232	1,31	<1	1,52	<1
Spatial-Task Complexity	1,43	4.99*	44118	1,31	2.48	1,59	1.65
Interaction	1,43	2.74	55747	1,31	1.23	1,57	<1
Region 2 (critical region containing the relative clause)							
Syntactic Complexity	1,43	34.86*	160378	1,31	45.87*	1,74	19.8*
Spatial-Task Complexity	1,43	117.82*	278444	1,31	318.85*	1,68	86.0*
Interaction	1,43	<1	105178	1,31	<1	1,72	<1
Region 3							
Syntactic Complexity	1,43	<1	296470	1,31	<1	1,64	<1
Spatial-Task Complexity	1,43	97.72*	1481332	1,31	587.36*	1,56	83.7*
Interaction	1,43	4.25*	266769	1,31	5.09*	1,74	2.31
Region 4							
Syntactic Complexity	1,43	1.26	122314	1,31	<1	1,57	<1
Spatial-Task Complexity	1,43	51.32*	1064008	1,31	584.14*	1,50	47.1*
Interaction	1,43	<1	171070	1,31	<1	1,53	<1

Note: Significant effects are marked by asterisk.

Discussion

The pattern of results in Experiment 1B-5 was very similar to the pattern of results in Experiment 1B-4. Specifically, we found a main effect of linguistic complexity, a main effect of spatial integration complexity, but no trace of an interaction. The attention-switching account of the interaction between syntactic and arithmetic complexity that was observed in Experiments 1B-1 – 1B-3 predicted a similar interaction between syntactic and spatial integration task complexity in Experiments 1B-4 and 1B-5. No such interaction was observed in either Experiment 1B-4 or 1B-5. Therefore, we conclude that the interactions observed in Experiments 1B-1 – 1B-3 cannot be accounted for in terms of the attention-switching account.

Experiment 1B-6: The spatial-rotation task with verbal instructions

To make the comparison between the tasks involving arithmetic and spatial integration processes even more minimal, Experiment 1B-6 was conducted. This experiment used the same dual task

paradigm as Experiments 1B-1 – 1B-5, with the secondary task from Experiment 1B-4 (the “pie” task). Critically, the participants were instructed to perform the task using a verbal strategy (treating the “pie” as a clock and estimating the number of hours at each step, adding the numbers along the way). The shared working memory resource pool hypothesis predicts that with the verbal instructions the “pie” task from Experiment 1B-4 should reveal the same pattern of results as the experiments involving arithmetic integrations.

Methods

Participants Eighty participants from MIT and the surrounding community were paid for their participation. All were native speakers of English and were naive as to the purposes of the study. None participated in Experiments 1B-1 – 1B-5.

Design and materials The experiment had a 2 x 2 design, crossing syntactic complexity (subject- / object-extracted RCs) with the complexity of the secondary task, which involved visual displays like those used in Experiment 1-B4 (circles with a sector of some size filled in in blue) but requested that the participants use a verbal strategy. In particular, the participants were told to think of the circle as a clock displaying a certain number of hours and were instructed to add the number of hours as they went along (let’s call this task “the hours task”).

The language materials were the same as those used in the other experiments.

The displays for the “hours” task were randomly generated on-line for each participant. The sectors always fell in a range of +/-10 degrees off the “hour” (e.g., 350-10, 20-40, 50-70, 80-100, 110-130, etc.), and the participants were told to approximate the number to the nearest hour. For the complexity manipulation, we had participants add numbers in the range of 1 – 5 hours in the easy condition, and in the range of 4 – 8 hours in the hard condition.

Procedure The procedure was identical to that of the other experiments, except for the new “hours” task.

To assure that the participants performed the task, at the end of each trial a forced-choice question appeared where one of the answers corresponded to the correct answer and the other answer was off by two hours. Participants were instructed to press one of two buttons to indicate their answer. If the correct answer was chosen, the word “Right!” flashed briefly on the screen; if the incorrect answer was chosen, the word “Wrong” flashed briefly. As in the other experiments, this task was followed by two comprehension questions about the content of the sentences.

Results

The “hours” task accuracy On average, participants answered the “hours” forced-choice question correctly 84.7% of the time. Table 17 presents the mean accuracies across the four conditions of Experiment 1B-6. A two-factor ANOVA crossing the “hours” task complexity (easy, hard) and syntactic complexity (easy, hard) revealed a main effect of extraction

($F(1,79)=6.55$; $MSe=0.1125$; $p < .02$; $F(1,31)=5.87$; $MSe=0.045$; $p < .05$), but no other significant effects ($F_s < 2.6$).

<u>The “hours task complexity</u>	<u>Syntactic complexity</u>	
	Subject-extraction (Easy)	Object-extraction (Hard)
Easy integrations	88.3 (1.7)	82.3 (2.1)
Hard integrations	84.8 (1.9)	83.3 (2.2)

Table 17. The “hours” task accuracies in percent correct, as a function of the “hours” task complexity and syntactic complexity in Experiment 1B-6 (standard errors in parentheses).

Comprehension question performance There were two comprehension questions following each experimental trial. Participants answered the first question correctly 73.9% of the time, and the second question 71.4% of the time. As in the other experiments, we collapsed the results in our analyses. Table 18 presents the mean accuracies across the four conditions of Experiment 1B-6. A two-factor ANOVA crossing the “hours” task complexity (easy, hard) and syntactic complexity (easy, hard) on the responses to the comprehension questions revealed a main effect of the “hours” task complexity in the participants analysis only ($F(1,79)=7.28$; $MSe=0.0762$; $p < .01$; $F(1,31)=1.73$; $MSe=0.0305$; n.s.), but no other effects ($F_s < 2.4$).

<u>The “hours” task complexity</u>	<u>Syntactic complexity</u>	
	Subject-extraction (Easy)	Object-extraction (Hard)
Easy integrations	75.6 (1.7)	72.8 (1.8)
Hard integrations	71.6 (1.7)	70.6 (1.6)

Table 18. Comprehension accuracies in percent correct, as a function of the “hours” task complexity and syntactic complexity in Experiment 1B-6 (standard errors in parentheses).

Reaction times As in the other experiments, we analyzed all trials, regardless of how the comprehension questions were answered and how the “hours” task was performed. Figure 12 presents the mean residual reaction times per region across the four conditions in Experiment 1B-6.

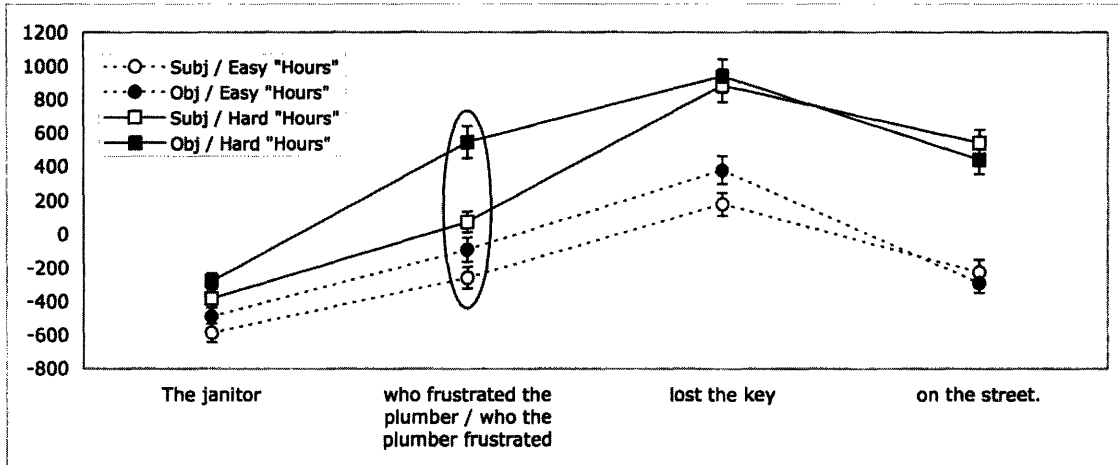


Figure 12: Reaction times per region in the four conditions of Experiment 1B-6. Error bars indicate standard errors. The critical region is circled.

We conducted a 2 x 2 ANOVA crossing syntactic complexity and the complexity of the “hours” task for each of the four regions. The results are presented in Table 19. We first present the analysis of the critical region, Region 2, which included the relative clause (“who frustrated the plumber” / “who the plumber frustrated”), followed by the analyses of the other three regions. At the critical region, the hard “hours” task conditions were read slower than the easy “hours” task conditions (307.7 ms vs. -176.2 ms), and the syntactically more complex object-extracted RC conditions were read slower than the subject-extracted RC conditions (226.1 ms vs. -94.5 ms). Critically, there was a significant interaction, such that in the hard “hours” task conditions, the difference between subject- and object-extracted RCs was larger (473.6 ms) than in the easy arithmetic conditions (167.6 ms). This interaction is predicted by the hypothesis whereby sentence processing and arithmetic processing rely on overlapping pools of WM resources, but not by the hypothesis whereby the pools of resources are independent.

In Region 1, there was an effect of the “hours” task complexity. There was also an unpredicted effect of syntactic complexity (marginal in the items analysis). There is no reason to expect this effect here, as the linguistic materials were exactly the same in the subject- and object-extracted conditions. In Region 3, there was a main effect of the “hours” task complexity. There was also a marginal effect of syntactic complexity in the participants analysis. Finally, in Region 4, there was again a main effect of the “hours” task complexity.

Table 19. Analysis of Variance results for Experiment 1B-6.

Source of variance	By participants			By items	
	<i>df</i>	F1 value	MSe	<i>df</i>	F2 value
Region 1					
Syntactic Complexity	1,79	4.67*	831341	1,31	3.42
The “hours” Task Complexity	1,79	14.9*	3406045	1,31	20.7*
Interaction	1,79	<1	560	1,31	<1
Region 2 (critical region containing the relative clause)					
Syntactic Complexity	1,79	18.3*	8223248	1,31	19.1*
The “hours” Task Complexity	1,79	47.2*	18733073	1,31	37.9*
Interaction	1,79	5.24*	1871937	1,31	5.42*
Region 3					
Syntactic Complexity	1,79	3.5	1350800	1,31	2.47
The “hours” Task Complexity	1,79	67.8*	32201687	1,31	54.2*
Interaction	1,79	1.12	430620	1,31	<1
Region 4					
Syntactic Complexity	1,79	1.52	534422	1,31	1.02
The “hours” Task Complexity	1,79	78.7*	44849741	1,31	207.9*
Interaction	1,79	<1	26624	1,31	<1

Note: Significant effects are marked by asterisk.

Discussion

The pattern of results in Experiment 1B-6 provides even stronger support for the shared working memory resource pool hypothesis whereby linguistic integrations and arithmetic integrations rely on overlapping pools of resources. The experiment was designed to use the same materials as one of the spatial-integration experiments and to differ only in the instructions given to the participants. We observed that when the participants were using the verbal strategy, linguistic complexity interacted super-additively with the complexity of the secondary task (the “hours” task).

Conclusions for Experiments 1B-1 through 1B-6

We reported the results of six dual-task experiments which were aimed at investigating the nature of working memory resources in linguistic integrations. The way we approached this question was by crossing syntactic complexity with the complexity of another task, which

involved similar integration processes. This secondary task either involved arithmetic integration processes and therefore relied on the use of verbal working memory, or it involved spatial integration processes. Experiments 1B-1 – 1B-3 crossed syntactic complexity and arithmetic complexity. These experiments showed two main effects and a super-additive interaction during the critical region of the linguistic materials, such that in the condition where both syntactic and arithmetic complexity was high the reaction times were longer than would be expected if the two complexity effects were additive. This pattern of results suggests that linguistic and arithmetic integrations rely on overlapping pools of verbal working memory resources.

To account for a potential confound in terms of attention-switching costs in dual-task paradigms, Experiments 1B-4 and 1B-5 crossed syntactic complexity and the complexity of a spatial integration task. The attention-switching account predicts a similar interaction regardless of the nature of the tasks involved, as long as the tasks are matched for complexity. In contrast, the hypothesis whereby linguistic and arithmetic integrations rely on overlapping pools of verbal working memory resources predicts no interaction in cases when one of the tasks does not rely on verbal working memory resources. Both Experiment 1B-4 and Experiment 1B-5 revealed two main effects, but no suggestion of an interaction comparable to the interactions observed in Experiments 1B-1 – 1B-3. These results therefore provide evidence against the attention-switching account.

Finally, in Experiment 1B-6 we used the secondary task from Experiment 1B-4 and instructed the participants to use a verbal strategy. We observed that linguistic complexity interacted super-additively with the complexity of the secondary task, as in the Experiments with the arithmetic task.

We discussed three alternative hypotheses which could account for the lack of an interaction in the spatial task experiments: (1) the spatial tasks may have been too easy, (2) the spatial tasks may have been too hard, and (3) the spatial task experiments may not have had enough power to detect an interaction similar to the one observed in Experiments 1B-1 – 1B-3. We presented arguments against each of these hypotheses. By comparing the arithmetic task used in Experiment 1B-1 and the spatial integration task used in Experiment 1B-4, we first established that the spatial integration task was not too easy (1) by showing that in the Norming Study – where each of these tasks was performed in isolation by an independent group of participants – the spatial integration task took longer to perform than the arithmetic task and the accuracies in the spatial integration task were plausibly lower; and (2) by showing that in Experiments 1B-1 and 1B-4, the ranges of raw reaction times were very similar across all regions and at the critical region. Second, we established that the spatial integration task was not too difficult by showing that in Experiment 1B-4, a significant main effect of syntactic complexity was observed, indicating that the spatial integration task was not swamping the syntactic complexity effect. Finally, we established that it was likely that Experiments 1B-4 and 1B-5 had sufficient power to detect an interaction of the size observed in Experiments 1B-1 and 1B-2: the power analysis revealed that the power in both Experiments 1B-4 and 1B-5 was $>.80$. We therefore argued that in the spatial task experiments, the lack of an interaction similar to the interaction observed in the arithmetic task experiments was plausibly due to the fact that whereas linguistic and arithmetic

integration processes rely on overlapping pools of verbal working memory resources, linguistic and spatial integration processes do not, at least not to the same degree.

In the reaction time analyses for the experiments described in this section, we focused on the critical region (Region 2 in all the experiments) where linguistic complexity was manipulated. It is worth noting, however, that the overall data patterns differ, to some extent, across the experiments. Whereas there is a main effect of secondary task complexity on Regions 2, 3 and 4 in all the experiments, the reaction times peak at different regions. Let us examine the patterns of RTs in Experiments 1B-1 – 1B-5. Reaction times peak at Region 2 in Experiments 1B-1, 1B-3 and 1B-4, and at Region 3 in Experiments 1B-2 and 1B-5. Importantly though, these differences in the peak point of reaction times do not correlate with the type of the secondary task: specifically, the secondary task in Experiments 1B-1 and 1B-3 is the *arithmetic* addition task, while in Experiment 1B-4 it is the *spatial* “pie” task; similarly, the secondary task in Experiment 1B-2 is the *arithmetic* addition-subtraction task, while in Experiment 1B-5 it is the *spatial* grid task. Because (1) we attempted to generalize over the three arithmetic tasks and the two spatial tasks, and distinguish between the arithmetic and spatial tasks, and (2) the critical interaction was observed during Region 2 in Experiments 1B-1, 1B-2 and 1B-3 and not in any of the other regions in any of the four experiments, the differences in the peak reaction times between Experiments 1B-1, 1B-3 and 1B-4 on one hand and Experiments 1B-2 and 1B-5 on the other hand do not seem relevant to the interpretation of the critical contrast between the presence of a super-additive interaction in Experiments 1B-1 – 1B-3 (with the arithmetic secondary tasks) and the absence of such an interaction in Experiments 1B-4 and 1B-5 (with the spatial secondary tasks). We hypothesize that the differences in the peak reaction times across experiments may be resulting from the differences in the difficulty of the secondary tasks across the experiments: the secondary tasks in Experiments 1B-2 (addition/subtraction) and 1B-5 (the grid spatial task) are plausibly more difficult than the secondary tasks in Experiments 1B-1, 1B-3 and 1B-4.

In our experimental logic we relied on the assumption that verbal and visuo-spatial working memory resource pools are independent, based on the earlier studies. The evidence for the independence of these two working memory resource pools comes from several kinds of studies: (1) dual-task experiments showing selective interference effects, such that a verbal memory task interferes to a larger extent with another verbal memory task, compared to a spatial memory task, and vice versa (e.g., Baddeley, 1986; Logie, 1986, 1995); (2) individual-differences studies showing that the correlations in people’s performance are higher within domains (verbal or visuo-spatial), than across domains (e.g., Shah & Miyake, 1996); (3) neuropsychological case studies of patients who are selectively impaired on verbal memory tasks or spatial memory tasks (Vallar & Shallice, 1990; Hanley et al., 1991); and (4) neuroimaging studies suggesting that different neural substrates underlie verbal memory tasks and spatial memory tasks (Jonides et al., 1993; Paulesu et al., 1993). All these different lines of evidence converge in their conclusions that there exist separate resource pools for verbal vs. visuo-spatial memory. It is worth noting, however, that in some of the previous studies the verbal and the visuo-spatial memory tasks were quite different in terms of the cognitive processes they involve, i.e. the tasks differed in more respects than the use of verbal vs. visuo-spatial resources. For example, a standard manipulation in dual-task experiments comparing the degree of interference produced by tasks from different domains has involved tapping the four corners of a square with a finger continuously for the spatial distractor task, and repeatedly pronouncing a word for the verbal distractor task. Even

though we did not intend to test the hypothesis that verbal and visuo-spatial working memory resource pools are independent (we assumed this to be the case, based on the earlier evidence), our results can be taken as additional strong evidence for the independence of these two resource pools. Specifically, in our experiments the arithmetic tasks (which rely on verbal working memory resources) and the spatial integration tasks (which rely on spatial working memory resources) were qualitatively very similar in terms of the cognitive processes they involved (combining simple representations into more complex representations over time), and yet they showed differential interference with respect to the language-processing task, which relies on verbal working memory resources.

As discussed in Section 1A, there exists behavioral and ERP evidence for two different types of working memory costs in online language processing: working memory resources for processing incomplete syntactic dependencies (Chomsky & Miller, 1963; Wanner & Maratsos, 1978; Gibson 1991, 1998, 2000; Kluender & Kutas, 1993; King & Kutas, 1995; Lewis, 1996; Chen et al., 2005), and working memory resources for integrating words to earlier words/positions in the sentence (Gibson, 1998, 2000; Kaan et al., 2000; Gordon et al., 2001; Warren & Gibson, 2002; Phillips et al., 2005; Grodner & Gibson, 2005). We argued that it might be necessary to take this evidence into consideration when investigating the extent of domain-specificity of working memory resources for online language processing. Specifically, we suggested that the two pools of working memory resources used in online language processing – the one involved in keeping track of incomplete syntactic dependencies and the one involved in integrating structural elements over long distances – may differ in the extent of their domain-specificity and in the extent of their overlap with other working memory systems. In this paper, we focused on investigating the nature of working memory resources in linguistic integrations by examining the relationship between linguistic integrations and similar integration processes which either involve or don't involve verbal working memory resources. We provided evidence for an overlap in resource pools used for linguistic and arithmetic integration processes. This suggests that future investigations aimed at understanding the nature of working memory resources in language processing may in fact benefit from examining the two different resource pools used in language processing independently.

The results reported here may be used to suggest that the verbal working memory resource pool is divided along the lines of the qualitative nature of the cognitive processes involved, rather than along the domains to which the tasks belong (see Mitchell, 2007, for a similar proposal). For example, Caplan and Waters (1999) argued that the verbal working memory resource pool is divided into resources used for online language processing and resources used for non-linguistic verbally-mediated tasks. However, it is possible that different pools of verbal working memory resources are used (1) for tasks which involve storing verbal representations in memory over time, and (2) for tasks which involve combining verbal representations into more complex representations. The behavioral and ERP evidence from the language processing literature discussed above is consistent with this line of reasoning, such that even within the resource pool for online language processing there appear to exist two independent sub-pools of resources – for keeping track of incomplete syntactic dependencies and for integrating structural elements with one another. Therefore, in conjunction with the results reported here, it is plausible that the resource pool that any given verbal task may rely on is determined by the nature of the processes involved in the task.

Background for Experiments 1B-7a and 1B-7b

In Experiments 1B-1 – 1B-3 we examined the relationship between linguistic integrations and a verbal working memory task involving similar integration processes (arithmetic additions). We have shown that linguistic and arithmetic integrations rely on overlapping pools of verbal working memory resources. In Experiments 1B-7a and 1B-7a we will examine the relationship between linguistic integrations and a non-verbal task involving similar integration processes. In particular, we will investigate the relationship between linguistic and musical integrations.

The relationship between language and music has been pondered for centuries. The two domains share a number of important characteristics (e.g., Lerdahl & Jackendoff, 1983). First, both language and music involve temporally unfolding sequences of sounds – where pitch and rhythm are important characteristics of the signal – produced by humans. Second, both language and music are rule-based systems where basic elements (words in language, tones and chords in music) are combined into an infinite number of high-order structures (sentences in language, melodies in music). Finally, both appear to be universal cognitive abilities and both have been argued to be unique to humans (see McDermott & Hauser, 2005, for a recent review of this literature). It is worth noting that there are a number of important differences, as well. First, while language is symbolic and referential, such that strings of sounds have meanings associated with them and there is a mapping between the strings of sounds and objects / events in the world, music is not referential in the same way. Second, while linguistic sequences convey meaning, musical sequences have been argued to convey, or evoke, emotions. Finally, while language serves an obvious communicative function, there is no obvious function for music. Despite these differences, however, the fact that language and music share some key properties has led to a number of proposals arguing for some shared cognitive and neural system(s) underlying both abilities. Some have even argued for a common adaptation. Many of the earlier proposals, however, left the exact nature of the hypothesized overlap underspecified and thus difficult to evaluate.

Several methods have been used to investigate the relationship between language and music, and different patterns of results have been reported across different methods. In neuropsychology, cases of double dissociations between language and music have been reported, such that there have been reports of patients who have a deficit in linguistic abilities without a deficit in any other cognitive ability (including musical ability) (e.g., Luria et al., 1965), and conversely, there have been reports of patients who have a deficit in musical abilities without a deficit in any other cognitive ability (e.g., Peretz 1993, Peretz & Coltheart, 2003). These patient studies have been interpreted as evidence for domain-specificity of language and music. In contrast to the neuropsychological investigations, different experimental methods (e.g., ERPs, MEG and functional MRI, among others) have revealed patterns of results inconsistent with the domain-specific view. I will here summarize a few key findings.

Patel et al. (1998) presented participants with sentences and chord sequences and varied the difficulty of integration at some point in the sentence/chord. It was demonstrated that difficult integrations in both language and music are associated with a very similar ERP component (the P600) with a similar scalp distribution. Patel et al. concluded that the P600 component reflected

processes of structural integration by language and music. There have been a number of subsequent studies across different research groups demonstrating similar responses to structural manipulations in language and music (e.g., Maess et al., 2001, Koelsch et al., 2002, 2005). There have also been several functional neuroimaging studies showing that structural manipulations in music appear to activate neural areas in and around Broca's area and its right hemisphere homologue (e.g., Levitin & Menon, 2003)⁷.

Patel (2003) attempted to reconcile the neuropsychological and experimental data by proposing that while linguistic and musical representations may be separate, language and music may rely on a shared system for structural integration. This non-domain-specific system was argued to be involved in integrating incoming elements (words/notes) into evolving structures (sentences/musical sequences). Patel hypothesized that the difficulty of structural integration in both language and music is influenced by the distance between the incoming element and the element it needs to connect to (in language; adopting Gibson's (1998) Dependency Locality theory, 1998), or between the incoming element and the tonic in the tonal pitch space (in music; adopting Lerdahl's (2001) Tonal Pitch Space theory): the greater the distance the more difficult the integration. Figure 13 schematically summarizes Patel's proposal.

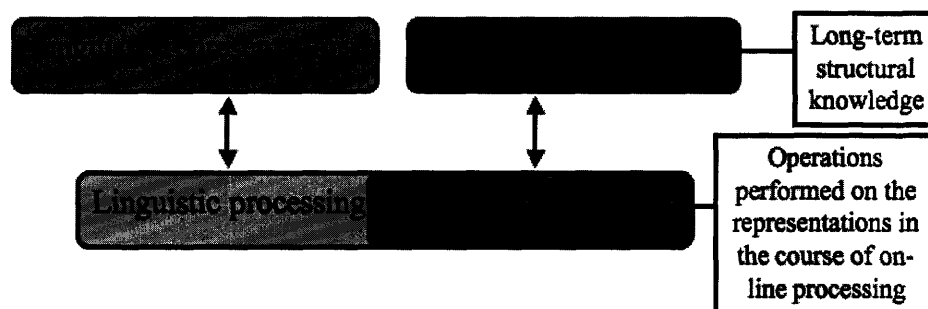


Figure 13: A schematic summary of Patel's (2003) hypothesis.

Patel's proposal predicts that taxing the shared processing system with concurrent difficult linguistic and musical integrations should result in super-additive processing difficulty. Experiments 1B-7a and 1B-7b are aimed at testing the predictions of Patel's proposal using a dual-task paradigm, similar to the one used in Experiments 1B-1 – 1B-6.

Experiments 1B-7a and 1B-7b: Linguistic vs. Musical integrations
[Fedorenko, Patel, Casasanto, Winawer, & Gibson, in preparation]

There are at least two different ways to combine linguistic and musical stimuli in a dual-task paradigm, and both have now been used to investigate a variety of questions with regard to the relationship between language and music. One paradigm involves presenting linguistic stimuli visually and accompanying each word / phrase with a note or a chord (e.g., Koelsch et al., 2005,

⁷ To the best of my knowledge, there have been no studies to date comparing structural processing in language and music within the same individuals. In order to claim that shared neural structures underlie linguistic and musical processing, within-individual comparisons are critical.

Slevc et al., 2007).

Another paradigm involves sung stimuli where the linguistic and the musical signals are combined in a single auditory stream. The past studies using this paradigm focused on the relationship between *semantic* processing and musical harmonic processing (e.g., Besson et al., 1998, Bonnel et al., 2001). Consequently, the results of these studies are somewhat orthogonal to the question of whether structural processing in language and music rely on a shared resource pool.

Two experiments using sung sentences were conducted. In Experiment 1B-7a, we manipulated the difficulty of linguistic integrations (in two different ways) and the difficulty of musical integrations. First, the difficulty of linguistic integrations was manipulated by using subject- and object-extracted relative clauses, similar to Experiments 1B-1 – 1B-6. Second, the difficulty of linguistic integrations was manipulated by using grammaticality violations. The difficulty of musical violations was manipulated by using melodies with and without distant (out-of-key) notes.

In Experiment 1B-7b, we manipulated the difficulty of linguistic integrations by using subject- and object-extracted relative clauses. And in addition to the original musical manipulation, we added an auditory oddball control condition: an increase in loudness at the critical position in the sentence. In order to argue that linguistic and musical integrations rely on the same / shared pool of resources for structural integration, it is important to rule out an explanation whereby the musical effect is driven by a lower-level perceptual unexpected event (an out-of-key note).

Experiment 1B-7a

In this experiment participants listened to sung sentences. As mentioned above, the difficulty of the linguistic integrations was manipulated in two different ways. The reason for including two different manipulations is that difficult integrations – in both language and music – can be conceptualized in different ways. First, the processing of an incoming element may be difficult because it is necessary to retrieve the element's structural dependent(s) from memory (in language), or because upon encountering a harmonically distant element it is necessary to activate the key structure associated with the incoming element because this key has not been active in the preceding context (in music). Second, the processing of an incoming element may be difficult because this incoming element does not satisfy the expectations generated by the preceding linguistic / musical context. Grammaticality violations are more conceptually similar to the latter way of thinking about difficult integrations. In summary, we wanted to investigate how difficult linguistic integrations of different kinds relate to difficult musical integrations. If only one kind of difficult linguistic integrations interacts with musical integrations, this would be informative with regard to the underlying nature of the difficulty in musical integrations.

Patel's (2003) shared integration resource hypothesis predicts that linguistic integrations should interact with musical integrations, such that when both integrations are difficult super-additive processing difficulty should ensue.

Methods

Participants Sixty-four participants from MIT and the surrounding community were paid for their participation. All were native speakers of English and were naive as to the purposes of the study.

Design and materials The experiment had a 2 x 2 x 2 design, crossing syntactic complexity (subject- / object-extracted RCs), grammaticality (no tense error in the RC / tense error in the RC), and musical complexity (easy / hard).

The language materials consisted of 40 sets of sentences, with four versions as in (10). Each sentence consisted of 12 mostly monosyllabic words, so that each word corresponded to one note in a melody. The ungrammatical conditions were created by using the infinitival form of the verb instead of the past tense form or the present tense form.

(10a) *Subject-extracted, grammatical:*

The boy / that helped the girl / got an “A” / on the test.

(10b) *Object-extracted, grammatical:*

The boy / that the girl helped / got an “A” / on the test.

(10c) *Subject-extracted, ungrammatical:*

The boy / that help the girl / got an “A” / on the test.

(10d) *Object-extracted, ungrammatical:*

The boy / that the girl help / got an “A” / on the test.

Each of these four versions was paired with two different versions of a melody (easy and hard).

In addition to the 40 experimental items, 50 filler sentences with a variety of syntactic structures were created. The sentences were 10-14 words in length, and like the experimental items, they consisted mostly of monosyllabic words, so that each word corresponded to one note in a melody. Each filler sentence was divided into 3-5 regions to resemble the experimental items. Half of the filler sentences contained a tense or a number violation.

The musical materials were created in two steps: (1) 40 target melodies (with two versions each) and 50 filler melodies were composed by a professional composer (Jason Rosenberg), and (2) each condition of every item and the filler items was recorded by one of our collaborators – a former opera singer – Daniel Casasanto.

Melody creation

Target melodies

All the melodies consisted of 12 notes, were strongly tonal and ended in a tonic note with an authentic cadence in the implied harmony. All the melodies were isochronous: all quarter notes except for the last note. The first five notes established a strong sense of key. Both versions of each melody were in the same key and differed by one note. The critical (6th) note – falling on

the last word of the relative clause – was either in-key or out-of-key. It always was on the downbeat of second full bar. When the note was out-of-key, it was one of the five possible non-diatonic notes (C#, D#, F#, G#, A# in C major). Sometimes out-of-key notes were only different by a semi-tone (e.g., C vs. C#).

The size of pitch jumps leading to and from the critical note were matched for the in-key and out-of-key notes, so that out-of-key notes were not odd in terms of voice leading compared to the in-key notes. Out-of-key notes were occasionally associated with tritone jumps, but for every occurrence of this kind there was another melody where the in-key note had a jump of a similar size.

All 12 major keys were used three times ($12 \times 3 = 36$ melodies), and 4 other keys were randomly selected for the remaining 4 melodies. The lowest pitch used was C#4 (277 Hz), and highest was F5 (698 Hz). The range was designed for a tenor.

Filler melodies

All the melodies consisted of 10-14 notes, were strongly tonal and resembled the target melodies in style. Half of the filler melodies contained an out-of-key note at some point in the melody. The out-of-key note occurred at least five notes into the melody. Half of the melodies containing out-of-key notes were paired with the filler sentences containing grammatical errors, and the other half with the filler sentences without grammatical errors. Furthermore, for the cases where the melodies containing out-of-key notes were paired with the filler sentences containing grammatical errors, half of the time the grammatical error coincided with the out-of-key note.

All 12 major keys were used four times ($12 \times 4 = 48$ melodies), and 2 other keys were randomly selected for the remaining 2 melodies. The pitch range used was the same as that used for creating the target melodies.

Recording the stimuli

The target and the filler stimuli were recorded in a soundproof room in the Psychology Department at Stanford University. For each experimental item, Regions 1-4 of Condition 1 (in-key melody, subject-extracted, grammatical) were recorded first, with each region recorded separately. Then, recordings of Region 2 of the remaining seven conditions were made. (Regions 1, 3 and 4 were only recorded once, since they were exactly the same across the eight conditions of the experiment.) For each filler item, every region was also recorded separately. After the recording process was completed, all the recordings were normalized for intensity (loudness) levels.

Procedure The task was self-paced phrase-by-phrase listening. The experiment was run using the Linger 2.9 software by Doug Rohde. Each experimental sentence had four regions (as shown in (10a)-(10d)): (1) a noun phrase, (2) an RC (subject-/object-extracted), (3) a main verb with a direct object, and (4) an adjunct prepositional phrase. The stimuli were presented to the participants via headphones. Each participant heard only one version of each sentence, following a Latin-Square design (see Appendix D for a complete list of linguistic materials).

The stimuli were pseudo-randomized separately for each participant, with at least one filler separating the target sentences.

Each trial began with a fixation cross. Participants pressed the spacebar to hear each region of the sentence. The amount of time the participant spent listening to each region was recorded as the time between key-presses.

To assure that the participants processed the sentences for meaning, a yes/no question was presented visually after the last region of the sentence. Participants pressed one of two keys to respond “yes” or “no”. After an incorrect answer, the word “INCORRECT” flashed briefly on the screen. Participants were told to answer the questions as quickly and accurately as they could, and to take wrong answers as an indication to be more careful. There was no musical task. Participants were instructed to listen to the sentences carefully.

Before the self-paced listening task each participant completed a short task that was aimed at identifying tone-deaf individuals, and a questionnaire about his/her musical background. The tone-deafness evaluation task – taken from The Montreal Battery of Evaluation of Amusia (Peretz et al., 2003) – consisted of 32 pairs of short melodies. Participants were instructed to mark on the answer sheet whether they thought the two melodies in each pair were the same or not. We set the threshold for exclusion from the study at below 60% accuracy.

The Music Background questionnaire included a series of questions about different aspects of the participants’ musical experience (e.g., formal training in music theory, instruments played, number of hours per day spent listening to the music, etc.) See Appendix E for the complete list of questions. The questionnaire was aimed at getting a more detailed profile of each participant’s musical experience.

Participants took approximately 45 minutes to complete the three parts of the experiment (the tone-deafness scale, the questionnaire and the critical self-paced listening task).

Results

Due to testing errors, several participants did not complete the Tone-Deafness Task and/or the Questionnaire. The data for the Tone-Deafness Task comes from 54 out of 64 participants. The questionnaire data comes from 51 out of 64 participants.

The Tone-Deafness Task performance On average, participants were 87.3% accurate. None of the participants were below the threshold of performance we set prior to running the study (60%). Therefore, all the participants were included in the analysis.

The Music Background questionnaire Below we present a summary of the data from the questionnaire.

- 86.3% of participants played a musical instrument at some point in their lives. 60.8% played two musical instruments. 23.5% played three musical instruments. 11.8% played four musical instruments. And 3.9% played more than four musical instruments. The

participants who played at least one musical instrument at some point in their lives played an average of 6.08 years, taking into account the number of instruments (calculated by dividing the sum of total years/instrument by the number of instruments).

- 31.4% of participants had some training in music theory. These participants had, on average, 4.5 years of training.
- 45% of participants sang in a choir at some point in their lives. These participants had, on average, 4.6 years of choir experience.
- Participants listened to an average of 3.26 hours of music daily.

In summary, participants had a wide range of musical experiences, but all the participants have been exposed to Western tonal music and therefore should be sensitive to the musical manipulation in the experiment.

Comprehension question performance Participants answered the comprehension question correctly 90.9% of the time. For ease of presentation⁸ and because the data patterns differed between the grammatical and ungrammatical conditions, we will first present the data for the grammatical conditions, followed by the presentation of the ungrammatical conditions.

One item (#29) contained a recording error and therefore was omitted from all the analyses. Furthermore, another item (#6) did not have the last region recorded due to an error, and therefore, in the listening time analyses, it was not present in the analyses of the last region.

Figure 14 presents the mean accuracies in the grammatical conditions of Experiment 1B-7a. A two-factor ANOVA crossing syntactic complexity (easy, hard) and musical complexity (easy, hard) on the responses to the comprehension questions revealed a marginal interaction in the items analysis, such that the condition both syntactic and musical integrations were difficult was less accurate than the other three conditions ($F(1,63)=2.55$, $MSe=452$, $p=.115$; $F(1,38)=3.54$, $MSe=289$, $p=.068$). This interaction is as predicted by Patel's shared structural integration resource hypothesis.

⁸ This way of presenting the data will also make it easier to compare the results in this Experiment with the results of Experiment 1B-7b.

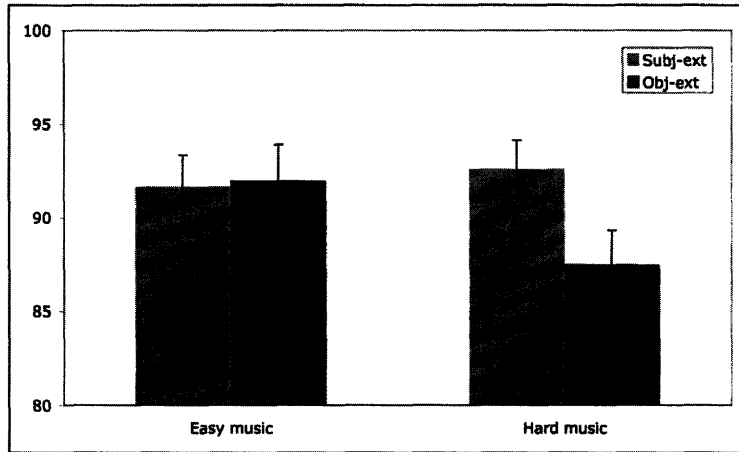


Figure 14: Comprehension accuracies in percent correct in the grammatical conditions of Experiment 1B-7a.

Figure 15 presents the mean accuracies in the ungrammatical conditions of Experiment 1B-7a. A two-factor ANOVA syntactic complexity (easy, hard) and musical complexity (easy, hard) on the responses to the comprehension questions revealed an effect of syntactic complexity, such that the object-extracted conditions were less accurate than the subject-extracted conditions ($F(1,63)=6.75$, $MSe=1833$, $p<.02$; $F(1,38)=6.85$, $MSe=1158$, $p<.02$), and a marginal effect of musical complexity in the items analysis, such that the hard musical conditions were more accurate than the easy musical conditions ($F(1,63)=2.75$, $MSe=520$, n.s.; $F(1,38)=1.99$, $MSe=325$, $p=.092$). There was no interaction ($F_s < 1$).

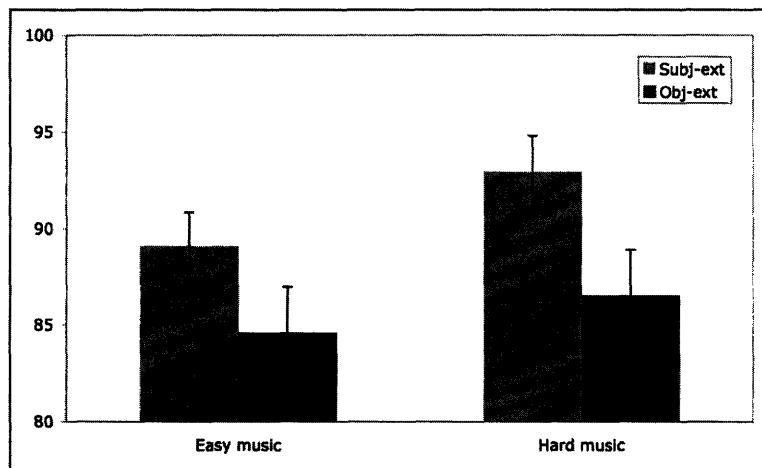


Figure 15: Comprehension accuracies in percent correct in the ungrammatical conditions of Experiment 1B-7a.

Reaction times As in the other experiments, we analyzed all trials, regardless of how the comprehension question was answered. As with the Comprehension data, for ease of presentation and because the data patterns differed between the grammatical and ungrammatical conditions, we will first present the data for the grammatical conditions, followed by the presentation of the ungrammatical conditions.

Figure 16 presents the mean listening times for the critical region (Region 2) in the four grammatical conditions of Experiment 1B-7a.

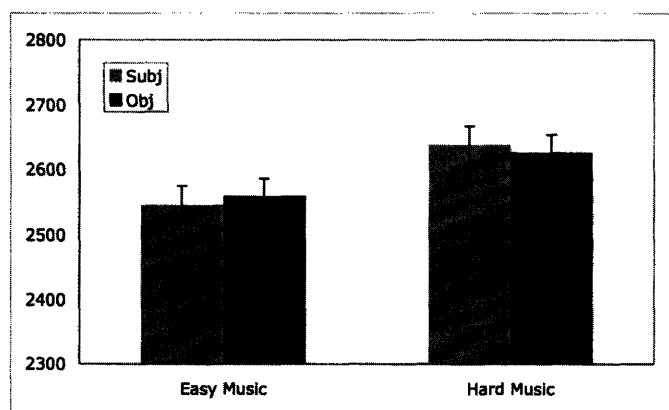


Figure 16: Listening times at the critical region in the four grammatical conditions of Experiment 1B-7a. Error bars indicate standard errors.

We conducted a 2 x 2 ANOVA crossing syntactic complexity and musical complexity for each of the four regions. The results are presented in Table 20. We first present the analysis of the critical region, Region 2, which included the relative clause (“that helped the boy” / “that the boy helped”), followed by the analyses of the other three regions. At the critical region, the listening times for the hard-musical-integration conditions were slower than the listening times for the easy-musical-integration conditions (2637.0 ms vs. 2558.6 ms). There were no other effects ($F_s < 1$).

In Region 1, there were no significant effects ($F_s < 1.7$). In Region 3, there was a marginal effect of syntactic complexity. Finally, in Region 4, there was an unpredicted marginal interaction such that in the easy-musical-integration conditions the subject-extracted condition was slower than the object-extracted condition, and in the hard-musical-integration condition, the subject-extracted condition was faster than the object-extracted condition. There were no other effects ($F_s < 1$).

Table 20. Analysis of Variance results for the grammatical conditions of Experiment 1B-7a.

Source of variance	By participants			By items	
	<i>df</i>	F1 value	MSe	<i>df</i>	F2 value
Region 1					
Syntactic Complexity	1,63	<1	5634	1,38	<1
Musical Complexity	1,63	<1	15254	1,38	<1
Interaction	1,63	1.68	24977	1,38	1.06
Region 2 (critical region containing the relative clause)					
Syntactic Complexity	1,63	<1	9074	1,38	<1
Musical Complexity	1,63	19.03*	392689	1,38	8.39
Interaction	1,63	<1	9770	1,38	<1
Region 3					
Syntactic Complexity	1,63	3.92	95740	1,38	3.13
Musical Complexity	1,63	<1	4083	1,38	<1
Interaction	1,63	<1	9973	1,38	<1
Region 4					
Syntactic Complexity	1,63	<1	7219	1,38	<1
Musical Complexity	1,63	<1	12259	1,38	<1
Interaction	1,63	3.40	124528	1,38	3.81

Note: Significant effects are marked by asterisk.

We will now present the data for the ungrammatical conditions. Figure 17 presents the mean listening times for the critical region (Region 2) in the four ungrammatical conditions.

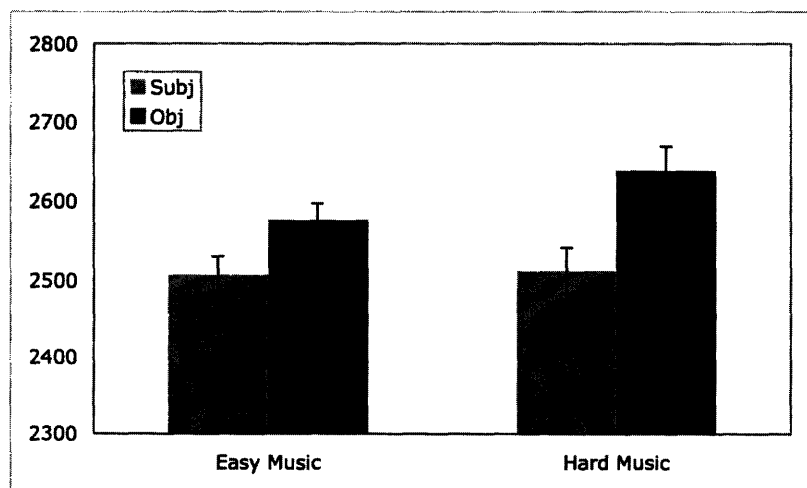


Figure 17: Listening times at the critical region in the four ungrammatical conditions of Experiment 1B-7a. Error bars indicate standard errors.

We conducted a 2 x 2 ANOVA crossing syntactic complexity and musical complexity for each of the four regions. The results are presented in Table 21. We first present the analysis of the critical region, Region 2, which included the relative clause (“that helped the boy” / “that the boy helped”), followed by the analyses of the other three regions. At the critical region, the listening times for the object-extracted conditions were slower than the listening times for the subject-extracted conditions (2624.9 ms vs. 2517.0 ms). Furthermore, the listening times for the hard-musical-integration conditions were marginally slower than the listening times for the easy-musical-integration conditions in the participants analysis (2589.8 ms vs. 2552.1 ms). Finally, there was a trend for an interaction such that in the hard-musical-integration conditions the difference between the subject- and the object-extracted conditions was larger than in the easy-musical-integration conditions. This interaction is as predicted by Patel’s shared structural integration resource hypothesis.

In Region 1, there was an unpredicted interaction in the participants analysis – which was marginal in the items analysis – such that in the easy-musical-integration conditions the object-extracted condition was slower than the subject-extracted condition, and in the hard-musical-integration condition, the object-extracted condition was faster than the subject-extracted condition. There is no reason to expect this effect here, as the materials were exactly the same across the four conditions. In Region 3, there was an unpredicted interaction, such that in the easy-musical-integration the subject-extracted condition was numerically slower than the object-extracted condition, in the hard-musical-integration conditions the object-extracted condition was slower than the subject-extracted condition. Finally, in Region 4, there was a similar interaction. There is no reason to expect an interaction of this sort either in Region 3 or 4, as the materials were exactly the same across the four conditions.

Table 21. Analysis of Variance results for the ungrammatical conditions of Experiment 1B-7a.

Source of variance	By participants			By items	
	<i>df</i>	F1 value	MSe	<i>df</i>	F2 value
Region 1					
Syntactic Complexity	1,63	<1	2161	1,38	<1
Musical Complexity	1,63	<1	1420	1,38	<1
Interaction	1,63	4.33*	59822	1,38	3.57
Region 2 (critical region containing the relative clause)					
Syntactic Complexity	1,63	21.8*	744152	1,38	18.02*
Musical Complexity	1,63	3.57	90638	1,38	2.22
Interaction	1,63	1.83	61974	1,38	1.17

Region 3					
Syntactic Complexity	1,63	1.103	25734	1,38	1.22
Musical Complexity	1,63	<1	5850	1,38	<1
Interaction	1,63	5.80*	133968	1,38	5.47*
Region 4					
Syntactic Complexity	1,63	1.63	52282	1,38	1.67
Musical Complexity	1,63	<1	702	1,38	<1
Interaction	1,63	5.603*	258623	1,38	5.19*

Note: Significant effects are marked by asterisk.

Discussion

The results of Experiment 1B-7a revealed different patterns of results across the grammatical and ungrammatical conditions. In particular, in the grammatical conditions, a trend for an interaction predicted by Patel's (2003) shared structural integration resource hypothesis was observed in the comprehension accuracy data. In the ungrammatical conditions, a trend for an interaction predicted by Patel's hypothesis was observed in the online listening time data.

While the results of Experiment 1B-7a are perhaps suggestive of some overlap between the pools of resources used for linguistic and musical integrations, they are certainly not conclusive. In Experiment 1B-7b we decided to focus on the grammatical conditions.

The motivation for conducting Experiment 1B-7b was two-fold. First, we wanted to increase the on-line processing demands in order to attempt to elicit a clearer pattern of results in the listening time data in the grammatical conditions. To increase the processing demands, we increased the playback speed of the audio-files. We reasoned that increasing the on-line processing demands may also result in a clearer pattern of results in the comprehension data. And second, we wanted to add an auditory oddball control condition. In particular, if linguistic and musical processing were shown to interact, in order to argue that linguistic and musical integrations rely on the same / shared pool of resources it would be important to rule out an explanation whereby the musical effect is driven by a lower-level perceptual unexpected event (an out-of-key note). To achieve this, we added a condition where the melodies without an out-of-key note had a perceptually salient increase in intensity (loudness) at the critical position.

Experiment 1B-7b

Patel's (2003) shared structural integration resource hypothesis predicts an interaction between linguistic and musical integrations for the structural manipulation in music, but not for the lower-level perceptual manipulation.

Methods

Participants Sixty participants from MIT and the surrounding community were paid for their participation. All were native speakers of English and were naive as to the purposes of the study. None have participated in Experiment 1B-7a.

Design and materials The experiment had a 2 x 3 design, crossing syntactic complexity (subject- / object-extracted RCs) and musical complexity (easy / hard / auditory oddball).

The materials for this experiment were created using a subset of 36 stimuli from the grammatical conditions of Experiment 1B-7a. The process of creating the materials involved two steps: (1) every audio file was sped up by 50%; (2) for the auditory oddball conditions, the intensity (loudness) level of the last word in the RC was increased by 10 dB.

In addition to the target items, 25 filler sentences (the grammatical filler sentences from Experiment 1B-7a) were used. Each filler file was also sped up by 50% and the intensity manipulation was added to 8 (roughly one third) of the fillers to reflect the distribution of the intensity increase occurrences in the target materials.

Procedure The procedure was exactly the same as that in Experiment 1B-7a, including the tone-deafness task and the Music Background questionnaire). Participants took approximately 35 minutes to complete the experiment.

Results

The Tone-Deafness Task performance On average, participants were 87% accurate. None of the participants were below the threshold of performance we set prior to running the study (60%). Therefore, all the participants were included in the analysis.

The Music Background questionnaire Below we present a summary of the data from the questionnaire.

- 81.4% of participants played a musical instrument at some point in their lives. 42.4% played two musical instruments. 15.3% played three musical instruments. 5.08% played four musical instruments. The participants who played at least one musical instrument at some point in their lives played an average of 6.01 years, taking into account the number of instruments (calculated by dividing the sum of total years/instrument by the number of instruments).
- 44% of participants had some training in music theory. These participants had, on average, 4.12 years of training.
- 46% of participants sang in a choir at some point in their lives. These participants had, on average, 5.7 years of choir experience.
- Participants listened to an average of 2.78 hours of music daily.

In summary, as in Experiment 1B-7a, participants had a wide range of musical experiences, but all the participants have been exposed to Western tonal music and therefore should be sensitive to the musical manipulation in the experiment.

Comprehension question performance Participants answered the comprehension question correctly 85.09% of the time. Figure 18 presents the mean accuracies across the six conditions of Experiment 1B-7b.

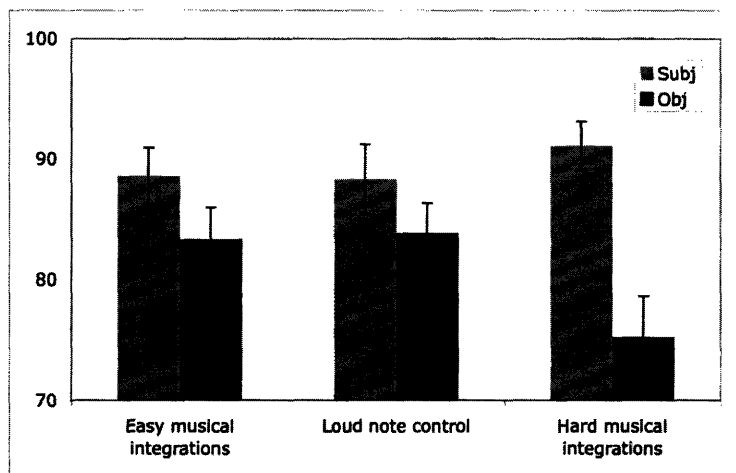


Figure 18: Comprehension accuracies in the six conditions of Experiment 1B-7b.

A two-factor ANOVA crossing syntactic complexity (easy, hard) and musical complexity (easy, hard, auditory oddball) and on the responses to the comprehension questions revealed a main effect of syntactic complexity and an interaction. First, participants were less accurate in the the object-extracted conditions, compared to the subject-extracted conditions ($F(1,59)=11.03$; $MSe=6531$; $p<.005$; $F(1,35)=14.4$; $MSe=3919$; $p<.002$). Second, the difference between the object- and the subject-extracted conditions was larger in the hard-musical-integration conditions, compared to the easy-musical-integration conditions or the auditory oddball conditions ($F(2,118)=6.62$; $MSe=1209$; $p<.005$; $F(2,70)=7.83$; $MSe=725$; $p<.002$). This interaction is as predicted by Patel's (2003) shared structural integration resource hypothesis.

Reaction times As in the other experiments, we analyzed all trials, regardless of how the comprehension question was answered. Figure 19 presents the mean listening times for the critical region (Region 2).

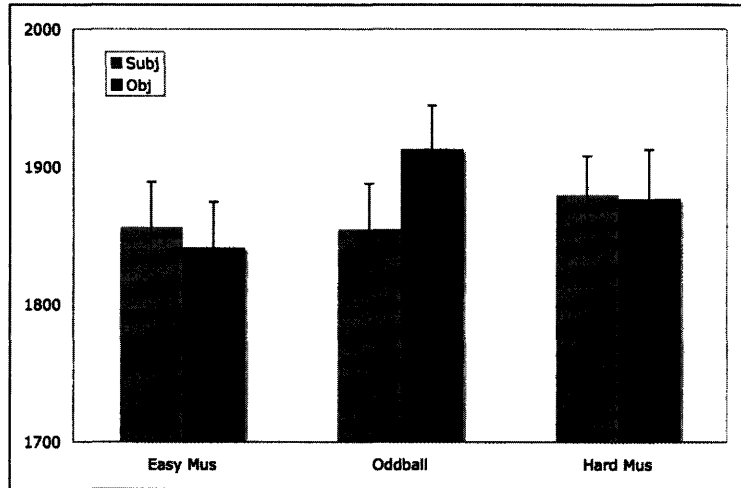


Figure 19: Listening times at the critical region in Experiment 1B-7b. Error bars indicate standard errors.

We conducted a 2 x 2 ANOVA crossing syntactic complexity and musical complexity for each of the four regions. The results are presented in Table 22. We first present the analysis of the critical region, Region 2, which included the relative clause (“that helped the boy” / “that the boy helped”), followed by the analyses of the other three regions. At the critical region, there was a marginal main effect of musical complexity in the participants analysis, such that the hard-musical-integration conditions and the oddball-conditions were slower than the easy-musical-integration conditions (1878.5 ms and 1884.2 ms vs. 1849.3 ms). There was also an unpredicted interaction that didn’t reach significance in the items analysis, such that the difference between the subject- and the object-extracted conditions was larger in the oddball conditions, compared to the easy- and the hard-musical integration conditions.

In Region 1, there were no significant effects ($F_s < 1.8$). In Region 3, there was a main effect of syntactic complexity, such that the object-extracted conditions were processed slower than the subject-extracted conditions. There were no other significant effects ($F_s < 2.08$). Finally, in Region 4, there was an unpredicted marginal interaction in the participants analysis. There is no reason to expect any differences among conditions at this region because the materials were exactly the same.

Table 22. Analysis of Variance results for Experiment 1B-7b.

Source of variance	By participants			By items	
	<i>df</i>	F1 value	MSe	<i>df</i>	F2 value
Region 1					
Syntactic Complexity	1,59	<1	4180	1,35	<1
Musical Complexity	2,118	1.50	15162	2,70	1.70
Interaction	2,118	<1	5079	2,70	<1

Region 2 (critical region containing the relative clause)					
Syntactic Complexity	1,59	1.18	16416	1,35	<1
Musical Complexity	2,118	3.06	42101	2,70	2.34
Interaction	2,118	3.53*	45998	2,70	3.07
Region 3					
Syntactic Complexity	1,59	14.3*	155500	1,35	11.4*
Musical Complexity	2,118	<1	2357	2,70	<1
Interaction	2,118	2.08	31200	2,70	1.64
Region 4					
Syntactic Complexity	1,59	<1	21350	1,35	<1
Musical Complexity	1,118	1.18	34490	2,70	1.18
Interaction	1,118	2.93	87327	2,70	2.17

Note: Significant effects are marked by asterisk.

Discussion

The motivation for Experiment 1B-7b was two-fold: (1) to elicit clearer patterns of results in listening time and comprehension accuracy data, and (2) to investigate how auditory oddball conditions compare to hard-musical-integration conditions.

With regard to the first goal, the pattern of results in the listening time data is still quite difficult to interpret. One possible reason for this is that the materials are presented in a region-by-region manner, which may not provide a sensitive enough measure for investigating the relationship between linguistic and musical integrations online. [We are currently conducting an ERP study in collaboration with Ina Bornkessel-Schlesewsky and Matthias Schlewsky using the design of Experiment 1B-7b.]

However, the pattern of results in the comprehension data is very interpretable in this experiment. In particular, we observed that participants were less accurate in the object-extracted conditions compared to subject-extracted conditions, and critically, that the difference between a subject- and an object-extracted condition was larger in the hard-musical-integration conditions, compared to the easy-musical-integration conditions and the auditory oddball conditions. This pattern of results is the same as that observed in the grammatical conditions of Experiment 1B-7a.

Conclusions for Experiments 1B-7a and 1B-7b

In both experiments we observed an interaction between linguistic and musical integrations in the comprehension accuracy data in the grammatical conditions: the difference between the subject- and the object-extracted conditions was larger in the conditions where musical integrations were hard, compared to the conditions where musical integrations were easy. The auditory oddball condition in Experiment 1B-7b further showed that this interaction is not due to a lower-level perceptual effect in the musical conditions: in particular, the accuracies in this control condition exhibited the same pattern as the easy-musical-integration conditions.

We think that this pattern of results is consistent with two interpretations. First, it is possible to interpret these data in terms of an overlap between linguistic and musical integrations in *on-line* processing. In particular, it is possible that (1) building more complex structural linguistic representations requires more resources, and (2) a complex structural integration in music interference with this process due to some overlap in the underlying resource pools. As discussed above, a possible reason for why we do not observe the interaction in the listening time data has to do with the region-by-region presentation, which is a rather crude measure of on-line processing.

Second, it is possible to interpret these data in terms of an overlap at the *retrieval* stage. In particular, it is possible that (1) there is no competition for resources in the on-line process of constructing structural representations in language and music, and (2) at the stage of retrieving the linguistic representation from memory, the presence of a complex structural integration in the accompanying musical stimulus makes the process of reconstructing the syntactic dependency structure more difficult.

In summary, while the results of Experiments 1B-7a and 1B-7b demonstrate that there are some aspects of structural integration in language and music that appear to be shared, the current data cannot determine the exact nature of these shared aspects.

General conclusions

The results of Experiments 1B-1 – 1B-3 and 1B-6 demonstrated that at least some aspects of linguistic integrations rely on a pool of working memory resources that is also used for arithmetic integrations. Experiments 1B-4 and 1B-5 ruled out an attention-based explanation of the interactions observed between linguistic and arithmetic integrations. Furthermore, the results of Experiments 1B-7a and 1B-7b suggested that there are some aspects of structural integration in language and music that also appear to be shared.

In summary, the data from the experiments described in section 1B have demonstrated that the pool of working memory resources underlying structural integrations in language is not entirely domain-specific. Another implication of the experiments in this section is that investigating linguistic integrations by examining how they relate to similar integration processes is a promising approach to investigating the extent of domain-specificity of the working memory system underlying language processing.

1C. Investigating verbal working memory tasks with *similar materials*

Background

Experiments in Section 1B investigated the extent of domain-specificity of the WM system underlying linguistic integrations by comparing linguistic integrations and tasks involving similar integration processes. This section will describe an experiment investigating the extent of domain-specificity of the WM system underlying linguistic integrations by comparing linguistic integrations and tasks involving similar materials.

Recently, Gordon, Hendrick & Levine (2002) argued that the load manipulation used in the previous dual-task experiments (e.g., increasing the number of memory-items in the digit-span task) was not the right one for the purposes of assessing the nature of verbal WM resources in sentence comprehension. They suggested that the critical characteristic of the memory load is its representational nature in relation to the representational nature of the linguistic materials. Gordon et al. tested the overlapping resource pools hypothesis of verbal WM for sentence comprehension using a novel dual-task paradigm, where participants read sentences of high and low syntactic complexity (subject- and object-extracted cleft sentences), which contained either occupations (e.g., “It was the dancer that the fireman liked before the argument began”), or personal names (e.g., “It was Tony that Joey liked before the argument began”). At the same time, participants were asked to remember a set of three words, which could also be either occupations (e.g., poet, cartoonist, voter), or personal names (e.g., Joel, Greg, Andy). This design resulted in two match conditions (memory-nouns and sentence-nouns from the same category) and two non-match conditions (memory-nouns and sentence-nouns from different categories). Gordon et al. hypothesized that the similarity between the memory-nouns and the sentence-nouns might affect the more complex sentences (object-extracted clefts) to a larger extent. The most interesting result of the experiment was a reliable off-line interaction between syntactic complexity and noun type match in response accuracy data, such that there was a larger difference between subject- and object-extracted clefts for the match conditions than for the non-match conditions. Gordon et al. argued that in cases where the memory traces of the memory-nouns are similar to the memory trace of the relevant antecedent, interference takes place, such that it is harder to identify the relevant antecedent among all the available memory traces. Therefore, the authors interpreted these results as evidence against an independent verbal WM resource pool for sentence comprehension.

Gordon et al.’s (2002) results suggest that the concurrent memory load tasks (e.g., digit span) used in most previous dual-task experiments – until now conceived of as storage tasks (in the memory literature sense of the word) – may, in fact, be capable of tapping the integration processes involved in sentence comprehension. The experiment reported by Gordon et al. is similar to the traditional dual-task experiments in that the secondary task involves remembering a list of memory items across the sentence-processing task, but in contrast to the earlier experiments, Gordon et al. made the memory items similar to the nouns in the sentences, creating a possibility of source-memory confusion during the critical region of the sentences. The integration of a new word can either be local – to the preceding word in the sentence – or non-local, requiring a retrieval of a word further back in the input stream. In the case of a local integration, the incoming word can be immediately connected to the preceding word, because there are no intervening potential attachment sites. Thus no search for an attachment site is

needed, and the existence of a set of stored memory items will probably not cause any interference in forming the connection. However, when the integration is non-local, the target attachment site needs to be retrieved from memory, and there may be interference from the intervening potential attachment sites. Specifically, in the case of an object-extracted relative clause, an embedded subject NP intervenes between the subject NP of the main clause and the embedded verb, and therefore may cause interference. Furthermore, when there is a list of stored items in working memory, the similarity of these items to the target attachment site (in this case, the subject NP of the main clause) may further increase the difficulty of an integration. Thus, in Gordon et al.'s (2002) experiment, there should be difficulty with both non-local integrations, but more in the condition where the memory items are similar to the item that needs to be retrieved from memory (the attachment site). The interaction between syntactic complexity and similarity between memory-nouns and sentence-nouns can therefore be accounted for in terms of integration difficulty: (a) there is little difficulty for *local* integrations, with either similar or dissimilar memory items; and (b) for *non-local* integrations, the condition with similar memory items is much more difficult to process than the condition with dissimilar memory items.

Applying the same assumptions to the previous dual-task experiments, which used digit-span as a secondary verbal WM task, it is possible to explain the lack of interactions between syntactic complexity and memory load. Because digits are qualitatively very different from the nouns in the sentence materials, they do not interfere with the integration processes involved in language comprehension. Thus, there is no reason to expect that the memory load should affect the more structurally complex sentences to a larger extent.

Gordon et al.'s (2002) results are the first report of an off-line interaction between syntactic complexity and memory load in a paradigm where the two tasks did not interrupt each other. Thus, these results are not likely to be attributable to shared attention costs associated with task-switching (as has been proposed by Caplan & Waters, 1999). However, Gordon et al.'s results do not speak directly to the nature of verbal WM resources in on-line (interpretive) linguistic processing, because the only significant interaction that Gordon et al. observed was an off-line effect in response accuracies to comprehension questions. Although the on-line reading time data showed a trend towards a similar interaction, it was short of significance ($p = 0.13$ in the subjects analysis; $p = 0.19$ in the items analysis). As discussed earlier, Caplan & Waters (1999) have argued that off-line tasks, like answering comprehension questions, tap into post-interpretive processing rather than interpretive processing. Thus the question of the nature of WM resources for on-line sentence comprehension is not yet resolved.

Experiment 1C-1

[Fedorenko, Gibson, Rohde (2006)]

This experiment was similar in design to Gordon et al.'s (2002) study. Participants read sentences phrase-by-phrase, and at the same time were required to remember one or three nouns, which were either similar to or dissimilar from the nouns used in the sentences. The design was different from that of Gordon et al.'s in several respects. First, we chose to use structures with subject- and object-extracted relative clauses, as opposed to clefts. Second, we only varied the noun type of the memory-nouns, keeping the nouns in the sentences the same. Third, we included a load manipulation in terms of the number of memory-nouns (one memory-noun vs.

three memory-nouns). As discussed above, Gordon et al. (2002) proposed that working memory capacity in language processing should be conceptualized in terms of the amount of interference produced by the items that must be kept active in memory. Gordon et al. manipulated the amount of interference produced by the memory items by varying the degree of similarity between the memory-nouns and the sentence-nouns. It is plausible that the amount of interference is a function of not only the representational characteristics of memory items but also the number of memory items. Syntactic complexity may therefore interact with memory load (manipulated in terms of the number of items) in the context of similar elements. If this is the case, then we would expect a three-way interaction among the three factors, such that an interaction between syntactic complexity and the number of memory-nouns should be observed in the conditions where the memory-nouns and the sentence-nouns are similar, but not in the conditions where they are dissimilar.

Furthermore, there was a difference in the procedure, such that unlike Gordon et al. who used center-screen presentation, we used a moving-window presentation, which is arguably more natural, with a stronger resemblance to normal reading (Just, Carpenter and Wooley, 1982). We reasoned that it was possible that part of the reason that Gordon et al. did not get an online interaction might be due to the procedure they used, which might not be sensitive enough.

Methods

Participants Forty-four participants from MIT and the surrounding community were paid for their participation. All were native speakers of English and were naive as to the purposes of the study.

Design and materials The experiment had a 2 x 2 x 2 design, crossing syntactic complexity (subject-extracted relative clause, object-extracted relative clause), memory load (one noun, three nouns), and memory-noun type (occupation, name). The nouns in the sentences were always occupations, and thus the memory-noun(s) could either match or not match the sentence-nouns in type.

The language materials consisted of 32 sets of sentences, having four different versions as in (11):

(11a) *Subject-extracted, version 1:*

The physician | who consulted the cardiologist | checked the files | in the office.

(11b) *Subject-extracted, version 2:*

The cardiologist | who consulted the physician | checked the files | in the office.

(11c) *Object-extracted, version 1:*

The physician | who the cardiologist consulted | checked the files | in the office.

(11d) *Object-extracted, version 2:*

The cardiologist | who the physician consulted | checked the files | in the office.

As described above, there were two levels of syntactic complexity – subject- and object-extractions – but there were four versions of each sentence in order to control for potential

plausibility differences between the subject- and object-extracted versions of each sentence. As a result, no independent plausibility control is needed in this design. Each participant saw only one version of each sentence, following a Latin-Square design (see Appendix F for a complete list of linguistic materials). The nouns for the memory task – the occupations and the names – were matched for frequency (using a Usenet corpus of 1.2 billion words) and length in syllables (the means are presented in Appendix G) and paired with the sentences, such that the memory-nouns were not related semantically (for occupations) or phonologically to the nouns in the sentence.

In addition to the target sentences, 40 filler sentences with various syntactic structures other than relative clauses were included. The filler sentences were preceded by one, two or three memory-noun(s), which were a combination of occupations and names. The length and syntactic complexity of the filler sentences was similar to that of the target sentences. The stimuli were pseudo-randomized separately for each participant, with at least one filler separating the target sentences.

Procedure The task was self-paced phrase-by-phrase reading with a moving-window display (Just, Carpenter & Woolley, 1982). The experiment was run using the Linger 2.85 software by Doug Rohde. Each experimental sentence had four regions (as shown in (11a)-(11d)): (1) a noun phrase, (2) a relative clause (subject- / object-extracted), (3) a main verb with a direct object (always an inanimate noun phrase), and (4) an adjunct prepositional phrase. The memory-noun(s) was / were presented in capital letters in the center of the screen. Each trial began with the memory-noun(s) appearing on the screen for 600 msec (one noun) or 1800 msec (three nouns). Participants were instructed to try to remember the noun(s) as well as they could. This was followed by a blank screen for 500 msec, which in turn was followed by a series of dashes marking the length and position of the words in the sentence. Participants pressed the spacebar to reveal each region of the sentence. As each new region appeared, the preceding region disappeared. The amount of time the participant spent reading each region was recorded as the time between key-presses. To make sure the participants performed the memory task, a box appeared on the screen after the last region of the sentence was read, and the participants were instructed to type in as many of the nouns that were presented at the beginning of the trial as possible in any order. If the noun(s) were typed in correctly, the word “RIGHT” flashed briefly on the screen. If two of the three nouns were typed in correctly (in the hard-load conditions), the words “ALMOST RIGHT” flashed briefly. Finally, if the noun was typed in incorrectly (in the easy-load conditions) or if less than two nouns were typed in correctly (in the hard-load conditions), the word “WRONG” flashed briefly on the screen.

To make sure the participants read the sentences for meaning, two true-or-false statements were presented sequentially after the memory task, asking about the propositional content of the sentence they just read. Participants pressed one of two keys to respond “true” or “false” to the statements. After a correct answer, the word “CORRECT” flashed briefly on the screen, and after an incorrect answer, the word “INCORRECT” flashed briefly.

Participants were instructed to read sentences silently at a natural pace and to be sure that they understood what they read. They were also told to take wrong answers as an indication to read more carefully.

Before the experiment started, a short list of practice items and questions was presented in order to familiarize the participants with the task. Participants took approximately 35 minutes to complete the experiment.

Results

Memory task The performance on the memory task was calculated using the following formula: hits / (hits + misses + false-alarms). This formula allowed us to give partial credit for partial responses in the hard-load conditions and to penalize participants for guessing. Minor spelling mistakes (deviations from the targets by up to 2 letters) were not taken into consideration and the words with such mistakes were counted as hits. Across conditions, participants performed at 86.7% correct. Table 23 presents the mean performance across the eight conditions of the experiment. A three-factor ANOVA crossing syntactic complexity (subject-extracted relative clause, object-extracted relative clause), memory load (one memory-noun, three memory-nouns), and memory-noun type (match, non-match) was performed. The analysis revealed a main effect of memory load ($F(1,43)=28.1$; $MSe=.618$; $p<.001$; $F(1,31)=28.6$; $MSe=.453$; $p<.001$; $\min F'(1,72)=14.1$; $p<.001$), such that people had higher accuracy rates in the easy-load conditions (90.9%), compared to the hard-load conditions (82.5%). There was also a main effect of memory-noun type ($F(1,43)=16.8$; $MSe=.381$; $p<.001$; $F(1,31)=17.8$; $MSe=.274$; $p<.001$; $\min F'(1,73)=8.64$; $p<.005$), such that names were recalled with higher accuracy (90.01%) than occupations (83.3%). No other significant effects were observed ($F_s<1.5$).

<u>Memory Load</u>	<u>Syntactic complexity</u>			
	Subject-extraction (Easy)		Object-extraction (Hard)	
<u>Memory-Noun</u>	Match	Non-match	Match	Non-match
One noun (Easy)	86.7 (3.0)	93.8 (1.7)	88.1 (3.0)	95.2 (1.7)
Three nouns (Hard)	80.8 (2.9)	85.8 (2.5)	78.2 (2.9)	85.3 (2.4)

Table 23. Memory task performance in percent correct, as a function of syntactic complexity, memory load, and memory-noun type in Experiment 1C-1 (standard errors in parentheses).

Comprehension question performance There were two comprehension questions following each experimental trial. Participants answered the first question correctly 79.4% of the time, and the second question 79.9% of the time. The percentages of correct answers by condition were very similar for the two questions, so we collapsed the results in our analyses. Table 24 presents the mean accuracies across the eight conditions of the experiment. A three-factor ANOVA crossing syntactic complexity (subject-extracted relative clause, object-extracted relative clause),

memory load (one memory-noun, three memory-nouns), and memory-noun type (match, non-match) on the responses to the comprehension questions revealed a main effect of memory load ($F(1,43)=42.02$; $MSe=1.12$; $p<.001$; $F(1,31)=63.4$; $MSe=.794$; $p<.001$; $\min F'(1,74)=25.2$; $p<.001$), such that people answered comprehension questions more accurately in the easy-load conditions (85.3%), compared to the hard-load conditions (74.04%), but no other significant effects. All Fs were less than 1.5, except for a marginal three-way interaction ($F(1,43)=3.38$; $MSe=.113$; $p=.073$; $F(1,31)=2.87$; $MSe=.0799$; $p=.10$; $\min F'(1,70)=1.55$; $p=.22$). The trend for this three-way interaction is likely due to the fact that numerically, the effect of memory load affected the object-extracted relative clause conditions more, and this trend appeared more pronounced in the match conditions. It is worth noting that we did not replicate Gordon et al.'s (2002) off-line interaction between syntactic complexity and memory-noun type. This, however, could be a result of the differences in the procedures used in our vs. Gordon et al.'s experiment. Specifically, in Gordon et al.'s study the comprehension questions were asked immediately after the sentences were read, whereas in our design, the reading of the sentences was followed by the memory recall task, which in turn was followed by the comprehension questions. The longer time lapse between the reading of the sentences and the comprehension questions in our design is likely to be responsible for overall lower comprehension accuracies (compared to Gordon et al.'s results), which could have potentially obscured some effects. Furthermore, by asking the participants to recall the memory items before answering the comprehension questions, we could have reduced the possible differential effect of memory-noun type on comprehension accuracies in subject- vs. object-extracted relative clause conditions, because after having attempted to select the memory-nouns from the set of available memory traces in memory, subjects might have better (although still quite poor, overall) access to the source-memory information – i.e. whether the noun came from the memory-list or from the sentence – about the sentence-nouns across the conditions.

<u>Memory Load</u>	<u>Syntactic complexity</u>			
	Subject-extraction (Easy)		Object-extraction (Hard)	
<u>Memory-Noun</u>	Match	Non-match	Match	Non-match
One noun (Easy)	83.2 (3.2)	86.7 (2.7)	86.9 (2.3)	84.4 (3.0)
Three nouns (Hard)	77.6 (3.0)	73.6 (3.0)	70.3 (3.8)	74.7 (3.5)

Table 24. Comprehension accuracies in percent correct, as a function of syntactic complexity, memory load, and memory-noun type in Experiment 1C-1 (standard errors in parentheses).

Reading times Because participants had to perform a memory task and to answer two comprehension questions for each sentence, the odds of getting all three correct were not very high overall (55%). As a result, we analyzed all trials, regardless of how the memory task was

performed and how the comprehension questions were answered. The data patterns were very similar in analyses of smaller amounts of data, in which we analyzed (1) trials in which one or both of the comprehension questions were answered correctly, or (2) trials in which the memory task was performed perfectly. To adjust for differences in region lengths as well as overall differences in participants' reading rates, a regression equation predicting reading times from region length was derived for each participant, using all filler and target items (Ferreira & Clifton, 1986; see Trueswell, Tanenhaus & Garnsey, 1994, for discussion). For each region, the reading time predicted by the participant's regression equation was subtracted from the actual measured reading time to obtain a residual reading time. The statistical analyses of the raw reading time data produced the same numerical patterns (see Appendix H for tables of the raw and residual reading times). Reading time data points that were less than 100 msec in the raw data (indicating erroneous key presses) or more than three standard deviations away from the mean residual RT for a position within a condition were excluded from the analysis, affecting 2.2% of the data.

We computed a three-factor ANOVA crossing syntactic complexity (subject-extracted relative clause, object-extracted relative clause), memory load (one memory-noun, three memory-nouns) and memory-noun type (match, non-match) on the critical region (Region 2) consisting of the relative clause ("who consulted the cardiologist" / "who the cardiologist consulted"). The results are presented in Table 25. Importantly, the ANOVA revealed a three-way interaction among the three factors (marginal in the items analysis), such that syntactic complexity and the number of memory-nouns interacted in the conditions where the memory-nouns and the sentence-nouns were similar, but not in the conditions where they were dissimilar. This interaction is consistent with the idea that the amount of interference is a function of not only the representational characteristics of memory items but also the number of memory items.

In addition to the three-way interaction, we observed the following effects: (1) a main effect of syntactic complexity, (2) a main effect of memory-noun type, (3) an interaction between syntactic complexity and memory-noun type, and (4) a marginal interaction between memory-noun type and memory load. For comparisons between means of conditions, we report 95% confidence intervals (CIs) based on the mean squared errors of the relevant effects from the participant analyses (Masson & Loftus, 2003). Syntactically more complex object-extracted relative clause conditions were read slower (278.2 ms) than subject-extracted relative clause conditions (-123.1 ms; 95% CI=164.4 ms). The matching conditions (where the memory-nouns were occupations) were read slower (129.04 ms) than the non-matching conditions (26.06 ms; 95% CI=146.4 ms). The effect of match affected only the more complex object-extracted relative clauses, but not the less complex subject-extracted relative clauses. Finally, the effect of match was only present in the hard-load conditions, but not in the easy-load conditions, although this interaction was marginal in the items analysis. Given that the three-way interaction revealed differences in the patterns of results between the easy-load and the hard-load conditions, we present the analyses for the easy-load and hard-load conditions separately. This will also allow us to compare our hard-load conditions, which are most similar to Gordon et al.'s (2002) design, to the four conditions of Gordon et al.'s experiment more easily.

Table 25. Analysis of Variance results for Experiment 1C-1.

Source of variance	By participants			By items		min F'	
	<i>df</i>	F1 value	MSe	<i>df</i>	F2 value	<i>df</i>	min F'
2 x 2 x 2 ANOVA							
Synt Complexity	1,43	48.47*	292390	1,31	88.98*	1,73	31.3*
Memory Load	1,43	1.11	486815	1,31	2.40	1,71	<1
Memory-NounType	1,43	4.03	231736	1,31	5.69*	1,74	2.35
Synt x Load	1,43	<1	178848	1,31	<1	1,63	<1
Synt x NounType	1,43	5.10*	183822	1,31	6.37*	1,74	2.83
Load x NounType	1,43	8.07*	109319	1,31	2.93	1,53	2.14
Synt x Load x NounType	1,43	4.80*	123237	1,31	3.19	1,65	1.91
2 x 2 ANOVA							
Easy-Load Conditions							
Synt Complexity	1,43	51.43*	151563	1,31	92.05*	1,73	32.9*
Memory-NounType	1,43	<1	136577	1,31	<1	1,31	<1
Synt x NounType	1,43	<1	77701	1,31	<1	1,42	<1
2 x 2 ANOVA							
Hard-Load Conditions							
Synt Complexity	1,43	20.06*	319676	1,31	35.90*	1,73	12.8*
Memory-NounType	1,43	8.88*	204478	1,31	5.64*	1,64	3.44
Synt x NounType	1,43	6.58*	229359	1,31	6.79*	1,72	3.34

Note: Significant effects are marked by asterisk.

We will now present the analyses for the easy-load conditions. Figure 20 presents the mean residual RTs per region across the four easy-load conditions. A 2 x 2 ANOVA crossing syntactic complexity (subject-extracted relative clause, object-extracted relative clause), and memory-noun type (match, non-match) in the critical region revealed a main effect of syntactic complexity, such that the object-extracted relative clause conditions were read significantly slower (248.9 ms) than the subject-extracted relative clause conditions (-172.05 ms; 95% CI=118.4 ms), but no other effects. The results are presented in Table 28. In the other three regions (Region 1, Region 3, and Region 4), there were no reliable effects, with the exception of an unpredicted interaction in Region 3, such that in the non-match conditions, the difference between the subject- and object-extracted relative clauses was larger than in the match conditions (F1(1,43)=4.07; MSe=424429; $p < .05$; F2(1,31)=5.78; MSe=324219; $p < .05$; minF'

(1,74)=2.38; $p=.13$). There is no reason to expect an interaction of this sort in this region, because there are no differences in the linguistic materials among the four conditions. Furthermore, the effect is not a spill-over effect from the previous region, because there is no trend for an interaction of this sort in Region 2. This effect is therefore likely to be spurious.

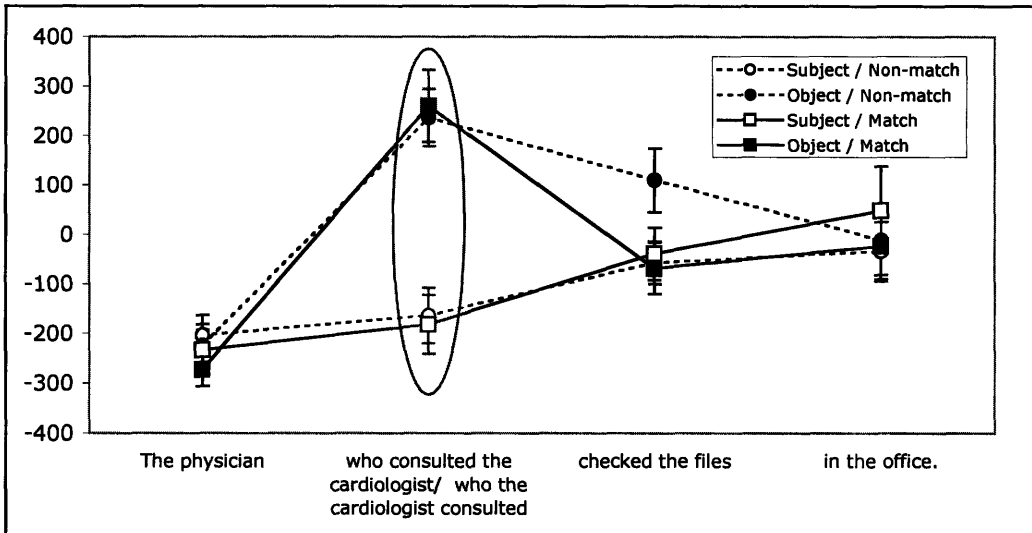


Figure 20: Reading times per region in the four easy-load (one memory-noun) conditions in Experiment 1C-1. Error bars indicate standard errors. The critical region is circled.

We will now present the analyses for the hard-load conditions. Figure 21 presents the mean residual RTs per region across the four hard-load conditions. A 2 x 2 ANOVA crossing syntactic complexity (subject-extracted relative clause, object-extracted relative clause), and memory-noun type (match, non-match) in the critical region revealed two significant main effects and a significant interaction. First, the object-extracted relative clause conditions were read significantly slower (307.6 ms) than the subject-extracted relative clause conditions (-74.2 ms; 95% CI=171.9 ms). Second, the match conditions were read significantly slower (218.2 ms) than the non-match conditions (15.2 ms; 95% CI=137.5 ms). Finally, there was a significant interaction, such that the effect of match was only present in the object-extracted relative clause conditions and not in the subject-extracted relative clause conditions. The results are presented in Table 3. In the other three regions (Region 1, Region 3, and Region 4), the only reliable effect was that of syntactic complexity in Region 3, such that object-extracted relative clause conditions were read slower (95.5 msec) than subject-extracted relative clause conditions (-30.8 msec; 95% CI=113.2 ms) ($F_1(1,43)=5.07$; $MSe=702610$; $p < .05$; $F_2(1,31)=6.56$; $MSe=462948$; $p < .02$; $\min F'(1,74)=2.85$; $p=.095$). This effect could be a possible spillover effect from Region 2.

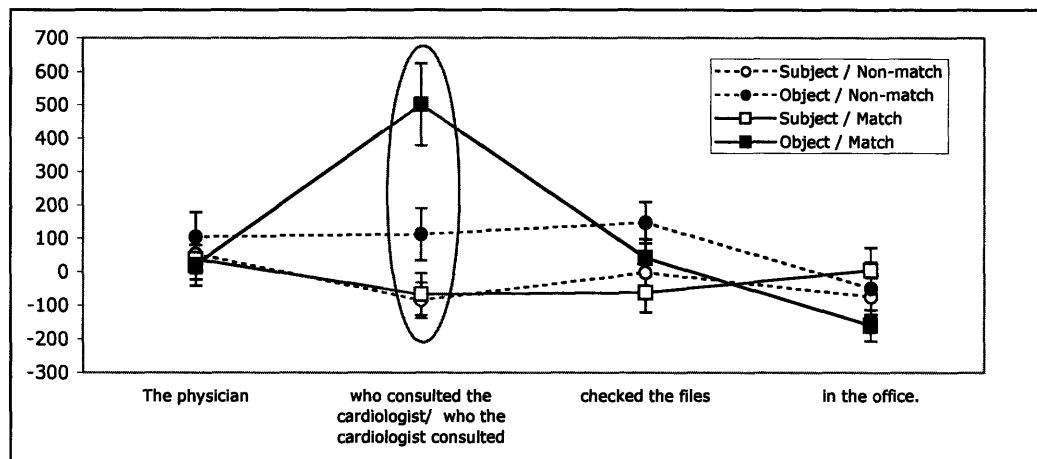


Figure 21: Reading times per region in the four hard-load (three memory-nouns) conditions in Experiment 1C-1. Error bars indicate standard errors. The critical region is circled.

Discussion

The most interesting result presented here is an interaction between syntactic complexity and the memory-noun / sentence-noun similarity during the critical region of the linguistic materials in the hard-load (three memory-nouns) conditions: people processed object-extracted relative clauses more slowly when they had to maintain a set of nouns that were similar to the nouns in the sentence than when they had to maintain a set of nouns that were dissimilar from the nouns in the sentence; in contrast, for the less complex subject-extracted relative clauses, there was no reading time difference between the similar and dissimilar memory load conditions. In the easy-load (one memory-noun) conditions no interaction between syntactic complexity and memory-noun / sentence-noun similarity was observed.

These results provide evidence against the hypothesis whereby there is a pool of domain-specific verbal working memory resources for sentence comprehension, *contra* Caplan & Waters (1999). Specifically, the results reported here extend the off-line results reported by Gordon et al. (2002) who – using a similar dual-task paradigm – provided evidence of an interaction between syntactic complexity and memory-noun / sentence-noun similarity in response-accuracies to questions about the content of the sentences. Although there was also a suggestion of an on-line reading time interaction in Gordon et al.'s experiment, this effect did not reach significance. Based on a visual examination of the graph in Gordon et al.'s paper (p. 429), the effect of memory-noun type in subject-extracted conditions is approximately 10 msec per word (40 msec over the four-word relative clause region), and the effect of memory-noun type in object-extracted conditions is approximately 35 msec (140 msec over the four-word relative clause region). In the hard-load conditions of our experiment, the effect of memory-noun type in subject-extracted conditions over the four-word relative clause region is 12 msec in raw RTs (18 msec in residual RTs), and the effect of memory-noun type in object-extracted conditions is 324 msec in raw RTs (388 msec in residual RTs). This difference in effect sizes of memory-noun type between Gordon et al.'s and our experiment is plausibly responsible for the interaction observed in the current experiment, and the lack of such an interaction in Gordon et al.'s study.

The current results therefore demonstrate that Gordon et al.'s results extend to on-line processing.

Furthermore, the interaction among the three factors provides evidence in support of the hypothesis whereby the amount of interference is a function of both the representational characteristics of memory items, and the number of memory items. Specifically, syntactic complexity interacted with memory load (manipulated in terms of the number of items) only in the context of similar elements. In light of these results, it is possible to explain the lack of interactions between syntactic complexity and number of memory items in the previous dual-task experiments, which used digit-span as a secondary verbal working memory task (e.g., Waters et al., 1987; Waters et al., 1995). Because digits are qualitatively very different from the nouns in the sentence materials (similar to the non-match conditions in our experiment), there is no reason to expect that the memory load should interact with syntactic complexity.

In order to account for the observed on-line interaction between syntactic complexity and the complexity of a non-linguistic verbally-mediated task, we would like to elaborate Gordon et al.'s proposal. One possible way to spell out the interaction between syntactic complexity and the memory-noun / sentence-noun similarity is in terms of local vs. non-local integrations, as suggested by Gordon et al. (p. 426). Gibson (1998, 2000) and Grodner and Gibson (2005) provide a framework to formalize this idea. The integration of a new word can either be local – to the preceding word in the sentence – or non-local, requiring a retrieval of a word further back in the input stream. In the case of a local integration, the incoming word can be immediately connected to the preceding word, because there are no intervening potential attachment sites. Thus, in subject-extracted relative clauses – at the point of integrating the embedded verb with the embedded subject – no search for an attachment site is needed, and the existence of a set of stored memory items will probably not cause any interference in forming the connection. However, when the integration is non-local, the target attachment site needs to be retrieved from memory, and there may be interference from the intervening potential attachment sites (e.g., McElree et al., 2003). Specifically, in object-extracted relative clauses, an embedded subject intervenes between the subject of the main clause and the embedded verb, and therefore may cause interference. Furthermore, when there is a list of stored items in working memory, the similarity of these items to the target attachment site (in this case, the subject of the main clause, which is also the object of the embedded verb) may further increase the difficulty of the integration. Thus, there should be some difficulty with both non-local integrations, but more difficulty in the condition where the memory items are similar to the item that needs to be retrieved from memory. The interaction between syntactic complexity and similarity between memory-nouns and sentence-nouns can therefore be accounted for in terms of integration difficulty: (a) there is little difficulty for *local* integrations, with either similar or dissimilar memory items; and (b) for *non-local* integrations, the condition with similar memory items is much more difficult to process than the condition with dissimilar memory items.

Although we have discussed the observed effects in terms of difficulty at the retrieval stage of the memory process, in theory it is possible that the effects occur during the encoding and/or during the maintenance stage of the memory process. For example, it is possible that the encoding process becomes increasingly more costly with an increase in the number of shared features between the to-be-remembered items. Specifically, when the sentence-nouns are being

encoded into memory, the difficulty of this process may be a function of the similarity of the nouns that have been recently encoded and are currently stored in the memory store. Similarly, it is possible that the difficulty of the storage process increases with an increase in the overlap in representational characteristics of the to-be-remembered items.

Research on short-term memory provides some suggestive evidence in favor of the hypothesis whereby proactive interference effects take place at the retrieval stage (Gardiner et al., 1972; as cited in Crowder, 1976). Gardiner et al. (1972; as cited in Crowder, 1976) conducted an experiment in order to identify the locus of proactive interference effects among the three different stages of the memory process, and provided evidence for placing these effects at the stage of retrieval. Specifically, Gardiner et al. conducted an experiment where subjects were presented with four lists, and tested for their memory of the items in the fourth list. The items in the first three lists all came from the same semantic category (e.g., garden flowers). The items in the fourth list came from a slightly different semantic category (e.g., wild flowers). Any version of a proactive interference hypothesis predicts interference effects from the first three lists on the memory for the items from the fourth list. In the attempt to identify the locus of the proactive interference effects, Gardiner et al. divided the subjects into three groups. The control group was not informed with regard to the difference in categories between the first three lists and the fourth list; the two experimental groups were informed about the difference in categories either before, or after the presentation of the fourth list. If the proactive interference effects took place at the stage of encoding or storage, then the group that was informed about the difference in categories between the first three lists and the fourth list *after* the presentation of the fourth list should behave similarly to the control group, as they would not be able to take advantage of the distinctiveness of the items in the fourth list until the retrieval stage. In contrast, if the proactive interference effects took place at the stage of retrieval, then this experimental group should behave similarly to the group that was informed about the difference in categories between the first three lists and the fourth list *before* the presentation. The results were consistent with the hypothesis whereby the interference effects take place at the retrieval stage. Thus, based on evidence from the short-term memory research, and given the fact that it is easiest to conceptualize the interference effects we observed in terms of relative ease/difficulty of retrieval in local/non-local integrations, we tentatively conclude that the locus of proactive interference effects in sentence processing is at the retrieval stage of the memory process.

Conclusions

In Section 1B we investigated the extent of domain-specificity of the WM system underlying linguistic integrations by comparing linguistic integrations to tasks involving similar integration processes. In Section 1C we described an experiment investigating the extent of domain-specificity of the WM system underlying linguistic integrations by comparing linguistic integrations and a involving similar materials. The data from Experiment 1C-1 provided further evidence against the strictly domain-specific view of the working memory system underlying language processing.

Chapter 2: The relationship between the working memory system and the long-term memory system in language processing.

General Background

In this Chapter, I will present six experiments investigating the relationship between the working memory system and the long-term memory system in language processing.

When a word is integrated into an evolving structure, the processing mechanism must retrieve (1) the lexical properties of the incoming word from long-term memory, and (2) the word's structural dependent(s) from working memory, if there are any (see Figure 22). The *lexical properties* of the word include both syntactic information and semantic information. *Syntactic information* includes the syntactic category of the word, the phrase structure rules associated with this category (which are plausibly retrieved probabilistically ranked by frequency), and different lexico-syntactic properties, such as gender (for languages that mark gender), number, tense, etc. *Semantic information* includes the meaning of the word and the thematic properties of the word. The *structural dependents* of the word include any word or words in the preceding context to which the incoming word is structurally related. Following Gibson (1998, 2000), I will assume that when an incoming word has to be integrated with an immediately preceding word, there is little/no cost associated with retrieving the immediately preceding word from memory.

Both of these retrieval operations – retrieving the word's lexical properties and retrieving the word's structural dependents – have been shown to affect the relative ease or difficulty of the integration process (e.g., MacDonald, 1994; Boland, 1997; Boland & Blodgett, 2001; Seidenberg & MacDonald, 1999; Gibson, 1998, 2000; Gordon et al., 2001, 2004; McElree et al., 2003; VanDyke & Lewis, 2003).

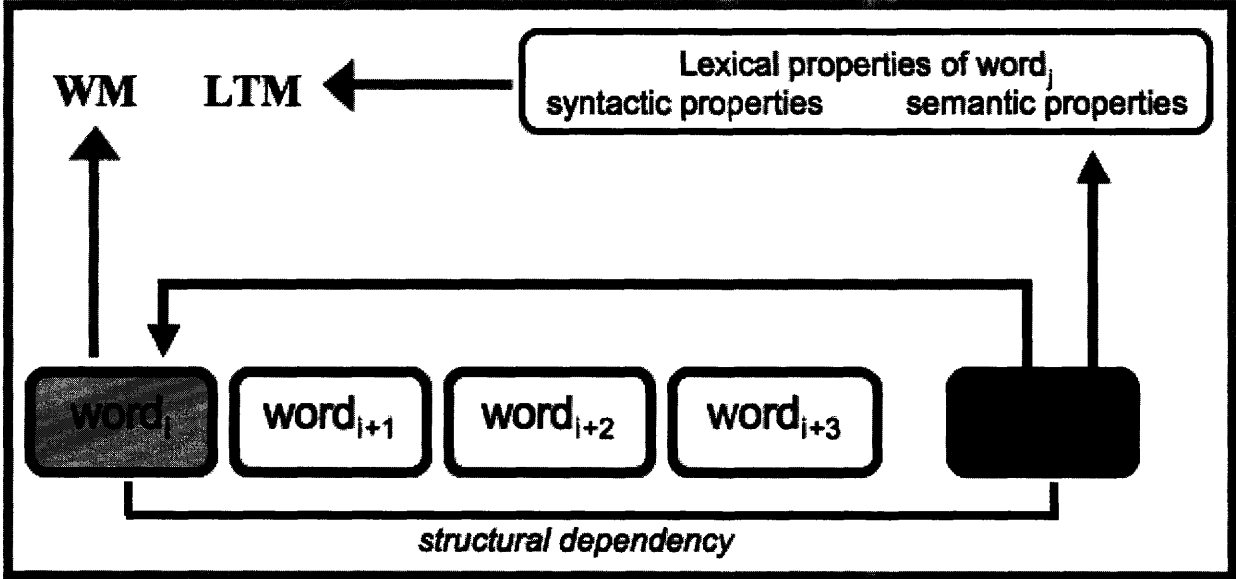


Figure 22: Retrieval operations (marked by red arrows) involved in integrating a word into a structure.

As discussed in the Introduction, Gibson (1998) proposed a model of linguistic complexity where integration difficulty was quantified in terms of the linear distance (measured in terms of new discourse referents) between the incoming word and the word/position earlier in the input to which this word needs to be connected. It was proposed that the further away the two elements of the dependency are from each other, the more difficult it is to integrate the second element of the dependency with the first element of the dependency because the first element of the dependency needs to be retrieved from memory. In Figure 22, this retrieval operation is illustrated by an arrow coming out of the word_j pointing to word_i, which needs to be retrieved from memory in order to connect word_j to it.

While Gibson's model accounts for a large amount of variance in reading times (e.g., Grodner & Gibson, 2005), it does not incorporate the other retrieval operation – the one involving accessing the lexical properties of the incoming word – which has been shown to also affect the difficulty of the integration process. Furthermore, recent research has begun to suggest that the two retrieval operations may not be independent (e.g., Gordon et al., 2001, 2004; Traxler et al., 2002, 2005). It therefore seems important to investigate how the two retrieval operations – the retrieval of lexical properties and the retrieval of structural dependents – relate to each other. I will attempt to do so in this Chapter.

I will evaluate two alternative hypotheses about the possible relationship between the two retrieval operations (the retrieval of lexical properties and the retrieval of structural dependents): (1) the *Separate Resource Pools hypothesis*, according to which the two retrieval operations rely on independent pools of working memory resources, and (2) the *Shared Resource Pool hypothesis*, according to which the two retrieval operations rely on the same pool of working memory resources. The Separate Resource Pools hypothesis predicts that manipulating the difficulty of one retrieval operation should have no effect on the difficulty of the other retrieval operation. In other words, it predicts two independent effects of the difficulty of each of the retrieval operations. In contrast, the Shared Resource Pool hypothesis predicts that the relative difficulty of one retrieval operation should affect the relative difficulty of the other retrieval operation: for example, the relative difficulty of retrieving the lexical properties of the incoming word should affect the difficulty of retrieving the word's structural dependents. I graphically present these predictions in more detail in Figures 23-25 below.

Figure 23 presents the predictions of the Separate Resource Pools hypothesis. Figure 23a presents a baseline where – at some fixed cost associated with retrieving the word's lexical properties – there is a difference between a condition where retrieving the word's structural dependents is easy (a local integration condition) and a condition where retrieving the word's structural dependents is hard (a non-local integration condition), such that the non-local condition is processed slower than the local condition. This difference between the non-local and the local conditions thus reflects the cost associated with retrieving the word's structural dependents. The baseline pair of bars is presented on the left in both Figures 23b and 23c. Figure 23b demonstrates a case where the cost associated with retrieving the word's lexical properties is *increased*, compared to 23a. The Separate Resource Pools hypothesis predicts that each of the conditions should suffer from this increase to the same extent (as shown in the pair of bars on the right), such that – while both the local and the non-local integration conditions

become slower – the difference between them remains unchanged. Figure 23c demonstrates a case where the cost associated with retrieving the word’s lexical properties is *decreased*, compared to 23a. The Separate Resource Pools hypothesis predicts that each of the conditions should benefit from this decrease to the same extent (as shown the in the pair of bars on the right), such that – while both the local and the non-local integration conditions become faster – the difference between them remains unchanged.

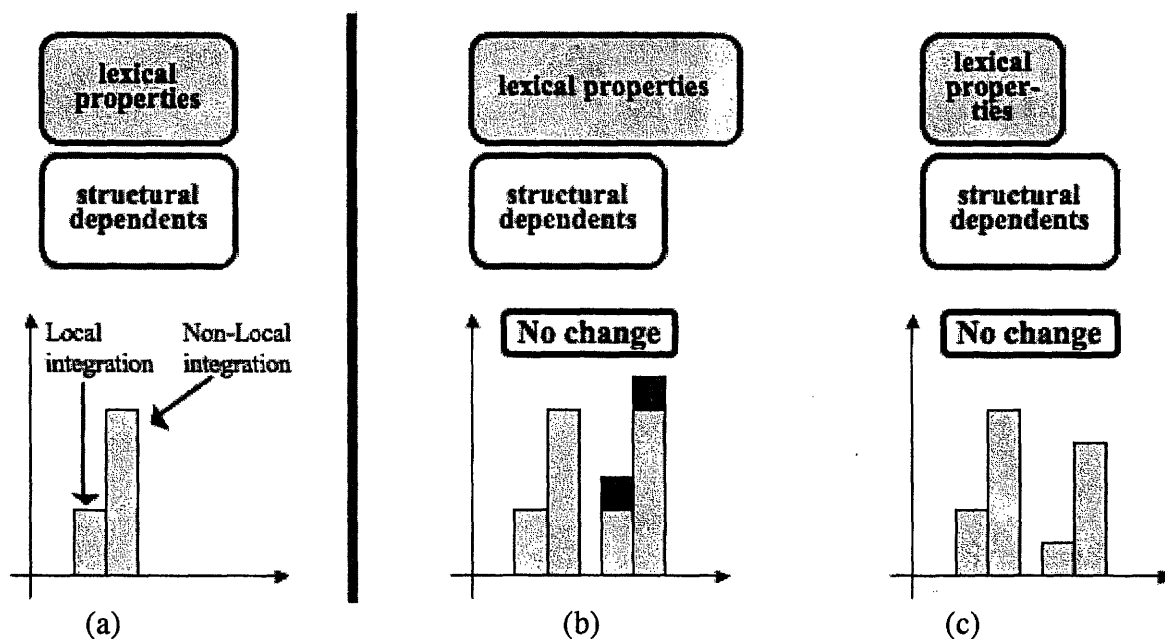


Figure 23: Predictions of the Separate Resource Pool hypothesis in cases where the retrieval of the word’s lexical properties is hard (b) vs. easy (c), compared to a baseline (a).

Figure 24 presents the predictions of the Shared Resource Pool hypothesis for the case where retrieving the word’s lexical properties is *hard*. As in Figure 23 above, Figure 24a presents a baseline where – at some fixed cost associated with retrieving the word’s lexical properties – there is a difference between a condition where retrieving the word’s structural dependents is easy (a local integration condition) and a condition where retrieving the word’s structural dependents is hard (a non-local integration condition). This difference between the non-local and the local conditions thus reflects the cost associated with retrieving the word’s structural dependents. The baseline is presented on the left in both Figures 24b and 24c. The Shared Resource Pool hypothesis predicts that the non-local condition should suffer from an increase in the cost of retrieving the word’s lexical properties to a larger extent than the local condition, because fewer resources are available for performing the retrieval of the structural dependents, as shown, for example, in Figure 24b. In other words, the Shared Resource Pool hypothesis predicts a superadditive processing difficulty in the non-local condition when retrieving the word’s lexical properties is hard.

However, the difficulty associated with retrieving the word’s structural dependents cannot increase indefinitely as a function of the increasing difficulty of retrieving the word’s lexical properties. In particular, in most natural language processing environments comprehenders do not have control over the speed of the linguistic input. Therefore, when processing difficulties

arise, comprehenders cannot spend as much time as may be necessary for them in order to fully process the difficult part of the input because they have to continue processing the incoming words. I hypothesize that, as a result, comprehenders may sometimes delay some aspects of processing in cases of the difficult input until some later time (plausibly until the point when the input becomes relatively easy leaving some resources to complete the processing of the previously encountered input). This incremental nature of language processing will plausibly have the following consequence: when the cost associated with retrieving the word's lexical properties reaches a certain threshold – such that very few resources are left for retrieving the word's structural dependents – retrieving the word's structural dependents may be partially or completely delayed until a later point in the sentence. This will translate into (1) a decrease in or an elimination of the difficulty associated with retrieving the word's structural dependents during the processing of the target incoming word, and (2) difficulty associated with retrieving the word's structural dependents following the target incoming word. An example of such an outcome is shown in Figure 24c.

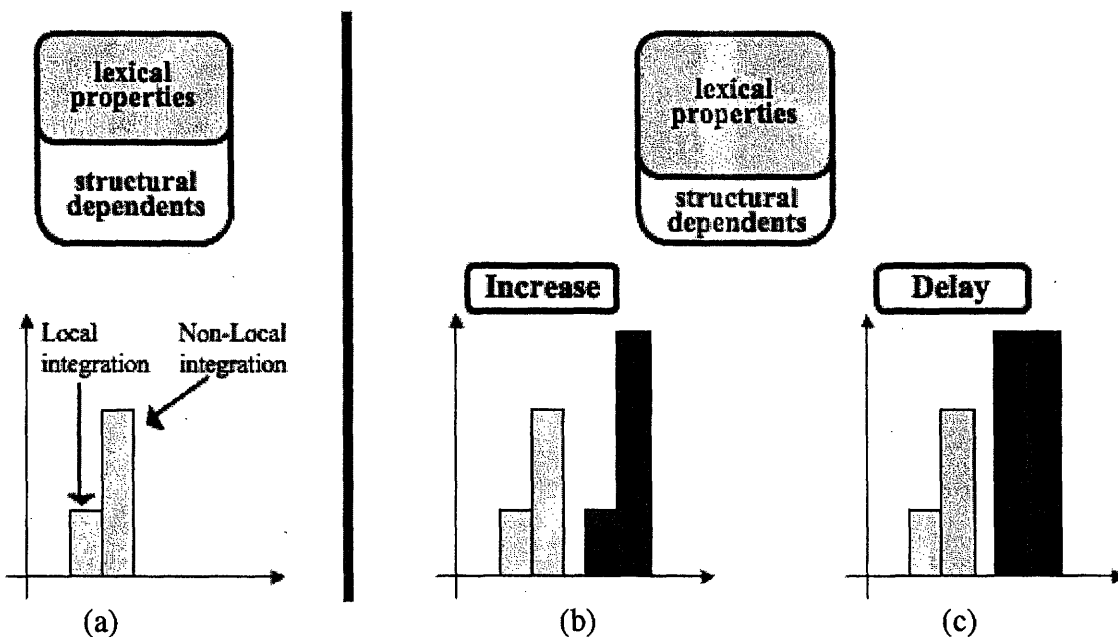


Figure 24: Predictions of the Shared Resource Pool hypothesis in cases where the retrieval of the word's lexical properties is hard ((b) or (c)), compared to a baseline (a).

Figure 25 presents the predictions of the Shared Resource Pool hypothesis for the case where retrieving the word's lexical properties is *easy*. As in Figures 23 and 24 above, Figure 25a presents a baseline. This baseline is presented on the left in both Figures 25b and 25c. The Shared Resource Pool hypothesis predicts that the non-local condition should benefit from a decrease in the cost of retrieving the word's lexical properties to a larger extent than the local condition, because the local condition already has no cost associated with retrieving the word's structural dependents, as shown in Figure 25b. Furthermore, at some point – as the cost associated with retrieving the word's lexical properties continues to decrease – the non-local condition may become as easy to process as the local condition because resources for retrieving the word's structural dependents become abundant. An example is shown in Figure 25c.

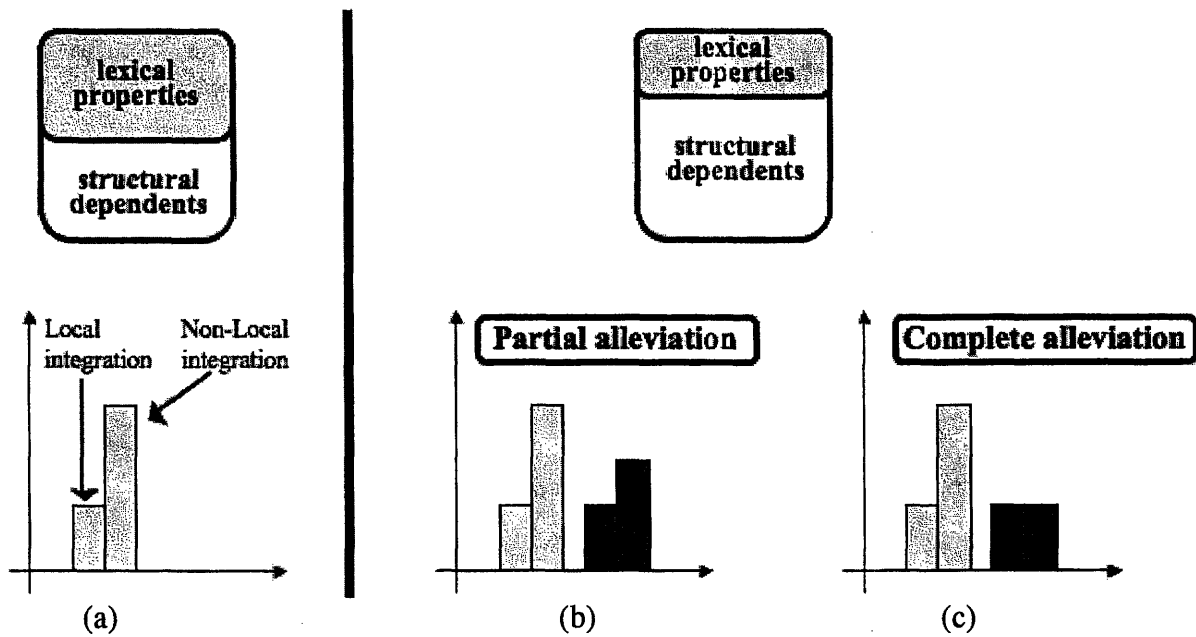


Figure 25: Predictions of the Shared Resource Pool hypothesis in cases where the retrieval of the word's lexical properties is easy ((b) or (c)), compared to a baseline (a).

The chapter will be divided into two sections: Section 2A will examine two cases where retrieving the lexical properties of the incoming word is difficult, and Section 2B will examine two cases where retrieving the lexical properties of the incoming word is easy. It will be shown that the relative difficulty of retrieving the lexical properties of the incoming word affects the difficulty of retrieving the word's structural dependents. The results will be argued to provide support for the Shared Resource Pool hypothesis, according to which retrieving the word's lexical properties and retrieving its structural dependents rely on the same pool of working memory resources.

2A. The difficulty of retrieving the lexical properties of an incoming word increases or delays the difficulty of retrieving the word's structural dependents.

In this section I will examine two cases where retrieving the lexical properties of an incoming word is difficult. In particular, I will examine how the difficulty of retrieving (1) the word's thematic properties, and (2) the word's meaning interacts with the difficulty of retrieving the word's structural dependents. I will show that the difficulty of retrieving the word's lexical properties can *increase* or *delay* the difficulty of retrieving the word's structural dependents.

Background for Experiments 2A-1 through 2A-3

Most common types of events we observe in the world involve animate entities acting upon inanimate entities (e.g., *a girl eating an apple* or *a woman pushing a baby-stroller*) or upon animate entities (e.g., *a girl kissing a boy*, or *a man telephoning a woman*). Previous research in language processing has shown that comprehenders are sensitive to this information in the course of on-line language understanding. Specifically, comprehenders prefer the agents to be animate⁹.

In this section, we will investigate how retrieving the thematic properties of a verb – which we assume include information about the verb's typical agents and patients – affects the difficulty of retrieving its structural dependents. We will first review a recent finding from the literature (Traxler et al., 2002) and show how the results can be accounted for by the Shared Resource Pool hypothesis. We will then discuss an alternative account of these results provided by Traxler et al. Finally, we will attempt to distinguish between the two accounts and we will attempt to show that the Shared Resource Pool hypothesis is a better account of these and other data.

Traxler et al. (2002, 2005) reported an interaction between syntactic complexity (subject-/object-extracted RCs) and the animacy of the RC subject (animate/inanimate) in sentences like (12a)-(12d), such that the inanimate-RC-subject object-extracted condition (12d) caused processing difficulty, compared to the other three conditions that were equally easy to process (see Mak et al., 2002, 2006, for similar findings in Dutch).

(12a) *Animate RC subject / Subject-extracted:*

The director that watched the movie received a prize at the film festival.

(12b) *Animate RC subject / Object-extracted:*

The movie that the director watched received a prize at the film festival.

(12c) *Inanimate RC subject / Subject-extracted:*

The movie that pleased the director received a prize at the film festival.

(12d) *Inanimate RC subject / Object-extracted:*

The director that the movie pleased received a prize at the film festival.

The Shared Resource Pool hypothesis can account for this pattern of results in the following way. As has been discussed in different sections of the thesis – we argue that the general preference for subject-extracted RCs over object-extracted RCs can be explained by the

⁹ A disclaimer: While there may be differences among animate nouns in terms of agent typicality, I will be only considering the broad distinction between animate nouns (typical agents) and inanimate nouns (atypical agents).

difference in the difficulty of the retrieval operation involved in retrieving the word's (specifically, the embedded verb's, in this case) structural dependents (e.g., Gibson, 1998, 2000). In subject-extracted RCs all the integrations within the RC are local and therefore cost-free. In contrast, in object-extracted RCs the embedded verb has to be integrated with the relativizer co-indexed with the head noun and thus, the head noun needs to be retrieved from memory, which is costly. Furthermore, the lexical properties of the verb include information about its thematic properties which, in turn, include information about typical agents and patients of the verb (e.g., Trueswell et al., 1993; MacDonald et al., 1994). Therefore, when a verb is encountered which prefers animate agents (this includes most of the verbs in English) it may be more difficult to access the representation of this verb in cases where the already encountered agent of the verb is inanimate, compared to cases where the agent is animate and thus satisfies the verb's thematic preference.

We will now go through the four conditions examined by Traxler et al. and show how the *Shared Resource Pool hypothesis* may provide an account of the observed results. Instead of using the example from Traxler et al.'s paper, however, we will use of our own examples (as will be discussed below, in Traxler et al.'s original study psychological verbs – such as *pleased* – were used which may confound the interpretation).

First, in the animate-RC-subject subject-extracted condition (*The mountaineer that moved the boulder...*), retrieving the structural dependent of the verb *moved* (the relativizer *that* co-indexed with the noun phrase *the mountaineer*) is associated with little/no cost because the integration is local. Retrieving the thematic properties of the verb *moved* is also not costly because the agent is animate and thus satisfies the verb's thematic preference. As a result, resources are abundant and no processing difficulty ensues (see Figure 26a).

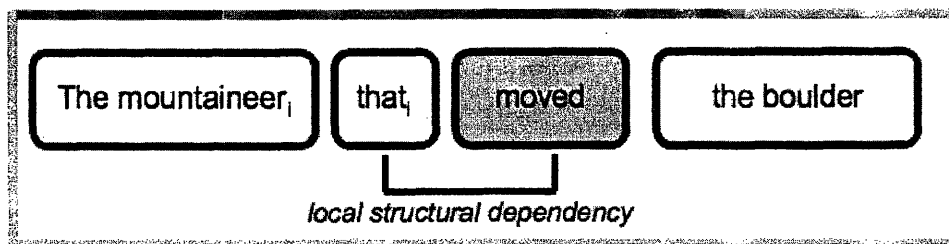


Figure 26a: The animate-RC-subject subject-extracted condition.

Second, in the animate-RC-subject object-extracted condition (*The boulder that the mountaineer moved...*), retrieving the structural dependent of the verb *moved* (the relativizer co-indexed with the noun phrase *the boulder*) is now costly because the integration is non-local and thus the relativizer has to be retrieved from memory. Retrieving the thematic properties of the verb *moved* is not costly – as in the animate-RC-subject subject-extracted condition – because the agent is animate and thus satisfies the verb's thematic preference. Because one of the retrieval operations (retrieving the word's lexical properties) is not costly, resources are abundant and little/no processing difficulty ensues (see Figure 26b).

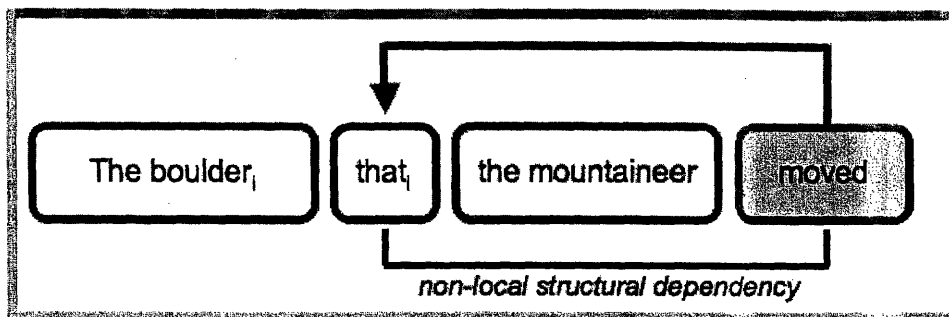


Figure 26b: The animate-RC-subject object-extracted condition.

Third, in the inanimate-RC-subject subject-extracted condition (*The boulder that hit the mountaineer...*), retrieving the structural dependent of the verb *hit* (the relativizer co-indexed with the noun phrase *the boulder*) is associated with little/no cost because the integration is local, as in the animate-RC-subject subject-extracted condition. However, retrieving the thematic properties of the verb *hit* is now costly because the agent is inanimate and thus does not satisfy the verb's thematic preference for an animate agent. Because one of the retrieval operations (retrieving the word's structural dependents) is not costly, resources are abundant and little/no processing difficulty ensues (see Figure 26c).

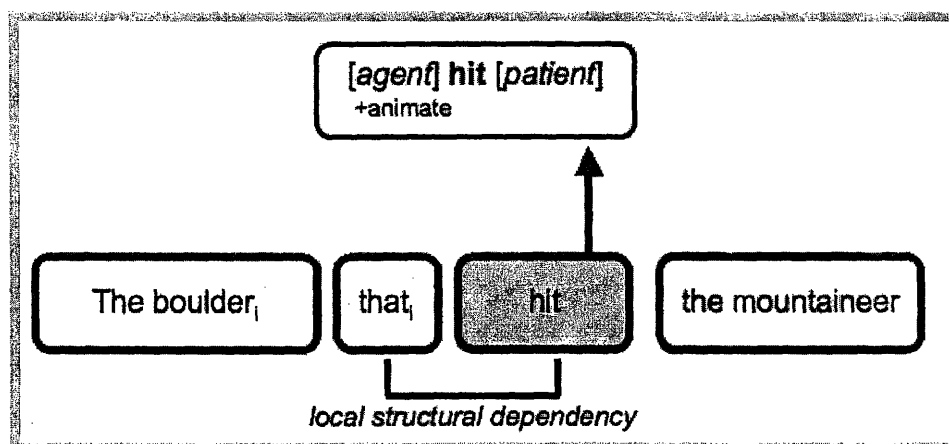


Figure 26c: The inanimate-RC-subject subject-extracted condition.

Finally, in the inanimate-RC-subject object-extracted condition (*The mountaineer that the boulder hit...*), retrieving the structural dependent of the verb *hit* (the relativizer co-indexed with the noun phrase *the mountaineer*) is costly because the integration is non-local and thus the relativizer has to be retrieved from memory. Furthermore, retrieving the thematic properties of the verb *hit* is also costly because the agent is inanimate and thus does not satisfy the verb's thematic preference for an animate agent. Because both of the retrieval operations are costly, resources are more scarce and processing difficulty ensues (see Figure 26d).

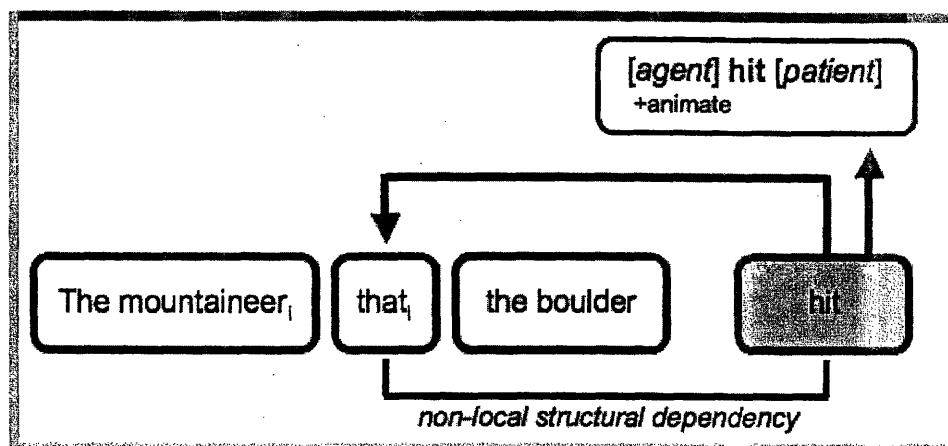


Figure 26d: The inanimate-RC-subject object-extracted condition.

Now that we have demonstrated how the Shared Resource Pool hypothesis can account for the pattern of results observed by Traxler et al. (2002), we will discuss Traxler et al.'s alternative account of their results. (Note that this pattern of results cannot be accounted for by the Separate Resource Pools hypothesis whereby the two retrieval operations rely on independent pools of working memory resources.)

Traxler et al. provided a *Reanalysis-based hypothesis* to account for their pattern of results. First, Traxler et al. argued that the general preference for subject-extracted RCs over object-extracted RCs can be explained by postulating that upon encountering the relativizer *that* comprehenders always adopt a subject-extracted RC interpretation treating the head-noun as the subject of both the main and the relative clause. Upon encountering the RC-subject, reanalysis is required to interpret the head-noun as the object of the RC. This reanalysis was argued to be responsible for the cost associated with processing object-extracted RCs. Traxler et al. further argued that the reanalysis cost is higher when the RC subject is inanimate and thus a poor agent. To illustrate these ideas using the example illustrated in Figures 26a – 26d: first, sentences in 26b (*The boulder that the mountaineer...*) and 26d (*The mountaineer that the boulder...*) require reanalysis upon encountering the subject of the relative clause (*the mountaineer* and *the boulder*, respectively); and second, the reanalysis process is more difficult in 26d (*The mountaineer that the boulder...*), compared to 26b because *boulder* is not a good agent and thus it is harder to interpret it as the subject of the RC.

Before presenting the experiments in this section, it is worth noting that interpreting the results of Traxler et al.'s original (2002) experiment is complicated by the fact that psychological verbs (in particular, theme-experiencer verbs) were used in the inanimate-RC-subject conditions (e.g., *The movie that pleased the director...* or *The director that the movie pleased...*). These theme-experiencer verbs (like *pleased*) – for which the experiencer is the object of the verb – have been argued to present processing difficulties, compared to regular active verbs (e.g., *kiss*) or experiencer-theme verbs (e.g., *love*), for which the experiencer is the subject of the verb (e.g., Cupples, 2002). To eliminate any potential differences among conditions in terms of the difficulty of the verb processing, we first conducted Experiment 2A-1 – which used actional verbs across all the conditions – to replicate the pattern of results reported by Traxler et al.

(2002)¹⁰.

We further conducted two additional experiments to evaluate the *Shared Resource Pool* hypothesis proposed here and Traxler et al.'s *Reanalysis-based* hypothesis.

Experiments 2A-1 – 2A-3: The effects of the difficulty of retrieving the word's thematic properties on the difficulty of retrieving its structural dependents
[Fedorenko & Gibson, in preparation]

Experiment 2A-1 was aimed at replicating Traxler et al.'s results using actional verbs. Experiment 2A-2 evaluated the Shared Resource Pool hypothesis by presenting the materials from Experiment 2A-1 with a concurrent digit-span task. We reasoned that if the interaction between retrieving the word's lexical properties (thematic properties, in this case) and retrieving its structural dependents is indeed responsible for the pattern of results reported by Traxler et al., then we should be able to observe a cost associated with retrieving the word's lexical properties in structures where the retrieval of structural dependents is not costly (e.g., subject-extracted RCs) if we find another way to decrease the amount of verbal working memory resources available to the comprehenders. Finally, Experiment 2A-3 was aimed at directly testing the predictions of Traxler et al.'s Reanalysis-based hypothesis.

Experiment 2A-1

This experiment was aimed at replicating the results reported in Traxler et al. (2002) using the same class of verbs (actional verbs) across conditions.

Methods

Participants Forty participants from MIT and the surrounding community were paid for their participation. All were native speakers of English and were naive as to the purposes of the study.

Design and materials The experiment had a 2 x 2 design (following Traxler et al., 2002), crossing syntactic complexity (subject-extracted RCs, object-extracted RCs) and the animacy of the RC subject (animate, inanimate). (Using the terms of the Shared Resource Pool and the Separate Resource Pools hypotheses, we crossed the difficulty of retrieving the word's structural dependents with the difficulty of retrieving the word's lexical properties – in particular, the verb's thematic properties.)

The materials consisted of 20 sets of sentences with four different versions as in (13):

(13a) *Animate RC subject / Subject-extracted:*

The mountaineer | that | moved the boulder | was | experienced | and strong.

¹⁰ Traxler et al. (2005) have also conducted an experiment using only active verbs and obtained similar results. However, Experiment 2A-1 was designed and run prior to the publication of Traxler et al.'s (2005) results.

(13b) *Animate RC subject / Object-extracted:*

The boulder | that | the mountaineer moved | was | large | and heavy.

(13c) *Inanimate RC subject / Subject-extracted:*

The boulder | that | hit the mountaineer | was | large | and heavy.

(13d) *Inanimate RC subject / Object-extracted:*

The mountaineer | that | the boulder hit | was | experienced | and strong.

As stated above, in this experiment the predictions of the Shared Resource Pool hypothesis and Traxler et al.'s Reanalysis-based hypothesis are the same.

In addition to the target sentences, the experiment included 40 sentences from two unrelated experiments and 24 fillers. The length and syntactic complexity of the sentences in the two unrelated experiments and in the fillers was similar to that of the target sentences. The stimuli were pseudo-randomized separately for each participant, with at least one item from another experiment or one filler separating the target sentences. Each participant saw only one version of each sentence, following a Latin-Square design (see Appendix I for a complete list of linguistic materials).

Procedure The task was self-paced word-by-word reading with a moving-window presentation (Just, Carpenter & Woolley, 1982). The experiment was run using the Linger 2.85 software by Doug Rohde. Each experimental sentence was structured, as follows: (1) a definite noun phrase, (2) a relativizer *that*, (3) a subject-/object-extracted relative clause, and (4) the main verb phrase which consisted of 4-6 words. As shown in (13), the main verb phrase was divided into three regions for the purposes of the analyses.

Each trial began with a series of dashes marking the position and length of the words in the sentence. Participants pressed the spacebar to reveal each word of the sentence. The amount of time the participant spent reading each word was recorded as the time between key-presses.

To assure that the participants read the sentences for meaning, a comprehension question was presented at the end of each trial, asking about the propositional content of the sentence. Participants pressed one of two keys to respond “yes” or “no”. After a correct response, the word “CORRECT” flashed briefly on the screen, and after an incorrect response, the word “INCORRECT” flashed briefly.

Before the experiment started, a short list of practice items and questions was presented in order to familiarize the participants with the task. Participants took approximately 25 minutes to complete the experiment.

Results

Comprehension question performance Across the conditions, participants answered the comprehension question correctly 97.5% of the time. Table 26 presents the mean accuracies across the four conditions. A two-factor ANOVA crossing syntactic complexity (subject-extracted RC, object-extracted RC) and the animacy of the RC subject (animate, inanimate) on the responses to the comprehension questions revealed a marginal effect of syntactic complexity

in the participants analysis, such that participants were less accurate in the object-extracted conditions ($F(1,39)=3.06$; $MSe=160$; $p=.088$; $F(1,19)=2.23$; $MSe=74$; $p=.15$). There were no other effects ($F_s < 1$).

<u>Syntactic complexity</u>	<u>RC subject animacy</u>	
	Animate	Inanimate
Subject-extracted RC	98.5 (.84)	98.5 (1.1)
Object-extracted RC	96.5 (1.2)	96.5 (1.2)

Table 26. Comprehension accuracies in percent correct, as a function of syntactic complexity and RC subject animacy in Experiment 2A-1 (standard errors in parentheses).

Reaction times We analyzed all the trials, regardless of whether the comprehension question was answered correctly. The statistical data patterns were very similar in the analysis of only the trials where the comprehension question was answered correctly. Reading times more than three standard deviations away from the mean for a position within condition were removed from the analyses, excluding 1.8% of the data.

We present the analyses of the critical regions first, followed by the analyses of the remaining regions. We defined two critical regions: (1) the relative clause, and (2) the first word the main verb phrase. Figure 27 presents the mean reading times per region across the four conditions.

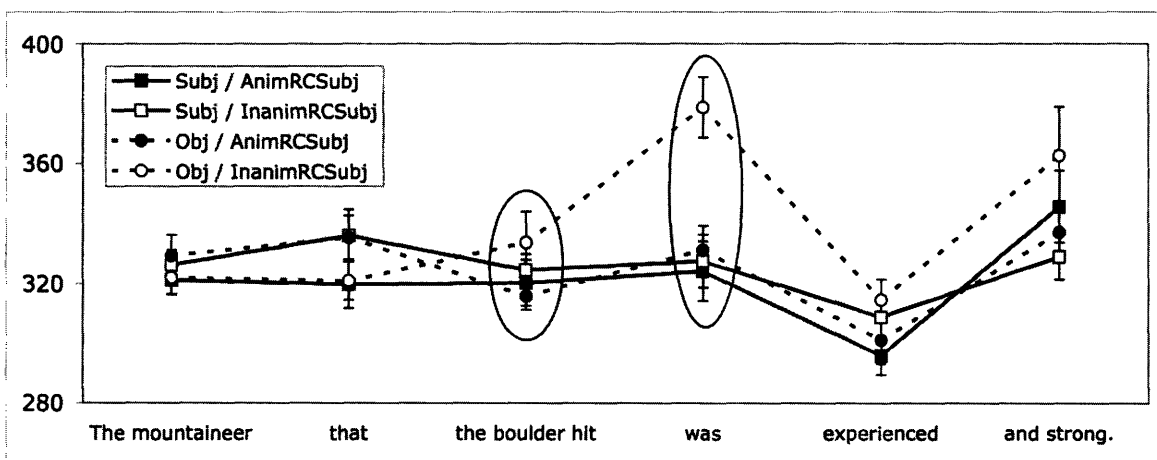


Figure 27: Reading times per region in Experiment 2A-1. Error bars indicate standard errors. The critical regions are circled.

A 2 x 2 ANOVA (subject-extracted RC / object-extracted RC, animate RC Subject / inanimate RC Subject) in the first critical region (the RC) revealed a marginal effect of RC subject animacy, such that the inanimate-RC-subject conditions were read slower than the animate-RC-subject conditions ($F(1,39)=3.97$; $MSe=3206$; $p=.053$; $F(1,19)=3.11$; $MSe=2429$; $p=.094$). There was also a trend for an interaction, such that the difference between the animate-RC-subject condition and the inanimate-RC-subject condition was larger in the object-extracted conditions than in the subject-extracted conditions ($F(1,39)=2.43$; $MSe=2694$; $p=.127$; $F(1,19)=1.31$; $MSe=902$; n.s.).

In the second critical region (the first word of the main verb phrase), a 2 x 2 ANOVA revealed two main effects and an interaction. The two main effects appear to be driven by the interaction. First, the inanimate-RC-subject conditions were read slower than the animate-RC-subject conditions ($F(1,39)=15.3$; $MSe=30238$; $p<.001$; $F(1,19)=13.8$; $MSe=21919$; $p<.002$). Second, the object-extracted conditions were read slower than the subject-extracted conditions ($F(1,39)=18.6$; $MSe=33268$; $p<.001$; $F(1,19)=5.46$; $MSe=16875$; $p<.05$). Finally, there was an interaction between the two factors, such that the inanimate-RC-subject object-extracted condition was read slower than the other three conditions ($F(1,39)=7.80$; $MSe=21131$; $p<.01$; $F(1,19)=8.05$; $MSe=9777$; $p<.02$). The pattern of results at the critical regions replicates the pattern of results reported by Traxler et al. (2002, 2005).

We will now present the results for the remaining regions. In the first region (the noun phrase), a 2 x 2 ANOVA revealed no effects ($F_s < 2$). In the second region (the relativizer), the ANOVA revealed an interaction – which did not reach significance in the items analysis – such that in the animate-RC-subject conditions the object-extracted condition was read slower than the subject-extracted condition, and in the inanimate-RC-subject conditions, the subject-extracted condition was read slower than the object-extracted condition ($F(1,39)=8.16$; $MSe=9905$; $p<.01$; $F(1,19)=2.77$; $MSe=4720$; $p=.113$). There were no other effects ($F_s < 1$). During the second word of the main verb phrase, the ANOVA revealed a main effect of RC subject animacy, such that the inanimate-RC-subject conditions were read slower than animate-RC-subject conditions ($F(1,39)=5.89$; $MSe=6361$; $p<.05$; $F(1,19)=6.23$; $MSe=3442$; $p<.05$). There were no other effects ($F_s < 1$). Finally, during the remainder of the sentence, the ANOVA revealed an interaction – which did not reach significance in the participants analysis – such that the difference between the animate-RC-subject condition and the inanimate-RC-subject condition was larger in the object-extracted conditions than in the subject-extracted conditions ($F(1,39)=3.89$; $MSe=13223$; $p=.056$; $F(1,19)=4.704$; $MSe=8809$; $p<.05$). There were no other significant effects ($F_s < 2.7$).

Discussion

In Experiment 2A-1 we successfully replicated the pattern of results reported in Traxler et al. (2002, 2005). In particular, during the critical regions of our materials we observed an interaction, such that the inanimate-RC-subject object-extracted condition was read slower than the other three conditions. The trend for this interaction emerged during the RC region, and the interaction peaked during the following region (the first word of the main verb phrase).

As discussed above, this pattern is consistent with both the Shared Resource Pool hypothesis and Traxler et al.'s Reanalysis-based hypothesis. To further evaluate the predictions of the Shared Resource Pool hypothesis Experiment 2A-2 was conducted.

Experiment 2A-2

According to the Shared Resource Pool account, retrieving the word's lexical properties and retrieving its structural dependents rely on the same pool of working memory resources. Specifically, we argue that object-extracted constructions with an inanimate RC subject are difficult because (1) retrieving the word's lexical properties (in this case, the verb's thematic properties) is costly, and (2) retrieving the word's structural dependents is costly. The fact that these two retrieval operations rely on the same resource pool results in the interaction observed by Traxler et al. (2002, 2005) and replicated here in Experiment 2A-1.

We reasoned that if the interaction between retrieving the word's lexical properties and retrieving its structural dependents is indeed responsible for the pattern of results reported in Traxler et al., then we should be able to observe a cost associated with retrieving the word's lexical properties in structures where the retrieval of structural dependents is not costly (e.g., subject-extracted RCs) if we find another way to decrease the amount of verbal working memory resources available to the comprehenders. One way to do this is by adding a concurrent digit-span task to the sentence-processing task from Experiment 2A-1. If – under the conditions of decreased verbal working memory resources – we see a cost associated with retrieving the verb's thematic properties in subject-extracted relative clauses, then it is plausible that the reason we see this cost in the object-extracted relative clauses has to do with the fact that the two retrieval operations overtax the working memory resource pool.

Methods

Participants Thirty-two participants from MIT and the surrounding community were paid for their participation. All were native speakers of English and were naive as to the purposes of the study. None participated in Experiment 2A-1.

Design and materials The experiment had the same design and used the same materials as Experiment 2A-1. The only difference was that each sentence was preceded by a string of four digits, and participants had to report the digits after the sentence.

In addition to the target sentences, the experiment included 32 sentences from an unrelated experiment and 78 fillers. The length and syntactic complexity of the sentences in the unrelated experiment and in the fillers was similar to that of the target sentences. The items in the unrelated experiment were presented with either zero or four digits, and the fillers were presented with 0-5 digits.

The stimuli were pseudo-randomized separately for each participant, with at least one item from the other experiment or one filler separating the target sentences. Each participant saw only one version of each sentence, following a Latin-Square design.

Procedure The procedure for the sentence-processing task was exactly the same as the one used in Experiment 2A-1.

The procedure for the digit-span task was, as follows. At the beginning of each trial the participants saw a randomly-generated list of four digits, presented in the form of number-words (e.g., “one”, “two”, etc.). We chose to present the digits as number-words in order to prevent the participants from processing the string of digits as two-, three- or four-digit numbers (e.g., from processing “1 4 7 2” as “fourteen seventy-two” or “one four-seventy-two”). We reasoned that this way of presenting the digits would minimize any chunking strategies the participants may try to use. The string of number-words was presented in the center of the screen for 2500 msec. Participants were instructed to try to remember the digits as well as they could. The string of number words was followed by a blank screen for 500 msec, which in turn was followed by a series of dashes marking the length and position of the words in the sentence. After the last region of the sentence, a box appeared on the screen, and the participants were instructed to type in as many of the digits that were presented at the beginning of the trial as possible in any order. They could either type them in as number-words (as they were presented) or as Arabic numerals. If the number-words/digits were typed in correctly, the word “RIGHT” flashed briefly on the screen. If two or three out of four number-words/digits were typed in correctly, the words “PARTIALLY RIGHT” flashed briefly. Finally, if one or none of the number-words/digits were typed in correctly, the word “WRONG” flashed briefly on the screen.

To assure that the participants read the sentences for meaning, a comprehension question was presented after the digit recall task, asking about the propositional content of the sentence. Participants pressed one of two keys to respond “yes” or “no”. After a correct response, the word “CORRECT” flashed briefly on the screen, and after an incorrect response, the word “INCORRECT” flashed briefly.

Before the experiment started, a short list of practice items and questions was presented in order to familiarize the participants with the task. Participants took approximately 45 minutes to complete the experiment.

Results

Due to a scripting error one condition of item #20 was not presented. Therefore, item #20 has been omitted from all the analyses.

Digit-span task The performance on the memory task was calculated using the following formula: $\text{hits} / (\text{hits} + \text{misses} + \text{false-alarms})$. This formula allowed us to give partial credit for partial responses and to penalize participants for guessing. Minor spelling mistakes in the number-words (deviations from the targets by up to 2 letters) were not taken into consideration and the number-words with such mistakes were counted as hits. Across conditions, participants performed at 95.5% correct. Table 27 presents the mean accuracies across the four conditions of the experiment. A two-factor ANOVA crossing syntactic complexity (subject-extracted RC, object-extracted RC) and RC subject animacy (animate RC subject, inanimate RC subject) revealed no significant effects. There was a marginal interaction in the items analysis, such that

the animate-RC-subject subject-extracted condition was less accurate than the other three conditions ($F(1,31)=1.56$, $MSe=.0057$, $p=.22$; $F(1,18)=4.12$, $MSe=.0030$, $p=.058$).

<u>Syntactic complexity</u>	<u>RC subject animacy</u>	
	Animate	Inanimate
Subject-extracted RC	94.2 (1.3)	96.1 (1.1)
Object-extracted RC	96.2 (1.3)	95.4 (1.4)

Table 27. Digit-span task performance in percent correct, as a function of syntactic complexity and RC subject animacy in Experiment 2A-2 (standard errors in parentheses).

Comprehension question performance Across conditions, participants answered the comprehension question correctly 91.6% of the time. Table 28 presents the mean accuracies across the four conditions. A two-factor ANOVA crossing syntactic complexity (subject-extracted RC, object-extracted RC) and RC subject animacy (animate RC subject, inanimate RC subject) revealed no effects ($F_s < 1$).

<u>Syntactic complexity</u>	<u>RC subject animacy</u>	
	Animate	Inanimate
Subject-extracted RC	90.8 (3.0)	91.6 (2.3)
Object-extracted RC	93.1 (2.5)	90.8 (2.6)

Table 28. Comprehension question performance in percent correct, as a function of syntactic complexity and RC subject animacy in Experiment 2A-2 (standard errors in parentheses).

Reaction times We analyzed all the trials, regardless of the performance on the digit-span task and regardless of whether the comprehension question was answered correctly. The statistical data patterns were very similar in the analyses of only the trials where the comprehension question was answered correctly. Reading times more than three standard deviations away from the mean for a position within condition were removed from the analyses, excluding 2% of the data.

As in Experiment 2A-1, we present the analyses of the critical regions first, followed by the analyses of the remaining regions. We defined the same critical regions as those in Experiment 2A-1: (1) the relative clause, and (2) the first word the main verb phrase. Figure 28 presents the mean reading times per region across the four conditions.

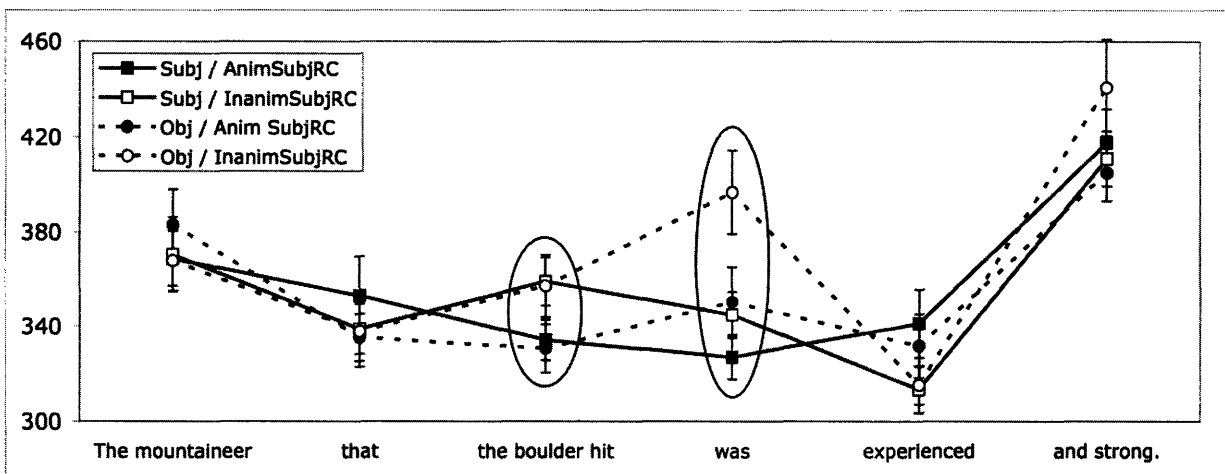


Figure 28: Reading times per region in Experiment 2A-2. Error bars indicate standard errors. The critical regions are circled.

A 2 x 2 ANOVA (subject-extracted RC / object-extracted RC, animate RC Subject / inanimate RC Subject) in the first critical region (the RC) revealed a main effect of RC subject animacy, such that the inanimate-RC-subject conditions were read slower than the animate-RC-subject conditions ($F(1,31)=17.5$, $MSe=26681$, $p<.001$; $F(1,18)=5.12$, $MSe=12392$, $p<.05$). There were no other effects ($F_s<1$).

In the second critical region (the first word of the main verb phrase), the ANOVA revealed two main effects and a trend for an interaction. First, the inanimate-RC-subject conditions were read slower than the animate-RC-subject conditions ($F(1,31)=11.2$; $MSe=38520$; $p<.005$; $F(1,18)=4.67$; $MSe=19523$; $p<.05$). Second, the object-extracted conditions were read slower than the subject-extracted conditions ($F(1,31)=9.51$; $MSe=45761$; $p<.005$; $F(1,18)=10.3$; $MSe=26890$; $p<.01$). Finally, there was a trend for an interaction between the two factors, such that the difference between the animate-RC-subject condition and the inanimate-RC-subject condition was larger in the object-extracted conditions than in the subject-extracted conditions ($F(1,31)=2.49$; $MSe=13317$; $p=.125$; $F(1,18)=1.05$; $MSe=3879$; n.s.).

We will now present the analyses for the remaining regions. In the first region (the noun phrase), a 2 x 2 ANOVA revealed no effects ($F_s<1.1$). Similarly, in the second region (the relativizer), the ANOVA revealed no effects ($F_s<1$). During the second word of the main verb phrase, the ANOVA revealed a main effect of RC subject animacy – which did not reach significance in the items analysis – such that the inanimate-RC-subject conditions were read slower than the animate-RC-subject conditions ($F(1,31)=9.84$; $MSe=15527$; $p<.005$; $F(1,18)=3.20$; $MSe=9338$; $p=.091$). There were no other effects ($F_s<1$). Finally, during the remainder of the sentence, the ANOVA revealed an interaction in the participants analysis, such

that the inanimate-RC-subject object-extracted condition was read slower than the other three conditions ($F(1,31)=4.57$; $MSe=14399$; $p<.05$; $F(1,18)=1.75$; $MSe=8622$; n.s.). There were no other effects ($F_s<2.5$).

At the first critical region (the RC region) we observed a similar pattern of results to Experiment 2A-1 for the *object-extracted conditions*, such that the inanimate-RC-subject condition was slower than the animate-RC-subject condition. In contrast to Experiment 2A-2, however, the pattern of results was different for the *subject-extracted conditions*: while in Experiment 2A-1 there was no difference between the animate-RC-subject and the inanimate-RC-subject conditions, in the current experiment the inanimate-RC-subject condition was slower than the animate-RC-subject condition (see Figure 29). Pair-wise comparisons support this interpretation. In Experiment 2A-1 $F_s<1$. In Experiment 2A-2, however, the difference between the inanimate-RC-subject subject-extracted condition and the animate-RC-subject subject-extracted condition is significant in the participants analysis and marginal in the items analysis ($F(1,31)=5.46$, $MSe=9762$, $p<.05$; $F(1,18)=3.08$, $MSe=5790$, $p=.096$).

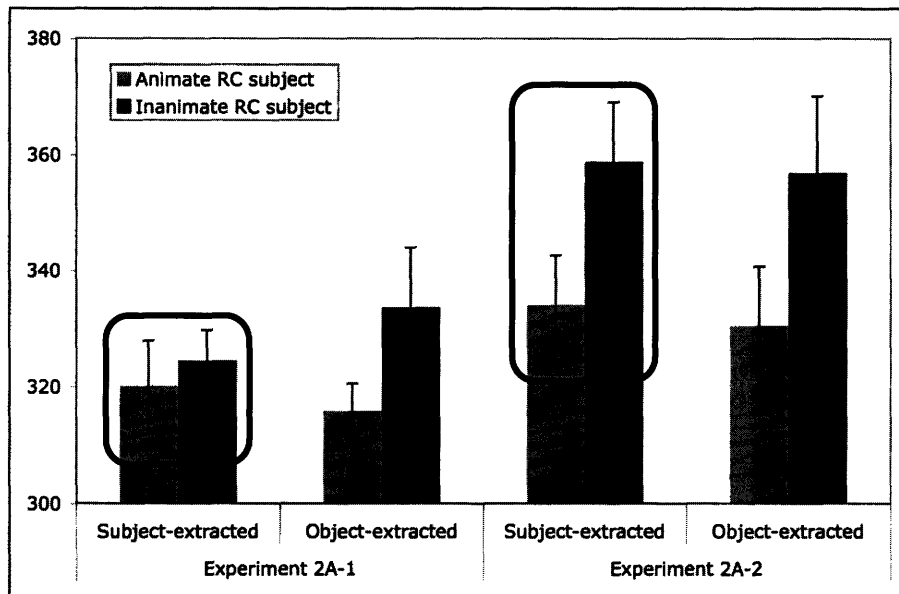


Figure 29: Reading times at the relative clause region in Experiments 2A-1 and 2A-2. Error bars indicate standard errors. The critical difference between the two experiments is circled.

Discussion

In Experiment 2A-2 – where the materials from Experiment 2A-1 were presented with a concurrent digit-span task – we observed a cost associated with retrieving the lexical properties of the word (the embedded verb’s thematic properties) in structures where retrieving the word’s structural dependents is associated with little/no cost (subject-extracted RCs). Specifically, the inanimate-RC-subject condition (*The boulder that hit the mountaineer...*) was processed slower than the animate-RC-subject condition (*The mountaineer that moved the boulder...*). This pattern of results supports the Shared Resource Pool hypothesis, according to which retrieving

the word's lexical properties and retrieving its structural dependents is performed by the same mechanism. Without the concurrent digit-span task, the cost associated with retrieving the lexical properties of the word is only observed in structures where retrieving the word's structural dependents is costly (object-extracted RCs). When the digit-span task is added – decreasing the amount of verbal working memory resources available to the comprehenders – the cost associated with retrieving the lexical properties of the word is also observed in structures where retrieving the word's structural dependents is not costly (subject-extracted RCs).

Whereas the pattern of results in Experiment 2A-2 is not predicted by Traxler et al.'s Reanalysis-based hypothesis, it is important to directly test the predictions of the Reanalysis-based hypothesis in order to evaluate it.

Experiment 2A-3

As discussed in the Background section, in order to explain the interaction between syntactic complexity and the animacy of the RC subject, Traxler et al. (2002) argued that the reanalysis cost is higher when the RC subject is inanimate and thus a poor agent. In the original materials, however, the animacy of the RC subject always co-varied with the animacy of the RC object, such that one was always animate and one was always inanimate, making it difficult to evaluate the claim about higher reanalysis cost for inanimate RC subjects because the materials did not form minimal pairs. For example, in comparing object-extracted conditions with an inanimate RC subject, like *The director that the movie...*, and conditions with an animate RC subject, like *The movie that the director...*, not only are RC subjects different in terms of animacy (*the movie* and *the director*, respectively), but also the RC objects (*the director* and *the movie*, respectively). Experiment 2A-3 was designed to address this issue and to evaluate the claim that the reanalysis cost is higher when the RC subject is inanimate.

Methods

Participants Seventy-two participants from MIT and the surrounding community were paid for their participation. All were native speakers of English and were naive as to the purposes of the study. None participated in Experiments 2A-1 or 2A-2.

Design and materials The experiment had a 2 x 2 design, crossing syntactic complexity (subject-extracted RCs, object-extracted RCs) and the animacy of the RC subject (animate, inanimate). Critically, the RC object was always animate.

The materials consisted of 24 sets of sentences with four different versions as in (14):

(14a) *Animate RC subject / Subject-extracted:*

The boy | that | upset the girl | during the class | was | often | quite mean.

(14b) *Animate RC subject / Object-extracted:*

The girl | that | the boy upset | during the class | cried | for | an hour.

(14c) *Inanimate RC subject / Subject-extracted:*

The grade | that | upset the girl | during the class | was | not | very important.

(14d) *Inanimate RC subject / Object-extracted*:

The girl | that | the grade upset | during the class | cried | for | an hour.

In addition to the target sentences, the experiment included 24 sentences from an unrelated experiment and 48 fillers. The length and syntactic complexity of the sentences in the unrelated experiment and in the fillers was similar to that of the target sentences. The stimuli were pseudo-randomized separately for each participant, with at least one item from another experiment or one filler separating the target sentences. Each participant saw only one version of each sentence, following a Latin-Square design (see Appendix J for a complete list of linguistic materials).

The two conditions of interest are the object-extracted conditions ((14b) and (14d)), where the sentence-initial NP is animate (*the girl* in (14)), and the RC subject is either animate (*the boy*) or inanimate (*the grade*). Traxler et al.'s Reanalysis-based account predicts that the condition with an inanimate RC subject (condition (14d)) should be more difficult to process than the condition with an animate RC subject (condition (14b)) because it should be easier to reanalyze the object-extracted condition when the RC subject is animate and thus a good agent, compared to when it's inanimate and thus a poor agent.

There is a class of resource-based accounts of linguistic complexity that make the opposite prediction. Specifically, accounts that argue that distance in non-local structural dependencies should be measured in terms of similar intervening elements (e.g., Lewis, 1996; Gordon et al., 2001, 2004; Lewis et al., 2006) predict that the condition where the RC subject is animate (condition (14b)) should be more difficult to process than the condition where the RC subject is inanimate (condition (14d)) because the two noun phrases in condition (14b) are more similar to each other (both are animate: *the girl* and *the boy*).

In summary, Traxler et al. (2002) Reanalysis-based hypothesis and a variety of similarity-based interference accounts make opposite predictions with regard to the relative difficulty of the two object-extracted conditions.

Because we are interested in comparing across different sets of nouns (which is inevitable for a direct comparison between animate and inanimate agents), we normed the sentences with the two sets of nouns for plausibility. A group of 20 participants (none of whom participated in Experiments 2A-1, 2A-2 or 2A-3) were presented with a series of statements and were asked to evaluate them based on the naturalness of the events they describe in the real world on a scale from 1 (Unnatural) to 7 (Natural). The statements were created from the experimental materials in the following way: the animate NP (e.g., *the boy* in (14)) or the inanimate NP (e.g., *the grade* in (14)) served as the subject of the sentence and the other NP from the materials (e.g., *the girl* in (14)) served as the object of the sentence. The verb and the post-verbal prepositional phrase were the same as those used in the experimental materials. For example, based on (14), the following two statements were formed:

(15a) *Animate*: The boy upset the girl during the class.

(15b) *Inanimate*: The grade upset the girl during the class.

Each participant saw only one version of each item, following a Latin-Square design, and the items were randomized for each participant. The results revealed that the Animate and the Inanimate statements were rated similarly: Animate – 5.7, Inanimate – 5.63. A t-test revealed no significant difference ($p=.63$).

Procedure The procedure was identical to that used in Experiment 2A-1 [in the future experiments, this procedure will be referred to as ‘the standard self-paced reading procedure’]. Each experimental sentence was structured, as follows: (1) a definite noun phrase, (2) a relativizer *that*, (3) a subject-/object-extracted relative clause, (4) a three-word long prepositional phrase (which was included in order to be able to evaluate the effects on the main verb region without concerns about possible spill-over from the RC region), and (5) the main verb phrase which consisted of 4 words. As shown in (14), the main verb phrase was divided into three regions for the purposes of the analyses.

Results

Comprehension question performance Across conditions, participants answered the comprehension question correctly 93.7% of the time. Table 29 presents the mean accuracies across the four conditions. A two-factor ANOVA crossing syntactic complexity (subject-extracted RC, object-extracted RC) and the animacy of the RC subject (animate, inanimate) on the responses to the comprehension questions revealed a main effect of syntactic complexity, such that participants were less accurate in the object-extracted conditions ($F(1,71)=29.5$; $MSe=2509$; $p<.001$; $F(1,23)=8.76$; $MSe=836$; $p<.01$). There were no other effects ($F_s<1.8$).

<u>Syntactic complexity</u>	<u>RC subject animacy</u>	
	Animate	Inanimate
Subject-extracted RC	97.7 (.68)	95.6 (1.1)
Object-extracted RC	90.3 (1.4)	91.2 (1.4)

Table 29. Comprehension accuracies in percent correct, as a function of syntactic complexity and RC subject animacy in Experiment 2A-3 (standard errors in parentheses).

Reaction times We analyzed all the trials, regardless of whether the comprehension question was answered correctly. The statistical data patterns were very similar in the analysis of only the trials where the comprehension question was answered correctly. Reading times more than three standard deviations away from the mean for a position within condition were removed from the analyses, excluding 1.5% of the data.

We present the analyses of the critical regions first, followed by the analyses of the remaining regions. As in Experiments 2A-1 and 2A-2, we defined two critical regions: (1) the relative clause, and (2) the first word of the main verb phrase. Figure 30 presents the mean reading times per region across the four conditions.

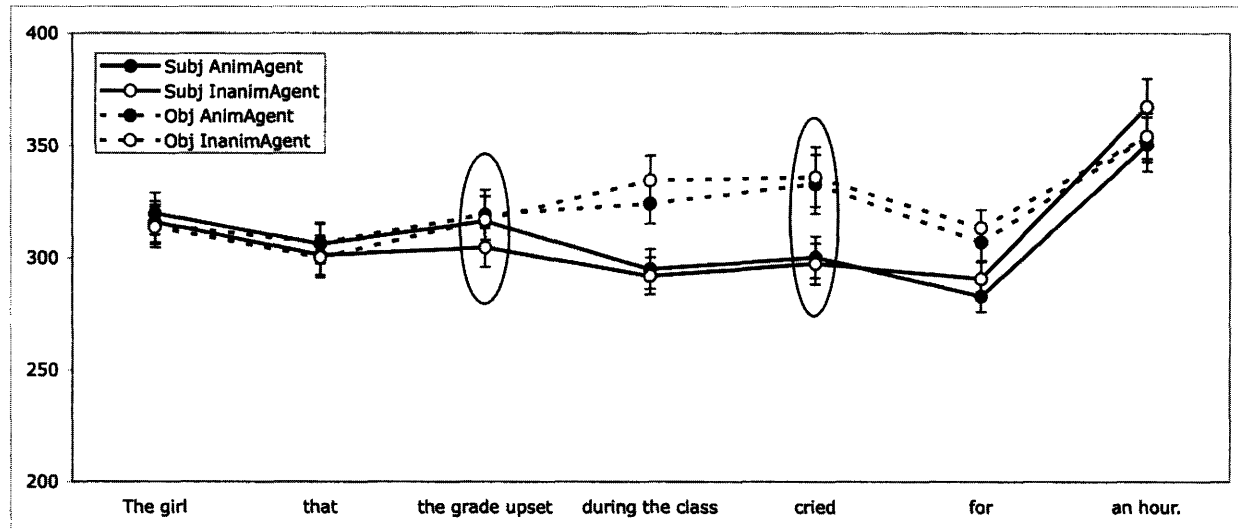


Figure 30: Reading times per region in Experiment 2A-3. Error bars indicate standard errors. The critical regions are circled.

A 2 x 2 ANOVA (subject-extracted RC / object-extracted RC, animate RC Subject / inanimate RC Subject) in the first critical region (the RC) revealed a main effect of extraction in the participants analysis, such that the object-extracted conditions were read slower than the subject-extracted conditions ($F(1,71)=5.61$, $MSe=4061$, $p<.05$; $F(1,23)=2.89$, $MSe=1069$, $p=.102$). It also revealed a marginal main effect of RC subject animacy in the participants analysis, such that the inanimate-RC-subject conditions were read slower than the animate-RC-subject conditions ($F(1,71)=3.78$, $MSe=3617$, $p=.056$; $F(1,23)=1.14$, $MSe=1123$, n.s.). There was also a non-significant trend for an interaction in the participants analysis, such that the inanimate-RC-subject subject-extracted condition was read faster than the other three conditions ($F(1,71)=2.45$, $MSe=1634$, $p=.122$; $F(1,23)<1$, n.s.).

Because the pattern of the reading times looking at the whole RC region did not distinguish between the two alternative accounts (the Shared Resource Pool hypothesis and Traxler et al.'s Reanalysis-based hypothesis), we examined the reading times at the RC region for the two object-extracted conditions (*The girl that the boy upset...* and *The girl that the grade upset...*) in more detail. Specifically, we compared the reading times at the embedded noun (*boy vs. girl*) and at the embedded verb (*upset*). At the embedded noun, the pair-wise comparison revealed an effect of RC subject animacy which did not quite reach significance in the items analysis, such that the animate-RC-subject condition was read slower than the inanimate-RC-subject condition ($F(1,71)=5.40$, $MSe=9828$, $p<.05$; $F(1,23)=3.71$, $MSe=3500$, $p=.067$). At the embedded verb, there was no difference between the two conditions ($F_s<1$). Because at the embedded noun region the comparison is made between two different sets of lexical items, we also compared the

reading times for the same sets of nouns in the subject-extracted conditions where these nouns appear sentence-initially (e.g., *The boy that...* and *The grade that...*). The pair-wise comparison revealed no difference between the two conditions ($F_s < 1.4$). See Figure 31.

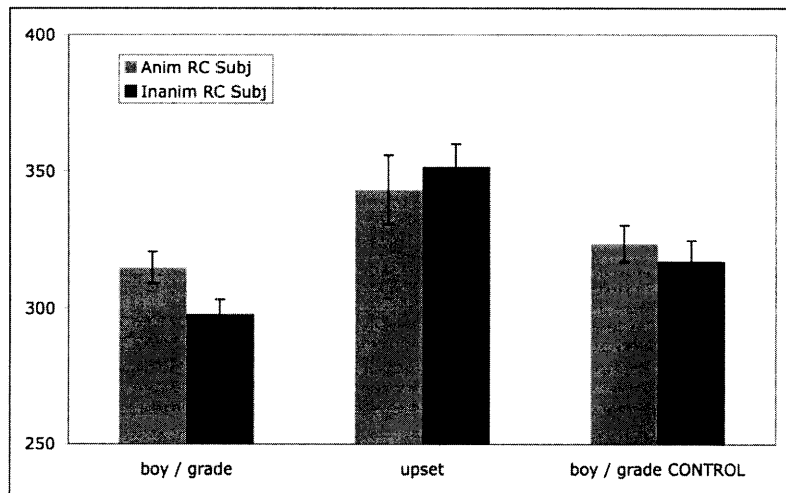


Figure 31: The first two pairs of bars demonstrate reading times for the object-extracted conditions at the embedded noun and embedded verb. Error bars indicate standard errors. The last pair of bars demonstrates reading times for the sentence-initial noun in the subject-extracted conditions, for control purposes.

The fact that the animate-RC-subject condition is processed slower than the inanimate-RC-subject condition at the embedded noun is consistent with similarity-based accounts, but not with Traxler et al.'s Reanalysis-based account, which predicts that the inanimate RC subject condition should be more difficult to process. The reading times at the embedded verb show that both object-extracted conditions are equally difficult to process at this point in the sentence.

In the second critical region (the first word of the main verb phrase), a 2 x 2 ANOVA revealed a main effect of syntactic complexity, such that the object-extracted conditions were processed slower than the subject-extracted conditions ($F(1,71)=33.9$, $MSe=91219$, $p < .001$; $F(1,23)=40.64$, $MSe=28685$, $p < .001$). There were no other effects ($F_s < 1$).

In the first region (the noun phrase), the ANOVA revealed no effects ($F_s < 1.7$). Similarly, in the second region (the relativizer *that*), the ANOVA revealed no effects ($F_s < 2.2$). In the post-RC region (the prepositional phrase), the ANOVA revealed a main effect of syntactic complexity, such that the object-extracted conditions were read slower than the subject-extracted conditions ($F(1,71)=78.1$, $MSe=92466$, $p < .001$; $F(1,23)=56.3$, $MSe=30760$, $p < .001$). The ANOVA also revealed an interaction in the participants analysis which did not reach significance in the items analysis, such that the difference between the animate-RC-subject condition and the inanimate-RC-subject condition was larger in the object-extracted conditions than in the subject-extracted conditions ($F(1,71)=5.045$, $MSe=3160$, $p < .05$; $F(1,23)=3.18$, $MSe=998$, $p = .088$). In the region following the main verb, the ANOVA revealed a main effect of syntactic complexity, such that the object-extracted conditions were processed slower than the subject-extracted conditions ($F(1,71)=36.6$, $MSe=39746$, $p < .001$; $F(1,23)=46.4$, $MSe=13560$, $p < .001$). The

ANOVA also revealed a main effect of RC subject animacy in the participants analysis which did not reach significance in the items analysis, such that the inanimate-RC-subject conditions were processed slower than the animate-RC-subject conditions ($F(1,71)=4.46$, $MSe=3761$, $p<.05$; $F(1,23)=2.63$, $MSe=1387$, n.s.). Finally, during the remainder of the sentence, the ANOVA revealed a main effect of RC subject animacy in the participants analysis, such that the inanimate-RC-subject conditions were processed slower than the animate RC subject conditions ($F(1,71)=4.18$, $MSe=5311$, $p<.05$; $F(1,23)=1.56$, $MSe=1436$, n.s.). The ANOVA also revealed a marginal interaction, such that the difference between the animate-RC-subject condition and the inanimate-RC-subject condition was larger in the subject-extracted conditions than in the object-extracted conditions ($F(1,39)=3.23$; $MSe=4806$; $p=.077$; $F(1,19)=3.41$; $MSe=1747$; $p=.078$).

Discussion

This experiment was aimed at directly testing the predictions of Traxler et al.'s Reanalysis-based hypothesis against the predictions of similarity-based interference accounts. While in the analysis of the whole RC region the two critical object-extracted conditions looked equally difficult, a more fine-grained analysis of the RC region revealed that at the embedded noun the animate-RC-subject condition was processed slower than the inanimate-RC-subject condition. This pattern is consistent with similarity-based accounts and inconsistent with Traxler et al.'s Reanalysis-based account.

It is worth reiterating that the right way to measure distance in non-local dependencies is still an open question. The results of Experiment 2A-3 (as well as e.g., Gordon et al., 2001, 2004; Fedorenko et al., 2006) provide some suggestive evidence for measuring distance in terms of similar intervening elements. However, as briefly discussed in the Introduction, there exists evidence for measuring distance in terms of linear distance (as suggested in Gibson's original 1998 decay-based proposal) (e.g., Grodner & Gibson, 2005; Levy, Fedorenko & Gibson, 2007). Additional research will be necessary in order to develop a distance metric that will account for the different results in the literature.

Conclusions for Experiments 2A-1 through 2A-3

In Experiments 2A-1 – 2A-3, we investigated how retrieving the verb's thematic properties – which we assumed include information about the verb's typical agents and patients – affects the difficulty of retrieving the word's structural dependents. We argued that the Shared Resource Pool hypothesis accounts for a recent finding by Traxler et al. (2002). We comparatively evaluated the Shared Resource Pool hypothesis and Traxler et al.'s Reanalysis-based hypothesis. According to the Reanalysis-based hypothesis, (1) reanalysis is required in object-extracted relative clauses (because object-extracted RCs are initially interpreted as subject-extracted RCs), and (2) the reanalysis process is more difficult when the RC subject is inanimate and thus a poor agent. According to the Shared Resource Pool hypothesis, retrieving the word's lexical properties and retrieving its structural dependents rely on the same pool of working memory resources. Specifically, according to the Shared Resource Pool hypothesis, object-extracted constructions with an inanimate RC subject are difficult because (1) retrieving the word's lexical

properties (in this case, the thematic properties of the verb) is costly, and (2) retrieving the word's structural dependents is costly. The fact that these two retrieval operations rely on the same pool of working memory resources results in the interaction observed by Traxler et al. (2002, 2005) and replicated here in Experiment 2A-1.

Experiment 2A-2 demonstrated that when the amount of verbal working memory resources available to comprehenders is decreased (by adding a concurrent digit-span task) the difficulty of retrieving the thematic properties of the verb is observed in structures where retrieving the word's structural dependents is associated with little/no cost. This pattern of results supports the Shared Resource Pool hypothesis, according to which the difficulty of retrieving the word's lexical properties interacts with the difficulty of retrieving its structural dependents to produce superadditive processing difficulty because the two retrieval operations overtax the working memory resource pool.

Experiment 2A-3 tested the predictions of Traxler et al.'s Reanalysis-based hypothesis against the predictions of similarity-based interference accounts. The pattern of results was inconsistent with Traxler et al.'s account. Therefore, we conclude that the Shared Resource Pool hypothesis provides a better account of the interaction between syntactic complexity and the animacy of the RC subject.

Background for Experiment 2A-4

In the previous section we examined the relationship between the difficulty of retrieving the thematic properties of a verb – which we assumed include information about typical agents and patients – and the difficulty of retrieving the word's structural dependents. We showed that the two retrieval operations interact to result in superadditive processing difficulty. We argued that this pattern of results provides support for the Shared Resource Pool hypothesis, whereby the two retrieval operations rely on the same pool of working memory resources.

In this section we will examine the relationship between the difficulty of retrieving a different aspect of the word's lexical properties (specifically retrieving the meaning of a word) and retrieving the word's structural dependents. In particular, we will investigate whether lexical frequency interacts with the difficulty of retrieving the word's structural dependents in a similar way to the difficulty of retrieving the word's thematic properties.

Experiment 2A-4: The effects of the difficulty of retrieving the word's meaning on the difficulty of retrieving its structural dependents *[Fedorenko, Gibson, & Jaramillo, in preparation]*

Experiment 2A-4 was designed to investigate the relationship between the difficulty of retrieving the meaning of a word and retrieving its structural dependents. The difficulty of retrieving the word's structural dependents was manipulated by using subject- and object-extracted constructions, as in Experiments 2A-1 – 2A-3. The difficulty of retrieving the meaning of a word was manipulated by varying the embedded verb's lexical frequency. (Lexical frequency

has been shown to affect processing difficulty at both the word level (e.g., Preston, 1935; Howes & Solomon, 1951; Rubenstein et al., 1970; Forster & Chambers, 1973) and the sentence level (e.g., Seidenberg & MacDonald, 1999; Tabor et al., 1997; Gibson 2006).

The Shared Resource Pool hypothesis predicts an interaction between the difficulty of retrieving the word's meaning and the difficulty of retrieving its structural dependents. In particular, as discussed at the beginning of Chapter 2, the Shared Resource Pool hypothesis predicts that increasing the cost associated with retrieving the word's lexical properties will lead to an increase in the difficulty of retrieving the word's structural dependents (leading to superadditive processing difficulty) up to some threshold level of difficulty of retrieving the word's lexical properties. When the cost associated with retrieving the word's lexical properties reaches a certain threshold, however, retrieving the word's structural dependents may be partially or completely delayed until a later point in the sentence. This will translate into (1) a decrease in or an elimination of the difficulty associated with retrieving the word's structural dependents during the processing of the target incoming word, and (2) difficulty associated with retrieving the word's structural dependents following the target incoming word.

In Experiment 2A-1 (where the retrieval of the verb's thematic properties – which include information about the typical agents and patients of the verb – was made difficult by presenting the verb in the context of an atypical agent) we observed superadditive processing difficulty in the condition where the two retrieval operations were difficult. In the case of the lexical frequency manipulation, the Shared Resource Pool hypothesis predicts either a similar pattern of results (an increase in the difficulty of retrieving the word's structural dependents), or a delay of the cost associated with retrieving the word's structural dependents. In contrast, the Separate Resource Pools hypothesis predicts two independent effects for the difficulty of the two retrieval operations.

Methods

Participants Forty-eight participants from MIT and the surrounding community were paid for their participation. All were native speakers of English and were naive as to the purposes of the study.

Design and materials The experiment had a 2 x 2 design, crossing syntactic complexity (subject-extracted clefts, object-extracted clefts) and the frequency of the embedded verb (high-frequency, low-frequency).

The materials consisted of 24 sets of sentences with four different versions as in (16):

(16a) *High Frequency / Subject-extracted:*

It was | Vivian | who | lectured Terrence | for always | being late.

(16b) *High Frequency / Object-extracted:*

It was | Vivian | who | Terrence lectured | for always | being late.

(16c) *Low Frequency / Subject-extracted:*

It was | Vivian | who | chided Terrence | for always | being late.

(16d) *Low Frequency / Object-extracted:*

It was | Vivian | who | Terrence chided | for always | being late.

We will now go through the four conditions and discuss the predictions of the Shared Resource Pool hypothesis. While it is difficult (and maybe not possible) to compare the relative difficulties of the two manipulations – the difficulty of retrieving the verb’s thematic properties (Experiment 2A-1) and the difficulty of retrieving the verb’s meaning (Experiment 2A-4) – we hypothesize that the lexical frequency manipulation may be more difficult because the low-frequency verbs were chosen from the very low end of the frequency spectrum.

First, in the high-frequency subject-extracted condition (*It was Vivian who lectured Terrence...*), retrieving the structural dependent of the verb *lectured* (the relativizer *who* co-indexed with the noun phrase *Vivian*) is associated with little/no cost because the integration is local. Retrieving the meaning of the verb *lectured* is also not costly because it is a high-frequency verb. As a result, resources are abundant and no processing difficulty should ensue (see Figure 32a).

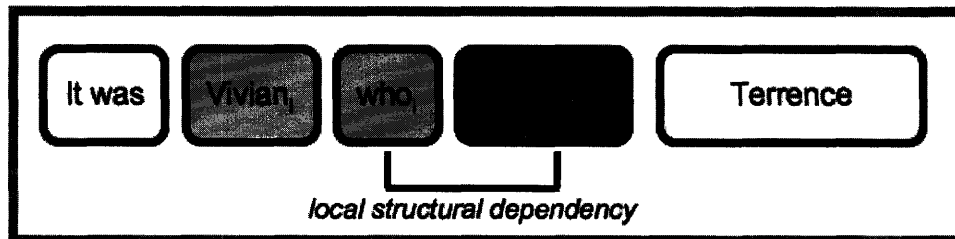


Figure 32a: The high-frequency subject-extracted condition.

Second, in the high-frequency object-extracted condition (*It was Vivian who Terrence lectured...*), retrieving the structural dependent of the verb *lectured* (the relativizer co-indexed with the noun phrase *Vivian*) is now costly because the integration is non-local and thus the relativizer has to be retrieved from memory. Retrieving the meaning of the verb *lectured* is not costly – as in the high-frequency subject-extracted condition – because it is a high frequency verb. Because one of the retrieval operations (retrieving the word’s lexical properties) is not costly, resources are abundant and only some processing difficulty should ensue (see Figure 32b).

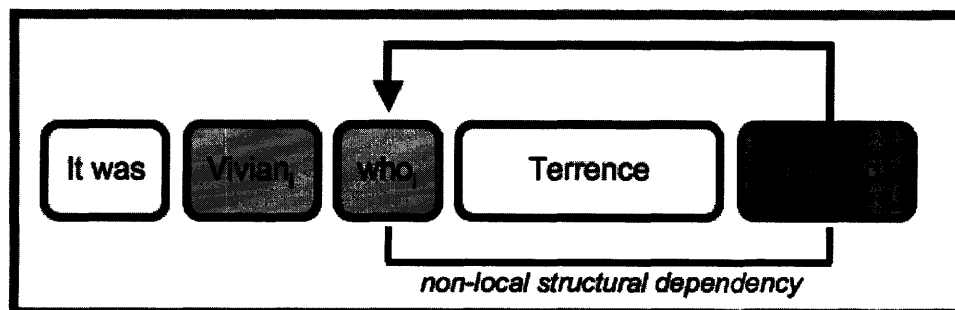


Figure 32b: The high-frequency object-extracted condition.

Third, in the low-frequency subject-extracted condition (*It was Vivian who chided Terrence...*), retrieving the structural dependent of the verb *chided* (the relativizer co-indexed with the noun phrase *Vivian*) is associated with little/no cost because the integration is local, as in the high-frequency subject-extracted condition. However, retrieving the meaning of the verb *chided* is costly because it is a low-frequency verb. Because one of the retrieval operations (retrieving the word's structural dependents) is not costly, resources are abundant and only some processing difficulty should ensue (see Figure 32c).

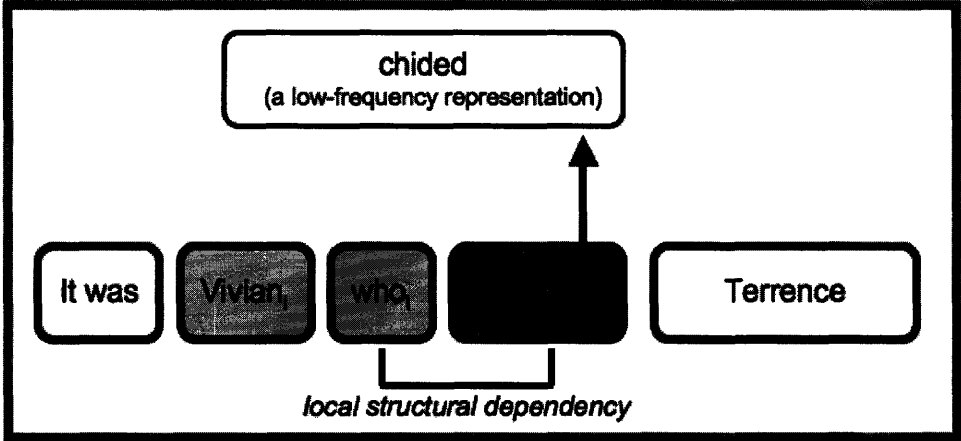


Figure 32c: The low-frequency subject-extracted condition.

Finally, in the low-frequency object-extracted condition (*It was Vivian who Terrence chided...*), retrieving the structural dependent of the verb *chided* (the relativizer co-indexed with the noun phrase *Vivian*) is costly because the integration is non-local and thus the relativizer has to be retrieved from memory. Furthermore, retrieving the meaning of the verb *chided* is also costly because it is a low-frequency verb. Because both of the retrieval operations are costly, resources are scarce and processing difficulty should ensue (see Figure 32d).

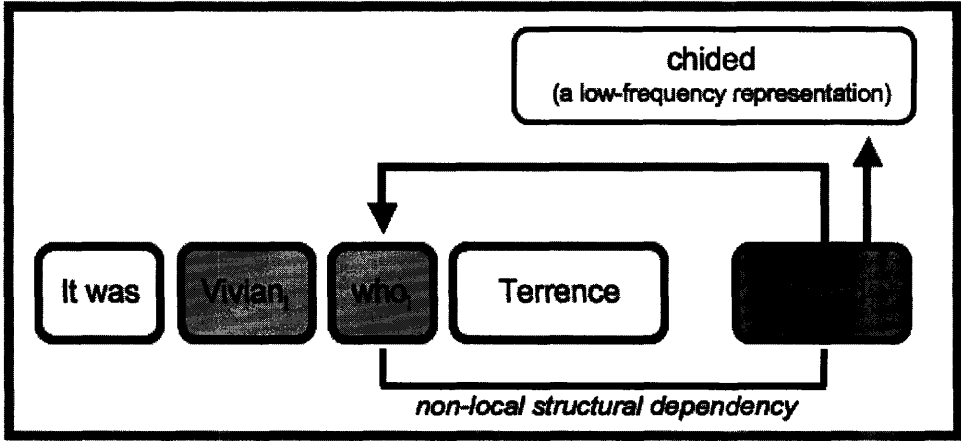


Figure 32d: The low-frequency object-extracted condition.

Clefts were used instead of relative clauses in order to allow the use of personal names instead of definite noun phrases. We reasoned that using personal names would be preferred in this experiment because the frequency of the definite noun phrases may potentially interact in some way with the frequency of the embedded verbs. Another advantage of using personal names is that no plausibility control is needed between the subject- and object-extracted conditions. The gender of the names was balanced: a quarter of the items (6 out of 24) had two male names, a quarter had two female names, a quarter had a male and a female name with the male name in the clefted position, and a quarter had a male and a female name with the female name in the clefted position.

The high- and low-frequency verb pairs were constructed such that the meanings of the verbs in each pair was as similar as possible. This was done in order to minimize the meaning differences between the high- and low-frequency-verb conditions.

Lexical frequencies were estimated using the Google search engine. Specifically, the past-tense forms of the verbs (e.g., lectured / chided) were entered as search terms. The high-frequency verbs had the average frequency of 90,739 thousand and the low-frequency verbs had the average frequency of 1,751 thousand. The t-test was significant ($p < .01$). The average ratio of high-frequency verbs to low-frequency verbs was 67, ranging from 5 to 573. The two groups of verbs were matched for length in number of letters ($p = .83$) and number of syllables ($p = .49$). See Appendix K for the detailed information on the two groups of verbs.

In addition to the target sentences, the experiment included 48 fillers. The length and syntactic complexity of the fillers was similar to that of the target sentences. The fillers were constructed to involve personal names to make them similar to the target sentences. The stimuli were pseudo-randomized separately for each participant, with at least one item from another experiment or one filler separating the target sentences. Each participant saw only one version of each sentence, following a Latin-Square design (see Appendix L for a complete list of linguistic materials).

Procedure The standard self-paced reading procedure was used. Each experimental sentence was structured, as follows: (1) the cleft structure *it was*, (2) a personal name, (3) a relativizer *who*, (4) a subject-/object-extracted clause, and (5) the ending consisting of four words. As shown in (16), the ending was divided into two regions for the purposes of the analyses.

Participants took approximately 20 minutes to complete the experiment.

Results

Comprehension question performance Across the conditions, participants answered the comprehension question correctly 80.5% of the time. Table 30 presents the mean accuracies across the four conditions. A two-factor ANOVA crossing syntactic complexity (subject-extracted cleft, object-extracted cleft) and the frequency of the verb (high, low) on the responses to the comprehension questions revealed a main effect of syntactic complexity, such that participants were less accurate in the object-extracted conditions ($F(1,47)=7.43$, $MSe=5346$, $p<.01$; $F(1,23)=19.6$, $MSe=2692$, $p<.001$). There were no other effects ($F_s<1$).

<u>Syntactic complexity</u>	<u>Verb frequency</u>	
	High	Low
Subject-extracted cleft	84.3 (3.0)	87.2 (2.8)
Object-extracted cleft	76.1 (3.4)	74.3 (3.9)

Table 30. Comprehension accuracies in percent correct, as a function of syntactic complexity and verb frequency in Experiment 2A-4 (standard errors in parentheses).

Reaction times We analyzed all the trials, regardless of whether the comprehension question was answered correctly. The statistical data patterns were very similar in the analysis of only the trials where the comprehension question was answered correctly. Reading times more than three standard deviations away from the mean for a position within condition were removed from the analyses, excluding 2.1% of the data.

We present the analysis of the critical region first, followed by the analyses of the other regions. We defined the critical region as the subject-/object-extracted cleft clause. Figure 33 presents the mean reading times per region across the four conditions.

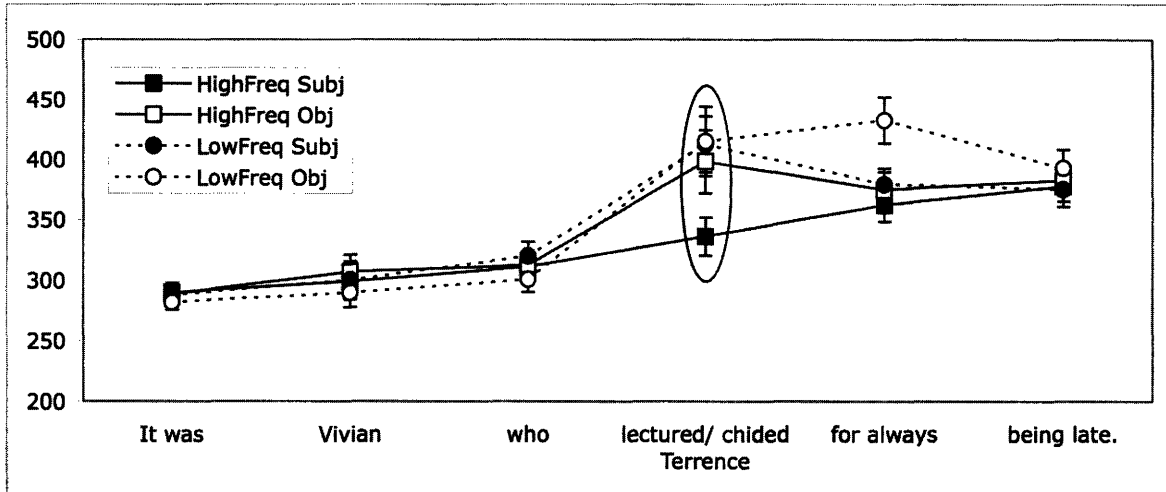


Figure 33: Reading times per region in Experiment 2A-4. Error bars indicate standard errors. The critical region is circled.

A 2 x 2 ANOVA (subject-extracted cleft / object-extracted cleft, high frequency verb / low frequency verb) in the critical region revealed two main effects and an interaction. The two main effects appear to be driven by the interaction. First, the object-extracted cleft conditions were read slower than the subject-extracted cleft conditions ($F(1,47)=5.13$, $MSe=50000$, $p<.05$; $F(1,23)=5.38$, $MSe=22044$, $p<.05$). Second, the low-frequency verb conditions were read slower than the high-frequency verb conditions ($F(1,47)=30.22$, $MSe=104722$, $p<.001$; $F(1,23)=28.9$, $MSe=54646$, $p<.001$). Finally, there was an interaction between the two factors, such that the difference between the object- and the subject-extracted conditions was larger in the high-frequency conditions than in the low-frequency conditions ($F(1,47)=12.1$, $MSe=42936$, $p<.002$; $F(1,23)=14.3$, $MSe=23394$, $p<.002$). In other words, while we observe the standard subject-/object-extraction difference in the high-frequency conditions, we don't observe this difference in the low-frequency conditions. Both of the low-frequency conditions appear to be comparable in difficulty to the object-extracted high-frequency condition. This pattern of results is different from that observed in Experiment 2A-1 where we manipulated the difficulty of retrieving the verb's thematic properties. We will discuss this below.

We will now present the results for the remaining regions. In the first region (*it was*), a 2 x 2 ANOVA revealed a marginal unpredicted effect of verb frequency in the participants analysis, such that low-frequency conditions were read slower than high-frequency conditions ($F(1,47)=3.40$, $MSe=925$, $p=.072$; $F(1,23)=2.81$, $MSe=515$, n.s.). We do not expect any differences among the conditions in this region, because the materials are exactly the same. In the second region (the personal name) the ANOVA revealed an unpredicted marginal interaction in the participants analysis, such that the object-extracted condition was read slower than the subject-extracted condition in the high-frequency conditions, but it was read faster than the subject-extracted condition in the low-frequency conditions ($F(1,47)=3.98$, $MSe=4028$, $p=.052$; $F(1,23)=2.901$, $MSe=1948$, n.s.). As in the first region, we do not expect any differences among the conditions in this region, because the materials are exactly the same. In the third region (the relativizer), the ANOVA revealed an unpredicted marginal effect of syntactic

complexity in the participants analysis, such that the subject-extracted conditions were read slower than the object-extracted conditions ($F(1,47)=2.903$, $MSe=3853$, $p=.095$; $F(1,23)<1$, $MSe=1657$, n.s.), and an unpredicted interaction in the items analysis – which was marginal in the participants analysis – such that while there was no difference between the subject- and the object-extracted conditions in the high-frequency conditions, in the low-frequency conditions the subject-extracted condition was read slower than the object-extracted condition ($F(1,47)=3.58$, $MSe=5113$, $p=.065$; $F(1,23)=12.8$, $MSe=3419$, $p<.005$). As in the first two regions, we do not expect any differences among the conditions in this region, because the materials are exactly the same.

In the region following the subject-/object-extracted clause, we observed two main effects and an interaction. The two main effects appear to be driven by the interaction. First, the object-extracted cleft conditions were read slower than the subject-extracted cleft conditions ($F(1,47)=8.94$, $MSe=52178$, $p<.005$; $F(1,23)=4.99$, $MSe=25035$, $p<.05$). Second, the low-frequency verb conditions were read slower than the high-frequency verb conditions ($F(1,47)=17.4$, $MSe=67638$, $p<.001$; $F(1,23)=15.5$, $MSe=30571$, $p<.002$). Finally, there was an interaction between the two factors – which didn't quite reach significance in the items analysis – such that the difference between the object- and the subject-extracted conditions was larger in the low-frequency conditions than in the high-frequency conditions ($F(1,47)=4.94$, $MSe=20120$, $p<.05$; $F(1,23)=3.77$, $MSe=9438$, $p=.064$). This interaction is the reverse of the interaction observed during the subject-/object-extracted clause where the difference between the object- and the subject-extracted conditions was larger in the high-frequency conditions than in the low-frequency conditions.

Finally, during the remainder of the sentence, the ANOVA revealed no significant effects ($F_s<2.2$)

Discussion

In Experiment 2A-4 we observed an interaction between the difficulty of retrieving the meaning of an embedded verb and retrieving its structural dependents in subject- and object-extracted cleft constructions. In particular, in the case of high-frequency embedded verbs we observed the standard subject-/object-extraction difference at the region where syntactic complexity was manipulated (the relative clause region). In contrast, in the case of low-frequency embedded verbs we observed that both subject- and object-extracted conditions were equally slow to process at the relative clause region. However, the difference between the subject- and the object-extracted conditions emerged at the following region.

As in the case of the thematic properties manipulation, we observed an interaction between the difficulty of retrieving the lexical properties of a word and retrieving its structural dependents. However, the observed interaction in the case of the lexical frequency manipulation was qualitatively different. In particular, in the case of the thematic properties manipulation, the difficulty of retrieving the lexical properties of a word and retrieving its structural dependents interacted to result in *superadditive processing difficulty* at the critical relative clause region. In the case of the lexical frequency manipulation, however, we observed a *delay* in the process of retrieving the word's structural dependents in cases where retrieving its lexical properties was

difficult. In particular, while both the low-frequency subject-extracted and the low-frequency object-extracted conditions were processed equally slowly (and as slow as the high-frequency object-extracted condition) at the relative clause region, the low-frequency object-extracted condition was processed slower than the low-frequency subject-extracted condition in the region following the relative clause region.

As discussed earlier, the Shared Resource Pool hypothesis predicts this pattern of results in cases where retrieving the word's lexical properties reaches a certain threshold of difficulty. The results are not consistent with the Separate Resource Pools hypothesis which predicts two independent effects of the difficulty of each retrieval operation.

In the ANOVA above we treated a continuous variable (frequency of the verb) as a discrete variable. In order to better understand the relationship between the difficulty of retrieving the meaning of the word and retrieving its structural dependents, we conducted an additional analysis using frequency as a continuous variable in order to see how well the verb's frequency can predict the reading times at the critical regions (the RC region and the post-RC region).

The *dependent measure* that captures the interaction we observed is the object-extracted subject-extracted difference across the two regions – the RC region and the post-RC region. Specifically, in the high-frequency-verb conditions we observed the difference between the subject-extracted and the object-extracted conditions at the *RC region*. In contrast, in the low-frequency-verb conditions we observed the difference between the subject-extracted and the object-extracted conditions at the *post-RC region*. We can capture this interaction by saying that the higher the frequency of the verb, the larger the difference should be between the subject-extracted and the object-extracted conditions at the RC region, and the lower the frequency of the verb, the larger the difference should be between the subject-extracted and the object-extracted conditions at the post-RC region. As a result we used the following dependent RT measure [(Object-extracted condition at the RC region – Subject-extracted condition at the RC region) – (Object-extracted condition at the post-RC-region – Subject-extracted condition at the post-RC region)]. We excluded ten items which had a negative value for the “Object-extracted minus Subject-extracted” difference at both the RC region and the post-RC region. The lack of a positive value for one of these regions suggested that these items failed to demonstrate the effect of difficulty of retrieving the word's structural dependents in the object-extracted conditions. Therefore, including these items in the analysis of the relationship between the difficulty of retrieving the word's meaning and the difficulty of retrieving the word's structural dependents would not be meaningful.

The *independent measure* is the verb frequency. The frequency values for the verbs used in the experiment were not normally distributed. In particular, most of the values clustered in the range between 0 and 5,000, with the rest of the values spread out between 5,000 and 625,000. Given the distribution of the frequency values, we transformed the frequency values into a common log scale and performed a linear regression using the log frequency values as the independent measure. The regression analysis revealed that log frequency is a highly significant predictor of the behavior ($F(1,36)=17.9, p<.001$) accounting for 33.2% of the variance (see Figure 34).

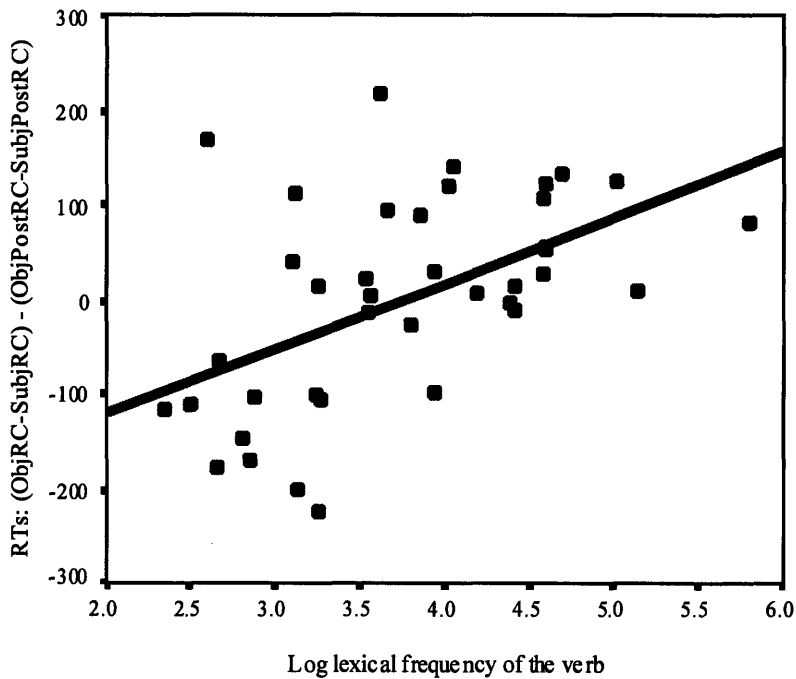


Figure 34: The relationship between the retrieval difficulty of the word’s meaning (log lexical frequency) and the retrieval difficulty of the word’s structural dependents.

Conclusions for Experiment 2A-4

The interaction observed in Experiment 2A-4 provides further support for the Shared Resource Pool hypothesis. In particular, the results demonstrated that when one of the retrieval operations (in this case, retrieving the word’s lexical properties) reaches a certain threshold of difficulty the other retrieval operation is delayed until a later point in the sentence.

General conclusions

In Section 2A we examined two cases where the retrieval of the lexical properties of a word (the verb’s thematic properties in one case, and the meaning of the verb in the other case) was difficult. It was shown that the difficulty of retrieving the lexical properties of the incoming word can *increase* or *delay* the difficulty of retrieving the word(s) to which the incoming word is connected in the dependency structure of the sentence. These patterns of interacting effects provide support for the Shared Resource Pool hypothesis whereby the two retrieval operations – retrieving the lexical properties of a word, and retrieving the word’s structural dependents – rely on the same pool of working memory resources.

2B. The ease of retrieving the lexical properties of an incoming word alleviates the difficulty of retrieving the word's structural dependents.

In Section 2A, we demonstrated that the difficulty of retrieving the lexical properties of the incoming word can *increase* or *delay* the difficulty of retrieving the word's structural dependents. The results were interpreted as supporting the Shared Resource Pool hypothesis, according to which (1) retrieving the lexical properties of a word, and (2) retrieving the word's structural dependents rely on the same pool of working memory resources.

As discussed at the beginning of Chapter 2, the Shared Resource Pool hypothesis also predicts that the difficulty of retrieving the word's lexical properties will interact with the difficulty of retrieving its structural dependents in cases where retrieving the word's lexical properties is easy. In particular, when the resources are abundant for performing the retrieval of structural dependents – which is the case when retrieving the word's lexical properties does not require many resources – the difficulty typically associated with retrieving the word's structural dependents may be alleviated.

Section 2B will examine two cases where the ease of retrieving the lexical properties of the incoming word partially or completely *alleviates* the difficulty of retrieving the word's structural dependents.

Experiments 2B-1 and 2B-2: Effects of contextual predictability on the difficulty of retrieving the word's structural dependents

In Experiments 2B-1 and 2B-2 we examine two cases where contextual information facilitates the processing of an incoming element by making it highly predictable: some or all of the word's lexical properties are highly active in working memory at the point of encountering the word thereby eliminating the need to retrieve these properties from long-term memory. According to the Shared Resource Pool hypothesis, eliminating the cost associated with retrieving the word's lexical properties should lead to abundant resources for retrieving the word's structural dependents, which should translate into a smaller or no cost associated with the retrieval of the word's structural dependents.

In the first experiment we manipulated the causal structure of the events described in the sentences, such that the material in the early part of the sentence either did or did not serve as a plausible cause for the event conveyed by the verb. We reasoned that in cases where the event described by the verb is preceded by a plausible cause, the processing of the verb should be facilitated because processing the preceding material plausibly activates a set of events that are likely to follow.

In the second experiment we manipulated the likelihood of a modifier which either did or did not involve a non-local dependency. We reasoned that in cases where the modifier with a non-local dependency is highly predictable the processing should be facilitated.

Experiment 2B-1

[Fedorenko, Gibson, & Frank, in preparation]

In this experiment we manipulated the causal structure of the events described in the sentences. Unlike all the experiments in Section 2A – where the difficulty of retrieving the word’s structural dependents was manipulated by using subject- and object-extracted constructions – in Experiment 2B-1 we will investigate local and non-local noun-verb dependencies (e.g., Grodner and Gibson, 2006). In particular, we varied whether the material in the early part of the sentence served as a plausible cause for the event conveyed by the verb. We reasoned that in cases where the event described by the verb is preceded by a plausible cause, the processing of the verb should be facilitated because processing the preceding material plausibly activates a set of events that are likely to follow according to world knowledge of the causal structure of events.

The Shared Resource Pool hypothesis predicts an interaction between the ease of retrieving the verb’s lexical properties and the difficulty of retrieving its structural dependents. In particular, the cost associated with retrieving the verb’s structural dependents in a non-local noun-verb dependency should be alleviated in cases where the verb is preceded by a description of an event which can serve as a plausible cause for the event described by the verb, thereby making some aspects of the verb’s meaning active in working memory.

Methods

Participants Thirty-two participants from MIT and the surrounding community were paid for their participation. All were native speakers of English and were naive as to the purposes of the study.

Design and materials The experiment had a 2 x 2 design, crossing the integration distance between the noun and the verb (local, non-local) and the extent to which the event described in the preceding material served as a plausible cause for the event described in the verb (plausible cause, less plausible cause).

The materials consisted of 20 sets of sentences with four different versions as in (17):

(17a) *Local / Plausible cause:*

After being shocked by the biopsy results | the woman | cried | for an hour | until her husband | tried to calm her down.

(17b) *Local / Less plausible cause:*

After being shocked by the airline service | the woman | cried | for an hour | until her husband | tried to calm her down.

(17c) *Non-local / Plausible cause:*

The woman | who the biopsy results shocked | cried | for an hour | until her husband | tried to calm her down.

(17d) *Non-local / Less plausible cause:*

The woman | who the airline service shocked | cried | for an hour | until her husband | tried to calm her down.

In the Local conditions, the integration between the noun *the woman* and the verb *cried* is local, and in the Non-local conditions, the noun and the verb are separated by a relative clause (*who the biopsy results / the airline service shocked*). This makes retrieving the noun at the point of processing the verb in the Non-local conditions costly. The information contained in the relative clause in the Non-local conditions is moved to the front of the sentence and appears in the form of an after-clause in the Local conditions.

In the Plausible-cause conditions the event described in the after-clause (for the Local conditions) or in the relative clause (for the Non-local conditions) served as a plausible reason for the event described in the verb. In the Less-plausible-cause conditions, the event described in the after-clause (for the Local conditions) or in the relative clause (for the Non-local conditions) provided a less plausible reason for the event described in the verb.

The critical region (the verb *cried* in (17)) was in the same sentence position in the Local and Non-local conditions, which is important given that in reading, people tend to speed up towards the end of the sentence making comparisons across different sentence positions difficult. Furthermore the amount of information present in the sentence before the critical verb is the same across the Local and Non-local conditions (c.f. Grodner & Gibson, 2005).

In addition to the target sentences, the experiment included 48 fillers. The length and syntactic complexity of the fillers was similar to that of the target sentences. The stimuli were pseudo-randomized separately for each participant, with at least one item from another experiment or one filler separating the target sentences. Each participant saw only one version of each sentence, following a Latin-Square design (see Appendix M for a complete list of linguistic materials).

Procedure The standard self-paced reading procedure was used. The experimental sentences in the Local conditions were structured, as follows: (1) an 'after'-clause, (2) a noun phrase, (3) a verb, (4) a prepositional phrase modifying the verb, and (5) the ending. The experimental sentences in the Non-local conditions were structured, as follows: (1) a noun phrase, (2) a relative clause, (3) a verb, (4) a prepositional phrase modifying the verb, and (5) the ending. As shown in (17), the ending was divided into two regions for the purposes of the analyses.

Participants took approximately 25 minutes to complete the experiment.

Results

Comprehension question performance Across conditions, participants answered the comprehension question correctly 89.7% of the time. Table 31 presents the mean accuracies across the four conditions. A two-factor ANOVA crossing the integration distance between the noun and the verb (local, non-local) and the plausibility of the cause in the context preceding the verb (plausible, less plausible) on the responses to the comprehension questions revealed a main effect of integration distance, such that participants were less accurate in the Non-local conditions ($F(1,31)=6.102$, $MSe=1250$, $p<.02$; $F(1,19)=6.33$, $MSe=781$, $p<.05$). There were no other effects ($F_s < 2$).

<u>Integration distance</u>	<u>The plausibility of the cause for the event described in the verb</u>	
	Plausible cause	Less plausible cause
Local	95.0 (1.8)	90.6 (2.7)
Non-local	85.6 (3.5)	87.5 (2.7)

Table 31. Comprehension accuracies in percent correct, as a function of integration distance and the plausibility of the cause for the event described in the verb in Experiment 2B-1 (standard errors in parentheses).

Reaction times We analyzed all the trials, regardless of whether the comprehension question was answered correctly. The statistical data patterns were very similar in the analysis of only the trials where the comprehension question was answered correctly. Reading times more than two standard deviations away from the mean for a position within condition were removed from the analyses, excluding 3.6% of the data.

We present the analysis of the critical region first, followed by the analyses of the other regions. We defined the verb (where the two retrieval operations take place) as the critical region. Figure 35 presents the mean reading times per region across the four conditions.

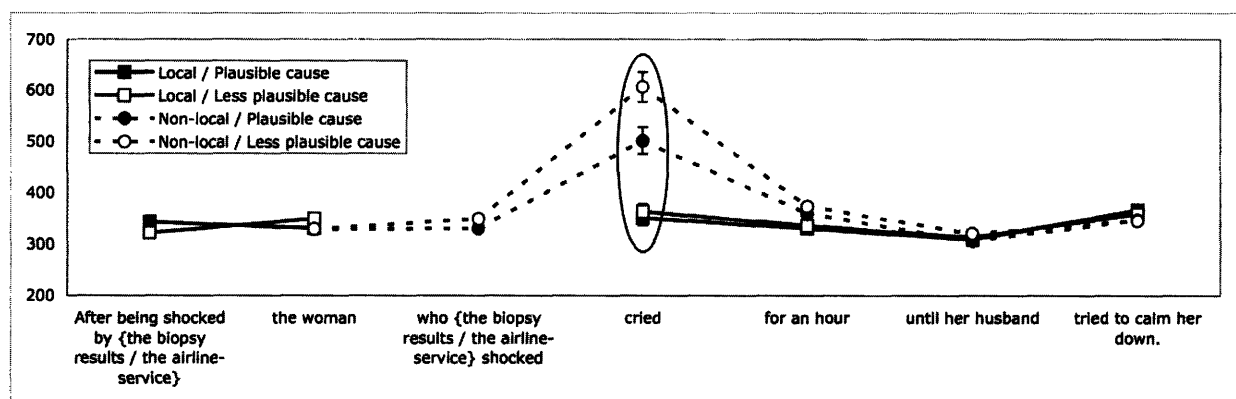


Figure 35: Reading times per region in Experiment 2B-1. Error bars indicate standard errors. The critical region is circled.

A 2 x 2 ANOVA (Local integration / Non-local integration, Plausible cause / Less plausible cause) in the critical region revealed two main effects and an interaction. First, the Non-local conditions were read slower than the Local conditions ($F(1,31)=70.82$, $MSe=1253406$, $p<.001$; $F(1,19)=78.4$, $MSe=775866$, $p<.001$). Second, the Plausible-cause conditions were read slower than the Less-plausible-cause conditions ($F(1,31)=10.49$, $MSe=88924$, $p<.005$; $F(1,19)=8.28$,

MSe=69322, $p < .02$). Finally, there was an interaction between the two factors, such that the difference between the Local and the Non-local conditions was larger in the Less-plausible-cause conditions, compared to the Plausible-cause conditions ($F(1,31)=5.23$, MSe=59774, $p < .05$; $F(1,19)=5.16$, MSe=42432, $p < .05$). This pattern of results is consistent with the Shared Resource Pool hypothesis, such that the presence – in the preceding context – of a plausible cause for the event described in the verb decreases the cost associated with retrieving the word's structural dependent (in this case, the noun) in the Non-local condition.

We will now present the results for the remaining regions. In the after-clause in the Local conditions, a one-factor ANOVA revealed a difference between the two conditions – which did not reach significance in the items analysis – such that the congruent condition was read slower than the incongruent condition ($F(1,31)=11.7$, MSe=7668, $p < .005$; $F(1,19)=3.54$, MSe=4343, $p = .075$). The materials across the two conditions involve different lexical items in some positions, so this difference is not unexpected. In the noun phrase region (e.g., *the woman*), a 2×2 ANOVA (Local integration / Non-local integration, Plausible cause / Less plausible cause) revealed a marginal effect of integration locality in the items analysis, such that the Non-local conditions were read slower than the Local conditions ($F(1,31)=1.025$, MSe=3993, n.s.; $F(1,19)=3.88$, MSe=2320, $p = .064$). There were no other effects ($F_s < 2.8$). The observed trend is not unexpected because the noun phrase is located in different positions across the Local and Non-local conditions: it is sentence-initial in the Non-local conditions, and it follows the 'after'-clause in the Local conditions. Previous research has shown that people tend to speed up towards the later positions in the sentence and, therefore, it is not surprising that the noun phrase is read slower in the Non-local conditions where it is the first sentence region. In the relative clause in the Non-local conditions, a one-factor ANOVA revealed no difference between the two conditions ($F_s < 2$). At the region following the critical verb region, the ANOVA revealed a main effect of integration locality such that the Non-local conditions were read slower than the Local conditions ($F(1,31)=21.9$, MSe=33074, $p < .001$; $F(1,19)=15.5$, MSe=21655, $p < .002$). There were no other effects ($F_s > 1.3$). In the following region, the ANOVA revealed an effect of cause plausibility in the participants analysis ($F(1,31)=7.081$, MSe=2776, $p < .02$; $F(1,19)=1.52$, MSe=1710, n.s.), such that the Less-plausible-cause conditions were read slower than the Plausible-cause conditions. There were no other effects ($F_s < 1.2$). Finally, during the remainder of the sentence, the ANOVA revealed a main effect of integration locality, such that Local conditions were read slower than the Non-local conditions ($F(1,31)=6.95$, MSe=6735, $p < .02$; $F(1,19)=4.69$, MSe=5118, $p < .05$). There were no other effects ($F_s < 2.3$).

Discussion

In Experiment 2B-1 we manipulated the causal structure of the events described in the sentences, such that the context preceding the verb in a noun-verb dependency either did or did not serve as a plausible cause for the event conveyed by the verb. We reasoned that in cases where the event described by the verb is preceded by a plausible cause, the processing of the verb should be facilitated because processing the preceding material plausibly activates a set of events that are likely to follow.

We observed an interaction between the difficulty of retrieving the lexical properties of a verb (in this case, aspects of the verb's meaning) and the difficulty of retrieving its structural

dependents. In particular, when retrieving the word's lexical properties was relatively easy (in the Plausible-cause conditions) the difference between the condition where retrieving the word's dependents was costly (the Non-local condition) and the condition where retrieving the word's dependents was cost-free (the Local condition) was smaller, compared to cases where retrieving the word's lexical properties was more difficult (the Less-plausible-cause conditions).

These results provide further support for the Shared Resource Pool hypothesis, according to which retrieving the word's lexical properties and retrieving the word's structural dependents rely on the same pool of working memory resources.

Experiment 2B-2

[Gibson, Fedorenko, & Ishizuka, in preparation]

In the previous experiment, we examined the effects of the plausibility of the cause for the event described in the verb in a noun-verb dependency on the difficulty of retrieving the verb's structural dependents and we observed an interaction between the two factors, such that the presence of a plausible cause decreased the cost associated with retrieving the word's structural dependents. In this section, we will examine another case where the ease of retrieving the lexical properties of the word alleviates the difficulty of retrieving its structural dependents. In particular, we manipulated the likelihood of a modifier which either did or did not involve a non-local dependency (object-extracted relative clause vs. subject-extracted relative clause). We reasoned that in cases where the modifier with a non-local dependency is highly predictable processing should be facilitated, compared to cases where it is less predictable.

The Shared Resource Pool hypothesis predicts an interaction between the ease of retrieving the verb's lexical properties and the difficulty of retrieving its structural dependents. Therefore, the cost associated with retrieving the verb's structural dependents should be alleviated in cases where the type of modifier (a subject- or an object-extracted relative clause) is highly predictable based on the preceding context.

Methods

Participants Thirty-two participants from MIT and the surrounding community were paid for their participation. All were native speakers of English and were naive as to the purposes of the study.

Design and materials A context supporting both a subject- and an object-extracted relative clause was presented prior to the target RCs. The critical sentences were presented in the form of an exchange between two speakers (Mary and John), such that the first speaker would produce an utterance involving a subject-extracted RC and the second speaker would produce an utterance involving an object-extracted RC, or vice versa. 24 items were constructed. A sample item is presented in (18).

(18)

At the press-conference, a senator and two reporters got into an argument.

The senator attacked one of the reporters and then the other reporter attacked the senator.

Mary: / I heard that / the reporter / that {attacked the senator} | {the senator attacked} / admitted to / making an error.

John: / I'm not sure about that. / I heard that / the reporter / that {the senator attacked} | {attacked the senator} / admitted to / making an error.

The critical comparison was between the subject- and the object-extracted RCs presented in the *first* utterance vs. the subject- and the object-extracted RCs presented in the *second* utterance. In the first utterance, at the point of encountering the definite noun phrase (*the reporter*), the reader knows that a modifier of some kind must follow in order to pick out the relevant reporter from the set of two reporters introduced in the discourse context. In the second utterance, at the point of encountering the definite noun phrase, the reader knows (1) that a contrast will be present, and (2) the contrast must concern the modifier type, because if the contrast concerned any other part of the sentence, a pronoun would be used instead of a definite NP.

For example:

Mary: I heard that the reporter that attacked the senator admitted to making an error.

John: I'm not sure about that. I heard that he never acknowledges his mistakes.

The following factors were counter-balanced:

- The structure of the second sentence in the context with regard to whether the description of the events started with the unique referent or one of the two referents of the same kind: half of the items started with the unique noun (*The NOUN VERBed one of the NOUNS and then the other NOUN VERBed the NOUN*), and the other half started with a non-unique noun (*One of the NOUNS VERBed the NOUN and then the NOUN VERBed the other NOUN*)
- Half of the items had the subject-extracted RC in the first (Mary's) utterance, and the other half had the object-extracted RC in the first utterance.
- Each item had two versions such that the identity of the unique noun was changed. For example, there was another version of the item shown in (17), which started as follows: *At the press-conference, a reporter and two senators got into an argument.*

In addition to the target sentences, the experiment included 48 fillers. The fillers were carefully constructed to resemble the target sentences. Specifically, the first context sentence always introduced three entities (all were unique nouns: e.g., *the model, the photographer and the hairdresser*). The second sentence described two actions taking place among the three participants (e.g., the first character doing something to the second character and the second character doing something to the third character; the order in which the nouns from the first sentence were used in these actions was balanced). The two actions were connected by a variety of connectives, such as *because, however, when, while, after*, etc. Mary's utterance started with *I heard that*, like in target sentences. This was followed by a statement about one of the three characters (16 fillers – 1st character, 16 fillers – 2nd character, and 16 fillers – 3rd character). John's utterance started with a contradiction in 12/48 fillers: this was similar to the target sentences, but a variety of different expressions preceded *I heard that*, such as "*Hmm, that's weird*" or "*I am not sure*". In the remaining 36/48 fillers John's utterance started with an expression of agreement, such as "*Yeah, that's right*" or "*I think you're right*". This way,

overall, half of the sentences involved a disagreement between Mary and John, and the other half – agreement. Critically, Mary’s and John’s utterances did not involve relative clauses. A sample filler is presented in (19).

(19)

During the photo-shoot, the model, the photographer and the hairdresser discussed possible hairstyles.

The model liked her hair straight; however, the photographer and the hairdresser wanted her hair to be curly.

Mary: I heard that the model is stubborn and hard to work with.

John: Yeah, that’s right, and I also heard that she gets paid ten thousand dollars for every photo-shoot.

See Appendix N for a complete list of materials.

Procedure The standard self-paced reading procedure was used, except that the context sentences were presented all-at-once (one sentence after the other), and the critical sentences were presented region-by-region, as marked in (18). The fillers were also broken down into regions in a similar fashion.

Each experimental item was structured, as follows: (1) the first context sentence introducing the three characters, (2) the second context sentence describing two actions that took place among the three characters, (3) *Mary:*, (4) *I heard that*, (5) the head noun phrase of the first RC, (6) the subject-/object-extracted RC, (7) main verb of the first RC, (8) ending, (9) *John:*, (10) *I am not sure about that.*, (11) *I heard that*, (12) the heard noun phrase of the second RC, (13) the object-/subject-extracted RC, (14) main verb of the second RC, and (15) ending.

The questions for the target sentences were about the content of the utterances. Specifically, they either asked about the RCs and were of the form *According to Mary, was it the reporter that attacked the senator / the senator attacked that admitted to making an error?*, or they asked about the main verb and were of the form *Did one of the reporters / senators admitted to making an error?* Half of the questions for the filler sentences were about the information contained in the context sentences, and the other half was about the information contained in the utterances.

Participants took approximately 30 minutes to complete the experiment.

Results

Comprehension question performance Participants answered the comprehension question correctly 75.7% of the time. The order of the RC-type did not affect performance ($F_s < 1$).

Reaction times We analyzed all the trials, regardless of whether the comprehension question was answered correctly. The statistical data patterns were very similar in the analysis of only the trials where the comprehension question was answered correctly. Reading times more than two standard deviations away from the mean for a position within condition were removed from the analyses, excluding 4.2 % of the data.

We present the analysis of the critical regions first, followed by the analyses of the other regions. We defined the critical regions as the two occurrences of the relative clauses: in Mary's utterance (early occurrence), and in John's utterance (late occurrence). Figure 36 presents the mean reading times per region across the two orders of the RC-type.

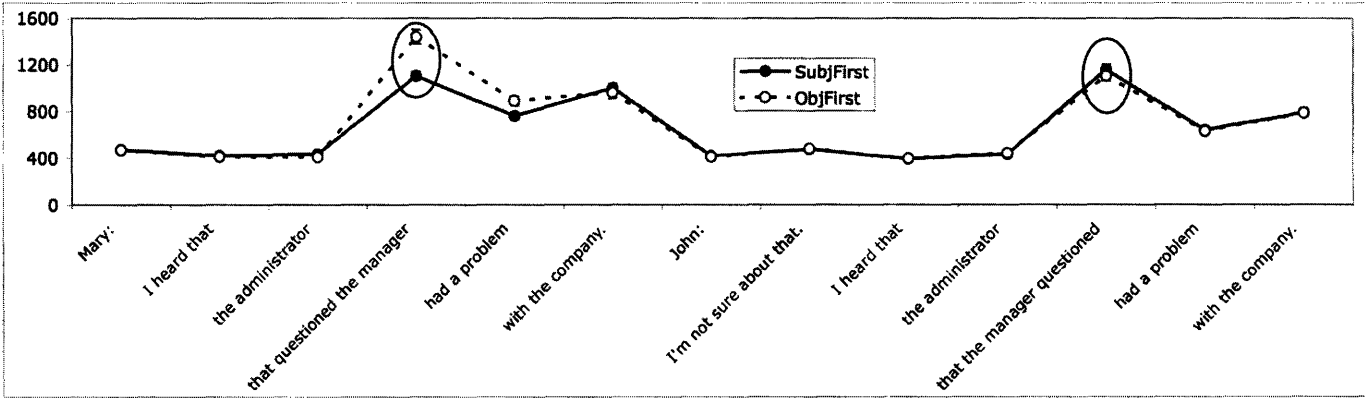


Figure 36: Reading times per region in Experiment 2B-2. Error bars indicate standard errors. The critical regions are circled.

An ANOVA in the first critical region revealed a significant effect, such that the condition where the object-extracted RC was presented in the early position was read significantly slower than the condition where the subject-extracted RC was presented in the early position ($F(1,31)=16.6$, $MSe=2147706$, $p<.001$; $F2(1,23)=31.3$, $MSe=1345296$, $p<.001$). In the second critical region the ANOVA revealed no significant effect ($F_s<1.1$), although numerically, the condition where the object-extracted RC was presented in the late position was read slower than the condition where the subject-extracted RC was presented in the late position. In summary, while we see the standard subject-/object- difference in the early position, this difference is not present in the late position. Figure 37 presents the RTs for the two critical regions.

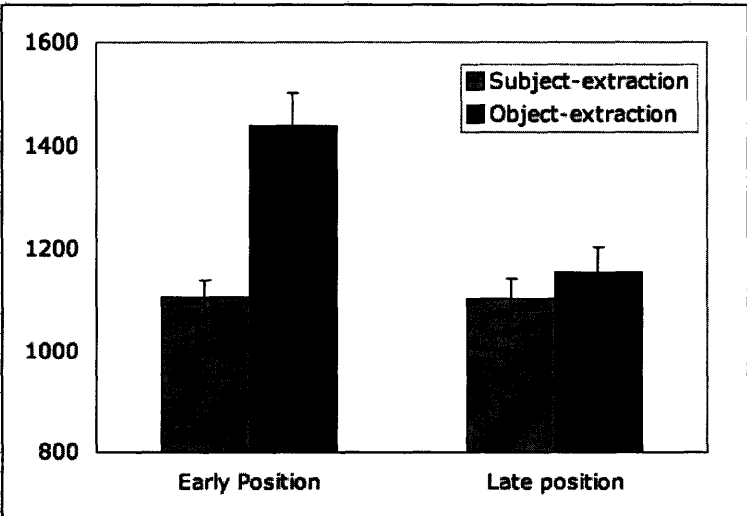


Figure 37: Reading times at the two critical regions in Experiment 2B-2. Error bars indicate standard errors.

We will now present the results for the remaining regions. In the first region (the first context sentence) and in the second region (the second context sentence), there were no significant differences between the two conditions (all $F_s < 1$). [These two data-points were not included in the graph in Figure 36 because the whole-sentence reading times were long (on average, 4237 ms for the first region and 4161 ms for the second region) and thus would skew the scale so that the differences at the critical regions would not be very visible.] In the third region (*Mary:*) and in the fourth region (*I heard that*), there were no significant differences between the two conditions ($F_s < 1.5$). In the fifth region (the head noun phrase in the first RC), there was an unpredicted difference, such that the condition where the subject-extracted RC was presented in the early position was read significantly slower than the condition where the object-extracted RC was presented in the early position ($F_1(1,31)=7.01$, $MSe=10303$, $p < .02$; $F_2(1,23)=4.97$, $MSe=7908$, $p < .05$). There is no reason to expect a difference in this region because the materials were exactly the same across the two conditions. In the region following the first RC (main verb of the first RC), there was a significant difference, such that the condition where the object-extracted RC was presented in the early position was read significantly slower than the condition where the subject-extracted RC was presented in the early position ($F_1(1,31)=11.2$, $MSe=292653$, $p < .005$; $F_2(1,23)=14.7$, $MSe=203791$, $p < .002$). This difference is plausibly a result of spill-over from the previous (RC) region. In the next region (the ending), there was no difference between the two conditions ($F_s < 1$). In regions 9-11 (*John.;* *I am not sure about that.;* and *I heard that*) there were no significant differences between the two conditions (all $F_s < 1.3$). Similarly, in region 12 (the head noun phrase in the first RC), there was no difference between the conditions ($F_s < 2.2$). Finally, in regions 14-15 (main verb of the second RC and the ending) there were no significant differences between the two conditions (all $F_s < 1$).

Discussion

In Experiment 2B-2 – similar to Experiment 2B-1 – we observed the ease of retrieving the lexical properties of a word facilitating the process of retrieving its structural dependents.

The results of Experiment 2B-2 provide further support for the Shared Resource Pool hypothesis.

Conclusions for Experiments 2B-1 and 2B-2

In Experiments 2B-1 and 2B-2 we examined two cases where contextual information facilitated the processing of an incoming element by making it highly predictable. In the first experiment we manipulated the causal structure of the events described in the sentences, such that the material in the earlier part of the sentence either did or did not serve as a plausible cause for the event conveyed by the verb. In the second experiment we manipulated the likelihood of a modifier which either did or did not involve a non-local dependency. In both experiments we observed that the cost associated with retrieving the word's structural dependents in the non-local integration conditions was alleviated (partially in Experiment 2B-1 and completely in Experiment 2B-2) in cases where the word was highly predictable from the preceding context.

The results in Section 2B provide further support for the Shared Resource Pool hypothesis, according to which (1) retrieving the word's lexical properties, and (2) retrieving its structural dependents rely on the same pool of working memory resources. Specifically, when one of the retrieval operations (in this case, retrieving the lexical properties of the word) is easy, resources are abundant for performing the other retrieval operation and therefore the cost associated with it is reduced.

Summary and Conclusions

This thesis investigated two questions about the working memory system underlying language processing:

1. To what extent is the working memory system underlying language processing domain-specific?
2. What is the relationship between the working memory system and the long-term memory system in language processing?

In Chapter 1, I presented a series of experiments that used a novel approach to the question of the domain-specificity of the working memory system underlying language processing. I have argued that there may be two possible reasons for why the previous attempts to find an interaction between linguistic complexity and non-linguistic verbally-mediated tasks have failed. First, the cognitive processes involved in the language processing task and in the digit-/word-span tasks – which have been traditionally used in investigating this question – are qualitatively different. Second, the materials involved in the language processing task and in the digit-/word-span tasks are qualitatively different. I addressed these problems by comparing language processing to verbal working memory tasks that (1) are similar in the types of mental processes they involve, and (2) involve similar materials. First, I demonstrated that in dual-task experiments where participants have to perform linguistic integrations and arithmetic integrations simultaneously, superadditive processing difficulty ensues in the condition where both linguistic and arithmetic integrations are difficult. I concluded that linguistic and arithmetic integrations rely on overlapping pools of verbal working memory resources. I further examined the relationship between linguistic and musical integrations. The results provided suggestive evidence for an overlap between the two domains in terms of the working memory resource pools. Second, I demonstrated (following Gordon et al., 2002) that in dual-task experiments where participants have to perform a sentence-processing task and a word-span task which involves materials similar to or dissimilar from those used in the sentences, superadditive processing difficulty ensues in the condition where linguistic integrations are difficult and the materials in the word-span task are similar to the ones used in the sentences. Based on the evidence presented in Chapter 1, I argued that at least some aspects of the working memory system used for linguistic integrations are not domain-specific, (1) being involved in arithmetic, and possibly, musical processing; and (2) being involved in some verbal working memory tasks involving storing verbal representations over time.

In Chapter 2, I presented a series of experiments investigating the relationship between two retrieval operations necessary when an incoming word is processed: (1) the retrieval of the lexical properties of the word from long-term memory, and (2) the retrieval of the word's structural dependents from working memory. I evaluated the predictions of two alternative hypotheses: the Separate Resource Pool hypothesis and the Shared Resource Pool hypothesis. According to the Separate Resource Pools hypothesis, the two retrieval operations rely on independent pools of working memory resources. According to the Shared Resource Pool hypothesis, the two retrieval operations rely on the same pool of working memory resources. I manipulated the relative difficulty of retrieving the lexical properties of the incoming word and observed that in the cases where retrieving the word's lexical properties was difficult, the

difficulty of retrieving its structural dependents was increased or delayed, and in the cases where retrieving the word's lexical properties was easy, the difficulty of retrieving its structural dependents was partially or completely alleviated. In particular, I showed that when retrieving the verb's thematic properties or the verb's meaning was difficult, the difficulty of retrieving the verb's structural dependents was increased or delayed in the non-local integration conditions. I further showed that when retrieving some aspects of verb's representation was easy due to the fact that the verb was highly predictable from the preceding context, the difficulty of retrieving the verb's structural dependents was partially or completely alleviated. Based on the evidence presented in Chapter 2, I argued that retrieving the word's lexical properties and retrieving its structural dependents rely on the same pool of working memory resources.

Figure 38 presents a graphic summary of the main findings of this thesis. The verbal working memory system was shown to contain at least two separate pools of resources: one for keeping track of incomplete syntactic dependencies (linguistic storage), and another for integrating incoming words to earlier words/positions in the sentence (linguistic integrations). (I leave open the possibility that there exist other pools of resources within the general verbal WM system underlying language processing.) I focused on investigating the second resource pool – the one underlying linguistic integrations – in the work described here. I demonstrated that this resource pool is not strictly domain-specific: evidence from Chapter 1 suggested that there exists some overlap between this resource pool and the resource pool underlying arithmetic integration processes. I further demonstrated that the two retrieval operations involved in integrating incoming words to earlier words/positions in the sentence – retrieving the word's lexical properties from long-term memory and retrieving its structural dependents from working memory – rely on the same pool of resources.

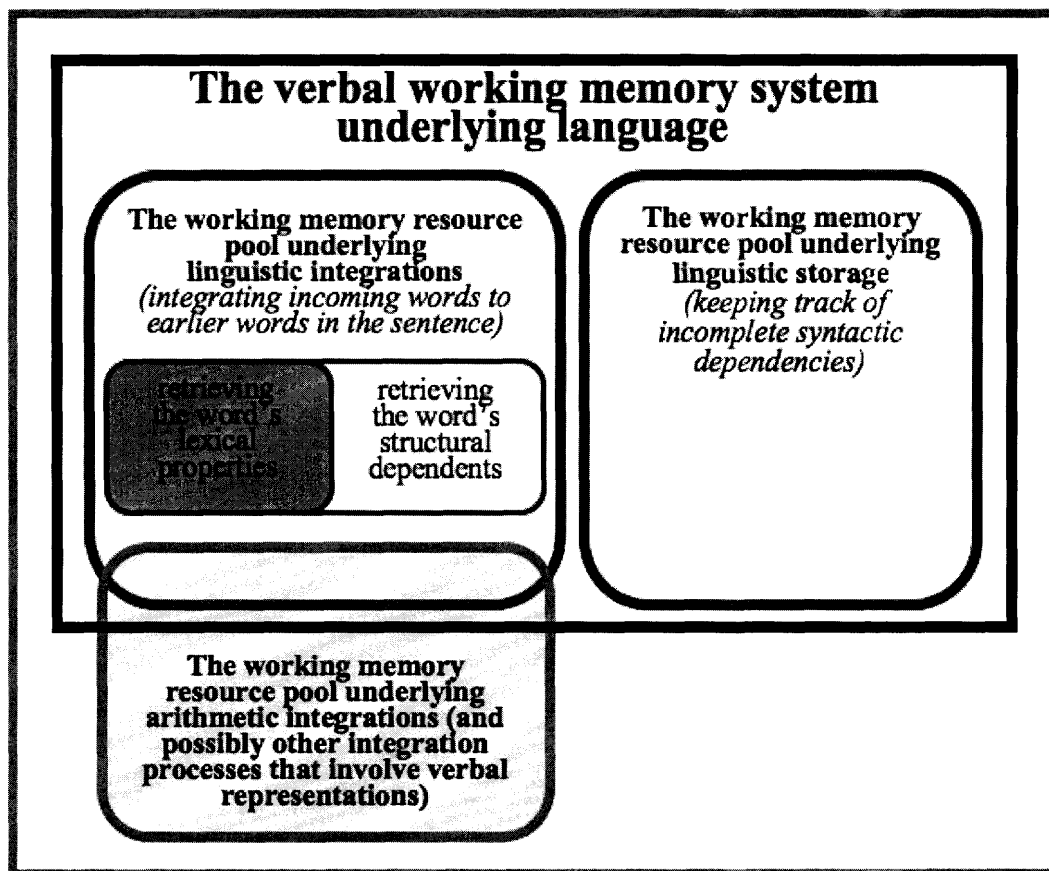


Figure 38: A graphic summary of the findings from Chapters 1 and 2.

The research described here has several implications. First, given recent evidence for the heterogeneity of the working memory resource pool underlying language processing, it seems important to investigate the nature of each resource pool separately. The two (and possibly more) resource pools underlying language processing may have different properties, and therefore, not distinguishing between these pools may lead to difficult-to-interpret patterns of results.

Second, the results reported in Chapter 1 suggest that verbal working memory (and perhaps working memory more generally) may be organized not along the lines of the domains to which different tasks belong (e.g. Caplan & Waters, 1999), but rather along the lines of the qualitative nature of the cognitive processes involved in the tasks: for example, storing verbal representations over time vs. connecting verbal representations into more complex representations.

Third, the fact that the results reported in Chapter 2 provided support for the Shared Resource Pool hypothesis – according to which retrieving the word’s lexical properties and retrieving its structural dependents rely on the same pool of working memory resources – suggests that linguistic complexity models need to incorporate both retrieval processes in modeling the process of linguistic integrations.

References

- Ashcraft, M. H. (1992). Cognitive Arithmetic: A review of data and theory. *Cognition*, 44, 75-106.
- Ashcraft, M. H. (1995). Cognitive psychology and simple arithmetic: A review and summary of new directions. *Mathematical Cognition*, 1, 3-34.
- Baddeley, A. D. & Hitch, G. (1974). Working Memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 8, pp. 47-89). New York: Academic Press.
- Baddeley, A. D. (1986). *Working Memory*. New York: Oxford University Press.
- Besson, M., Faita, F., Peretz I., Bonnel, A.J.R. (1998). Singing in the brain: independence of lyrics and tunes. *Psychological Science*, 9, 494-498.
- Boland, J. E. (1997). Resolving syntactic category ambiguities in discourse context: Probabilistic and discourse constraints. *Journal of Memory and Language*, 36, 588-615.
- Boland, J. E., & Blodgett, A. (2001). Understanding the constraints on syntactic generation: Lexical bias and discourse congruency effects on eye movements. *Journal of Memory and Language*, 45, 391-411.
- Bonnel, A.-M., Faita, F., Peretz, I., & Besson, M. (2001). Divided attention between lyrics and tunes of operatic songs: Evidence for independent processing. *Perception and Psychophysics*, 63:1201-1213.
- Campbell, J. I. D. & Xue, Q. (2001). Cognitive arithmetic across cultures. *Journal of Experimental Psychology: General*, 130 (2), 299-315.
- Caplan, D. & Waters, G. S. (1999). Verbal working memory and sentence comprehension. *Brain & Behavioral Sciences*, 22, 77-126.
- Chen, E., Gibson, E., & Wolf, F. (2005). Online syntactic storage costs in sentence comprehension. *Journal of Memory and Language*, 52(1), 144-169.
- Chomsky, N. & Miller, G.A. (1963). Introduction to the formal analysis of natural languages. In: Luce, R.D., Bush, R.R., Galanter, E. (Eds.), *Handbook of Mathematical Psychology*, vol. 2. Wiley, New York, pp. 269-321.
- Chomsky, N. (1986). *Knowledge of Language: Its Nature, Origin, and Use*. New York: Praeger Publishers.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd Edition). Lawrence Erlbaum Associates, Inc. Publishers. Hillsdale, NJ.

- Cosmides, L. & J. Tooby (1994). Origins of Domain Specificity: The Evolution of functional organization” In L. A. Hirschfeld & S. A. Gelman (eds), *Mapping the Mind: Domain specificity in cognition and culture*, New York : Cambridge University Press, 85-116.
- Cowan, N. (2001) The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, 24, 1, 87-114. Crowder, R.G. (1976). *Principles of Learning and Memory*. Lawrence Erlbaum Associates, Publishers: Hillsdale, New Jersey.
- Cupples, L. (2002). The structural characteristics and on-line comprehension of experiencer-verb sentences. *Language and Cognitive Processes*, 17, 125-162.
- Daneman, M. & Carpenter, P. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, 19, 450-466.
- Donders, F. C. (1868-1869). Over de snelheid van psychische processen. Onderzoekingen gedann in het Physiologisch Laboratorium der Utrechtsche Hoogeshool. Tweede Reeks, II, 92-120. In W. G. Koster (Ed. and transl.), *Attention and performance II*. Amsterdam: North-Holland, 1969. (Reprinted from *Acta Psychologica*, 1969, 30, 412-431.)
- Fedorenko, E., Gibson, E., & Rohde, D. (2006). The nature of working memory capacity in sentence comprehension: Evidence against domain-specific working memory resources. *Journal of Memory and Language*, 54.
- Fedorenko, E., Gibson, E., & Rohde, D. (2007). The nature of working memory in linguistic, arithmetic and spatial integration processes. *Journal of Memory and Language*, 56.
- Felser, C., Clahsen, H., & Munte, T. F. (2003). Storage and Integration in the processing of filler-gap dependencies: An ERP study of topicalization and wh-movement in German. *Brain & Language*, 87.
- Ferreira, F. & Clifton, C., Jr. (1986). The independence of syntactic processing. *Journal of Memory and Language*, 25, 348-368.
- Fiebach, C. J., Schlesewsky, M., & Friederici, A. D. (2002). Separating syntactic memory costs and syntactic integration costs during parsing: The processing of German WH-questions. *Journal of Memory and Language*, 47(2), 250-272.
- Fodor, J. (1983). *The Modularity of Mind: An Essay on Faculty Psychology*. Cambridge, MA: MIT Press.
- Ford, M. (1983). A method for obtaining measures of local parsing complexity throughout sentences. *Journal of Verbal Learning and Verbal Behavior*, 22, 203-218.
- Forster and Chambers. (1973). K.I. Forster and S.M. Chambers, Lexical access and naming time. *Journal of Verbal Learning and Verbal Behavior* 12 (1973), pp. 627–635

- Gardiner, J. M., Craik, F. I. M., & Birtwistle, J. (1972). Retrieval cues and release from proactive inhibition. *Journal of Verbal Learning and Verbal Behavior*, 11, 778–783.
- Gibson, E. (1991). *A computational theory of human linguistic processing: Memory limitations and processing breakdown*. Unpublished doctoral dissertation, Carnegie Mellon University, Pittsburgh, PA.
- Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, 68, 1-76.
- Gibson, E. (2000). The dependency locality theory: A distance-based theory of linguistic complexity. In Miyashita, Y., Marantz, A., & O’Neil, W. (Eds.), *Image, language, brain*. MIT Press, Cambridge, MA.
- Gibson, E. (2006). The interaction of top-down and bottom-up statistics in the resolution of syntactic category ambiguity. *Journal of Memory and Language*, 54, 363-388.
- Gillund, G., & Shiffrin, R.M. (1984). A retrieval model for both recognition and recall. *Psychological Review*, 91, 1-67.
- Gordon, P. C., Hendrick, R., & Johnson, M. (2001). Memory interference during language processing. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 27(6), 1411-1423.
- 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*, Vol. 29 (2), 261-290.
- Hanley, J. R., Young, A., & Pearson, N. A. (1991). Impairment of the visuo-spatial scratchpad. *Quarterly Journal of Experimental Psychology*, 43A, 101-125.
- Hintzman, D.L. (1986). “Schema abstraction” in a multiple-trace memory model. *Psychological Review*, 93, 411-428.
- Howes, D.H. and Solomon, R.L. Visual duration thresholds as a function of word-probability, *J. Exp. Psychol.* 41 (1951), pp. 401–410.
- Jonides, J., Smith, E. E., Koeppel, R. A., Awh, E., Minoshima, S., & Mintun, M. A. (1993). Spatial working memory in humans as revealed by PET. *Nature*, 363, 623-625.
- Just, M.A. & Carpenter, P.A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, 99, 122-149.

- Just, M.A., Carpenter, P.A., & Woolley, J.D. (1982). Paradigms and processing in reading comprehension. *Journal of Experimental Psychology: General*, 111, 228-238.
- Kaan, E., Harris, A., Gibson, E., & Holcomb, P. (2002). The P600 as an index of syntactic integration difficulty. *Language and Cognitive Processes*, 15(2), 159-201.
- Kanwisher, N., McDermott, J., & Chun, M. (1997). The Fusiform Face Area: A Module in Human Extrastriate Cortex Specialized for the Perception of Faces. *Journal of Neuroscience*, 17, 4302-4311.
- King, J. & Just, M. A. (1991). Individual differences in syntactic processing: the role of working memory. *Journal of Memory and Language*, 30, 580-602.
- King, J. & Kutas, M. (1995). Who did what to when?: Using word- and clause-level ERPs to monitor working memory usage in reading. *Journal of Cognitive Neuroscience*, 7(3), 376-395.
- Kluender, R & Kutas, M. (1993). Bridging the gap: Evidence from ERPs on the processing of unbounded dependencies. *Journal of Cognitive Neuroscience*, 5(2), 196-214.
- Konieczny, L. (2000). Locality and parsing complexity. *Journal of psycholinguistic research*, 29(6), 627-645.
- Lerdahl, F., & Jackendoff, R. (1983). *A Generative Theory of Tonal Music*. Cambridge, MA: MIT Press.
- Lewis, R. (1996). A theory of grammatical but unacceptable embeddings. *Journal of Psycholinguistic Research*, 25, 93-116.
- Lewis, R., Vasishth, S. and Van Dyke, J. (2006). Computational principles of working memory in sentence comprehension. *Trends in Cognitive Science*, 10(10):447-454.
- Loftus, G. (1978). On interpretation of interactions. *Memory and Cognition*, 6(3), 312-319.
- Logie, R. H. (1986). Visuo-spatial processing in working memory. *Quarterly Journal of Experimental Psychology*, 38A, 229-247
- Logie, R. H. (1995). *Visuo-spatial working memory*. Hove, UK: Erlbaum.
- Luria, A. R., Tsvetkova, L. S. and Futer, D. S. (1965) Aphasia in a composer (V. G. Shebalin). *J Neurol Sci* 2, 3, 288-92.
- Masson, M.E. & Loftus, G.R. (2003). Using confidence intervals for graphically based data interpretation. *Canadian Journal of Experimental Psychology*, 57, 202-220.

- McClelland, J. L. (1979). On the time relations of mental processes: An examination of systems of processing in cascade. *Psychological Review*, 86(4), 287-330.
- McDermott, J. & Hauser, M. (2006). The origins of music: Innateness, development, and evolution. *Music Perception*.
- MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (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 clause processing. *Journal of Memory and Language*, 47, 50-68.
- McElree, B., Foraker, S., & Dyer, L. (2003). Memory structures that subserve sentence comprehension. *Journal of Memory and Language*, 48, 67-91.
- Patel, A.D. (2003). Language, music, syntax, and the brain. *Nature Neuroscience*, 6, 674-681.
- Patel, A.D., Gibson, E., Ratner, J., Besson, M. & Holcomb, P. (1998). Processing syntactic relations in language and music: An event-related potential study. *Journal of Cognitive Neuroscience*, 10:717-733.
- Paulesu, E., Frith, C. D., & Frackowiak, R. S. J. (1993). The neural correlates of the verbal component of working memory. *Nature*, 362, 342-345.
- Peretz, I. (1993). Auditory atonalia for melodies. *Cognitive Neuropsychology*, 10, 21-56.
- Peretz, I., & Coltheart, M. (2003) Modularity of music processing. *Nature Neuroscience*, 6, 688-691.
- Peretz, I. & Hyde, K. (2003). What is specific to music processing? Insights from congenital amusia. *Trends in Cognitive Science*, 8(7), 362-367.
- Phillips, C., Kazanina, N., & Abada, S. (2005). ERP effects of the processing of syntactic long-distance dependencies. *Cognitive Brain Research*, 22, 407-428.
- Rohde, D., Gonnerman, L., & Plaut, D. (in preparation). An improved method for deriving word meaning from lexical co-occurrence.
- Rosenthal, R. & Rosnow, R.L. (1991). *Essentials of behavioral research: Methods and data analysis* (2nd Edition). McGraw-Hill Series in Psychology. New York, NY.
- Seidenberg, M. S., & MacDonald, M. C. (1999). A probabilistic constraints approach to language acquisition and processing. *Cognitive Science*, 23, 569-588.

- Shah, P. & Miyake, A. (1996). The Separability of Working Memory Resources for Spatial Thinking and Language Processing: An Individual Differences Approach. *Journal of Experimental Psychology: General*, 125 (1), 4-27.
- Shepard, R. N. & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, 171, 701-703.
- Siegler, R. S. (1987). Strategy choices in subtraction. In J. A. Slobada & D. Rogers (Eds.), *Cognitive processes in mathematics*. Oxford: Clarendon Press.
- Sperber, D. (1994). The modularity of thought and the epidemiology of representations. In L. A. Hirschfeld & S. A. Gelman (eds), *Mapping the Mind: Domain specificity in cognition and culture*, New York : Cambridge University Press, 39-67.
- Sternberg, S. (1969). The discovery of processing stages: Extensions of Donders' method. In W. G. Koster (Ed.), *Attention and performance II*. Amsterdam: North-Holland. (Reprinted from *Acta Psychologica*, 1969, 30, 276-315.
- 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.
- 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.
- Traxler, M., Williams, R., Blozis, S. and Morris, R. (2005). Working memory, animacy, and verb class in the processing of relative clauses. *Journal of Memory and Language*, Volume 53, Issue 2, Pages 204-224.
- Trueswell, J. C., Tanenhaus, M. K., & Kello, C. (1993). Verb-specific constraints in sentence processing: Separating effects of lexical preference from garden-paths. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19, 528-553.
- 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.
- Vallar, G. & Shallice, T. (Eds.). (1990). *Neuropsychological Impairments of short-term memory*. New York: Cambridge University Press.
- 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.
- Vasishth, S. (2003). *Working memory in sentence comprehension*. New York: Garland Press.

- Wanner, E. & Maratsos, M. (1978). An ATN approach to comprehension. In Halle, M., Miller, G. & Bresnan, J. (Eds.), *Linguistic theory and psychological reality*. MIT Press. Cambridge, MA.
- Warren, T. & Gibson, E. (2002). The influence of referential processing on sentence complexity. *Cognition*, 85, 79-112.
- Waters, G. & Caplan, D. (1999). The effect of a digit-load on syntactic processing in subjects with high and low working memory capacity. Unpublished manuscript.
- Waters, G., Caplan, D. & Hildebrandt, N. (1987). Working memory and written sentence comprehension. In Coltheart, M (Ed.), *Attention and performance XII: The psychology of reading*. Erlbaum.
- Waters, G., Caplan, D. & Rochon, E. (1995). Processing capacity and sentence comprehension in patients with Alzheimer's disease. *Cognitive Neuropsychology*, 12, 1-30.

Appendix A

Materials used in Experiment 1A.

1. After the appraiser notified the dealer (from a small town in Greece), the thief knew (the thief's knowledge) that the painting which the millionaire inherited could sell for millions of dollars (would prove to be useless).
2. After the officer caught the mobster (from a mafia network in Sicily), the detective suspected (the detective's suspicion) that the company which the manager (from a mafia network in Sicily) ran was trafficking young prostitutes from abroad (was verified within a month).
3. When the animator warned the executive (from a studio outside of Hollywood), the producer decided (the producer's decision) that the film which the cinematographer (from a studio outside of Hollywood) presented was far too expensive to make (made sense to everyone).
4. Before the physician examined the patients (from a hospital in North London), the firm doubted (the firm's doubt) that the drug which the scientist (from a hospital in North London) patented would cure the new epidemic disease (concerned the investors).
- 5 a After the magistrate reprimanded the prosecutor (from a town in North Carolina), the lawyer predicted (the lawyer's prediction) that the case which the agency (from a town in North Carolina) handled would be settled outside of court (ended up being correct).
6. When the CIA captured the infiltrator (at a base outside the country), the government denied (the government's denial) that the money which the troops (at a base outside the country) utilized was spent on a missile system (caused an uproar).
7. Before the judge consulted the lawyer (from the capital city of Columbia), the defendant claimed (the defendant's claim) that the narcotics which the cop (from the capital city of Columbia) located belonged to a dealer in Peru (seemed absurd).
8. After the husband harassed the seller (from a well-known region of France), the manufacturer promised (the manufacturer's promise) that the lotion which the woman (from a well-known region of France) bought would work wonders within a month (justified the cost).
9. Before the police arrested the employee (with strong views on North Korea), the editor believed (the editor's belief) that the article which the writer (with strong views on North Korea) submitted would be popular with the readers (seemed perfectly reasonable).
10. While the busboy consulted the waiter (from a small island near Cuba), the customer demanded (the customer's demand) that the dish which the chef (from a small island near Cuba) prepared should be made without any garlic (offended the other patrons).
11. When the cop informed the lieutenant (from a secret airbase in Nevada), the investigator theorized (the investigator's theory) that the plane which the pilot (from a secret airbase in Nevada) crashed must have spiraled out of control (failed to be supported).

12. While the commissioner notified the investigator (from the police department in Canada), the immigrant asserted (the immigrant's assertion) that the documents which the official (from the police department in Canada) rejected followed all the necessary requirements precisely (was falling on deaf ears).

13. Before the government contacted the spy (aboard the passenger airliner from India), the journalist commented (the journalist's comment) that the cargo which the ambassador (aboard the passenger airliner from India) accompanied contained some highly hazardous radioactive materials (appeared totally unfounded).

14. After the economist congratulated the mayor (from the town of Niagara Falls), the analyst calculated (the analyst's calculation) that the hotel which the corporation (from the town of Niagara Falls) owned would soon become even more profitable (convinced many doubters).

15. After the president addressed the organization (with an agenda promoting African Americans), the reporter remarked (the reporter's remark) that the decree which the senator (with an agenda promoting African Americans) defended would certainly make a real difference (was considered by the politicians).

16. While the apprentice assisted the mechanic (right across from the Town Hall), the biker asserted (the biker's assertion) that the garage which the firm (right across from the Town Hall) operated was the best in the area (pleased the owner).

17. After the executive thanked the employees (in the highest skyscraper on Broadway), the director announced (the director's announcement) that the project which the company (in the highest skyscraper on Broadway) completed was the biggest success in years (was met with applause).

18. After the cardiologist informed the nurse (from the medical school in Dallas), the doctor hoped (the doctor's hope) that the instrument which the surgeon (from the medical school in Dallas) borrowed could save a lot of children (proved to be justified).

19. When the conservative attacked the screenwriter (of the controversial movie about Afghanistan), the actor desired (the actor's desire) that the movie which the producer (of the controversial movie about Afghanistan) released would cause people to take action (was stronger than ever).

20. After the designer complimented the editor (from the popular fashion magazine Vogue), the model believed (the model's belief) that the photo which the artist (from the popular fashion magazine Vogue) took would promptly lead to international recognition (was finally justified).

21. Before the supervisor called the specialist (from the frozen woodlands of Siberia), the zookeeper feared (the zookeeper's fear) that the tiger which the trainer (from the frozen woodlands of Siberia) brought would scare little boys and girls (was very strong).

22. After the customer praised the chef (from a southern province in Spain), the waiter dreamed (the waiter's dream) that the menu which the owner (from a southern province in Spain) introduced would soon bring many new customers (could become a reality).

23. When the associate called the statistician (from the consulting firm in Japan), the executive predicted (the executive's prediction) that the stock which the analyst (from the consulting firm in Japan) recommended would go up significantly within months (received some support).

24. Before the assistant consulted the administrator (from the Physics Department at Amherst), the librarian advised (the librarian's advice) that the book which the student (from the Physics Department at Amherst) wanted should always be reserved in advance (sounded very uninformed).

25. When the expert criticized the judge (from the county court in Alabama), the journalist concluded (the journalist's conclusion) that the thief which the jury (from the county court in Alabama) convicted should be pardoned given the circumstances (was riddled with doubt).

26. After the parishioner met the lobbyist (with some money from the government), the priest prayed (the priest's prayer) that the shelter which the nun (with some money from the government) founded would not have to close down (seemed most justifiable).

27. After the commentator interviewed the decorator (from the countryside in Southern Scotland), the florist suggested (the florist's suggestion) that the roses which the child (from the countryside in Southern Scotland) gathered could be made into a bouquet (interested the child's mother).

28. After the analyst approached the governor (with many active supporters from Maine), the committee stated (the committee's statement) that the legislation which the congressman (with many active supporters from Maine) penned would improve the lives of many (received a big round of applause).

29. When the apprentice recognized the merchant (from a famous village in China), the tailor believed (the tailor's belief) that the silk which the customer (from a famous village in China) chose would make a lovely festive gown (became stronger still).

30. Before the manager queried the teller (from a small bank in Georgia), the landlord assumed (the landlord's assumption) that the check which the tenant (from a small bank in Georgia) sent would probably bounce and cause problems (was resting on a pure intuition).

31. When the dictator executed the dissident (from the mountain region up North), the guerilla warned (the guerilla's warning) that the rebels which the soldiers (from the mountain region up North) imprisoned would seek revenge sooner than expected (scared a lot of civilians).

32. After the reporter interviewed the meteorologist (from a resort town in Florida), the anchorman announced (the anchorman's announcement) that the news which the correspondent

(from a resort town in Florida) delivered should be taken into serious consideration (started to worry the residents).

34. After the researcher consulted the pollster (from the humor newspaper 'The Onion'), the comedian doubted (the comedian's doubt) that the joke which the columnist (from the humor newspaper 'The Onion') shared would be popular with the audience (became a certainty).

35. When the physicist challenged the salesman (from the equipment store "Starry Sky"), the astronomer realized (the astronomer's realization) that the telescope which the assistant (from the equipment store "Starry Sky") described would not have sufficient magnifying power (was confirmed by an expert opinion).

36. When the organizer notified the choreographer (from the most acclaimed ballet Giselle), the director confirmed (the director's confirmation) that the piece which the dancer (from the most acclaimed ballet Giselle) performed should appear later in the program (was expected by everyone).

37. While the historian introduced the excavator (from the recent expedition to Egypt), the linguist asserted (the linguist's assertion) that the manuscript which the archeologist (from the recent expedition to Egypt) recovered could be decoded given sufficient time (excited the scientists).

38. When the secretary overheard the spokesman (from the Department of Foreign Affairs), the translator warned (the translator's warning) that the letter which the minister (from the Department of Foreign Affairs) signed could be later used against him (was no longer relevant).

39. After the conductor confronted the representative (from the Association for Women's Rights), the pianist argued (the pianist's argument) that the concert which the activist (from the Association for Women's Rights) organized should have been planned more carefully (was finally taken seriously).

40. When the director consulted the agriculturalist (from the experimental farm in Ukraine), the farmer asserted (the farmer's assertion) that the carrots which the salesman (from the experimental farm in Ukraine) marketed would not grow in this region (was found to be correct).

Appendix B

Reading times for every region across the four conditions of Experiment 1A.

Region 1:	After the appraiser notified the dealer	
Region 2:	from a small town in Greece,	[low-integration conditions]
Region 3:	the thief knew / the thief's knowledge	[low-storage / high-storage]
Region 4:	that the painting	
Region 5:	which the millionaire	
Region 6:	from a small town in Greece	[high-integration conditions]
Region 7:	inherited	
Region 8:	could sell for millions of dollars	
Region 9:	would prove to be useless	[high-storage conditions]

	<u>Region</u>			
<u>Condition</u>	1	2/6	3	4
Low storage / Low integration	1107 (150)	402 (47)	48 (34)	-172 (22)
Low storage / High integration	1037 (128)	-39 (37)	287 (40)	-80 (25)
High storage / Low integration	912 (99)	333 (47)	29 (33)	-170 (25)
High storage / High integration	1226 (138)	-64 (43)	291 (44)	-88 (25)

	<u>Region</u>			
<u>Condition</u>	5	7 (critical)	8	9
Low storage / Low integration	-248 (23)	-133 (23)	-266 (48)	
Low storage / High integration	-225 (23)	-72 (23)	-267 (49)	
High storage / Low integration	-228 (28)	-105 (24)	-295 (44)	-186 (35)
High storage / High integration	-158 (24)	-46 (21)	-291 (45)	-168 (34)

Appendix C

Language materials used in Experiments 1B-1 – 1B-6.

One of the four subject- / object-extracted RC versions is shown below for each of the 32 items. The other three versions can be generated as exemplified in (1) below.

1. a. Subject-extracted, version 1:
The janitor who frustrated the plumber lost the key on the street.
b. Subject-extracted, version 2:
The plumber who frustrated the janitor lost the key on the street.
c. Object-extracted, version 1:
The janitor who the plumber frustrated lost the key on the street.
d. Object-extracted, version 2:
The plumber who the janitor frustrated lost the key on the street.
2. The hairdresser who hired the beautician transformed the salon for the better.
3. The lecturer who provoked the dean left the university in the summer.
4. The trumpeter who loved the drummer formed the band two years ago.
5. The intern who distrusted the boss disregarded the messages on her voicemail.
6. The roommate who annoyed the landlord slammed the door of the apartment.
7. The player who avoided the coach entered the room at the gym.
8. The mayor who called the advisor requested an update on the project.
9. The librarian who angered the teacher misplaced the book from the depository.
10. The pharmacist who helped the assistant placed the order for the drug.
11. The waitress who hugged the bartender dropped the tray on the floor.
12. The client who contacted the retailer offered a deal of the century.
13. The celebrity who admired the athlete won the award at the ceremony.
14. The detective who recognized the spy crossed the street at the light.
15. The journalist who complimented the editor revised the article for the newspaper.
16. The employee who praised the executive finished the project right on time.
17. The legislator who visited the senator falsified the documents for the trip.
18. The soldier who shot the enemy received a medal for the battle.
19. The officer who described the murderer told a lie about the past.
20. The reporter who followed the cameraman damaged the equipment during the trip.
21. The understudy who telephoned the agent shared the news about the suicide.
22. The consultant who confronted the programmer broke the computer in a rage.
23. The supervisor who deceived the owner kept the money in the end.
24. The entrepreneur who hated the stockbroker sold the shares after the merger.
25. The mole who revealed the defector rejected the offer on the spot.
26. The singer who blamed the organizer cancelled the concert in Los Angeles.
27. The acrobat who mocked the clown performed the trick at the show.
28. The customer who upset the seller forgot the receipt on the counter.
29. The partner who introduced the businessman presented the report at the meeting.
30. The messenger who summoned the knight read the letter from the king.
31. The linguist who ridiculed the historian proposed the hypothesis for the problem.
32. The biker who ignored the driver made the turn at the crossing.

Appendix D

Language materials used in Experiments 1B-7a and 1B-7b.

One of the four subject- / object-extracted RC versions is shown below for each of the 40 items (items #6, #37, #38, #39 were not used in Experiment 1B-7b). The other three versions can be generated as exemplified in (1) below.

1. a. Subject-extracted, grammatical:
 The boy that helped the girl got an “A” on the test.
 b. Object-extracted, grammatical:
 The boy that the girl helped got an “A” on the test.
 c. Subject-extracted, ungrammatical:
 The boy that help the girl got an “A” on the test.
 d. Object-extracted, ungrammatical:
 The boy that the girl help got an “A” on the test.

2. The clerk that liked the boss had a desk by the window.
3. The guest that kissed the host brought a cake to the party.
4. The priest that thanked the nun left the church in a hurry.
5. The thief that saw the guard had a gun in his holster.
6. The prince that mocked the lord spread a lie [recording error]
7. The crook that warned the thief fled the town the next morning.
8. The knight that helped the king sent a gift from his castle.
9. The cop that met the spy wrote a book about the case.
10. The nurse that blamed the coach checked the file of the gymnast.
11. The count that knew the queen owned a castle by the lake.
12. The scout that punched the coach had a fight with a manager.
13. The cat that fought the dog licked its wounds in the corner.
14. The whale that bit the shark won the fight in the end.
15. The maid that loved the chef quit the job at the house.
16. The bum that scared the cop crossed the street at the light.
17. The man that phoned the nurse left his pills at the office.
18. The priest that paid the cook signed the check at the bank.
19. The dean that heard the guard made a call about the matter.
20. The friend that teased the bride told a joke about the past.
21. The fox that chased the wolf hurt its paws on the way.
22. The groom that charmed the aunt raised a toast to the parents.
23. The nun that blessed the monk lit a candle on the table.
24. The guy that thanked the judge left the room with a smile.
25. The king that pleased the guest poured the wine from the jug.
26. The girl that pushed the nerd broke the vase with the flowers.
27. The owl that scared the bat made a loop in the air.
28. The car that pulled the truck had a scratch on the door.
29. The rod that bent the pipe had a hole in the middle.
30. The hat that matched the skirt had a bow in the back.
31. The niece that kissed the aunt sang a song for the guests.

32. The boat that chased the yacht made a turn at the boathouse.
33. The desk that scratched the bed was too old to be moved.
34. The cook that hugged the maid had a son yesterday.
35. The boss that mocked the clerk had a crush on the intern.
36. The fruit that squashed the cake made a mess in the bag.
37. The road that crossed the street had a house with a garden.
38. The dog that pulled the man strained the leash to the limit.
39. The tree that touched the bush was in bloom at this time.
40. The dean that called the boy had a voice full of anger.

Appendix E

The questionnaire used in Experiments 1B-7a and 1B-7b.

Music Background Questionnaire

Date of Birth: _____

Are you right-handed _____ or left-handed _____?

Did you ever play any musical instruments? Y ___ N ___

If so, which instrument(s) and how many years?

Instrument	Start year	End year
1.		
2.		
3.		
4.		
5.		

Did you ever sing in a choir? Y ___ N ___

If yes, during which time period:

From _____ to _____

Have you had any formal training in music theory? Y ___ N ___

If yes, during which time period:

From _____ to _____

How much music do you listen to currently (hours per day)? _____

What style(s) of music do you listen to? (e.g., classical, rock) _____

Anything else we should know about your musical background?

Appendix F

Language materials used in Experiment 1C-1.

One of the four subject- / object-extracted RC versions is shown below for each of the 32 items. The other three versions can be generated as exemplified in (1) below.

1. a. Subject-extracted, version 1:
The physician who consulted the cardiologist checked the files in his office.
 - b. Subject-extracted, version 2:
The cardiologist who consulted the physician checked the files in his office.
 - c. Object-extracted, version 1:
The physician who the cardiologist consulted checked the files in his office.
 - d. Object-extracted, version 2:
The cardiologist who the physician consulted checked the files in his office.
2. The babysitter who liked the parents planned a trip to Puerto Rico.
 3. The banker who informed the chairman invested a million in a start-up.
 4. The violinist who flattered the cellist played a piece from the symphony.
 5. The burglar who wounded the policeman reloaded the revolver in a hurry.
 6. The carpenter who punched the electrician quit the job a week later.
 7. The accountant who advised the statistician calculated the costs of the project.
 8. The model who approached the artist signed the contract for a year.
 9. The student who trusted the professor answered the question about the experiment.
 10. The mobster who attacked the dealer organized some crimes in New York.
 11. The investigator who overheard the cop closed the case without an arrest.
 12. The actor who respected the starlet forgot the lines during the scene.
 13. The defendant who misled the lawyer blamed the system for the conviction.
 14. The count who adored the princess brought a gift to the reception.
 15. The bachelor who pursued the socialite owned a company in the area.
 16. The councilman who kissed the secretary covered the expenses for the party.
 17. The contestant who offended the host ruined the show for the audience.
 18. The mathematician who addressed the physicist offered the proof at the conference.
 19. The diplomat who insulted the congressman ended the negotiations on the spot.
 20. The priest who thanked the nun founded the shelter near the church.
 21. The analyst who queried the governor proposed some changes to the plan.
 22. The farmer who questioned the expert promoted the product at the fair.
 23. The official who harassed the manager questioned the policy of lowering wages.
 24. The clerk who disliked the director typed the letter to the administration.
 25. The guitarist who recommended the band recorded the song for the album.
 26. The salesman who resented the cashier mislabeled the products in the brochure.
 27. The waiter who invited the cook tasted the sauce for the meat.
 28. The medic who assisted the doctor borrowed the instrument for the surgery.
 29. The passenger who befriended the stewardess remembered the flight across the Atlantic.
 30. The cheerleader who bothered the quarterback attended the game at the college.
 31. The animator who criticized the producer offered a solution to the problem.
 32. The dictator who despised the dissident gave a speech about the protests.

Appendix G

Memory-nouns statistics for Experiment 1C-1.

	Lexical Frequency ¹¹	Length (in syllables)
Occupation memory-nouns	10717	2.18
Names memory-nouns	10715	2.24
T-test	.98	.16

¹¹ Lexical frequencies were matched using a Usenet corpus of 1.2 billion words (Rohde et al., in preparation).

Appendix H
Residual reading times for Experiment 1C-1.

<u>Memory Load</u>	<u>Syntactic complexity</u>			
	Subject-extraction (Easy)		Object-extraction (Hard)	
<u>Memory-Noun</u>	Match	Non-match	Match	Non-match
One noun (Easy)	-233 (40)	-204 (40)	-273 (33)	-225 (44)
Three nouns (Hard)	38.6 (61)	54.03 (57)	20.7 (60)	106 (73)

Table H1. Residual reading times in milliseconds, as a function of syntactic complexity, memory load, and memory-noun type (standard errors in parentheses) for Region 1.

<u>Memory Load</u>	<u>Syntactic complexity</u>			
	Subject-extraction (Easy)		Object-extraction (Hard)	
<u>Memory-Noun</u>	Match	Non-match	Match	Non-match
One noun (Easy)	-181 (59)	-163 (56)	261 (73)	237 (58)
Three nouns (Hard)	-65.2 (63)	-83.1 (53)	502 (123)	113 (78)

Table H2. Residual reading times in milliseconds, as a function of syntactic complexity, memory load, and memory-noun type (standard errors in parentheses) for Region 2.

<u>Memory Load</u>	<u>Syntactic complexity</u>			
	Subject-extraction (Easy)		Object-extraction (Hard)	
<u>Memory-Noun</u>	Match	Non-match	Match	Non-match
One noun (Easy)	-38.8 (53)	-56.9 (43)	-68.02 (52)	110 (65)
Three nouns (Hard)	-61.2 (58)	-0.47 (57)	43.4 (54)	148 (61)

Table H3. Residual reading times in milliseconds, as a function of syntactic complexity, memory load, and memory-noun type (standard errors in parentheses) for Region 3.

<u>Memory Load</u>	<u>Syntactic complexity</u>			
	Subject-extraction (Easy)		Object-extraction (Hard)	
<u>Memory-Noun</u>	Match	Non-match	Match	Non-match
One noun (Easy)	49.1 (89)	-33.7 (59)	-23.1 (67)	-10.9 (71)
Three nouns (Hard)	6.48 (68)	-72.6 (57)	-159 (46)	-47.1 (78)

Table H4. Residual reading times in milliseconds, as a function of syntactic complexity, memory load, and memory-noun type (standard errors in parentheses) for Region 4.

Appendix I

Language materials used in Experiments 2A-1 and 2A-2.

The subject-extracted versions are shown below for each of the 20 items. The object-extracted versions can be generated as exemplified in (1) below.

1. a. Animate RC subject, subject-extracted:
The mountaineer that moved the boulder was experienced and strong.
b. Animate RC subject, object-extracted:
The boulder that the mountaineer moved was large and heavy.
c. Inanimate RC subject, subject-extracted:
The boulder that injured the mountaineer was large and heavy.
d. Inanimate RC subject, object-extracted:
The mountaineer that the boulder injured was experienced and strong.
2. The electrician that repaired the radio was skilled and hard-working.
The radio that woke up the electrician was loud and old.
3. The woman that bought the pastries was in her forties.
The pastries that fattened the woman were imported from France.
4. The patient that ordered the wheelchair was weak and fragile.
The wheelchair that transported the patient had many advanced functions.
5. The tourist that visited the exhibition was a passionate art-lover.
The exhibition that attracted the tourist had paintings by Picasso.
6. The child that caught the illness missed two weeks of school.
The illness that weakened the child could have serious consequences.
7. The farmer that voted for the senate-bill hoped for the best.
The senate-bill that benefited the farmer passed by three votes.
8. The cyclist that caused the accident violated the traffic rules.
The accident that slowed down the cyclist was all over evening news.
9. The surfer that enjoyed the sun wished for a cold drink.
The sun that burnt the surfer was harsh and dangerous.
10. The rower that postponed the surgery wanted to compete in the Olympics.
The surgery that saved the rower was serious and risky.
11. The criminal that forged the evidence could get twenty years.
The evidence that implicated the criminal persuaded the members of the jury.

12. The driver that followed the directions was in a rush.
The directions that helped the driver were detailed and clear.
13. The boy-scout that made the fire set up a tent.
The fire that warmed-up the boy-scout kept the bears away.
14. The sailor that observed the stars headed for the island.
The stars that guided the sailor were bright and beautiful.
15. The traveler that took the trip was happy to be back.
The trip that toughened the traveler was a real challenge.
16. The cab-driver that anticipated the traffic-jam made a phone call.
The traffic-jam that delayed the cab-driver lasted for an hour.
17. The soldier that fired the bullet was hiding behind the ruins.
The bullet that missed the soldier hit the moving truck.
18. The dancer that requested the stage-lights had a spectacular performance.
The stage-lights that illuminated the dancer added a lot to the performance.
19. The runner that grabbed a drink wanted to keep going.
The drink that energized the runner was cold and refreshing.
20. The princess that inherited the painting married a rich duke.
The painting that depicted the princess was worth a lot of money.

Appendix J

Language materials used in Experiments 2A-3.

1. The girl that the boy upset during the class cried for an hour.
The boy that upset the girl during the class was often quite mean.
The girl that the grade upset during the class cried for an hour.
The grade that upset the girl during the class was not very important.
2. The professor that the student annoyed quite a bit yelled at the class.
The student that annoyed the professor quite a bit made very strong claims.
The professor that the book annoyed quite a bit yelled at the class.
The book that annoyed the professor quite a bit made very strong claims.
3. The boss that the secretary irritated quite a lot embezzled from the account.
The secretary that irritated the boss quite a lot was much too talkative.
The boss that the meeting irritated quite a lot embezzled from the account.
The meeting that irritated the boss quite a lot was much too long.
4. The child that the clown mesmerized at the show stole from the gift-shop.
The clown that mesmerized the child at the show had a big unicycle.
The child that the trick mesmerized at the show stole from the gift-shop.
The trick that mesmerized the child at the show had no obvious explanation.
5. The critic that the virtuoso amazed at the concert lied about his credentials.
The virtuoso that amazed the critic at the concert was really quite unbelievable.
The critic that the performance amazed at the concert lied about his credentials.
The performance that amazed the critic at the concert was really quite unbelievable.
6. The baby that the nanny distracted for ten minutes had an upset stomach.
The nanny that distracted the baby for ten minutes played a funny game.
The baby that the TV distracted for ten minutes had an upset stomach.
The TV that distracted the baby for ten minutes played some funny cartoons.
7. The woman that the comedian offended on the television complained to the FCC.
The comedian that offended the woman on the television talked about obese people.
The woman that the ad offended on the television complained to the FCC.
The ad that offended the woman on the television talked about obese people.
8. The girl that the monster frightened in the theater had a weak heart.
The monster that frightened the girl in the theater had big sharp teeth.
The girl that the movie frightened in the theater had a weak heart.
The movie that frightened the girl in the theater had a horrifying scene.
9. The invalid that the speaker inspired to take action wanted to help others.
The speaker that inspired the invalid to take action talked about cancer survivors.
The invalid that the story inspired to take action wanted to help others.

- The story that inspired the invalid to take action talked about cancer survivors.
10. The teenager that the instructor bored in first period wrote on his desk.
The instructor that bored the teenager in first period dealt with discipline issues.
The teenager that the topic bored in first period wrote on his desk.
The topic that bored the teenager in first period dealt with discipline issues.
 11. The nurse that the patient worried during her shift didn't write any useful notes.
The patient that worried the nurse during her shift had a stroke recently.
The nurse that the side-effect worried during her shift didn't write any useful notes.
The side-effect that worried the nurse during her shift had an unusual time-course.
 12. The baby that the mother comforted after the noise slept through the night.
The mother who comforted the baby after the noise was so very caring.
The baby that the toy comforted after the noise slept through the night.
The toy that comforted the baby after the noise was soft and pink.
 13. The adolescent that the toddler exhausted in the afternoon snapped at his father.
The toddler that exhausted the adolescent in the afternoon was way too active.
The adolescent that the run exhausted in the afternoon snapped at his father.
The run that exhausted the adolescent in the afternoon was way too long.
 14. The spectator that the poet charmed at the reading smiled quietly to himself.
The poet that charmed the spectator at the reading had a pleasant voice.
The spectator that the poem charmed at the reading smiled quietly to himself.
The poem that charmed the spectator at the reading had a nice rhythm.
 15. The agent that the writer impressed with his originality contacted the publishing house.
The writer that impressed the agent with his originality touched the readers' hearts.
The agent that the novel impressed with its originality contacted the publishing house.
The novel that impressed the agent with its originality touched the readers' hearts.
 16. The patient that the masseuse relaxed in the evening improved after the treatment.
The masseuse that relaxed the patient in the evening was very highly trained.
The patient that the massage relaxed in the evening improved after the treatment.
The massage that relaxed the patient in the evening was prescribed by doctors.
 17. The hippie that the philosopher intrigued during the lecture signed up for a class.
The philosopher that intrigued the hippie during the lecture talked about the past.
The hippie that the riddle intrigued during the lecture signed up for a class.
The riddle that intrigued the hippie during the lecture talked about the past.
 18. The visitor that the lecturer perplexed during the seminar asked a difficult question.
The lecturer that perplexed the visitor during the seminar posed a fascinating question.
The visitor that the lecture perplexed during the seminar asked a difficult question.
The lecture that perplexed the visitor during the seminar posed a fascinating question.

19. The model that the photographer embarrassed at the restaurant spoke to her lawyer.
The photographer that embarrassed the model at the restaurant was always very obnoxious.
The model that the picture embarrassed at the restaurant spoke to her lawyer.
The picture that embarrassed the model at the restaurant was from some party.
20. The audience that the actor impressed during the play might return for another performance.
The actor that impressed the audience during the play imitated a famous politician.
The audience that the special effects impressed during the play might return for another performance.
The special effects that impressed the audience during the play imitated a real hurricane.
21. The babysitter that the twins overwhelmed on Friday night quit the next week.
The twins that overwhelmed the babysitter on Friday night had too many requests.
The babysitter that the job overwhelmed on Friday night quit the next week.
The job that overwhelmed the babysitter on Friday night had too many requirements.
22. The policeman that the protester maddened at the rally fired at the crowd.
The protester that maddened the policeman at the rally was yelling very loudly.
The policeman that the chant maddened at the rally fired at the crowd.
The chant that maddened the policeman at the rally was really quite offensive.
23. The boy that the vampire scared at the movies left with his mother.
The vampire that scared the boy at the movies looked much too realistic.
The boy that the explosion scared at the movies left with his mother.
The explosion that scared the boy at the movies looked much too realistic.
24. The victim that the criminal terrified during the incident required some psychological care.
The criminal that terrified the victim during the incident took out a gun.
The victim that the shooting terrified during the incident required some psychological care.
The shooting that terrified the victim during the incident took place earlier today.

Appendix K
Verb statistics for Experiment 2A-4.

	Lexical Frequency (in thousands)	Length (in letters)	Length (in syllables)
High-frequency verbs	90,739	6.75	2.04
Low-frequency verbs	1,751	6.71	2.17
T-test	<.01	.94	.63

Appendix L

Materials for Experiment 2A-4.

The subject-extracted versions with the two verb versions are shown below for each of the 24 items. The object-extracted versions can be generated as exemplified in (1) below.

1. It was Brandon who felicitated Dustin when the commencement ended.
It was Brandon who Dustin felicitated when the commencement ended.
It was Brandon who congratulated Dustin when the commencement ended.
It was Brandon who Dustin congratulated when the commencement ended.
2. It was Armando who criticized / belittled Jacob during the presentation.
3. It was Jared who impressed / awed Elvin with a wonderful performance.
4. It was Cedric who insulted / affronted Julius when the game was over.
5. It was Zachary who bothered / badgered Carlton in the library reading room.
6. It was Hector who praised / lauded Dillon for getting an A in the difficult math exam.
7. It was Marcela who calmed / placated Dora before the show started.
8. It was Pauline who aggravated / peeved Natalie because of a previous dispute.
9. It was Tara who hated / disdained Stefanie since early childhood.
10. It was Mollie who humiliated / abased Lauren in front of everyone in the audience.
11. It was Kylie who mesmerized / allured Selena during the prom dance.
12. It was Janelle who paid / remunerated Gladys after an argument.
13. It was Daniel who charmed / beguiled Kendra from the first date.
14. It was Jeremy who disciplined / chastised Francesca because of unacceptable behavior.
15. It was Brendan who pleased / assuaged Amanda with a radiant gift.
16. It was Andrew who punished / castigated Shawna following an inadmissible mistake.
17. It was Timothy who signaled / beckoned Samantha the directions to the park.
18. It was Tyler who murdered / annihilated Kathryn that dreadful night.
19. It was Sofia who strangled / asphyxiated Joshua in the first scene of the play.
20. It was Genevieve who teased / pestered Sebastian because of the weird shoes.
21. It was Tania who confused / vexed Byron explaining the procedure.
22. It was Vivian who lectured / chided Terrence for always being late.
23. It was Isabel who informed / apprised Garrett of the importance of the exam.
24. It was Chloe who rejected / spurned Edgar for lack of confidence.

Appendix M

Materials for Experiment 2B-1.

The Non-local versions are shown below for each of the 24 items. The Local versions can be generated as exemplified in (1) below.

1. The woman who the biopsy results shocked cried for an hour until her husband tried to calm her down.
The woman who the airline-service shocked cried for an hour until her husband tried to calm her down.
After being shocked by the biopsy results the woman cried for an hour until her husband tried to calm her down.
After being shocked by the airline-service the woman cried for an hour until her husband tried to calm her down.
2. The policeman who the gunshot / the spider injured fell on the ground and then his partner called for an ambulance.
3. The accountant who the boss / the massage-therapist mistreated quit two days ago and so the human resources office hired a temp.
4. The citizen who the death penalty / the construction disturbed protested in the street but everyone ignored her.
5. The athlete who the marathon / the discussion overwhelmed sweated a lot and so she drank some water.
6. The teacher who the principal / the security-guard insulted was fired without any warning and so he sued the school board to get his job back.
7. The man who the terrorist / the plumber unnerved fainted on the floor and had to be revived with smelling salts.
8. The businessman who the government / the telemarketers harassed went bankrupt some time ago and moved to a new state.
9. The lawyer who the surgeon / the neighbor examined died shortly thereafter and the family was in mourning.
10. The visitor who the class / the conversation bored fell asleep for a few minutes and did not wake up until someone shook him.
11. The child who the meal / the story disgusted threw up and then she quickly left the room.
12. The boy who the accident / the playground distressed screamed very loudly until his mother rushed to help him.

13. The passenger who the poor service / the magazine dismayed complained to the flight attendant but the man sitting next to her ignored her.
14. The biker who the taxi-driver / the waiter surprised crashed onto the road but fortunately he was wearing a helmet.
15. The student who the professor / the doctor warned failed in school and his parents were very concerned.
16. The girl who the clown / the hairdresser pleased laughed out loud and then she told her friends.
17. The analyst who the traffic / the electrician aggravated was late for the important meeting and the executive seemed very annoyed.
18. The writer who the editor / the dog ignored became depressed after a while and it took him weeks to start writing again.
19. The resident who the landlord / the girlfriend lied to sued a month ago but never got a response.
20. The politician who the mobster / the reporter threatened got kidnapped last Friday but the police arrived soon after to rescue him.

Appendix N

Materials for Experiment 2B-2.

Each of the 24 items had two versions, as shown in (1). The second version of each item can be generated in a similar way.

1. At the press-conference a senator and two reporters got into an argument. The senator attacked one of the reporters and then the other reporter attacked the senator.
Mary: I heard that the reporter that attacked the senator admitted to making an error.
John: I am not sure about that. I heard that the reporter that the senator attacked admitted to making an error.

At the press-conference a reporter and two senators got into an argument. The reporter attacked one of the senators and then the other senator attacked the reporter.
Mary: I heard that the senator that attacked the reporter admitted to making an error.
John: I am not sure about that. I heard that the senator that the reporter attacked admitted to making an error.

2. During the interview a newscaster and two musicians had a brief discussion. The newscaster insulted one of the musicians and then the other musician insulted the newscaster.
Mary: I heard that the musician that insulted the newscaster left the building after the interview.
John: I am not sure about that. I heard that the musician that the newscaster insulted left the building after the interview.

3. In the store a customer and two cashiers talked about recent events. The customer thanked one of the cashiers and then the other cashier thanked the customer.
Mary: I heard that the cashier that thanked the customer supported the new bill about immigration rules.
John: I am not sure about that. I heard that the cashier that the customer thanked supported the new bill about immigration rules.

4. After the lecture a scientist and two interns went over the problem set. The scientist confused one of the interns and then the other intern confused the scientist.
Mary: I heard that the intern that confused the scientist worked at a famous lab at Harvard University.
John: I am not sure about that. I heard that the intern that the scientist confused worked at a famous lab at Harvard University.

5. After the incident an officer and two detectives talked about possible suspects. The officer approached one of the detectives and then the other detective approached the officer.
Mary: I heard that the detective that approached the officer had a good record in solving similar cases
John: I am not sure about that. I heard that the detective that the officer approached had a good record in solving similar cases

6. In the office a secretary and two co-workers got along very well. The secretary welcomed one

of the co-workers and then the other co-worker welcomed the secretary.

Mary: I heard that the co-worker that welcomed the secretary brought some flowers to the office.

John: I am not sure about that. I heard that the co-worker that the secretary welcomed brought some flowers to the office.

7. Over the weekend a dentist and two pediatricians talked about various medications. The dentist called one of the pediatricians and then the other pediatrician called the dentist.

Mary: I heard that the pediatrician that the dentist called left a message about the recommended dosage.

John: I am not sure about that. I heard that the pediatrician that called the dentist left a message about the recommended dosage.

8. Before the surgery a physician and two neurologists had a long discussion. The physician helped one of the neurologists and then the other neurologist helped the physician.

Mary: I heard that the neurologist that the physician helped worked at MGH for the last ten years.

John: I am not sure about that. I heard that the neurologist that helped the physician worked at MGH for the last ten years.

9. During the event an author and two critics argued about the point of a book. The author annoyed one of the critics and then the other critic annoyed the author.

Mary: I heard that the critic that the author annoyed had strong opinions about many things.

John: I am not sure about that. I heard that the critic that annoyed the author had strong opinions about many things.

10. Before the class a teacher and two students went over the homework. The teacher greeted one of the students and then the other student greeted the teacher.

Mary: I heard that the student that the teacher greeted gave an interview to the school newspaper recently.

John: I am not sure about that. I heard that the student that greeted the teacher gave an interview to the school newspaper recently.

11. After the interview a politician and two journalists argued about the new law. The politician criticized one of the journalists and then the other journalist criticized the politician.

Mary: I heard that the journalist that the politician criticized left the room around 3pm.

John: I am not sure about that. I heard that the journalist that criticized the politician left the room around 3pm.

12. At the convention a researcher and two inventors displayed innovative technology. The researcher praised one of the inventors and then the other inventor praised the researcher.

Mary: I heard that the inventor that the researcher praised patented several inventions over the last five years.

John: I am not sure about that. I heard that the inventor that praised the researcher patented several inventions over the last five years.

13. At the debate a congressman and two governors argued about the national election. One of the governors antagonized the congressman and then the congressman antagonized the other governor.

Mary: I heard that the governor that antagonized the congressman apologized for being too aggressive.

John: I am not sure about that. I heard that the governor that the congressman antagonized apologized for being too aggressive.

14. Earlier this month an interpreter and two ambassadors planned a trip. One of the ambassadors contacted the interpreter and then the interpreter contacted the other ambassador.

Mary: I heard that the ambassador that contacted the interpreter lived in Africa for many years.

John: I am not sure about that. I heard that the ambassador that the interpreter contacted lived in Africa for many years.

15. After the meeting an administrator and two managers examined the accounting books. One of the managers questioned the administrator and then the administrator questioned the other manager.

Mary: I heard that the manager that questioned the administrator had a problem with the company.

John: I am not sure about that. I heard that the manager that the administrator questioned had a problem with the company.

16. During the lecture an anthropologist and two historians discussed the article. One of the historians challenged the anthropologist and then the anthropologist challenged the other historian.

Mary: I heard that the historian that challenged the anthropologist published a famous book on the same topic.

John: I am not sure about that. I heard that the historian that the anthropologist challenged published a famous book on the same topic.

17. Before the meeting an instructor and two counselors talked about teaching methods. One of the counselors offended the instructor and then the instructor offended the other counselor.

Mary: I heard that the counselor that offended the instructor regretted the comment after the presentation.

John: I am not sure about that. I heard that the counselor that the instructor offended regretted the comment after the presentation.

18. In the boardroom a manufacturer and two analysts talked about the deal. One of the analysts consulted the manufacturer and then the manufacturer consulted the other analyst.

Mary: I heard that the analyst that consulted the manufacturer was involved in a scandal not long ago.

John: I am not sure about that. I heard that the analyst that the manufacturer consulted was involved in a scandal not long ago.

19. At the laboratory a mathematician and two engineers talked about the project. One of the engineers impressed the mathematician and then the mathematician impressed the other

engineer.

Mary: I heard that the engineer that the mathematician impressed received an award at a recent conference.

John: I am not sure about that. I heard that the engineer that impressed the mathematician received an award at a recent conference.

20. A week ago a writer and two artists met at the museum. One of the artists upset the writer and then the writer upset the other artist.

Mary: I heard that the artist that the writer upset left the museum in a bad mood.

John: I am not sure about that. I heard that the artist that upset the writer left the museum in a bad mood.

21. At the gym a gymnast and two wrestlers were training for an upcoming meet. One of the wrestlers observed the gymnast and then the gymnast observed the other wrestler.

Mary: I heard that the wrestler that the gymnast observed attended the college on a scholarship.

John: I am not sure about that. I heard that the wrestler that observed the gymnast attended the college on a scholarship.

22. Before the reunion a chef and two caterers talked about the food. One of the caterers frustrated the chef and then the chef frustrated the other caterer.

Mary: I heard that the caterer that the chef frustrated was famous for his mushroom soup recipe.

John: I am not sure about that. I heard that the caterer that frustrated the chef was famous for his mushroom soup recipe.

23. At a construction-site a bricklayer and two carpenters worked near one another. One of the carpenters assisted the bricklayer and then the bricklayer assisted the other carpenter.

Mary: I heard that the carpenter that the bricklayer assisted had twenty years of experience.

John: I am not sure about that. I heard that the carpenter that assisted the bricklayer had twenty years of experience.

24. Before the press-conference a lawyer and two legislators talked among themselves. One of the legislators cautioned the lawyer and then the lawyer cautioned the other legislator.

Mary: I heard that the legislator that the lawyer cautioned was reported to have considered taking bribes.

John: I am not sure about that. I heard that the legislator that cautioned the lawyer was reported to have considered taking bribes.