Eye Movements as a Function of Spoken Sentence Comprehension and Scene Perception

Julia C. Di Nardo

A Thesis

in

The Department

of

Psychology

Presented in Partial Fulfillment of the Requirements
for the Degree of Master of Arts (Psychology) at
Concordia University
Montreal, Quebec, Canada

August 2005

Library and
Archives Canada

Published Heritage
Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et
Archives Canada

Direction du
Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence
ISBN: 0-494-10180-6
Our file Notre référence
ISBN: 0-494-10180-6

Abstract

Eye Movements as a Function of Spoken Sentence Comprehension and Scene Perception

Julia C. Di Nardo

The purpose of this research was to examine how linguistic and visual processes interact, and to determine the locus of this interaction. Three studies were conducted to this aim. The first, a normative study, asked participants to list the objects they saw after viewing 17 naturalistic scenes for a short time in order to establish which objects were most salient. The second and third studies used an eye-tracking methodology to record participants' eye movements as they looked at static and dynamic scenes (the same as those in the Normative Study) and listened to related sentences. The sentences differed with respect to the class of their main verbs (causative *vs.* perception/psychological verbs), and the scenes differed according to the positioning and motion (or apparent motion in the static scenes) of the human agent (towards, away, or neutral) relative to the object referent of the verb's direct object (the target object). Results indicate that the direction of apparent motion of the agents in the scenes did affect how frequently the target object was listed by participants in the Normative Study. For the two experiments, the only consistent finding was that motion context had a significant main effect on how quickly eye movements were launched towards the target object after the verb was uttered. The effect of verb type was found to be less consistently significant, depending on the statistical procedure employed, which fails to support the notion that verb-specific information can guide eye movements towards the target object.

Acknowledgments

First and foremost, I am indebted to my supervisor, Dr. Roberto de Almeida, for his support, both intellectual and otherwise. This work would not have been possible without his brilliance and insight into the workings of the human mind. For their acting talents, thanks go to Ioana Constantinescu, Jake Godfrey, Caroline de Almeida and Daniel de Almeida. Thanks also to Zinnia Madon for lending her voice to the spoken sentences, as well as for her help in preparing the stimuli and running the experiment, and to Caroline de Almeida for filming and editing the movies. For allowing me the use of his eye-tracker, I would like to thank Dr. Michael von Grünau. My gratitude also extends to the laboratory's research assistants, for their help in recruiting and running participants.

I would also like to thank my parents, without whose support and encouragement I would never had made it this far. Their lessons have taught me that hard work, persistence and dreaming big are what bring you satisfaction in life. Steve, thanks for being there throughout it all.

Last, but not least, I am grateful for the financial support I have received, in the form of a graduate fellowship from NSERC, and for the opportunity to make my work known to the scientific community through the support of departmental travel grants and Dr. de Almeida's grants.

Table of Contents

List of Figures

List of Tables

Eye Movements as a Function of Spoken Sentence Comprehension and Scene Perception

How is the mind organized? A broad but important topic in the field of cognitive science, formulating a blueprint of the mind's cognitive architecture can help draw the lines among its various subsystems, pointing to its basic and fundamental organization. Fodor's (1983) modularity theory provides one such blueprint, focusing on the cognitive processes that give rise to our uniquely human experience of the world. If we take cognition to be a theory of information processing, then an account of the different processes and their structure, relation to each other, timing and sequence of execution can inform us as about the structure of the mind. The collection of studies reported here attempt to clarify this organization by examining how linguistic and visual variables computed by both the language and visual systems affect eye movement behaviour. By examining the pattern and timing of eye movements across static and dynamic scenes, we can better understand how two of our most important cognitive faculties, language and vision, interact with each other to govern our experience of the world and how to behave within it.

Three studies were conducted to this aim. The first, a normative study, was conducted to establish which objects in various naturalistic scenes were most salient by asking participants to list the objects they saw after viewing the scenes for a short time. In addition, they were asked to generate sentences regarding present and future events within those scenes. The second and third studies used an eye-tracking methodology to record participants' eye movements as they looked at static and dynamic scenes (the same as those used in the Normative Study) and listened to sentences related to those scenes. The sentences differed with respect to the class of their main verbs (causative *vs.* perception/psychological verbs), and the scenes differed according to the positioning and motion (or apparent motion in the static scenes) of the human agent, who was the referent of the subject of the sentences. Results are discussed in the context of the interaction of the language and vision modules, and conceptual representations computed within conceptual short-term memory (CSTM; Potter, 1993, 1999).

Modularity of Mind

Many theorists describe the mind as modular, a computational theory of cognitive architecture based on Fodor's (1983) modularity of mind hypothesis. At its most basic level, the theory posits that the mind is organized along two levels of hierarchy: at the bottom, or input level, there are a series of modules, or specialized processors, that deal with only certain types of perceptual information, such as visual or linguistic input. These mental modules operate in parallel, outside of conscious awareness, and their main function is to match the incoming sensory information with stored representations, thus functioning as perceptual recognition machines. Although this type of cognitive processing occurs outside of conscious awareness, its products, or outputs, are available to consciousness (e.g., hearing your own name in a crowd). It is the recognition process itself that is not available to consciousness and therefore is not subject to conscious control.

The modules are not involved in low-level sensory processing (e.g., colour, pitch), but rather take the sensory representations (e.g., visual objects, spoken words) as their input. An important distinction to make is that between cognitive and noncognitive perceptual processes. Cognitive processes make reference to information or representations stored in memory while noncognitive processes do not, making the processing computed by mental modules inferential. Noninferential processes include the early stages of visual or auditory processing, which do not refer to stored representations.

At the top of the hierarchy are the central processors, which deal with higher level cognitive and conceptual processing. These include functions such as reasoning, planning, goal-directed attention, and long-term memory retrieval and encoding. In contrast to the processing executed by the modules, the computations performed by the central processors are under conscious control. Although both modular and central processes are purported to be cognitive processes, the main distinction between them is whether or not they are open to conscious control.

Fodor (1983) is mainly interested in describing the properties of the inferential but nonconscious cognitive processes, the so-called modules. The first defining feature of a module is domain-specificity, meaning that the processing they perform is limited to a single domain, such as linguistic or visual information. Second, their operation is said to be mandatory; one cannot help "hearing" a sentence in one's own language, just as one cannot help "seeing" a visual stimulus. Third, they are informationally encapsulated, such that they are not subject to the influences of either other modules or higher-level central processes (a feature also known as cognitive impenetrability). Visual illusions such as the Müller-Lyer illusion illustrate this property (see Figure 1); possessing the knowledge that one line is not in fact longer than another does not affect the visual processing of the lines, allowing the illusion to persist. Fourth, because modules are domain-specific and informationally encapsulated, their processes are fast, so that they do not require extra

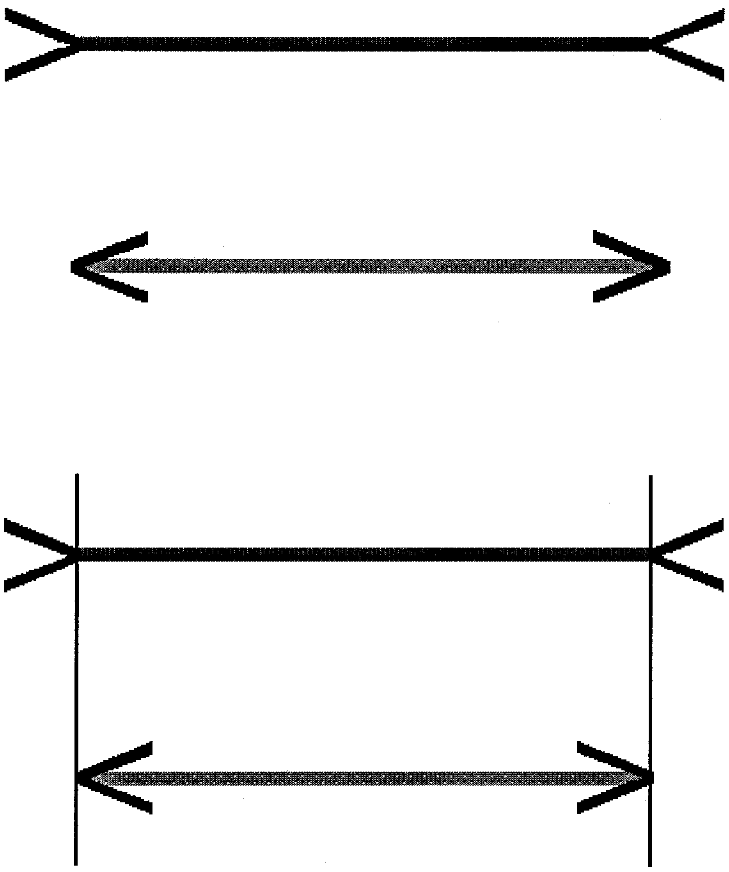


Figure 1. The Müller-Lyer illusion. In the top half, the first line is perceived as being longer than the second, although the second half shows that both parallel lines are exactly the same length. This is an illustration of the notion of cognitive impenetrability, where "knowing" that something is not true (that one line is not longer than the other) does not suppress the visual processing that allows the illusion to persist.

time referencing extramodular information. Fifth, and finally, modules have shallow outputs, which undergo more complex transformations at the central level.

The central systems are thought to be mainly involved in conceptual processing (Fodor, 1983). At a general level, it comprises what most of us consider "conscious thought," and many of its processes are open to conscious control. Unlike the computations performed by the modules, which are said to be mandatory (e.g., the perceptual recognition of a word), those that fall under the rubric of the central processors are not (e.g., the calculation of an arithmetic problem). Although the computations performed by the modules are cognitive, in the sense that they are not involved in basic sensory processing, they can be distinguished from the central processes on the basis of an important feature: encapsulation. Modular functions are encapsulated, while central processors by hypothesis are not. This means that higher-level functions such as memory and attention can have information from many different modules as their input; their processes are not dedicated to any one particular information type. They integrate sensory and perceptual information into a stable, coherent whole. The importance of this point is paramount: this implies that, according to Fodor, any interaction between two or more sources of input must take place at the central level. The central processors have access to both the outputs of the modules and information stored in long-term memory. Any integration, therefore, of information from the environment must be fully processed by the modules (i.e., be matched to some other information already stored in memory) before it outputs to the central level. Because of this, the processing performed by the central systems is relatively slow and highly flexible, allowing us to purposely plan and direct our behaviour. Because central processes are slow and limited in capacity, not all outputs from all modules can be integrated or held in conscious awareness at once; therefore, selective attention is one of the important functions (mechanisms) performed by the central processors, selecting information which is most relevant to the task at hand.

In summary, then, there are several types of processes that can take place: sensory processing, which is noninferential and encapsulated; perceptual or modular processing, which is encapsulated but inferential; and central processing, which is inferential but not encapsulated. Although most researchers accept that the various modules interact at some level (usually taken to be the central level, according to Fodor, 1983), the exact locus at which these interactions occur is under debate. The debate mostly focuses on whether certain findings are congruent with Fodor's specifications for a module (e.g., that it be fast, mandatory, etc.) or not, as well as whether interactions can occur both between modules and among the various levels within a module. In the next section, I will outline some of the objections to Fodor's conceptualization of the cognitive architecture, as well as some alternatives, and present some hypotheses as to the nature of mental representations and the locus of their interaction.

The Modularity Debate

One of the major objections to modularity theory came from a body of evidence in linguistic research conducted by Marslen-Wilson and Tyler (1989). Their central argument is that the processes that are involved in discourse analysis, which Fodor would subsume under the central level, in fact share many of the properties considered exclusive to modular processes. Specifically, they argue that pragmatic information (real-world knowledge) interacts with low-level linguistic parsing, and that the distribution of properties typical of modular processes is not as neatly separated from central processes as Fodor claims. One claim that Marslen-Wilson and Tyler (1989) make, pertinent to the work presented here, is that mandatoriness is not always limited to modular processing. Some processes typically considered central, e.g., pronoun assignation in discourse analysis, can in fact occur automatically. Therefore, by logical extension, if mandatory processing is not diagnostic of a module, then one could make the argument that a process that does not occur automatically could be considered exclusive to the central level. In other words, if a process does occur automatically, it can belong to either the modular or central level, but if it does not, it cannot belong to the modular level and must necessarily be part of the central level.

Marslen-Wilson and Tyler (1989) make a similar point with regards to the property of speed: very fast processes are likely mandatory, but not all mandatory processes are fast (their example is of getting old; unfortunately mandatory, but thankfully slow). Taken with the point made above, if a process is not fast, and not mandatory, it is therefore a central process, although if it is fast and mandatory it may be either central or modular. However, the aging process is not a cognitive process, and therefore this argument does not hold. In addition, there are no guidelines regarding "how fast is fast" when it comes to modular processes. This lack of a definite quantitative threshold cannot allow speed, alone, to be diagnostic of a modular process. A central level process may still occur within a very short amount of time and still be post-modular. Therefore, only an obviously slow process can point towards a central level process, while a relatively fast one may point towards either a modular or central level process.

With regards to informational encapsulation, arguably the most central property of modularity, Marslen-Wilson and Tyler (1989) point out a loophole in Fodor's conceptualization: top-down processing cannot occur between modular and central levels (such that central processes cannot influence the stepwise bottom-up processing conducted by the modules), but no claims are made towards top-down influences on modular processing. They refer to this as autonomous processing, where the processing within a single informational domain (e.g., language) can occur in an interlevel interactional, autonomous fashion, so long as real-world or conceptual knowledge contained within the central level does not interact with the representations that serve as inputs to modular processes. However, Fodor does concede that top-down processing can occur within a single modular system, with the exception of conceptual knowledge (including beliefs or expectations).

A theory consistent with this idea was put forth by Jackendoff (1987a), a theory that is alternative to, but not entirely incompatible with, Fodor's modularity hypothesis. Jackendoff has proposed a language processing model wherein the various levels of linguistic analysis (acoustic, phonological, syntactic and conceptual) interact in both a bottom-up and top-down manner. However, while the interlevel influences can occur both adjacently and nonadjacently (e.g., the first level can interact with the third level, skipping over the second), these influences can only occur between adjacent incoming acoustic segments (taken to be individual phonemes). Jackendoff terms this the interactive parallel processing hypothesis, because the various levels (within the language module) interact, but in a parallel fashion (i.e. the levels interact for any two adjacent acoustic segments in parallel). This conceptualization explains the mapping of acoustic information onto conceptual representations (bottom-up processing), as well as the mapping of conceptual structure onto phonological information (top-down processing), in addition to the integration of newly available phonological, syntactic and conceptual information into a unified structure. In summary, if we consider the conceptual level to be part of the central level, Jackendoff's model can account for the notion that representations being combined at the central level are in a constant state of flux, depending on the changing input the modules receive. Note that this model does not speak to the interaction of modules at the modular level, and hence is not incompatible with Fodor's modularity hypothesis.

Mental Representations

A discussion of modular interactions at the central level begs the question of what language each system uses. What is the nature of the outputs of the various modules, linguistic or otherwise? Fodor (1975) proposes that there is a "language of thought" wherein there are compositionally complex symbols that function as mental representations. Thinking constitutes the combination of these representations in much the same way as the combinatorial syntactic structure of language. Central to the work presented here is the assumption that thought takes place with some form of mental representations, whatever their nature may be, and regardless of whether the representations computed by each module differs across the modules. For our purposes, we assume these to take the form of concepts which get activated by stimuli in the environment (of course, these representations may be activated without external stimulation, as would occur when generating a spontaneous thought, but this point is not relevant to the work presented here). In addition, these representations all take the same form *at the central level*, regardless of the modality from which they were activated (i.e., whether linguistic, visual or otherwise).

Once activated, these concepts combine to influence thought and behaviour. The format that these structurally complex combinations may take has been outlined by Jackendoff's Conceptual Semantics (1987b). He proposes that there are a number primitive conceptual categories, which include objects, events, states, place, as well as others (see Jackendoff, 1987b and 1983 for more on this theory). These combine according to formation rules and can be represented in the form of propositional structures (first proposed by Kintsch, 1974 but consistent with Jackendoff's Conceptual Semantics). For example, the sentence (and its mental equivalent), "The woman is eating dinner," which describes an event, can be expressed in the form shown in (1), where EAT is the event being described, WOMAN is the entity performing the event (the agent), and DINNER is the entity upon which the event is being performed (the patient).

(1) $[_{\text{EVENT}}$ EAT (WOMAN, DINNER)]

Similarly, a given state of affairs, such as "The woman is in the kitchen," can be expressed in the form shown in (2), where BE represents the state, WOMAN is the entity is the state, and KITCHEN is the place constituting the state.

(2) $[_{\text{STATE}}$ BE (WOMAN, KITCHEN)]

Each constituent of this propositional structure can refer to specific objects in the environment but also necessarily belong to some conceptual category. This is why a sentence such as, "The dinner is eating the woman" is syntactically possible but conceptually ill-formed: the conceptual properties of the verb do not licence the entities in the agent and patient roles. Dinners have no agentive properties and therefore cannot eat, and although women can be eaten, this is such an infrequently occurring event that they do not make typical patients for the verb "eat" (here, "typicality" refers to semantic selectional restrictions, not categorical typicality).

To summarize, mental representations that serve as outputs from the various modules, by hypothesis, take the form of concepts are combined in manner analogous to the formation rules proposed by Jackendoff (1983, 1987b; note that this view does not necessarily endorse Jackendoff's lexical-decompositional view). The resulting propositional structures can refer to states or events, among others, but for events, it is important to note that the conceptual properties of a given event descriptor (a verb, within the language domain) restrict which subjects and patients can be associated with it. At this point, it is useful to review how verb meanings are constructed, because of the central role they play in the studies reported here, and because of their unique ability to semantically refer to a range of object and event/states.

Verb Structure

Verbs have complex sets of features (Levin, 1993), which include the range of meanings and argument expressions that a given verb can have. Rappaport Hovav and Levin (1998) have proposed a theory that can account for this variation. For example, the verb "wipe" can take one or two arguments (e.g., "The woman wiped the table," "The woman wiped the crumbs into the sink), or none at all (e.g., "The woman wiped"). According to Rappaport Hovav and Levin (1998), this variation is not random, but rather can be accounted for by the class to which a verb belongs. Each verb class displays certain predictable patterns, as can be illustrated by comparing "manner verbs" and "result verbs." Manner verbs specify *how* an action is carried out, while result verbs specify the *result* of an action. With regards to the arguments they can take, one example is that manner verbs allow for the omission of the direct object (e.g., "The woman wiped"), while result verbs do not (e.g. * "The woman broke").

Of interest to us is one particular subcategory of result verbs: causative verbs, also known as accomplishment verbs, which denote an activity and its resulting state, such as the verb "spill." Their lexical structure, according to Rappaport Hovav and Levin, takes the form [[x ACT] CAUSE [BECOME [y <*STATE*>]]], where some "x ACT" causes a given object to become "y <*STATE*>," or spilled. This decomposition of the causative class of verbs is consistent with Jackendoff's (1983, 1987b, 1993) view, although they disagree as to the representational level at which this type of decomposition takes place. Rappaport Hovav and Levin believe it occurs at a linguistic level of semantic representation, while Jackendoff believes it occurs within a central conceptual structure.

These causative verbs are of interest to us for two reasons. First, they denote actions ("x ACT") which can be visually represented (or alluded to, in the case of the work presented here) in the form of both static (Experiment 1) and dynamic (Experiment 2) events. Second, they are semantically restrictive, in the sense that the direct objects they can take are restricted by the semantic properties inherent to the verb concept. In other words, "spill" can only take patients that are spillable, such as milk and other liquid substances, and cannot take patients that do not possess the property of being spillable, such as various solid state objects. This is in contrast to a class of verbs composed of perception and psychological verbs, which refer to the perception of some object by an experiencer (e.g., "see"), or a psychological event (e.g., "inspect"). Perception/psychological verbs are much less semantically restrictive in the sense that they can refer to any perceptible object.

McRae's view of thematic roles also fits with this idea of verb-specific concepts (McRae, Ferretti, & Amyote, 1997). Thematic roles "represent the roles that participants play in the events described by verbs" (Tanenhaus, Carlson, & Trueswell, 1989, as cited by McRae et al., 1997, p. 138). These participants include agents and patients, as described above, and can be termed "role fillers" according to this theory. McRae and his colleagues believe that thematic roles are conceptual in nature, containing information about the typical entities that participate in the events described by verbs (an idea originally proposed by Dowty, 1991). In particular, this knowledge can guide the assignment of noun phrases (referents of real-world people and objects) to the various roles licensed by a verb during language comprehension (Carlson & Tanenhaus, 1998), a point crucial to our thesis. For a discussion of the nature of these role concepts, and the evidence supporting such a formulation, see McRae et al. (1997).

In sum, verbs can be seen as having a conceptual structure similar to that proposed by Jackendoff (1983, 1987b), which refer to certain events and their likely subjects and patients. In addition, different verb classes licence different argument structures, and vary with respect to their semantic restrictiveness. According to Fodor (1983), the language module processes syntactic information and determines the permissible fillers for the various syntactic categories. However, given that conceptual information may be activated from various modalities (e.g., visual and linguistic), this information must be integrated at some point to allow for a coherent experience of the world. We propose that this occurs at a post-modular conceptual short-term memory store, as outlined by Potter (1993, 1999), and discussed in the section below.

Conceptual Short-Term Memory

Conceptual short-term memory (CSTM) is a form of very short-term memory where conceptual representations are activated, for a very brief period of time, at the early stages of perceptual processing, memory retrieval and thought (Potter, 1993). This form of memory is distinct from iconic or echoic memory as well as the form of short-term memory measured by memory span. Potter's (1993, 1999) hypothesis holds that once a stimulus is identified, its meaning is activated within CSTM but rapidly lost if it is not integrated within long-term memory (LTM) and other ongoing processes. An important feature of CSTM is a matching process that occurs between the contents of CSTM and LTM, analogous to the inferential processing described by Fodor's (1983) modularity hypothesis, making it central to cognitive but not perceptual processing. CSTM conceptually structures the activated representations, linking them with semantically related information stored within LTM, thus simultaneously activating this information. Because of the multitude of representations that are activated at any given time, only those that are meaningfully structured and organized will be retained by LTM. For a review of the evidence supporting this hypothesis, see Potter (1999).

The notion of a CSTM store is important for the work presented here because it provides a framework for understanding where the outputs of the various modules may interact. If concepts are activated by various stimuli from several modules, then CSTM seems a likely candidate for the locus of the integration of these concepts both to create a LTM store and a coherent experience of the world. Given that the contents of CSTM are rapidly changing as information from our dynamic environment is processed, only that which is conceptually relevant and well-structured is selected for further processing.

How does CSTM relate to modularity? If it is assumed that the representations that activated at any given time within CSTM are conceptual in nature, it seems likely that the processing that occurs at this level is post-modular. However, at the present time, no claim can be made regarding whether CSTM is strictly a central level process. In fact, based on Potter's (1999) claim that CSTM activation and selection are unconscious (in Fodor's sense) and extremely rapid, it seems unlikely that CSTM is a central process. O'Conner and Potter (2002) also suggest that CSTM is unlike other forms of STM in that it is not under conscious control. Therefore, based on these claims, and for the purposes of the present study, we hypothesize that CSTM is a post-modular but pre-central process that is partly responsible for the interaction and integration of conceptual representations outputted by the modules. In fact, CSTM may be a horizontal faculty (as discussed by Fodor, 1983), but one that serves to integrate rapidly activated representations. Attention (whether stimulus-driven or goal-directed) may serve to bond these representations within CSTM. In addition, the nature of the representations computed by CSTM also has relevance to modularity. If these representations, which we purport to be conceptual in nature (as conceptualized by Rappaport Hovav and Levin's (1998) and Dowty's (1991) view that the information contained within a verb is conceptual and not linguistic), are in fact so, then they are not module-specific and are computed at the central, and not the modular, level.

To summarize, some alternatives to Fodor's strict differentiation between the modular and central levels of cognitive processing have been proposed. One view held by Jackendoff (1987) is that interactions can occur between the various levels within a single module. He also proposes that the mental representations of states and events can be decomposed into proposition-like structures referring to the event/state and its participants, an idea consistent with current beliefs about the representational structure of verbs (Levin, 1993; McRae et al., 1997). Importantly, these representations likely contain semantic conceptual information regarding typical participants (agents and patients, or role fillers; Dowty, 1991). During cognitive processing, we propose that the concepts activated by our environment interact within CSTM (Potter, 1999). What remains unclear, however, is the evidence supporting modularity as described here. In the next section, the framework for the research reported here is outlined, and studies examining the interaction of language and visual processing, as they pertain to the modularity debate, will be reviewed.

The Interaction of Vision and Language

One approach to studying intermodular interactivity is by examining how linguistic and visual processing influence each other. Because these two domains are two of the main ways in which we understand the world, they are central to any study of cognitive processing. Language has the important property of referring to entities in the real world, which can be perceived by the visual system. Thus, they are ideally suited to study using the eye-tracking paradigm (Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). Because most work in the domain of modularity research has focused primarily on studies of language and vision, basic linguistic and visual processes pertinent to the methodology of the debate will also be reviewed. The next section reviews the eye-tracking methodology employed in research of a similar nature to that presented here.

Eye-Tracking Paradigm

One way of attempting to resolve the modularity debate is by using an eye-tracking methodology. At its core, the “visual world” paradigm consists of the recording of eye movements as participants are presented with auditory and visual stimuli, typically spoken sentences accompanied by static scenes. It is based on the assumption that eye movements reflect on-line cognitive processing, specifically the interaction between spoken language and visual processing (Tanenhaus & Spivey-Knowlton, 1996). The advantages of this methodology are that language processing can be examined non-invasively and with a very precise temporal resolution. In fact, it has been found that fixations to objects are closely time-locked to the auditory recognition of their names (Allopenna, Magnuson, & Tanenhaus, 1998). In addition, language processing can be studied within a realistic visual context, which can aid in the comprehension of spoken language unfolds, especially if the spoken language directly relates to the visual context (Eberhard, Spivey-Knowlton, Sedivy, & Tanenhaus, 1995). Finally, as Eberhard and colleagues (Eberhard et al., 1995, p. 435) point out, “studying language in impoverished situations encourages, rather than challenges, a modular information-encapsulation view of the system” – impoverished in the sense that there is no relevant visual context. This has important implications for the modularity debate.

One of the earliest studies to employ an eye-tracking technique to demonstrate the role of eye movements as visible indices of cognitive events was conducted by Yarbus (1967), which showed that eye movement patterns across paintings differed depending on the instructions given to participants. The fact that eye movement patterns were not determined by the stimulus demands alone, but were dependent on the task demands, suggested that eye movements are coupled to high-level cognitive processes.

The first eye-tracking study to investigate eye movements during spoken language comprehension was conducted by Cooper (1974). This work demonstrated that eye movements are directed towards line drawings of objects semantically related linguistic elements (such as words and phrases) presented in spoken sentences. Furthermore, objects were often fixated prior to the offset of their noun referents, providing evidence for the anticipatory abilities of eye movements.

These two early studies point to the importance of eye movements in understanding cognitive processing, and specifically language processing. Prior to a discussion of the literature pertaining to this methodology, a brief review of the basic processes involved in visual perception and cognition and their relevance to the eye-tracking paradigm is in order.

Eye Movements and Visual Processing

The centrality of eye movements and fixations in the eye-tracking paradigm warrants a brief discussion of their properties. Saccades are one form of voluntary eye movements (although they can sometimes occur involuntarily to highly visually salient features); they are the sudden, rapid and jumpy eye movements made as we scan our visual environment. Saccades have two phases to their execution: the saccade itself, and the fixation, where the eyes are held constant in position so that an object of interest can be projected onto the fovea. Approximately 100-300 ms are required to initiate a saccade (from the time a stimulus is presented until the eye begins its movement), and another 30-

120 ms to complete the saccade, depending on the visual angle (distance across the visual field) traversed (Henderson & Ferreira, 2004).

Saccadic eye movements are the primary means for selective visual attention to be allocated to our environment; they are how we focus our attention to important aspects of our visual world. Therefore, recording eye movements can shed light on the underlying cognitive processes important for selecting aspects of the visual environment (Henderson & Ferreira, 2004). Importantly, however, this relationship is not entirely one-to-one: shifts of attention can be accomplished without a corresponding shift in gaze (Fischer & Breitmeyer, 1987). In addition, the standard 200 ms that has been assigned for the length of time needed to initiate a saccade (the so-called "express saccade") only applies when the attentional system is disengaged: when it is engaged, longer times are required to initiate a new saccade because the attentional system must disengage from the currently fixated location (Fischer & Breitmeyer, 1987). However, according to Kowler (1999), the programming of saccadic eye movements requires little attention on the whole. This suggests that attentional resources can be freed up for high-level cognitive processing, such as deciding upon relevant saccadic targets – which is the very process we are measuring in the two experiments reported below.

An understanding of how eye movements in the visual world paradigm informs visual processing requires a consideration of scene processing. Scene processing can be regarded as a special form of visual processing in general. A scene can be defined as a "semantically coherent (and often nameable) human-scaled view of a real-world environment comprising background elements and multiple discrete objects arranged in a spatially licensed manner" (Henderson and Hollingworth, 1999, as cited by Henderson &

Ferreira, 2004, p. 5). Scene recognition, and in fact the extraction of a scene's "gist" (defined as the semantic information related to the scene category, such as a kitchen) can occur very rapidly, in less than the time required for a single eye fixation (Henderson & Ferreira, 2004). Interestingly, this process is thought not to require attentional processes.

A series of experiments conducted by Potter (1969, 1975, 1976; see a review in Potter, 1999) indicated that scene identification can occur very quickly, at presentation rates of up to 125 ms per scene, when the task is to identify a scene as specified by either a verbal label or a picture. However, this scene identification information is not retained without some kind of conceptual framework to structure the material (as specified by the CSTM model). In addition, similar work by Intraub (1999) suggests that rapidly presented scenes (at rates from 110 ms to 333 ms) can still allow for the momentary extraction of the scenes' gists, even though this information is almost as rapidly forgotten. These results indicate that gist detection can occur within approximately 125 ms. More recent work has shown that scenes can be categorized in as little time as 20 ms (VanRullen & Thorpe, 2001). ERP studies of object identification within natural scenes (Thorpe, Fize, & Marlot, 1996; VanRullen & Thorpe, 2001) indicate that this seemingly complex cognitive process can occur in less than 150 ms. Taken as a whole, these results suggest that early scene processing, including gist extraction and object identification occurs very rapidly.

There are two types of useful information that can be gleaned from eye movement patterns across scenes: where the fixations are directed and how long the fixations last. With regards to fixation durations, Henderson and Ferreira (2004) report that these are longer for more visually and semantically informative regions. With regards to fixation location, early studies using the eyetracking methodology (e.g., Yarbus, 1967; Buswell, 1935; as reported by Henderson & Ferreira, 2004) found that the more informative a particular region of a scene is, the more fixations it will receive. However, how to define "informativeness" depends on several factors. Henderson and Ferreira (2004) report various low-level (bottom-up) visual features such as local contrast (large variability of intensity in a given region) and a large difference between the intensity of a given point and its surroundings tend to attract more fixations. However, overall, features such as colour, contrast, intensity, symmetry and colour density do not appear to correlate well with fixation position, and in fact these correlations tend to increase when the visual patterns are meaningless.

Top-down influences in fixation location determination have not been well-investigated. Henderson and Ferreira (2004) speculate the role that these factors play is rather important, given that fixation targets are likely selected very early on in scene viewing, as it has been shown that semantic information about a scene can be gleaned within a fraction of a second. See Henderson and Ferreira (2004) for a review of different forms of cognitive (top-down) factors that influence fixation location. Of particular relevance is what they term "scene schema knowledge," which is the generic knowledge about scenes that arises from semantic regularities across multiple encounters with those scenes. These semantic regularities can refer to spatial location and objects typically found together in a given scene type. This type of knowledge can implicitly influence visual attention.

In summary, then, scene identification can occur very quickly without requiring a full scan of the entire scene, nor the identification of any objects within it (Henderson &

Ferreira, 2004). In addition, fixation location early in scene viewing is determined mainly by visual saliency (Henderson & Ferreira, 2004), although cognitive factors also play a role in the determination of fixation location (a conclusion also reached by Kowler, 1999).

Given that scene gist identification occurs at such an early stage, what are the implications for the interaction of scene processing and language comprehension? The next section will review the eye-tracking studies that have examined the effects of visual context on language processing.

Context Effects on Language Processing

The main premise of the visual world eye-tracking paradigm, as it relates to modularity, is that the encapsulation of the language processing module can be examined by determining whether language comprehension is affected by the visual context at the early stages of processing (Tanenhaus et al., 1995). The belief in psycholinguistics is that syntactic processing is encapsulated from other cognitive and perceptual modules, due to a body of evidence suggesting that syntactic ambiguities can be resolved without reference to contextual factors. However, there is a second line of research, employing the eye-tracking methodology, which suggests that natural behavioural contexts do influence language comprehension at very early stages of comprehension (Tanenhaus et al., 1995).

In a series of recent studies, Tanenhaus and his colleagues have shown that static scene perception and sentence comprehension are interactive (Chambers, Tanenhaus, Eberhard, Filip, & Carlson, 2002; Eberhard et al., 1995; Sedivy, Tanenhaus, Chambers, & Carlson, 1999; Spivey, Tanenhaus, Eberhard, & Sedivy, 2002; Spivey, Tyler, Eberhard, & Tanenhaus, 2001; Tanenhaus, Magnuson, Dahan, & Chambers, 2000; Tanenhaus et al., 1995). These studies mainly focused on the recording of eye movements as participants were asked to manipulate real objects arrayed on a table. Syntactic ambiguity, with respect to the visual referents of the objects named, was manipulated. For example, eye movement behaviour in response to the two instructions, "Put the apple on the towel in the box," and "Put the apple that's on the towel in the box," was compared when there was an apple already on a towel, an apple on a napkin, and an empty towel in the visual display (Tanenhaus et al., 1995; see also Eberhard et al., 1995; Spivey et al., 2002). The first sentence is ambiguous at the point between "towel" and "in" because it is not clear which apple should be moved, whereas the location conveyed by the second is not ambiguous. They found that eye movements were launched towards the object referents approximately 250 ms after the nouns were uttered. In addition, the time course of eye movements was closely locked to the utterance of words that would aid in resolving this ambiguity, in order to establish this reference as early as possible in the presence of the syntactic ambiguity. They take these results as support for the interaction of linguistic processing with context, as visual context affected language interpretation at its early stages by helping to establish reference in real-world displays.

In addition, certain preferences in the resolution of syntactic ambiguities disappeared with this methodology. Specifically, studies of syntactic ambiguity resolution (e.g., Clifton, Frazier & Connine, 1984) have found that the language parser interprets ambiguous sentence structures according to *a priori* syntactic preferences, even when the context is incongruent with that structure. When this incongruity is detected by the parser, it reinterprets the sentence correctly, thus suggesting that the language system works independently and does not integrate contextual information, a view consistent with modularity. On the other hand, other work (e.g., Altmann & Steedman, 1988) suggests that the parser processes information incrementally, allowing it to interact with the context in selecting the most plausible syntactic structure. This view is consistent with what they term “weak modularity.” However, the work by Tanenhaus et al. (1995) provides support for the notion that the visual context can provide strong influences on syntactic preferences, and is thus incompatible with the modularity hypothesis.

In a similar series of studies, using the same methodology (tracking eye movements as participants are asked to manipulate objects in a real-world display), Sedivy and her colleagues (Sedivy et al., 1999) attempted to extend these findings to the resolution of *semantic* ambiguities. In order to examine how and when visual context mediates the resolution of semantic ambiguities, participants were asked to touch the visual referents of a named object in an array as their eye movements were recorded. They manipulated the number and order of adjectives preceding the nouns, as well as the objects in the array; for example, participants were asked to “Touch the yellow comb,” when there were two yellow objects in the array, one of which was the comb. They were able to demonstrate that the interpretation of adjectives was incremental, as indicated by the pattern of participants’ eye movements; namely, that ambiguities were resolved as the sentence unfolded. Again, they take this to indicate that contextual and linguistic information interact at the very early stages of sentence comprehension, as also demonstrated by the previously reported studies of syntactic resolution.

As further support for the interaction of semantic and contextual factors, Tanenhaus and colleagues (Chambers et al., 2002) demonstrated that semantic constraints in the form of spatial prepositions (such as "inside") constrain visual attention to objects in the visual environment. Participants were asked to manipulate these objects with instructions such as, "Put the cube *inside* the can." They found that after hearing the spatial preposition, participants limited their eye movements to objects that were consistent with that preposition. They take the results to mean that "referential domains," the scene regions that could possibly refer to the noun phrase contained in the sentences, are constantly being restricted as language interpretation unfolds.

In order to demonstrate that visual attention can be constrained by language processing, this same group of researchers took a different approach to the previous studies (Spivey et al., 2001). In a study of attention (without eye-tracking), they presented participants with arrays of vertical and horizontal green and red bars, and asked them whether a certain conjuction of these features was present (e.g., "Is there a red vertical?"). They manipulated whether the target was present or absent, as well as the set size (total number of bars). They also manipulated the point at which the array became visible to participants in the sentence: either at the end of the question or at the onset of the colour or orientation name (the order in which these were named was counterbalanced across two conditions). RT was measured (by the pressing of a "yes" or "no" button) in response to the question asked.

Results indicated that there was an interaction between array onset time and set size: reaction times increased more slowly as set size increased in the condition where the array became visible after the utterance of the disambiguating adjective than in the condition where the array become visible only after the end of the question. In other words, when the display became visible at the same time the first feature was mentioned, participants required less time to identify whether the target was present as the set size increased than when the display became visible at the end of the question, when the target had already been fully specified. This indicated that linguistic information can be used to constrain the domain of visual attention in an incremental fashion; that is, as language processing unfolds. This further strengthened their belief that visual perception (typically considered modular) can interact with cognitive processes such as attention (a central process).

Taken together, these studies indicate that visual and linguistic processes are mutually interactive at the earliest stages of processing. Tanenhaus and his colleagues (Tanenhaus et al., 2000) dispute the notion that perceptual and encapsulated systems can produce context-independent representations on the basis of these results. However, the conclusion of interactivity of this body of research relies primarily on the fact that the context-dependent ambiguity resolution occurs within a short time, approximately 250 ms after the critical linguistic stimulus. Although modular processing is assumed to be fast, speed in and of itself cannot be considered diagnostic of modularity or interactivity, as discussed above. That the visual context has an effect on language processing at its earliest stages does not necessarily point to modular interactivity. These perceptual representations may in fact be context-independent until they interact with other sources of conceptual information in CSTM, which as Potter (1993, 1999) has shown, occurs very early in visual and linguistic processing (within approximately 125 ms).

Altmann and his colleagues (Altmann & Kamide, 1999; Kamide, Altmann, & Haywood, 2003) report research consistent with that of Tanenhaus and colleagues (as reviewed above). Our work employs a methodology very similar to that of Altmann and colleagues. However, instead of focusing on the ability of syntactic or semantic ambiguities to circumscribe referential domains, the work of Altmann and his colleagues has examined the role of semantic information contained by verbs. Specifically, they tested the hypothesis that verb-specific semantic information can guide visual attention towards the object referent of the verb's direct object, *before* the noun itself is uttered.

In one experiment, Altmann and Kamide (1999) presented participants with line drawings containing a person surrounded by several objects. For example, in one scene a boy was found sitting on the floor surrounded by several objects, one of which was a cake. Participants' eye movements were recorded as they scanned these scenes and listened to sentences related to those scenes; for example, "The boy will eat/move the cake." The sentences differed with respect to the verb used: one was more semantically restrictive ("eat") than the other ("move"). Crucially, the direct object of the semantically restrictive verb only had one possible visual referent in the scene; that is, in the example given here, the cake was the only edible object in the array. Contrast this with "move," which could have referred to any of the objects within the scene. At the end of each trial, participants were asked whether the sentence could apply to the scene presented.

Altmann and Kamide (1999) were interested in two aspects of participants' eye movement behaviour: first, what proportion of saccades were launched towards the visual referent of direct object of the verb (cake) before the onset of the noun, and second, at which point in time participants first launched a saccade towards the target object relative to the onset of the verb. They found that participants fixated the target object in 90% of the trials, and that the first saccade to the target object was launched prior to the onset of the noun in 38% of the semantically non-restrictive trials and in 54% of the semantically restrictive trials. However, there was no significant effect of verb type on this data. With regards to the saccade onset time, they found that the first saccade after the verb occurred 127 ms after the onset of the noun in the semantically nonrestrictive condition, and 85 ms before the onset of the noun in the semantically restrictive condition. Here, the effect of verb type was significant. These data suggest that verb-specific information does direct eye movements towards objects that are semantically consistent with the selectional restrictions of the verbs.

However, because task demands may have affected the speed and pattern of eye movements, in that the judgment required at the end of every trial may have induced anticipatory eye movements, a second experiment was carried out without such a task (Altmann & Kamide, 1999). Here, the findings were similar: the target object was fixated in 93% of the trials, the first saccades towards the target object was initiated before the noun in 18% of the nonrestrictive trials and in 32% of the restrictive trials (this difference was statistically significant), and the saccade onset time was 536 ms in the nonrestrictive condition and 591 ms in the restrictive condition (this difference was also statistically significant). Notably, saccade onset times were much longer in this experiment, as might have been expected due to the lack of the decision task. Nevertheless, the data still point to the ability of verb-specific information to guide eye movements to the appropriate depictions of complement noun referents.

At a theoretical level, Altmann and Kamide (1999) speculate that these results suggest that the linguistic processor can predict possible fillers of a verb's patient role prior to the utterance of the noun phrase. This is consistent with McRae et al.'s (1997) and Dowty's (1991) thematic role assignment theory, where thematic roles are presumed to contain knowledge about typical agents and patients. Importantly, when role concepts are activated, they are compared to candidate noun fillers. In Altmann and Kamide's (1999) experiments, these candidate noun fillers may have been activated by the visual context, where the mental representations of the objects encountered in the scenes may have been evaluated for their compatibility in fulfilling the semantic restrictions of the verbs.

A second series of experiments conducted by the same group (Kamide et al., 2003) examined whether verb arguments other than themes (patients, or direct objects), such as goals (indirect objects), could guide eye movements in the same way. Using the same methodology described above (line drawings accompanied by spoken sentences), they found that more eye movements were directed towards objects that were consistent with the goals of the verbs uttered in the sentences than those that were not. This further strengthens the view that verbs contain semantic/syntactic information that can be used to predict, or at least constrain, possible role fillers. In addition, it also supports the idea that scene knowledge (in the form of the objects contained within it and its layout) is rapidly used and interacts with ongoing linguistic processes such as verb-semantic role assignment.

In sum, the bulk of the evidence brought forth by the visual world paradigm supports interactivity, especially from the point of view of the work conducted by

Tanenhaus and his colleagues. In addition, work by Altmann and Kamide (Altmann & Kamide, 1999; Kamide et al., 2003) suggests that the selectional restrictions of verbs can be used to predict possible role fillers. However, we argue that this process of matching arguments and grammatical object referents, as evidenced by eye movements that seek out these objects, occurs beyond the modular level, and not at the linguistic level as these researchers believe. As Dowty (1991) argues, if thematic roles are conceptual in nature, and if there are effects of context on argument fillers, then this process is occurring beyond the initial parsing conducted by the linguistic system, which deals only with domain-specific (i.e., linguistic) representations. In fact, we take it one step further by suggesting that this process occurs within CSTM. We propose that this occurs via a matching process where conceptual representations activated by the visual context are compared with possible role fillers activated by verb-specific information within LTM (whether the entire concepts or just their features mediates this process is beyond the scope of this paper). The purpose of the studies reported here is to replicate the findings of the Altmann and Kamide (1999) study, and further explore the nature of the interaction of linguistic and visual contextual information and its relation to Fodor's (1983) modularity hypothesis.

Rationale of the Present Studies

Although the studies reviewed in the section above have shed much light on the modularity debate, they do have some methodological problems, most notably in the area of ecological validity. The two most significant problems with the Altmann and Kamide (1999) and Kamide et al. (2003) studies are that naturalistic visual contexts were not employed in the studies, nor did they include a dynamic component. Because the visual world paradigm is based on the assumption that language comprehension takes place within a real-world visual context, experiments using this paradigm should strive to use visual contexts that are as naturalistic as possible.

Although the studies conducted by Tanenhaus and his colleagues (Chambers et al., 2002; Eberhard et al., 1995; Sedivy et al., 1999; Spivey et al., 2002; Spivey et al., 2001; Tanenhaus et al., 2000; Tanenhaus et al., 1995) did use naturalistic contexts (in fact, they used actual real-world visual displays), the main difficulty with these studies is that they studied a very specific aspect of language comprehension within a visual world, namely, one where the participant was required to interact with the world in some way. Although we are certainly required to manipulate objects in our environment on command at times, this is not representative of language processing in general and not the ideal manner to study context effects on language. In addition, the task demands may have affected eye movement behaviour in a way similar to that shown by Altmann and Kamide (1999). Their procedure required participants to solve a particular kind of problem: to pick up an object and move it to a different location. With repeated instructions of the same nature, participants may have learned to anticipate upcoming demands. In addition, this places additional attentional demands on participants, which may interfere with the ecological validity of these studies. It is also important to note that although different contexts produced different scan paths, participants also tended to fixate the inappropriate object (in the example given above, the apple on the napkin) prior to fixating the appropriate one (the apple on the towel). By doing so, and by not waiting until the noun phrase was fully uttered, they were demonstrating the need to quickly solve the problem, a task demand which again compromised the ecological validity of the study.

In addition, although these studies clearly demonstrate effects of visual context on language processing, the locus of the interaction of these processes is not clear. Does it occur within the linguistic system or at a central, post-perceptual system, or somewhere in between – such as the post-modular conceptual level where CSTM is purported to exist?

The present experiments address these problems by using still photographs and short movie clips of naturalistic scenes in which an agent, the subject of the spoken sentence that accompanies each picture or clip, is seen performing various everyday activities (e.g. a woman baking some muffins or a child playing with toys). Previous studies by Altmann and Kamide (Altmann & Kamide, 1999; Kamide et al., 2003) employed line drawings of arrays of objects haphazardly placed (e.g., an entire lit birthday cake located on the floor), and often not proportional in size to each other (e.g., a lipstick that is half the size of a bottle of wine). The major problems with the visual stimuli used in these experiments are succinctly described by Henderson and Ferreira (2004). As they point out, the visual stimuli typically used in psycholinguistic studies cannot be considered true scenes, and it is therefore difficult to generalize these results to other findings regarding gist extraction. First, these scenes violated "spatial licensing" constraints, which specify that in order to be considered a scene, the objects in the array must be semantically consistent (i.e., cakes are not typically found on the floor, nor are bottles of wine and lipsticks nearly the same size). This is what they refer to as an "*ersatz* scene," which lacks some of the essential features of a real-world scene, as opposed to a "true scene," whose properties include naturally-occurring and appropriately arranged entities, proportionality, and sufficiently detailed backgrounds. In addition, arrays have no discernable gist, and the visual system can only use the linguistic input to guide eye movements, rather than semantic information that could be gleaned from a true scene. This is particularly important in the present studies because the interaction of vision and language at a cognitive level necessarily encompasses semantic processing. In order to determine how the semantics of language processing affects and is affected by the visual context, the context provided should have a semantically interpretable dimension. The present studies aim to overcome this difficulty, thus allowing viewers to obtain scene semantic information very quickly through the use of true scenes.

Another distinction Henderson and Ferreira (2004) make is between real environments and depictions. They point to the fact that researchers in this field need to make the assumption that a depiction of a real-world scene is equivalent to one, or at least that it makes a good trade-off between the essential properties of a real-world scene (e.g., scene complexity and semantics) and the factors that would complicate the empirical study of a real-world scene (e.g., motion and nonpictorial depth cues). Although depictions are often used for pragmatic reasons, namely the difficulty and expense associated with studying eye movements in real-world environments, the problem with the Altmann and Kamide (1999) and Kamide et al. (2003) experiments is the fact that the visual stimuli were line drawings. Our materials will be *depictions* of true scenes, thus improving over the *ersatz* scenes traditionally used in these experiments.

Although a photograph of a scene strikes a balance between the factors of ecological validity and pragmatic feasibility, a motion picture of an event adds factors that are difficult to "control" in real-world events. However, language comprehension occurs in parallel with the dynamic perception of visual events – we do not live in a static visual world. This may introduce an additional layer of complexity, namely a dynamic component, to scene perception, but it is a fair representation of the way our language and visual systems function nearly all of the time. The only psycholinguistic experiments that have studied eye movements in real-world contexts are those where participants directly interact with the world (e.g., those conducted by Tanenhaus and his colleagues), and do not merely observe it, as experiments in the language-vision interaction tradition have.

As Henderson and Ferreira (2004) point out, the majority of psycholinguistic studies have used arrays of objects, both real and depicted, as well as scene sketches; to their knowledge, nor to ours, there have been no studies that make use of photographs of naturalistic scenes, nor any that make use of dynamic scenes. The present series of studies is likely the first to have done so.

Research Questions and Hypotheses

The approach taken by the research presented here is to examine how linguistic and visual processing interact, and to try to determine whether this takes place at the central or modular level, or some point in between, by examining the influence of visual and linguistic variables on eye movement behaviour using an eye-tracking paradigm. The main question is not whether language and vision interact, but rather the locus at which this occurs. We are going to attempt to answer this by asking two main questions:

(1) Do visual variables (motion context, object saliency and event saliency, and indirectly, scene complexity and scene layout) affect language processing, as evidenced by eye movements?

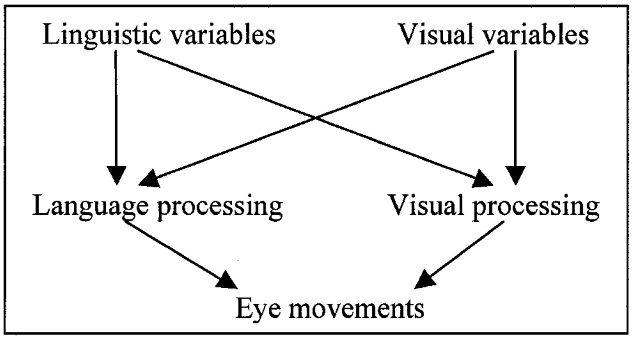
(2) Do linguistic variables (in this case semantic restrictiveness, manipulated by verb class) affect scene processing, as evidenced by eye movement behaviour?

Inherent to the methodology here are two additional questions, which are not being directly asked, but that are nonetheless relevant:

(3) Does verb class affect language processing, as evidenced by eye movements?

(4) Do the visual variables mentioned above affect scene processing, again as evidenced by eye movements?

In other words, there are four possible effects on eye movements, due to a combination of linguistic and visual variables on language and visual processing, as illustrated be the following schematic:



Because of the difficulty in teasing out these effects using the eye-tracking paradigm, in that we are studying both visual and linguistic effects simultaneously on behavioural output in the form of eye movements, no direct claims can be made as to the degree these four influences have on eye movements relative to each other. However, by relying on the manipulations of verb class and motion context, we can determine whether linguistic and visual variables have an effect on eye movements, and under which circumstances.

A related question pertains to the relative strength of verb class vs. motion context in facilitating sentence processing, and whether this differs if the motion is implied (as in Experiment 1) as opposed to actually visually presented (as in Experiment 2). Although, as mentioned above, the relative degree of these influences cannot be directly measured, the presence or absence of any main effects can point towards such a difference.

Two separate questions posed by the Normative Study concern the saliency of certain objects and events within the scenes used throughout this body of research. By asking participants to list the objects they saw after viewing these scenes, and to generate sentences related to those scenes, we attempted to obtain objective measures of saliency. See Figure 2 for an outline of the variables manipulated and measured by the experiment, and the hypothesized sequence of processing that occurs according to modularity.

The first study reported here, the Normative Study, employed still pictures taken from the scenes used in Experiment 1. These were briefly shown to participants in order to get some measure of the conceptual information derived from these scenes in the form of object and event representations. This study was conducted because participants in Experiments 1 and 2 viewed the scenes for a period of time before hearing the critical verb; certain conceptual representation would have already been activated, not only about what was in the scenes, but what was happening as well as expectations about what was about to happen. These are the visually-driven concepts that we propose will interact with the concepts activated by the sentences. These activated concepts would serve as working hypotheses as to the event taking place and the likely candidates involved in that event

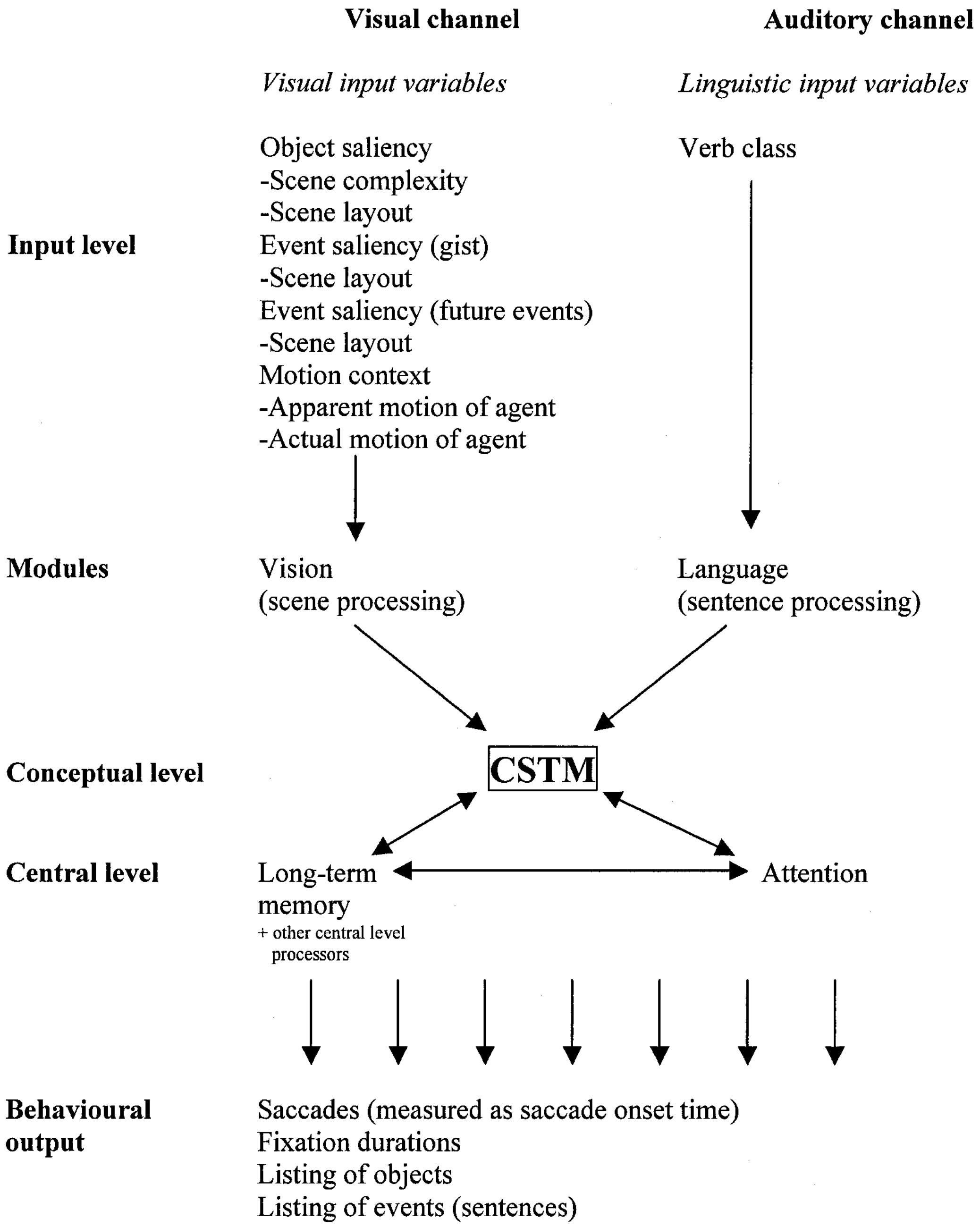


Figure 2. A model of the sequence of processing undergone by participants in the experiments presented here. The input is presented via two modalities, the visual and auditory channels, in the form of scenes (static and dynamic) and spoken sentences. Several variables were manipulated: in the visual channel, these included object and

event saliency, as well as motion context (both implied and actual motion of the agent), while in the auditory channel, this was the verb class of the main verb used in the sentences. We propose that the language and vision modules process the information relevant to each separately, such that the vision module processes all relevant scene variables while the linguistic system engages in phonological decoding and structuring as well as syntactic parsing. The outputs of these two modules are integrated in CSTM, at least for a very brief time, before undergoing more complex processing at the central level. Once a stable perception of the input has been established, we hypothesize that eye movements are initiated (at the behavioural level) to reflect the processing that has taken place, while in the case of the Normative Study, the listing of events and sentences is made in response to the relevant scene variables.

(both as agents and patients, or objects being acted open, taken to be the target object in each scene). In addition, because these indices of object and event saliency might have affected eye movement behaviour in the following two experiments, we correlated these two variables to control for the effects of visual saliency.

The purpose of Experiment 1 was to replicate previous studies that used still pictures with spoken sentences, but with the methodological improvements noted above (namely the use of naturalistic, true scene depictions). The purpose of Experiment 2 was to replicate Experiment 1 except with the added dimension of dynamic motion in order to determine how motion affected the pattern of eye movements.

These scenes all included a person, who we termed the "agent," taken to be the subject of the main clause of the accompanying sentence. This was done for two reasons. First, in Experiment 2, the inclusion of a moving person was necessary to introduce the dynamic component. However, with regards to Experiment 1, aside from the fact that including the agent was unavoidable as the pictures were taken directly from the movies, we wanted the scenes to be as ecologically valid as possible, and sentences in the real world about people engaging in activities are usually associated with or call up actual images of people doing those things.

These two experiments contained two types of manipulations: the type of verb used (perceptual vs. causative) and the apparent or actual motion of the agent (towards or away from the target object, or neither). As discussed above, verbs not only specify how agents and patients in scenes are related, but also provide information about event-related concepts. For the more semantically restrictive verbs (the causative verbs), we expected that eye movements would be launched more quickly towards the visual referents of the verb's direct object than for the non-restrictive verbs (the perception verbs) in both experiments.

In Experiment 1, we hypothesized that the apparent motion of the agents may have directed people's eye movements in a particular direction more quickly. This may be for two reasons: one, if somebody appears to be moving in a certain direction, we may be interested in knowing what they are moving towards. Interestingly, we may also be interested in knowing what they are moving away from, especially since this would be semantically inconsistent with the accompanying sentence. Two, the direction of an individual's gaze can also convey important information about which objects in the scene should be attended to (Henderson & Ferreira, 2004). Therefore, we hypothesized that saccades would be made more quickly towards the target object after the verb-onset in the towards condition than in the away or neutral conditions.

In Experiment 2, a similar pattern of results would be expected. Because in this experiment the direction of motion was not implied but actually observable, we predicted that motion direction would have an even stronger effect on eye movement behaviour. This is because the dynamic component is such a visually salient feature, which may have served to "grab" most of the participants' visual attention.

Normative Study

Introduction

The purpose of this study was twofold. First, we wanted to obtain information regarding the frequency with which the various objects in each scene were listed by participants, in addition to the frequencies of sentences generated by participants regarding what they believed was occurring in the scene as well as what might occur next. Second, we were interested in knowing whether these frequencies differed across the three motion conditions. With regards to the frequencies, we were specifically interested in knowing how frequently the target objects for each scene were listed, as well as in learning the how well the events conveyed by the sentences used in Experiments 1 and 2 could be predicted on the basis of the scenes alone. Because there were two different versions of the present study, where participants were either asked to write what they believed would happen next (version "a") or what they thought was presently happening in the scene (version "b"), there were in fact two types of target events. These were defined as the propositional structure corresponding to the main sentence clause of the other two experiments for version "a" of the study, and the propositional structure corresponding to the patch clause at the beginning of the sentences used in the other two experiments for version "b."

In addition, we were interested in knowing whether these frequencies differed across the three motion conditions. For the objects, we hypothesized that the frequencies would be higher when the scene's agent appeared to be moving towards the target object (or directing his or her gaze towards it) than if the agent appeared to be moving away or remained in a neutral position with respect to it. In addition, we expected the target object to be listed more frequently in the Neutral than the Away condition, as the layout in the Away scenes might draw attention away from the target objects. For the sentences, we also hypothesized that the target event in version "a" would be listed more frequently in the Towards version of the scenes than in the Neutral and Away versions, and more frequently in the Neutral than the Away conditions. This was because viewing the agent facing the target object should increase the probability that an event structure implicating this object would be activated, and facing away from it should decrease this likelihood. Similarly, a difference between motion conditions was expected for version "b" because the scene's gist also typically included the target object. In addition, different meanings might be extracted from the scenes based on whether the agent appeared to be entering, leaving or remaining within the scene. Therefore, we expected that all three motion conditions would differ significantly.

Method

Participants

Thirty-four Concordia University students participated in this experiment. There were 26 females and 8 males, ranging in age from 19 to 54. Inclusion criteria included being a native speaker of English (defined as having learned the language by the age of five), and having normal or corrected-to-normal vision. All either received course credit for their participation, or were given monetary compensation.

Materials and apparatus

The stimulus materials consisted of seventeen naturalistic scene triplets of everyday events taking place in their natural environment (see Appendix A for a complete list). These were frames extracted from the movies used in Experiment 2, and were identical to the still pictures used in Experiment 1. Typical scenarios included kitchen scenes of a woman preparing baked goods or a child playing with toys. Each scene contained one person, termed the "agent," because of his or her role as the subject of each sentence used in Experiments 1 and 2, and particularly because of his or her role as an entity in motion effectuating some action in the scenes. Scenes were constructed so that the agent and the target object were at opposite sides of the display; this was to ensure that the visual angle (or distance) between the two remained relatively constant across the scenes. In addition, the side on which each appeared was counterbalanced across the scenes, so that the position in which the agent and the target object appeared could not be predicted from scene to scene.

Each of the scene triplets differed with respect to the apparent motion direction of the agent relative to the location of the target object. Because the pictures used in this study were single frames taken from the film clips used in Experiment 2, where the agent either moved towards or away from the target object, or remained in a static position with respect to it, the three conditions of apparent motion were termed Away, Neutral and Towards. The frames chosen from the Away and Towards versions of each movie were the first in which it was obvious that the agent was on a path either away from or towards the target object, without having moved too close to it or having begun to leave the scene.

Procedure

The scenes were presented in pseudorandom order in a PowerPoint slideshow on a Macintosh G4 computer with a 17" screen in a laboratory testing room. There were three versions of each slideshow, with a relatively equal number of Away, Neutral and Towards scenes in each (each slideshow contained six scenes from two conditions and five from the other). Subjects were randomly assigned to one of the three versions of the slideshow. After providing informed consent (see Appendix B for a copy of the consent form) for participation in the study, participants were given both oral and written instructions (see Appendix C). The first slide in the slideshow consisted of a brief description of the participants' task (see Appendix D), which was supplemented by a detailed verbal explanation given by the experimenter.

The format of each trial was as follows: first, a fixation cross (a white cross on a black background) appeared on the screen for three seconds in order to orient participants' eyes to the centre of the screen, ensuring that all participants would begin viewing each scene from the same starting point. The actual scene was then presented for 2 s, during which time the participants were instructed to extract as much information as possible regarding the contents of the scene and the likely event taking place or about to take place. Scenes were presented for such a brief amount of time to maximize the likelihood that participants would report their initial impression of the scene, in terms of which entities "popped out" the most and what they believed was going on in the scene or what was about to happen. This is in keeping with the findings that the gist of a scene can be extracted in a very short amount of time (e.g., Intraub, 1999; Potter, 1999; Thorpe et al., 1996; VanRullen & Thorpe, 2001), thus indicating which object and event concepts were activated at the earliest viewing of the scene. This was followed by a message stating "List the objects in the scene" which appeared in the centre of the screen for 3 s, after which point the sound of a computerized tone indicated that participants should begin writing. Participants then had 15 s to list up to six of the objects they remembered seeing in the scene presented to them. Participants were provided with a response booklet in which to record their answers (see Appendix E for a copy of a sample page). Each scene had a full page dedicated to it, in order to minimize any interference between trials. The screen remained blank during this time, and a second, different computerized tone indicated the end of the 15 s response time. Next, a second message appeared on the screen asking participants to write a sentence related to the scene. In version "a" of the response booklet, the message took the form "What is about to happen in the scene?" and in version "b" it took the form "What is happening in the scene?" Again, participants had 15 s to record their responses, with the beginning and end times indicated by the same two tones. This sequence, fixation cross, picture, object listing, sentence generation, constituted a single trial which was repeated 17 times.

To ensure that participants understood what was required of them, three practice trials were administered after the instructions, during which time the experimenter remained in the testing room. At the end of the practice trials, the slideshow was paused, and the experimenter reviewed the participants' responses, making sure they were responding correctly (e.g., listing actual objects and not events in the object section, and providing complete sentences in the sentence section). The experimenter then left the room and participants resumed the presentation when they were ready. The entire experiment took approximately 15 minutes to complete. Participants were debriefed at the end of the study and given a short written explanation of the purpose of the study.

Coding and Analyses

Two separate sets of analyses were conducted, one for the objects listed and the second for the sentences generated by participants. For the objects, the frequency with which the target object in each scene was listed was calculated. This was computed by calculating the number of times the object was listed over the total number of objects listed across participants. This number was calculated for each experimental condition (Away, Neutral and Towards) and across all conditions. For the analysis, a repeated-measures one-way ANOVA was computed with motion type as the independent variable. We expected that the target object listing frequencies would be higher in the Towards condition than in the Away and Neutral conditions.

For the sentences, each was reduced to one or more propositional statements of the form VERB [AGENT , PATIENT], as proposed by Kintsch (1974) and compatible with Jackendoff (1987b), with AGENT being the subject of the sentence and PATIENT being the direct object of the verb used in the sentence. For instance, if asked to describe what was about to happen in the scene, and the participant were to generate the sentence, "The boy looks like he's going to drink some milk," the proposition DRINK [BOY, MILK] was recorded. Only phrases containing an actual event structure (a description of an event taking place, having taken place or about to take place) were retained in the analysis, as opposed to phrases describing a state, which describe the presence, location or states of various entities (e.g., "The boy is in the living room."). If participants were to compose a sentence containing more than one propositional event structure, both would be recorded (e.g., "The boy is going to get some milk and then drink it," χ GET [BOY, MILK] + DRINK [BOY, MILK]).

For each scene, there was a target propositional structure that corresponded to the actual sentences used in Experiments 1 and 2. The frequency with which these structures were generated by participants was calculated separately for the two question types asked to the participants (i.e., what is happening vs. what will happen next), again by condition and across conditions. As in the above analysis, two repeated-measures ANOVAs were calculated for the two sets of propositional structures. Again, we expected that the target event frequencies would be higher in the Towards condition than in the Away and Neutral conditions for versions “a” and “b.” An alpha level of .05 was used for all statistical tests, unless otherwise indicated.

Results

For each scene, the frequency with which each target object was listed was computed as described above. Table 1 lists the frequencies by condition and overall, across all conditions. The overall frequencies ranged from .054 (*egg*) to .221 (*vase*), meaning that the egg consisted of 5.4% of the objects listed by participants, whereas the vase consisted of 22.1% of the total number of objects listed by participants. Note that because participants were given six lines with which to record their responses, we would expect (if all six lines were completed) that if an object were always listed, it would have a maximum frequency of .167 (1.0 / 6). Therefore, the target object listing frequency was a function of both how often an object was listed as well as how many lines were completed in the response booklet. For any frequency greater than this number, not all of the six lines were completed, indicating that the scene was relatively noncomplex and contained few salient, nameable objects.

A similar computation was performed for the frequency with which the agent in each scene was listed. See Table 2 for a list of these frequencies, again by condition and overall, across all three conditions. Here the overall frequencies ranged from .153 (*plate* and *shirt*) to .221 (*vase*), a more restricted and higher range than that for the target objects, indicating that the person was almost always named by the participants. This has

Table 1

Target Object Listing Frequency

Target Object	Away	Neutral	Towards	Overall
Ball	.222	.207	.227	.213
Butter	.038	.097	.074	.071
Car Crash	.211	.189	.204	.201
Car Start	.204	.196	.194	.198
Chair	.138	.150	.170	.152
Cube[a]	.155	.217	.152	.179
Egg	.021	.037	.106	.054
Ice	.111	.033	.070	.068
Kite	.073	.038	.088	.067
Milk[b]	.128	.119	.143	.128
Oven[c]	.070	.042	.125	.084
Paper	.091	.080	.152	.106
Picture	.190	.157	.188	.178
Plate	.127	.158	.176	.153
Shirt	.123	.097	.158	.119
Shoes	.176	.140	.192	.169
Vase	.227	.200	.231	.221
Mean	.136	.127	.156	.139

[a] These numbers were computed by counting two separately listed objects: “box/square toy” (F[A] = .017, F[N] = .050, F[T] = .000, and F[Overall] = .026), which we took as

referring to the cube, and “toy(s)” (F[A] = .138, F[N] = .167, F[T] = .152, and F[Overall] = .152), which was included because a cube belongs to the “toy” category.

[b] These numbers were computed by totalling the frequencies listed for both “milk” (F[A] = .021, F[N] = .000, F[T] = .024, and F[Total] = .014) and “cup” (F[A] = .106, F[N] = .119, F[T] = .119, and F[Overall] = .115), which referred to the same object.

[c] These numbers were computed by totalling the frequencies listed for both “oven” (F[A] = .000, F[N] = .021, F[T] = .063, and F[Overall] = .032) and “stove” (F[A] = .070, F[N] = .021, F[T] = .063, and F[Overall] = .052), which referred to the same object and were listed separately by one subject, precluding the lumping together of the two synonyms.

Table 2

Human Agent Listing Frequency

Scene	Away	Neutral	Towards	Overall
Ball	.222	.207	.182	.200
Butter	.173	.161	.167	.167
Car Crash	.123	.170	.184	.157
Car Start	.204	.176	.161	.179
Chair	.155	.167	.189	.170
Cube	.190	.250	.152	.205
Egg	.208	.185	.191	.195
Ice	.222	.183	.233	.209
Kite	.182	.212	.158	.183
Milk	.213	.186	.214	.203
Oven	.233	.188	.188	.200
Paper	.182	.200	.196	.192
Picture	.155	.176	.229	.185
Plate	.145	.123	.196	.153
Shirt	.158	.161	.158	.153
Shoes	.157	.211	.173	.181
Vase	.227	.250	.192	.221
Mean	0.185	0.189	0.186	0.185

important implications for the interpretation of the results of the other two experiments, because the relative saliency of the target objects and human agents likely influenced where eye movements were directed. In order to examine whether the apparent direction of motion had an effect on how often the target object was listed across conditions, a repeated-measures one-way ANOVA was conducted. We expected that the target object would be listed more frequently in the Towards condition than in the Neutral condition, which in turn would have a higher mean than the Away condition. The results mostly confirmed these hypotheses, as shown in Figure 3 (see Appendix F for all ANOVA tables relevant to the analyses performed for this study). The analysis (*N* = 17) indicated that there was a main effect of motion type, *F*(2, 32) = 6.32, *p* < .01. A modified Bonferroni/Dunn test, with adjusted alpha levels of .03 per pairwaise comparison, was conducted to determine which conditions differed from each other. These comparisons revealed that the Away (*M* = .136, *SD* = .064) and Towards (*M* = .156, *SD* = .050) conditions differed significantly (*p* = .02), as did the Neutral (*M* = .127, *SD* = .065) and Towards conditions (*p* < .01). The Away and Neutral conditions were not significantly different *p* = .31).

The next series of analyses examined the propositional structures extracted from the sentences generated by participants. The analyses were conducted separately for versions "a" and "b," which concerned future and present events, respectively. For version "a" of the experiment, 415 propositional structures were generated, of which 410 were event structures and 5 were state structures, not counted in the analysis and constituting 1.2% of the total propositional structures. Table 3 lists the frequencies of the target structures (corresponding to those equivalent to the main clauses of the sentences

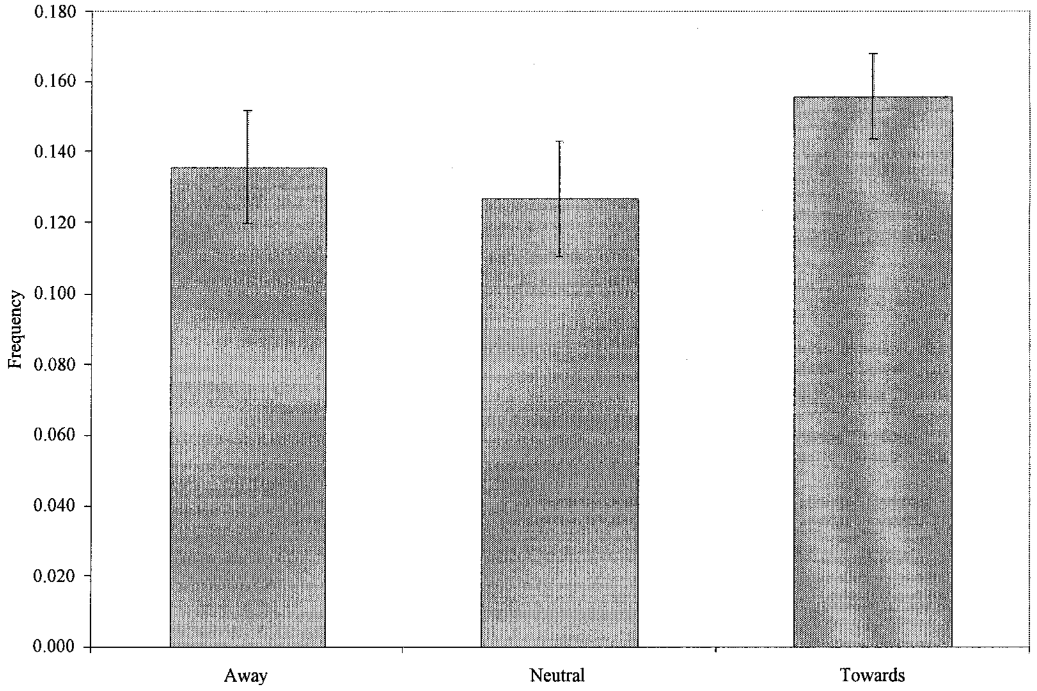


Figure 3. Mean frequency ratings (± *SE*) of target objects as a function of motion condition.

Table 3

Target Event Propositional Structure Listing Frequency – Version "a" ("What will happen next?")

Scene	Away	Neutral	Total	Overall
Ball	.000	.000	.000	.000
Butter	.000	.000	.000	.000
Car Crash	.000	.000	.000	.000
Car Start	.000	.000	.000	.000
Chair	.000	.000	.000	.000
Cube	.000	.000	.000	.000
Egg	.000	.000	.000	.000
Ice	.000	.000	.500	.150
Kite	.000	.091	.167	.083
Milk	.000	.000	.000	.000
Oven	.000	.000	.000	.000
Paper	.000	.000	.000	.000
Picture	.143	.000	.222	.107
Plate	.333	.000	.000	.136
Shirt	.000	.000	.000	.000
Shoes	.000	.000	.000	.000
Vase	.000	.000	.000	.000
Mean	0.028	0.005	0.052	0.028

containing a causative verb), both overall and by condition. Frequencies ranged from 0 (*ball, butter, car crash, car start, chair, cube, egg, milk, oven, paper, shirt, shoes* and *vase*) to .136 (*plate*). We expected that the target event would be listed more frequently in the Towards condition than in the Neutral condition, which in turn would have a higher mean than the Away condition, as this possible event would be more likely to be activated if the agent appeared to be moving towards the target object and less likely if he or she appeared be moving away. This hypothesis was not confirmed. A repeated-measures one-way ANOVA ($N = 17$) indicated that there was no main effect of motion type ($F(2, 32) = 1.15, p = .33$).

For version "b" of the experiment, 312 propositional structures were generated, of which 296 were event structures and 16 were state structures, not counted in the analysis and constituting 5.1% of the total propositional structures. Because the first clauses from the sentences for the *paper* and *plate* scenes referred to past and future events respectively (see Appendix A), no current event could be targeted and data from these two scenes was omitted. Table 4 lists the frequencies of the target structures (corresponding to the first patch clause used in the sentences, and referred to as the "gist" of the scene). These frequencies ranged from 0 (*picture*) to .882 (*cube*). We did not expect there to be a difference in frequencies between motion conditions, as the direction in which the agent appeared to be moving should not affect what the overall gist of the scene was. This hypothesis was confirmed. A repeated-measures one-way ANOVA ($N = 15$) did not indicate a significant main effect of motion type ($F(2, 28) < 1, p = .91$), thus confirming the hypothesis.

Table 4

Target Event Propositional Structure Listing Frequency – Version "b" ("What is happening now?")

Scene	Away	Neutral	Total	Overall
Ball	.000	.667	.000	0.222
Butter	.429	.167	.500	.333
Car Crash	.091	.100	.000	.077
Car Start	.000	.143	.100	.087
Chair	.400	.167	.125	.211
Cube	1.000	1.000	.600	.882
Egg	.333	.286	.500	.368
Ice	.750	.667	.000	.786
Kite	.333	.000	.125	.167
Milk	.286	.000	.200	.150
Oven	.400	.167	.200	.238
Picture	.000	.000	.000	.000
Shirt	.500	.667	.571	.600
Shoes	.429	.667	.500	.526
Vase	.800	.667	1.000	.824
Mean	0.383	0.357	0.361	0.365

Discussion

The results indicated that the direction of apparent motion of the agents, or their direction of gaze, in the various scenes affected how frequently the target object was listed by participants across the three conditions. Furthermore, results indicated that the target object was listed more frequently in the Towards condition than in the Away and Neutral conditions, which did not differ significantly. However, the direction of apparent motion did not affect how frequently the two target events were listed across these conditions, both those regarding future events and those regarding current events, or the gist of the scene.

The presence of a significant main effect for motion type on target object listing confirmed our hypothesis, namely that when the agent was facing towards the target object, the frequency with which it was listed by participants increased. This is consistent with other studies that indicate that gaze direction attracts attention to visually salient scene regions (Henderson & Ferreira, 2004). However, it failed to support the notion that when the agent faced or appeared to be moving away from the target object, it was less likely to be listed. This can be interpreted to indicate that not facing or appearing to move towards an object has the same effect of not increasing attention towards a particular object as when that same person is simply not looking at it at all (in the Neutral condition). In other words, facing neither towards nor away from an object is equivalent to facing away from it.

With regards to the target event frequencies, it is worth noting that the frequencies were higher for the version "b" propositions (which referred to events occurring presently) than for the version "a" propositions (which referred to likely future events). In other words, version “a” informs us as to whether the conceptual event structure of the main sentence clause was activated early in scene viewing while version “b” tells us whether the gist of the scene extracted early in scene viewing matches that which was intended in the scene layout. These higher means imply that the gist of the scene was extracted more easily than the predictiveness of the scene.

The lack of a main effect for the version “a” frequencies did not confirm our hypothesis that the target event (corresponding to the main clause of the sentences used in Experiments 1 and 2) would be listed more frequently in the Towards condition than in the Away and Neutral conditions. Because no studies have been done employing this methodology and examining scene “gists” in terms of propositional codes, we cannot contrast these results with those of the literature on scene processing. However, we can speculate that this lack of effect may be due to several reasons. First, from a statistical point of view, the sample size may not have been large enough. As there were two versions of the experiment, only approximately half of our participants (*N* = 17 for version “a’ and *N* = 15 for version “b”) were asked to generate a sentence regarding what they believed would happen next in the scene or what was happening presently. As such, power was relatively low (.226 for version “a” and .063 for version “b”) and having a larger sample may have detected a significant effect. This is reasonable to assume because the mean for the Towards group is higher than the means for the other two groups for version “a,” although the same cannot be said for version “b.”

Another possible reason that might account for this pattern of results is that by the time participants were asked to write the sentences (which would not have occurred before 21 s after the scene disappeared, due to the object listing task that came prior) they no longer remembered what the scenes looked like, or what they thought was happening/what would happen next, or both. In other words, the event structure concepts originally activated by the scene may have disintegrated in (or faded out of) their short-term memory stores. Thus, it may be that this task did not sufficiently tap into the nature of the events encoded in the scenes as they were perceived.

Despite this lack of effect for the target events, the frequencies computed for each scene by condition, not the overall frequencies, were used to interpret the results of Experiments 1 and 2. The same applied for the target object frequencies. The main purpose of the Normative Study was to provide this basic information regarding the relative saliency of the target objects and events, which we hypothesized would have an effect on one of the dependent variables used in the two experiments (i.e., how quickly saccades were initiated towards target objects after the verb). Although these results do not speak directly to the totality of the object and event concepts activated in CSTM, the fact that the ones mentioned were retained in LTM (or long enough to be recorded on the answer sheet) indicated that they must have at some point passed through this processing stage, as Potter (1993, 1999) believes that all information in LTM undergoes some structuring process in CSTM.

Experiment 1

Introduction

The main purposes of this experiment were to add to an existing body of literature following the visual world paradigm and to achieve greater clarity about the nature of the interaction between the visual and language systems. Specifically, it aimed to replicate previous findings that eye movements are driven by linguistic constraints (in this case, verb-specific information) but within a more naturalistic visual context than has been used in similar studies (e.g., Altmann & Kamide, 1995). In addition, it aimed to test the hypothesis that the apparent motion of agents in these scenes would have an effect on the time course of post-verbal eye movements. Finally, the relationship between target object and event saliency, and post-verbal eye movement behaviour was examined.

Method

Participants

Thirty-two participants took part in this study, drawn from the Concordia University student community. None of these participants took part in either the Normative Study or Experiment 2. There were 21 females and 11 males, ranging in age from 18 to 60 (the median age was 22). Inclusion criteria were the same as in the Normative Study, except that participants with glasses were excluded due to possible interference with the eye-tracking device. All participants either received course credit for their participation or were given monetary compensation.

Materials and apparatus

Stimuli. The visual materials were the same as those used in the Normative Study, and the sentences employed all had the same format (see Appendix A for complete list of the stimulus sentences and their corresponding visual scenes): a patch clause, a main clause, and a second patch clause. For example, in the sentence, "While playing with the lid, the boy will spill the milk that is on the table," the main clause is "the boy will spill the milk," and the two patch clauses are "While playing with the" and "that is on the table." The patch clauses were added to increase the total length of the utterance, which was especially of importance in Experiment 2, where the movies were much longer than the main clauses alone. In addition, the visual referent of the direct object of the verb used in the main clause of each sentence (e.g., "milk" is the direct object of the verb "spill"), the target object, was present in each scene. The main verb was of one of two verb classes, either causatives or perception/psychological verbs (which will be referred to as perception verb from hereon in). All sentences were spoken by a female speaker and recorded on an Apple Macintosh using SoundEdit at 16 bits and 44.1 kHz. The film clips were of naturalistic scenes and recorded digitally by a film student using a JVC mini-DV camera and saved at a resolution of 720 X 480.

Apparatus. The still pictures were presented on a Macintosh G4 with a Sony Trinitron Multiscan E500 21" monitor placed 41 cm in front of the participant. Participants wore clip-on earphones through which the sentences were presented binaurally. The experiment was run with PsyScope software (Cohen, MacWhinney, Flatt, & Provost, 1993). Participants' eye movements were recorded using the EyeLink I (SR Research) at a sampling rate of 250 Hz from the left eye only (viewing was binocular). This device is head-mounted, and although head movements were minimized with the use of a chinrest, any minor movements that did occur were corrected by a system of four sensors affixed to the monitor's four corners, which allowed for the continuous measurement of the head's position in relation to the screen.

Procedure

Prior to the experiment, participants gave their informed consent (see Appendix G for a copy of the consent form) and were given written instructions regarding the study (see Appendix H). Participants were assigned to one of the six experimental lists in consecutive order (i.e., the first participant was assigned to list one, the second to list two, and so on). The first phase of the experiment consisted of the manual calibration of the eye-tracker, which lasted approximately five to ten minutes. Participants were then shown a short version of the instructions, reminding them of how to proceed during the experiment (see Appendix I). Participants were asked to look at the pictures and to pay attention to the sentences, although no specific task was required of them. Each trial began with a fixation cross that appeared on the screen for two seconds, and was followed by the static scene. The fixation cross was included to orient participants' eyes at the beginning of each trial and to ensure a uniform starting point for all participants. Each new trial was initiated by the press of the spacebar, and the 17 trials were randomly presented. The calibration and experimental phases were conducted in the dark to minimize glare for the eye-tracking camera and to help participants focus their attention on the screen. At the end of the experiment, participants were given a short quiz, a 12-item cued recognition task that consisted of six pictures and six written sentences, half of which were foils and the other half of which were taken from the experimental stimuli (the pictures were frames taken from the films). This task was to ensure that participants were paying attention during the experiment. The entire experiment lasted approximately 30 minutes.

Analyses

Several sets of analyses were conducted, the design of which will be outlined below. The first examined the effects of target object saliency on post-verbal eye movement behaviour. The second examined the effects of target event saliency on post-verbal eye movement behaviour. The third set constituted the main analyses, wherein the effects of verb type and motion type on post-verbal eye movements were investigated. These studies all employed a mixed factorial design. The fourth, and final, set explored the cumulative fixations made to the target object during the early moments after verb-onset. An alpha level of .05 was used for all statistical tests, unless otherwise indicated.

Anticipatory eye movements. In order to determine whether any anticipatory eye movements may have been made, SOTs were compared to two other time points in the sentences: the noun-onset and the noun-offset. SOTs were computed by subtracting the latency between verb-onset and noun-onset, as well as between verb-onset and noun-offset, then averaged by condition. This was done to examine whether eye movements may have been initiated after hearing the verb but prior to hearing the onset and offset of the noun. In addition, we calculated the proportion of trials (those spoilt due to corrupt data were not counted) in which a saccade was launched towards the target object *before* the onset of the noun. Finally, the difference in time between the offset of the noun and SOT was computed and subjected to a 2 (verb type) X 3 (motion type) repeated-measures ANOVA, in order to determine whether these first saccades were affected by verb type and motion type. We predicted that there would be a main effect of both variables, such that the Causative condition would yield a lower mean difference than the Perception condition, and the Towards condition would yield a lower mean difference than the Away and Neutral conditions. In addition, to compare the effects of verb type within the Towards condition, a planned comparison was conducted. This consisted of a paired one-tailed t-test, and it was hypothesized that the Causative-Towards condition would have a lower mean difference than the Perception-Towards condition.

Target object saliency. These analyses examined the effects of target object saliency on eye movement behaviour. The first explored the relation between target object saliency on the amount of time participants spent looking at target object. A Pearson's correlation was computed between target object saliency (taken as the frequency listings obtained in the Normative Study, by motion condition), the time spent fixating the target object before the verb-onset, after the verb-onset, and in total. Trials in which participants were already fixating the target object at verb-onset were excluded from this analysis because of the difficulty in clearly separating pre- and post-verbal fixations. We hypothesized that significant positive correlations would be obtained between the saliency ratings and the three fixation measurements.

The second analysis explored the relation between target object saliency and the speed of post-verbal saccade initiation. A Pearson's correlation was computed between target object saliency and a measure post-verbal eye movement behaviour, which we termed "saccade onset time" (SOT). This was the length of time taken by participants after verb-onset to initiate a saccade towards the target object. Saccades that followed a fixation that was already on the object at verb-onset were not counted because the effect of the verb could not be parsed out: it is impossible to determine whether participants continued fixating the object because they heard a verb potentially selecting the object, or because they already happened to be fixating it. The distance between the fixation point prior to the verb onset and the target object was not a factor in determining the size of the SOT because the time to fixation is not counted but rather the *onset* of the saccade; i.e., how long it takes for participants to initiate an eye movement to that target object, not how long it takes them to get there.

The third analysis examined the relationship between target object saliency and whether or not the target object was being fixated at verb-onset. A point biserial correlation was computed between these two variables. A significant correlation was expected, because the more salient an object is, the more fixations it should attract at any given time, including verb-onset.

Target event saliency. This set of analyses was identical to those described in the section above, except the target event saliency was employed in the place of the target object saliency. The hypothesized results were expected to be similar as well: positive significant correlations for all three correlations, because the more predictive a scene is of the target event structure, the more time the target object should be fixated. In addition, the higher the saliency rating, the more quickly saccades should be initiated towards the target object after verb onset, and the more likely it should be fixated at any given time.

Main analyses: first post-verb saccades to target object. These analyses examined the effect of verb type (VT) and motion type (MT) on post-verbal eye movement behaviour. All analyses constituted a mixed factorial design, as all participants were exposed to all six of the experimental conditions (Away-Causative, Neutral-Causative, Towards-Causative, Away-Perception, Neutral-Perception, Towards-Perception), but only to one condition for each scene. Out of a possible 102 stimuli combinations (three motion conditions X two verb conditions X 17 scenes), each participant was exposed to one of six lists, each containing 17 randomly selected items from the six conditions. In addition, trials in which participants were already fixating the target object at verb-onset were not included in the analyses, nor were those in which participants never fixated the target object after verb-onset.

The first analysis examined the effect of verb type and motion type on the SOTs. This therefore constituted a 2 X 3 ANOVA conducted by subjects, meaning that the cell means for each condition combination were computed for each subject, such that each subject had six such means. This analysis was also conducted by items.

The first analysis was to be repeated as an ANCOVA with target object saliency and target event saliency as the covariates, as long as these covariates correlated significantly with the SOTs. This was done to eliminate any variability in the data due to variations in scene complexity and object/event salience. Because some objects may have been fixated more quickly than others based on these variables, the ANCOVA would control for any potential effects of scene complexity and object/event salience on the dependent variable, which may have led to an underestimation of the effects of the independent variables of interest (verb type and motion type).

The hypotheses for both of the analyses in this section were identical to each other. We expected there to be a main effect of both verb type and motion type, but no interaction effect. More specifically, we expected SOTs to be lower for the causative condition than the perception condition, as well as in the towards condition than in the away and neutral conditions.

Planned comparisons were also conducted for each of these analyses to examine our hypotheses. These consisted of ANOVAs and one-tailed paired t-tests. Because the hypothesis that the SOTs would be lower in the causative condition than in the perception condition was tested by the ANOVAs described above, no planned comparison was conducted. However, to clarify whether the SOTs were lower in the towards condition than in the away and neutral conditions, two separate 2 X 2 ANOVAs were conducted comparing (1) the towards and away conditions and (2) the towards and neutral conditions. In addition, to examine the difference between causative and perception verbs without the moderating effect of apparent motion, we compared the Causative-Neutral and Perception-Neutral groups using a one-tailed paired t-test. Finally, we examined the difference between causative and perception verbs in the Towards condition, as it was expected that this motion context might serve to increase the speed at which saccades were initiated due to the semantic consistency between the linguistic and visual contexts.

Analysis of early post-verb cumulative saccades to the target object. This final analysis is based on a combination of data analysis techniques suggested by Altmann and Kamide (2004). The purpose was to determine if verb-guided eye movements would be closely time-locked to the utterance of the verb, especially before the offset of the noun, as has been shown in other studies (e.g., Altmann and Kamide, 1999). First, the cumulative number of saccades that were initiated towards the target object during each 50-ms bin following the onset of the verb was computed. In other words, for each trial, the number of new saccades made in the first 50 ms after the verb-onset was counted, and so on for every 50-ms interval following that point. Saccades were only counted if they were launched from a location on the screen other than the target object, not if they were launched from a position within the boundaries of the space occupied by the object. Next, for each critical point in the sentence (verb-offset, noun-onset and noun-offset), the cumulative number of saccades was divided by the total number of trials for each condition to arrive at a cell mean. Because the critical sentence points differed for each sentence, based on the length of the individual words for each as well as the speaker's rate of speech, the corresponding 50-ms bins were different for every sentence. Thus, the cumulative fixation proportions for each participant were taken from these different bins. Despite occurring at different points in time, they corresponded to the same linguistic markers, namely the end points of the verb and noun, as well as the onset of the noun. Finally, the effects of verb type, motion type and sentence point on the cumulative proportion of saccades to target object after verb-onset were examined using a 3 (sentence point: verb-offset, noun-onset, noun-offset) X 2 (verb type: causative *vs.* perception) X 3 (motion type: away, neutral, towards) ANOVA (by subjects).

We hypothesized that there would be a significant interaction effect between sentence point and verb type, such that the mean number of cumulative saccades to the target object would increase more at each sentence point for the Causative condition than the Perception condition, because the interpretation of the verb and the noun phrase should proceed incrementally, thus providing more restrictive information as time proceeds, especially in the more semantically restrictive Causative condition. This would indicate that verb-specific information can constrain visual attention at the very early stages of processing. We also expected there to be a significant effect of motion type.

Planned comparisons were also conducted for each of these analyses to test the same series of hypotheses. These consisted of repeated-measures ANOVAs. To clarify whether the mean number of cumulative saccades was lower in the towards condition than in the away and neutral conditions, two separate ANOVAs were conducted comparing (1) the towards and away conditions across all sentence points and (2) the towards and neutral conditions across all sentence points. In addition, to examine the difference between causative and perception verbs without the moderating effect of apparent motion, we compared the Causative-Neutral and Perception-Neutral groups across all sentence points, with the expectation that Causative-Neutral would have a significantly higher mean than the Perception-Neutral groups. Finally, we compared the Causative-Towards and Perception-Towards groups, with the expectation that the mean would be higher in the former than the latter.

Results

Manipulation Checks

In order to ensure that participants were paying attention to the experimental task, a short cued recall test was given at the end. All but one of the quiz scores ranged from 92% to 100%. One participant scored 50%, and was not included in the analyses. It is not clear if the participant did not understand the task at hand, or did not sufficiently attend to the scenes and sentences, but to eliminate the possibility of skewing the results, this participant's data was excluded from all of the following analyses.

A second manipulation check examined the effect of condition on the proportion of trials (by subject) where the participant looked at the target object before the onset of the verb. Trials where participants were already fixating the object at verb-onset were included because this had no bearing on whether they chose to look at the object before they heard the verb. We did not expect a difference based on verb type because the verb could not have possibly had any effect on eye movement behaviour prior to its utterance. However, we did expect a possible main effect of motion type based on the position and apparent motion of the agent, because this did differ across motion conditions, and may have drawn more or less attention towards the target object in the early moments of scene viewing. The results (shown in Figure 4) suggested that the apparent motion of the agent did not significantly affect pre-verb eye movement behaviour, which was contrary to our hypothesis, and confirmed that there was no difference in how often the target object was fixated before the verb, as expected. The repeated-measures 2 X 3 ANOVA (*N* = 32) did not reveal any significant main effects or interaction effect (verb type: *F*(1, 62) = 1.42, *p* = .24; motion type: *F*(2, 62) = 1.89, *p* = .16; interaction: *F*(2, 62) < 1, *p* = .61).

Missing Data

There were three sources of missing data in this experiment. The first was due to a computer error in which the data from the first trial of several experimental runs was corrupted. Out of the 528 trials presented to participants, 32 (6.1%) were due to corrupted data. This represents a small but substantial proportion of lost data considering it was due to experimental and not participant error, but because trials were presented randomly, these trials were distributed evenly across the various experimental conditions.

The second source of missing data (for some of the analyses reported below; see Analyses section above) was trials in which participants never fixated the target object after verb-onset. Twenty-eight (5.3%) such trials were recorded. In order to examine subject) where participants did not launch a saccade to the target object after verb-onset (and the utterance of its noun referent, to be exact), a repeated-measures 2 X 3 ANOVA whether verb type and motion type had an effect on the proportion of trials (computed

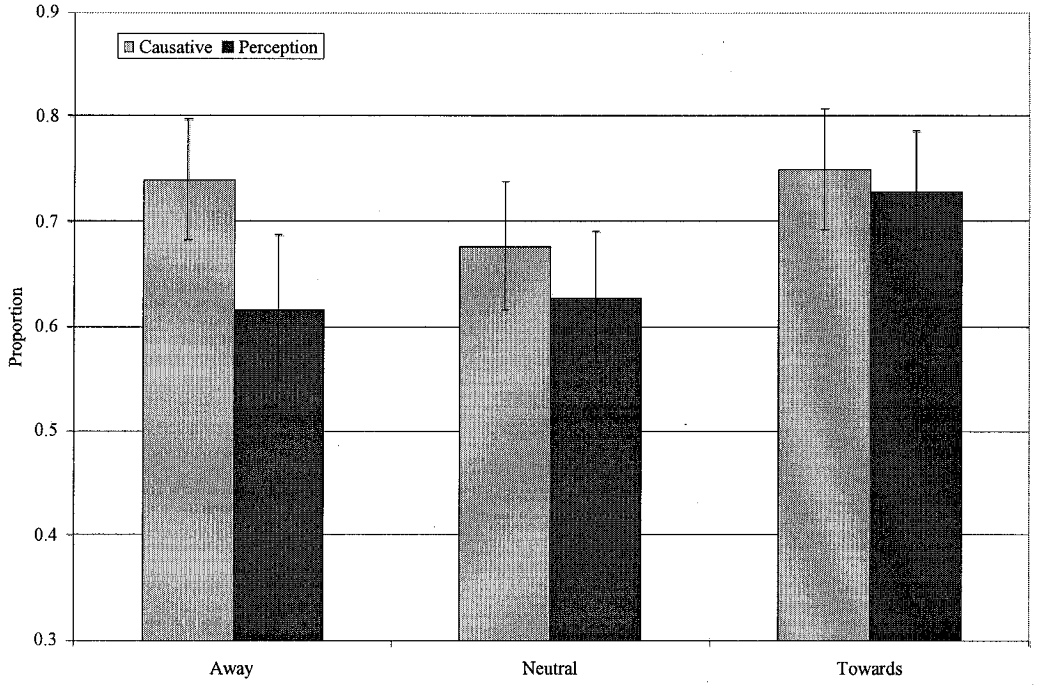


Figure 4. Mean proportion (± SE) of trials containing one or more fixations to the target object before verb-onset as a function of condition.

by ($N = 31$) was conducted. Neither verb type ($F(1, 60) < 1, p = .54$) nor motion type ($F(2, 60) < 1, p = .80$) had a significant main effect, nor was there a significant interaction effect between the two independent variables ($F(2, 60) < 1, p = .62$). This indicates that the apparent motion of the agents in the scenes and verb class did not affect whether participants did not fixate the target object after verb-onset; in other words, the number of trials with no post-verb fixations was evenly distributed across the six conditions.

The third source of missing data derived from trials in which participants happened to be fixating the target object at verb-onset. This occurred in 99 (18.8%) of the trials, a surprisingly high proportion. These trials had to be excluded from any analyses examining the effect of verb type on subsequent eye movement behaviour, because of the inability to discern whether participants continued to fixate the object because they heard the more semantically-restrictive verb that could have involved that object or not. A 2 X 3 repeated-measures ANOVA ($N = 31$) was conducted to examine the effect of verb type and motion type on the proportion of these trials (by subject). The results indicated that there were no main effects of verb type ($F(1, 60) < 1, p = .65$) or motion type ($F(2, 60) < 1, p = .73$), nor a significant interaction effect ($F(2, 60) < 1, p = .92$). These results confirm the idea that the agent's apparent direction of motion did not affect whether participants were fixating the target object at the time the verb was spoken. They also support the notion that the verb class did not affect whether participants were looking at the object at the time the verb was spoken, as would be expected given that this would have been a retroactive (and therefore impossible) effect. Finally, these results indicate that these trials were evenly distributed across the six conditions.

To further explore what may have led participants to be fixating the target object at verb-onset, apart from the stimulus variables, a chi-square analysis was conducted. Here, the effect of whether or not the target object was fixated before verb-onset on whether it was being fixated at verb-onset was examined. The analysis revealed that there was no significant effect, $\chi^2(1, N = 495) = .10, p > .05$. This is contrary to the expectation that having previously fixated on an object might elicit further fixations at some point in the future. However, it is consistent with the idea that having spent little time attending to it earlier might have elicited more fixations at a later time, either because its probability of being fixated increased as time went on or because it had been a previously unexamined region of the scene.

In summary, the pattern of missing data attributable to these three sources (corrupt data trials, trials where the participant never fixated the target object after the verb, and trials where the participant was already fixating the target object at verb-onset) was randomly distributed across the six conditions. This therefore can be taken to indicate that the effects of the two main variables of the main analyses can be interpreted meaningfully.

Raw Data Conversion

In order to convert the data generated by the EyeLink DataViewer, the following procedure was followed. The DataViewer produced a scan path of the saccades and fixations across the scene, as illustrated by Figure 5. Only saccades initiated towards and fixations positioned on the target object were retained for further analyses. For the main variable of interest, SOT, the difference between the two time points was calculated: the time of verb-onset (standardized to 3189.5 ms for all scene/sentence combinations) and

Figure 5. The scan path obtained from a typical trial. The yellow arrows correspond to saccades, and are numbered to indicate their sequence. The turquoise circles correspond to fixations; their diameter is proportional to their duration, indicated by the numbers located at their periphery. Note how most of the fixations are clustered around visually salient features of the scene, such as the agent's face, and various clearly defined objects (e.g., the cup of milk).

the time at which the first saccade was initiated towards the target object after the verb-onset. In addition, other variables recorded included the time spent fixating the target object before and after the verb onset (calculated by summing the durations of the saccades launched towards and fixations on the target object), as well as whether the target object was being fixated at verb onset.

Anticipatory Eye Movements

In order to determine whether any anticipatory eye movements may have been made, SOTs were compared to two other time points in the sentences: the noun-onset and the noun-offset. This was done to examine whether eye movements may have been initiated after hearing the verb but prior to hearing the onset and offset of the noun. In addition, we calculated the proportion of trials (those spoilt due to corrupt data were not counted) in which a saccade was launched towards the target object *before* the onset of the noun. No such evidence of anticipatory eye movements was found. On average, eye movements were initiated 771 ms after the noun-onset, and 486 ms after the offset of the noun, across all conditions. In addition, saccades were launched towards the target object before the onset of the noun in only 6.3% of the causative trials and 9.7% of the perception trials.

Finally, the difference in time between the offset of the noun and SOT was computed and subjected to a 2 (verb type) X 3 (motion type) repeated-measures ANOVA, in order to determine whether these first saccades were affected by verb type and motion type. Results indicated that only motion type had a significant main effect, $F(2, 48) = 9.3, p < .001$, as predicted. A modified Bonferroni/Dunn test, with corrected alpha levels of .03 for each of the three pairwise comparisons, was conducted to explore the significant main effect of motion type. This indicated that the Towards condition (*M* = 170, *SD* = 541) had a lower mean difference than the Away (*M* = 552, *SD* = 561) and Neutral (*M* = 626, *SD* = 618) conditions, $p1 < .01$, $p2 < .001$, which did not significantly differ from each other, $p = .52$, again, as predicted. The effect of verb type was not significant, $F(1, 48) < 1$, $p = .43$, contrary to our hypothesis, nor was the interaction, $F(2, 48) < 1$, $p = .82$. A paired one-tailed t-test between the Causative-Towards (*M* = 161, *SD* = 493) and Perception-Towards (*M* = 180, *SD* = 595) conditions did not reveal any significant differences, $t(24) = -.13$, $p = .45$, contrary to our hypothesis.

Target Object Saliency

The first analysis examined the correlation between target object saliency ratings, pre-verb fixation durations, post-verb fixation durations and total fixation durations. We hypothesized that the saliency ratings would correlate positively and significantly with the amount of time spent looking at the target objects. A Pearson's correlation (*N* = 396) indicated that target object saliency ratings did correlate significantly with the time spent looking at the target object before verb-onset ($r = .17$, $p < .01$) and the total time spent looking at it ($r = .14$, $p < .01$), but not the amount of time spent looking after verb-onset ($r = .06$, $p = .27$).

Next, the relationship between target object saliency and SOTs was computed. We hypothesized that there would be a significant positive correlation, which a Pearson's correlation (*N* = 368) did not confirm ($r = -.001$, $p = .98$). Because this correlation was not significant, target object saliency was not included as a covariate in the main analysis described below.

Finally, the relationship between target object saliency ratings and whether or not the target object was being fixated at verb-onset was examined. We expected that there would be a significant positive correlation, such that the higher the saliency rating, the more likely the target object would be fixated at verb-onset. A one-tailed point biserial correlation ($N = 495$) was computed, which indicated that there was no significant correlation ($r = -.05, p = .12$), contrary to our hypothesis.

Target Event Saliency

The same set of analyses described above was conducted with target event saliency instead of target object saliency. First, the relationship between target event saliency ratings and the three fixation durations was examined and was expected to yield significant positive correlations. However, a Pearson's correlation ($N = 396$) indicated that only the amount of time spent looking at the target object after verb-onset correlated significantly with target event saliency ($r = .10, p = .04$); time spent fixating before did not ($r = -.04, p = .47$), nor did the total time spent fixating ($r = .07, p = .16$). Because the correlation between target event saliency and time spent fixating the target object after verb-onset was so low, target event saliency was not included as a covariate in the main analysis.

Second, the relationship between target event saliency and SOTs was computed. We hypothesized that there would be a significant positive correlation, which a Pearson's correlation ($N = 368$) did not confirm ($r = -.11, p = .04$); although the correlation was significant, it was in the direction opposite to that expected, such that as target event saliency increased, SOTs decreased. For that reason, target event saliency was not included as a covariate in the main analysis described below.

Third, the relationship between target event saliency ratings and whether or not the target object was being fixated at verb-onset was examined. We expected that the more predictive a scene was in terms of the target event, the more likely the target object (implicated in the target event) would be fixated at verb-onset. A one-tailed point biserial correlation ($N = 495$) was computed, which indicated that there was a marginally significant correlation ($r = .08, p = .06$), such that the higher the target event saliency rating, the more likely that the target object would be fixated upon at verb-onset.

Main Analysis: First Post-Verb Saccades to Target Object

The first analysis examined the effect of verb type and motion type on post-verbal eye movement behaviour, namely saccade onset time (by subjects). The data from only 24 participants was included, due to missing data from seven of them (from the sources described above). We hypothesized that both verb type and motion type should have a main effect on SOT. However, this was not confirmed, as shown in Figure 6 (see Appendix J for all ANOVA tables relevant to the results of this experiment). The results indicated that only motion type had a significant effect on SOTs, $F(2, 46) = 8.95, p < .01$. Verb type did not have a significant main effect, $F(1, 46) < 1, p = .55$, nor was the interaction significant, $F(2, 46) < 1, p = .92$. A modified Bonferroni/Dunn test, with adjusted alpha levels of .03 for each of the three pairwaise comparisons, was conducted to explore the significant main effect of motion type. This indicated that both the Away (M= 1266, SD = 563) and Neutral (M=1316, SD = 617) groups differed significantly from the Towards group (M= 886, SD = 540), Away vs. Towards: $p < .01$; Neutral vs. Towards: $p < .001$, such that the Towards group had the lowest mean SOT. The Away and Neutral groups did not differ significantly ($p = .65$).

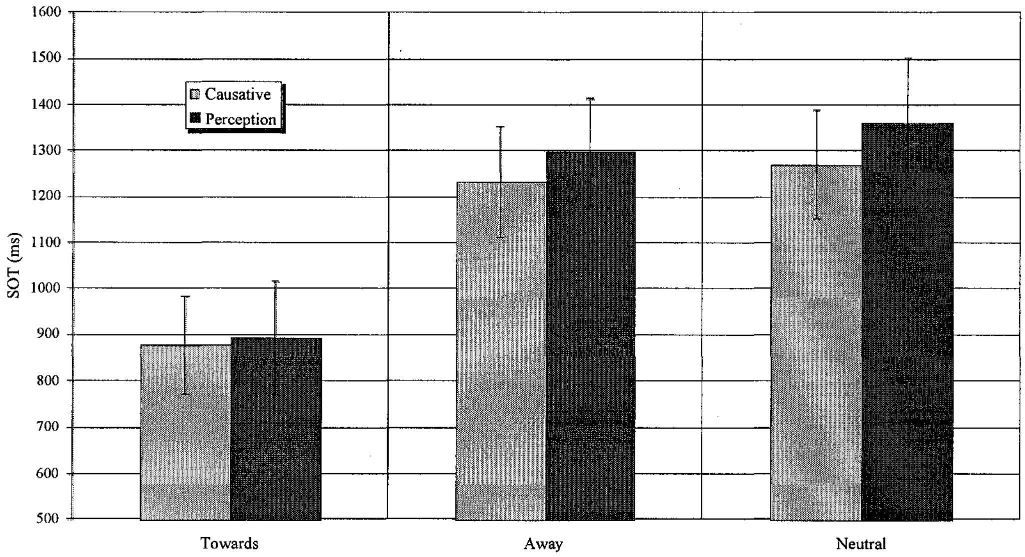


Figure 6. Mean SOTs (± *SE*) as a function of verb type and motion type, computed by subject.

In order to test the hypotheses outlined in the Analyses section, four planned comparisons were conducted. The first hypothesis was that the Away and Towards conditions would differ significantly, such that the Towards condition would have a lower mean SOT than the Away condition. A 2 X 2 (motion type X verb type) ANOVA was conducted to that effect ($N = 30$), and revealed that motion type did indeed have a significant main effect, $F(1, 29) = 8.93, p < .01$, with the Towards condition ($M = 994$, $SD = 608$) having a lower mean SOT than the Away condition ($M = 1332$, $SD = 651$), as predicted. The second hypothesis was that the Neutral and Towards conditions would differ significantly, such that the Towards condition would have a lower mean SOT than the Neutral condition. Again, this hypothesis was confirmed with a 2 X 2 repeated-measures ANOVA ($N = 25$), $F(1, 24) = 29.44, p < .0001$: the Towards condition ($M = 872$, $SD = 534$) had a lower mean SOT than the Neutral condition ($M = 1318$, $SD = 644$). Third, to compare the two verb types without the confounding effects of the motion context, the Causative-Neutral and Perception-Neutral groups were compared. In the absence of any apparent motion in the scenes, we predicted that the Causative condition would yield shorter SOTs than the Perception condition. A one-tailed paired t-test failed to support this prediction, $t(23) = -.72, p = .24$, although the trend was in the right direction, with the Causative-Neutral condition ($M = 1230$, $SD = 590$) having a lower mean than the Perception-Neutral condition ($M = 1406$, $SD = 695$). Finally, to compare the two verb types in the Towards condition, we compared the Causative-Towards and Perception-Towards conditions. We expected that the visual context (agent appearing to move towards the target object) would aid in the semantic interpretation of the verb, such that SOTs would be lower in the Causative-Towards than in the Perception-Towards condition. A one-tailed paired t-test failed to lend support to this notion, $t(30) = -.98$, $p = .17$, although the trend was in the right direction, with the Causative-Towards condition ($M = 918$, $SD = 529$) having a lower mean than the Perception-Towards condition ($M = 1041$, $SD = 674$).

The second analysis was identical to that described above, except that it was conducted by items and not by subjects. The same hypotheses held for this analysis. Because of missing cell means from two of the items, *shoes* and *vase*, these analyses were conducted on 15 of the 17 items, which does weaken the analysis. This analysis, contrary to that computed by subjects, supported the hypotheses. A 2 X 3 repeated-measures ANOVA revealed a marginally significant effect for verb type, $F(1, 28) = 4.32$, $p = .06$, with the Causative condition having a lower mean SOT than the Perception condition, and a significant effect for motion type, $F(2, 28) = 8.02$, $p = .02$ The interaction was not significant, $F(2, 28) < 1$, $p = .66$. A modified Bonferroni/Dunn test, with corrected alpha levels of .03 for each of the three pairwise comparisons, was conducted to explore the significant main effect of motion type. The Towards condition produced significantly shorter SOTs than the Away and Neutral conditions: Away vs. Towards ($p < .01$); Neutral vs. Towards ($p < .01$). The Away and Neutral conditions did not differ significantly ($p = .63$).

Planned comparisons for this data set also remained the same, as did their respective hypotheses: (1) Away < Towards; (2) Neutral < Towards; (3) Causative-Neutral < Perception-Neutral; and (4) Causative-Towards < Perception-Towards. The first hypothesis was supported by a 2 X 2 repeated-measures ANOVA; the main effect of motion type was significant, $F(1, 14) = 21.48$, $p < .001$, with the Towards condition having a shorter mean SOT than the Away condition, as expected. The second hypothesis was also confirmed; the main effect of motion type in a second 2 X 2 repeated-measures ANOVA was significant, $F(1, 14) = 7.10, p = .02$, such that the Towards condition also had a shorter mean SOT than the Neutral condition. The third hypothesis had some support; a one-tailed paired t-test indicated that the Causative-Neutral condition had a marginally significantly shorter mean SOT than the Perception-Neutral condition, $t(16) = -1.45, p = .08$. Finally, the fourth hypothesis was supported; the Causative-Towards condition had a significantly lower mean than the Perception-Towards condition, $t(14) = -7.11, p < .0001$.

As mentioned above, although SOT did correlate significantly with target event saliency, as reported above, these analyses were not repeated as ANCOVAs with target event saliency as the covariate. This was because the correlation was not sufficiently high to warrant this analysis.

Analysis of Early Post-Verb Cumulative Saccades to the Target Object

This final analysis examined the effects of verb type, motion type and sentence point on the cumulative proportion of saccades to the target object after verb-onset using a 3 (sentence point) X 2 (verb type) X 3 (motion type) ANOVA (by subjects). We expected that there would be a significant interaction between sentence point and verb type (such that the difference between the two verb types would increase as the sentence unfolded), as well as a significant main effect of motion type. The results (which are plotted in Figures 7 and 8) did not confirm our hypotheses. The results of the ANOVA indicate that the two-way interaction of sentence point and motion type was significant, $F(4, 120) = 2.58, p = .04$, contrary to our hypotheses. In addition, the interaction between

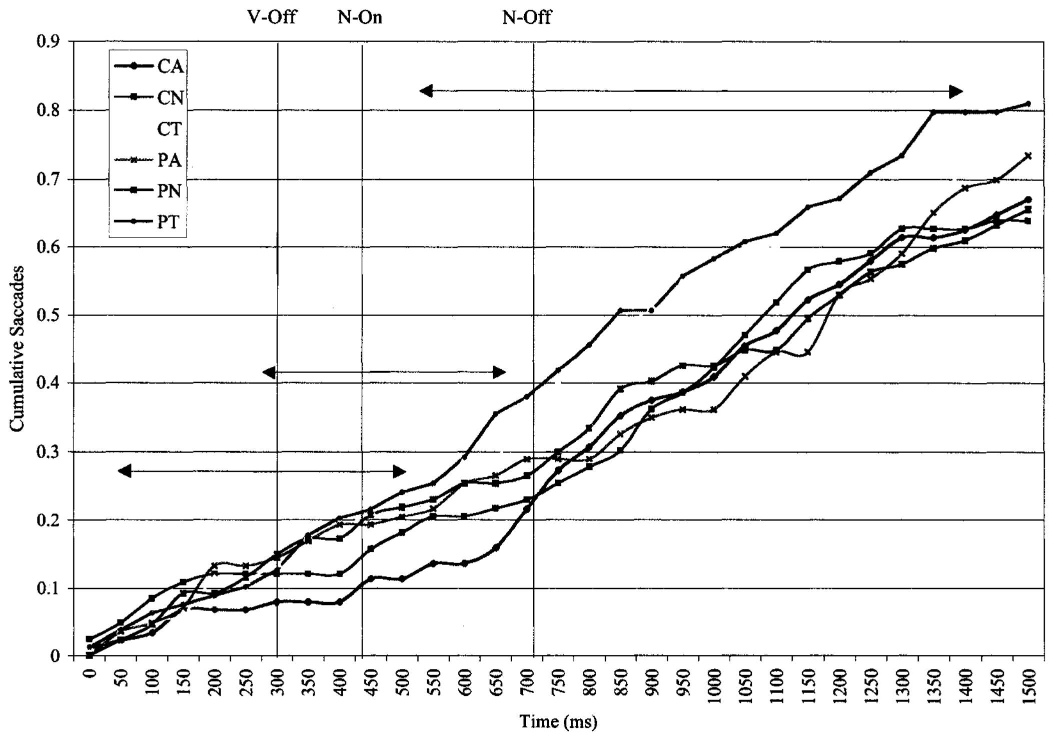


Figure 7. A plot of the mean cumulative number of saccades to the target object after verb-onset. Each line refers to a single condition, and each point to one 50-ms bin. The origin of the X-axis refers to the verb-onset, and the three vertical lines mark the temporal boundaries of the verb and noun (average onset and offset). The double-headed horizontal arrows on each boundary indicate the range of onsets and offsets at the points in time relative to the verb-onset. This was done to take into account the variable lengths of each of the sentence segments (i.e., the verbs and noun phrases). The point at which each of the coloured lines (referring to cumulative fixations for each condition) intersects with the three critical sentence points (verb-offset, noun-onset and noun-offset) were computed and compared in an ANOVA.

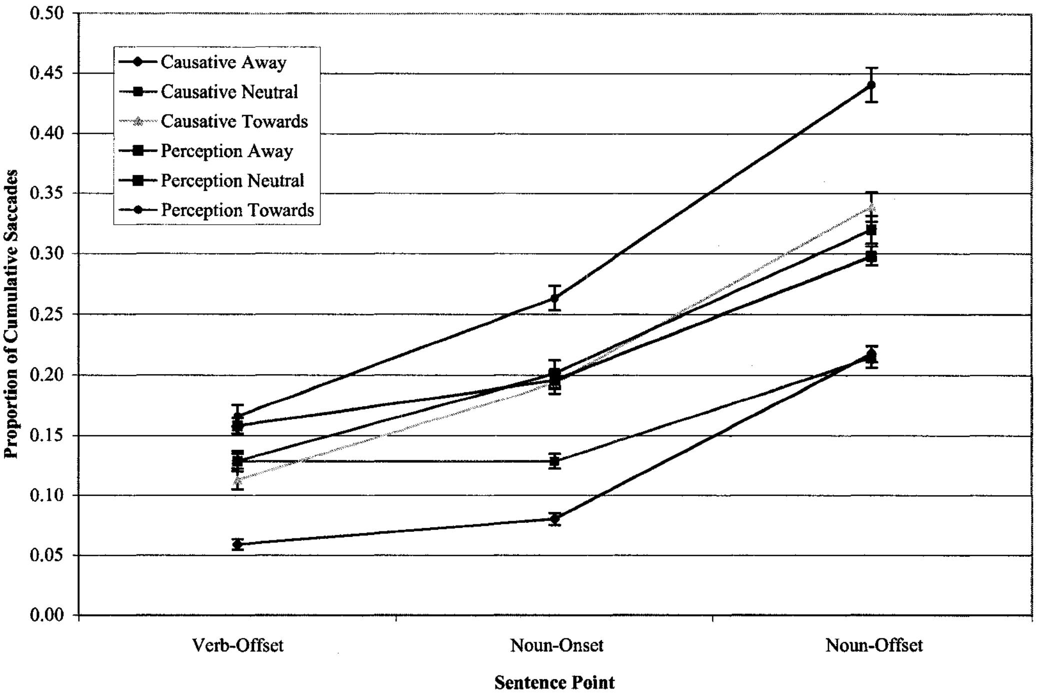


Figure 8. Mean number of cumulative saccades towards target object (± *SE*) at each of the three critical sentence points: verb-offset, noun-onset and noun-offset.

sentence point and verb type was not significant, $F(2, 120) = 1.42, p = .25$, again contrary to our hypotheses. However, both sentence point and verb type had a significant main effect; $F1(2, 120) = 53.16, p < .0001$, $F2(1, 120) = 4.22, p < .05$. As expected, the Causative group ($M = .164$, $SD = .248$) had a lower mean than the Perception group ($M = .242$, $SD = .311$). In addition, the Verb-Offset (V-Off) condition ($M = .126$, $SD = .221$) had a significantly lower mean than the Noun-Onset (N-On; $M = .177$, $SD = .258$) and Noun-Offset (N-Off; $M = .305$, $SD = .332$) conditions, $p1 < .01$, $p2 < .0001$; the N-On condition had a significantly lower mean than the N-Off group as well, $p < .0001$. Finally, the main effect of motion type was not significant, $F(2, 120) = 2.25, p = .11$, failing to provide support for our hypotheses.

To further explore the significant interaction of motion type and sentence point, a test of the simple effect of sentence point was conducted for each of the three levels of motion, which indicated that the effect of sentence point did differ at the various levels of motion type. Specifically, the simple effect of sentence point was significant at each level of motion type (Away: $F(2, 60) = 21.39, p < .0001$; Neutral: $F(2, 60) = 25.13, p < .0001$; Towards: $F(2, 60) = 18.55, p < .0001$).

Furthermore, a simple comparison indicated that the difference between V-Off and N-On at Away was significant, such that the mean number of cumulative fixations was lower at V-Off ($M = .094$, $SD = .208$) than at N-On ($M = .141$, $SD = .267$), $F(1, 123) = 6.13, p = .02$. In addition, the simple comparison between V-Off and N-Off ($M = .269$, $SD = .292$) at Away was also significant, $F(1, 30) = 27.31, p < .0001$. Finally, the simple comparison between N-On and N-Off at Away was significant as well, $F(1, 30) = 20.04$, $p = .0001$.

Next, to examine the simple of effect of sentence point at Neutral, a series of simple effects revealed that V-Off (*M* = .144, *SD* = .198) was not significantly different from N-On (*M* = .162, *SD* = .211), *F*(1, 30) = 3.08, *p* = .09, although V-Off was significantly different from N-Off (*M* =.228, *SD* = .287), *F*(1, 30) = 33.74, *p* < .0001. In addition, N-On was significantly different from N-Off, *F*(1, 30) = 23.25, *p* < .0001.

The last series of simple comparisons of sentence point at Towards revealed that V-Off (*M* = .140, *SD* = .253) was significantly different from N-On (*M* = .228, *SD* = .287), *F*(1, 30) = 11.23, *p* < .01. In addition, V-Off was significantly different from N-Off (*M* = .390, *SD* = .414) at Towards, *F*(1, 30) = 23.51, *p* < .0001. Finally, the simple effect of sentence point at Towards indicated that N-On was significantly different from N-Off, *F*(1, 30) = 14.17, *p* < .001.

In addition, there was a significant main effect of verb type, such that the mean number of fixations of the causative group (*M* = .164, *SD* = .248) was lower than that of the Perception group (*M* = .242, *SD* = .311), *F*(1, 120) = 4.22, *p* < .05, which is in the direction opposite to that which was predicted. The three-way interaction between sentence point, verb type and motion type was not significant, *F*(4, 120) < 1, *p* = .96.

Planned comparisons tested the hypotheses that the Away group would exhibit significantly higher means than the Neutral and Towards groups across all sentence points. These hypotheses were not supported: there only a marginally significant difference between the Away group (*M* = .168, *SD* = .267) and the Towards group (*M* = .253, *SD* = .340), *F*(1, 60) = 3.24, *p* = .08, although the trend was in the predicted direction. In addition, the difference between the Neutral (*M* = .188, *SD* = .228) and Towards groups was not significant, *F*(1, 60) = 2.39, *p* = .13, although again, the trend was in the predicted direction. Next, the hypothesis that the Causative-Neutral group would have a higher mean than the Perception-Neutral group across all sentence points was not supported. Results indicated that the Causative-Neutral condition (*M* = .158, *SD* = .226) was not significantly different from the Causative-Perception condition (*M* = .218, *SD* = .229), *F*(1, 60) = 1.55, *p* = .22, and in fact, the trend was in the direction opposite to that expected. Finally, the hypothesis that the Causative-Towards condition (*M* = .215, *SD* = .311) would have a lower mean than the Perception-Towards condition (*M* = .290, *SD* = .364) across all sentence points was not supported, *F*(1, 60) = 1.28, *p* = .27.

Discussion

The results of this experiment supported some, but not all, of our hypotheses. The saccade onset times reported here are not consistent with the notion that participants initiated these saccades towards the target object before the noun-onset or the noun-offset. In fact, on average, saccades were launched 486 ms after the end of the noun's utterance. In addition, this differed by motion context, such that in the Towards condition, saccades were launched sooner (170 ms after the offset of the noun) than in the Away (552 ms after) and Neutral (626 ms after) conditions. This did not differ by verb type, however, with the saccades being launched 412 ms after the offset of the noun in the Causative condition and 487 ms after in the Perception condition. This is in contrast with other results indicating that eye movements are closely locked to language input. For example, Tanenhaus et al. (1995) found that eye movements were launched towards the object referents approximately 250 ms after the nouns were uttered. Similarly, Altmann and Kamide (1999) report that the first saccade after the verb occurred 127 ms after the onset of the noun in the semantically nonrestrictive condition, and 85 ms before the onset of the noun in the semantically restrictive condition. However, note that when the visual context was consistent with the verb class, such that the agent appeared to be moving towards the target object, saccades were actually initiated during the lifetime of the noun, if we take the programming of a saccade to require 200 ms. This indicates that eye movements were closely locked to the utterance of the verb.

In the present experiment, saccades were launched towards the target object before the onset of the noun in only 6.3% of the causative trials and 9.7% of the perception trials, compared to 32% of the semantically restrictive trials and 18% of the semantically non-restrictive trials in the second Altmann and Kamide (1999) experiment. It might be that the initiation of saccades was delayed in this experiment because of the nature of the scenes; perhaps their complexity, similarity to real-life scenes, and the relatively low salience of the target objects (compared to the simplistic line drawings used in the Altmann and Kamide, 1999 study) contributed to the lag.

On the other hand, there was some support for a relationship between target object saliency and the amount of time spent fixating the target object. The correlation was significant for time spent fixating before verb-onset and the total time, but not for time spent fixating after verb-onset. The fact that fixation times before verb-onset increased as the saliency of the object increased, whereas fixation times after verb-onset did not, points to the possibility that saliency has more of an effect early in scene viewing. These objects may attract longer viewing times at first, but once they have been sufficiently inspected, they may no longer do so. However, given that the correlations were so low (approximately .1 or less), target object saliency may not have been a particularly important factor in determining fixation lengths.

With regards to target event saliency, a somewhat different pattern of findings was obtained, relative to target object saliency. Target event saliency correlated only with the amount of time spent fixating the target object after verb-onset, not with the total time or the time spent fixating before. This could indicate that the target event was not anticipated by participants early in the trial, but once the event and its participants had been named, fixations towards the object involved may have increased with increasing event saliency.

Contrary to our hypothesis, target event saliency did not significantly and positively correlate with SOT; in fact, the correlation was significantly negative. This indicates that the more predictive the scene was of the target event, the less quickly participants initiated a saccade towards the target object. Why this was so is not clear; scene factors, such as their layout and complexity, may have contributed to this finding, but because the correlation was so low (-.106), its significance may have been spurious. In addition, the saliency ratings were relatively low (most items had saliency ratings of 0), which may also have contributed to this lack of a correlation.

Finally, target event saliency did affect whether or not participants were fixating the target object at verb-onset. This indicates that the more predictive a scene was of the event described by the sentence (and, by extension, its likely participants or role fillers), the more likely a participant would be to fixate the target object at the point at which the verb was uttered. Interestingly, this may indicate that participants were anticipating which object would be involved in the sentence described, and were fixating it just prior to the utterance of the verb so as to confirm whether or not it would be involved as an event participant. This corresponds to our hypothesis that the information conveyed by the sentence is used to confirm expectations about likely events garnered from the scene (its gist).

The main analyses (as well as the analysis of early post-verb eye movement behaviour) constituted the principal means of investigating the major hypothesis of this research, namely that verb-specific information can constrain referential domains, such that the more semantically restrictive verbs should have led to saccades being initiated towards the appropriate object referent of the grammatical object of the verb. In addition, they investigated the previously unexplored effect of apparent motion direction on eye movement behaviour. Overall, the results indicate that verb type did not have an effect on how quickly saccades were launched towards the target object after verb-onset, except in the analyses conducted by items, where the difference was only marginally significant. In addition, the analysis was somewhat weakened by the fact that two items had to be excluded due to missing data. This lack of an effect was further corroborated by the finding that there was no difference between the Causative-Neutral and Perception-Neutral conditions, where it was expected that the neutrality of the agent's positioning could not mediate the effect of verb type. Even with this neutrality, however, the more semantically restrictive causative verbs did not cause participants to launch saccades towards the target objects more quickly. Furthermore, there was no difference between the Causative-Towards and Perception-Towards conditions (except in the by items analysis, which suffers from the statistical considerations discussed above), which indicates that the apparent motion of the agent towards the target object did not serve to enhance the interpretation of the semantically restrictive verb to constrain eye movements towards that object.

These findings fail to support the notion that verb-specific information, in the form of thematic or conceptual roles, can guide eye movements, as measured by the first saccade initiated to the object referent of a verb's patient role filler after the utterance of that verb. This is contrary to the bulk of the findings reported in the literature; all have found support for the notion that syntactic and semantic information do direct eye movements towards objects that are consistent with the sentence being interpreted, in an incremental fashion (e.g., Tanenhaus et al., 2000). Moreover, studies that have employed a very similar stimulus set and methodology to the work presented here (e.g., Altmann & Kamide, 1999; Kamide et al., 2003) have found that the semantic information contained within verbs constrains eye movements in a similar fashion. However, the present results do not support the same view.

One reason may be that not enough participants were tested; the power for verb type was only .088 (as opposed to motion type, which was .974). Therefore, the sample size may not have been large enough to detect a significant difference between the two verb types. However, aside from these statistical considerations, there may be other factors that led to this result. One possibility is that the low-level properties of the scenes themselves may have prevented any cognitive (top-down) influences from affecting eye movements. For example, the fact that the still pictures we used were depictions of true scenes, and not *ersatz* scenes, may have changed how participants responded in terms of their scan patterns. It may be that the linguistic variables only have an effect on the visual system (as measured by eye movements) when the visual input is relatively impoverished and non-complex. Here, the scenes were fairly complex, and included actual pictures of people, who are known to attract a larger proportion of fixations (Henderson & Ferreira, 2004).

Nevertheless, it must be noted that the analysis by items did reveal an effect of verb type, despite (or perhaps because of) the loss of two items, and although the differences between causative and perception verbs was not significant in the analysis by subjects, the trend was in the expected direction. This, coupled with the low power of the experiment, does indicate that verb-specific semantic information may serve to constrain eye movements to some extent.

On the other hand, motion type did have a significant effect on the speed with which saccades were initiated towards the target object after the verb was uttered. In addition, the scenes in which the agent was facing or appeared to be moving towards the target object led participants to look towards the target object more quickly than when the agent appeared to be moving or facing away, or when he or she remained in a neutral position with respect to the object. Furthermore, there was no difference in saccade onset time between the Away and Neutral conditions. These findings confirmed our hypotheses, and suggest that the position of a human within a scene can serve to orient eye movements when that person's position points to scene regions of interest, at least when that position is consistent with the semantics of the sentence being uttered. However, when the person's orientation is not consistent with the sentence's meaning (i.e., the apparent motion or position of the agent does not match the upcoming event), it may be that a the central processes involved in integrating the linguistic input and visual context require extra time to process the inconsistency between the two inputs.

Because other studies within the visual world paradigm have shown that eye movements are locked to the incremental process of sentence comprehension at the earliest moments of interpretation, a separate analysis was conducted that examined eye movement behaviour at three critical sentence points: verb-offset, noun-onset and noun-offset. Although the hypothesis that sentence point would interact with verb type (such that the cumulative number of saccades towards the target object would increase more with time in the Causative condition than the Perception condition) was not supported, main effects for each factor were detected. This indicates, as would be expected, that the number of saccades initiated towards the target object increased as the sentence unfolded.

In addition, the effect of verb type was significant, although in the direction opposite to that predicted. Similar to this finding were the results of the planned comparison between the Causative-Neutral and Perception-Neutral groups across sentence points. These indicated that there was no effect of verb type; in other words, there was no difference between these two conditions. This can be interpreted to mean that verb type did not affect eye movements in the Neutral condition, although when all three conditions were included in the analyses, verb class did have an effect. On the whole, the verbs' selectional restrictions did not have an effect early in sentence comprehension.

In addition, the interaction between sentence point and motion type was significant, such that the effect of sentence point was different at the various levels of motion type, which was not expected. While statistically significant, this finding is not of theoretical significance; more telling is that overall, the motion conditions did not significantly differ from each other. The planned comparisons also lent support to this notion, as the Away and Towards group did not differ from each other, and nor did the Neutral and Towards groups. This may have been because, for the few saccades that were initiated so early after the verb-onset, the linguistic context took precedence over the motion context. However, overall, when counting all trials, motion context had a stronger influence over the direction of eye movements.

In summary, the results of Experiments 1 show that there was only weak support for the hypothesis that verb-specific information can guide eye movements. Much stronger support was shown for the effect of the apparent motion of the agent on eye movements. It appears that when realistic depictions of everyday scenes are employed, the visual context takes precedence over the linguistic context in constraining the domain of subsequent reference. What remains to be seen is whether these weak verb effects would hold for a dynamic visual context. Experiment 2 was motivated by this question, and because language interpretation occurs mainly within a dynamic world, studying these effects with dynamic scenes would serve to increase the ecological validity of Experiment 1.

Experiment 2

Introduction

The purpose of Experiment 2 was to examine whether the findings from the visual world paradigm would extend to a dynamic visual context. We expected that verb-specific semantic information would constrain eye movements to the relevant visual referents. In addition, it was expected that motion type would also have an effect on eye movements, especially given the additional component of motion within the scenes.

Methods

Participants

Thirty-eight Concordia University students participated in this experiment. None of these participants took part in the previous two experiments. There were 32 females and 6 males, ranging in age from 19 to 33. Participants were all native speakers of English, and had normal or corrected-to-normal vision. All either received course credit for their participation, or were given monetary compensation.

Materials and apparatus

The stimuli used in this experiment were very similar to those used in Experiment 1. The sentences were identical, but the scenes presented to participants were in the form of short movie clips. As in Experiment 1, the film clips differed in terms of the direction of motion of the agent in relation to the visual referent of the verb's direct object (e.g., milk). In each movie, the agent either moved towards or away from the object, or remained in a neutral position. There were a total of 102 sentence/movie combinations distributed in 6 lists. The movies ranged in length from 1582 ms to 5656 ms, although each one had a "disambiguating point," defined as the time at which the agent began moving (in the Away and Towards conditions), which coincided with the onset of the main verb. All eye movements subsequent to this disambiguating point were included in the analyses. This point occurred between 1875 and 2689 ms after the beginning of the movie. The equipment was the same as that used in Experiment 1.

Procedure

The procedure in this experiment was identical to that for Experiment 1, except for some minor differences in the presentation and instructions. See Appendices K and L for copies of the consent forms and instructions given to the participants. Appendix M contains a copy of the short form of the instructions presented just prior to the initiation of the experiment.

Participants were asked to watch the movies and to pay attention to the sentences, and no specific task was required of them. Each trial began with a red fixation cross that appeared on a black screen for 1 s, followed by the first frame of the movie overlaid by the fixation cross for another 2 s, at which point the cross disappeared and the film clip was put into motion. The fixation cross was included to orient participants' eyes at the beginning of each trial and to ensure a uniform starting point for all participants, and the first frame of each movie was presented for two seconds, along with the fixation cross, to allow participants to extract some basic information about the scene prior to the beginning of the sentence.

Analyses

All analyses and predictions were identical to those in Experiment 1, except for one. In the analysis of early cumulative saccades to the target object after verb-onset, instead of a significant main effect of motion type, we expected there to be a significant interaction between sentence point and motion type, with there being a greater increase in the number of cumulative saccades in the towards condition than in the away and neutral conditions at increasing sentence points. This is because the number of saccades initiated towards the target object should increase with increasing sentence points, especially as the direction of motion towards the target becomes more apparent with time. This is in contrast to Experiment 1, where the agent never moved, and where the motion context was only implied.

Results

Manipulation Checks

As in Experiment 1, a short cued recall test was given at the end to ensure attention was paid during the experimental trials. Quiz scores ranged from 75% to 100%; therefore, all participants were retained in the analyses.

A second manipulation check examined the effect of condition on the proportion of trials (by subject) where the participant looked at the target object before the onset of the verb. Trials where participants were already fixating the object at verb-onset were included because this had no bearing on whether they chose to look at the object before they heard the verb. We did not expect a difference based on verb type because the verb could not have possibly had any effect on eye movement behaviour prior to its utterance. In addition, we did not expect a main effect of motion type because this did not differ across motion conditions, as all agents remained in more or less the same position prior to the disambiguating point. The results (shown in Figure 9) suggested that the apparent motion of the agent did significantly affect pre-verb eye movement behaviour, contrary to our hypothesis, and confirmed that there was no difference in how often the target object

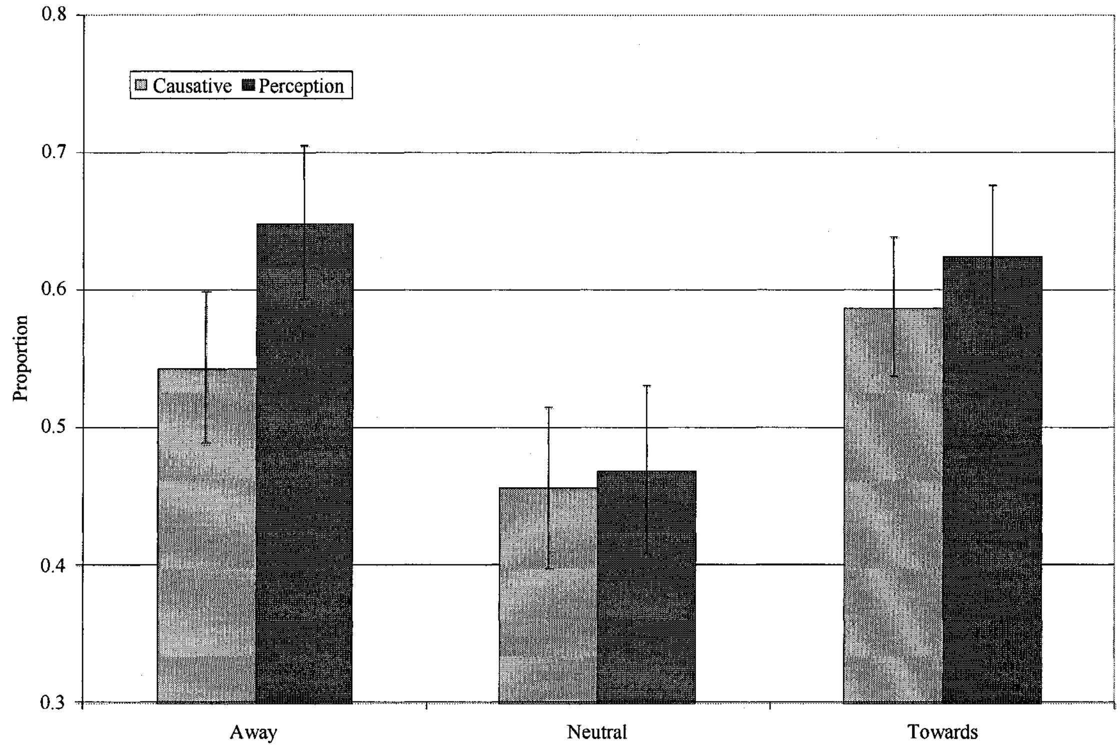


Figure 9. Mean proportion (± SE) of trials containing one or more fixations to the target object before verb-onset as a function of condition.

was fixated before the verb, as expected. The repeated-measures 2 X 3 ANOVA (*N* = 38) did not reveal a significant main effect of verb type, nor an interaction effect (verb type: *F*(1, 74) = 2.56, $p = .12$; interaction: *F*(2, 74) < 1, $p = .61$). However, there was a significant main effect of motion type, *F*(2, 74) = 6.20 $p < .01$, such that a Modified Bonferroni/Dunn test, with the alpha set to .03 for each pairwise comparison, indicated that the Away condition (*M* = 60, *SD* = 34) had significantly more fixations than the Neutral condition (*M* = 46, *SD* = 36), $p < .01$. In addition, the Neutral condition had a significantly more fixations than the Towards condition (*M* = 61, *SD* = 31), $p < .01$. There was no difference between the Away and Towards conditions, $p = .83$.

These results indicate that participants fixated the target object more often in the Towards condition than in the Away and Neutral conditions. Why this result was obtained is difficult to interpret, because of the lack of any clear path of motion prior to the disambiguating point. It is possible that the participants may have anticipated possible upcoming events and their participant object referents based on the gist extracted from the scenes early in the films, and may have participants may have launched anticipatory eye movements towards them. Because the pattern of eye movements was not equal across all motion condition, the interpretation of the results of the remainder of the analyses must be made cautiously. Having fixated the target object more in the Towards condition than in the other two conditions may have two possible consequences: subsequent eye movements may have occurred more quickly after the verb-onset due to participants having encoded the location of the relevant object, or may have occurred more slowly (or not at all) because of habituation to the relevant object.

Missing Data

The proportions of missing data from the same three sources as in Experiment 1 were computed. The first source of missing data was due to a computer error in which the data from the first trial of several experimental runs was corrupted. Out of the 637 trials presented to participants, 29 (4.5%) were due to corrupted data. These trials were distributed evenly across the various experimental conditions.

The second source of missing data (for some of the analyses reported below; see the Analyses section in Experiment 1) was trials in which participants never fixated the target object after verb-onset. Ninety-one (14.3%) such trials were recorded. In order to examine whether verb type and motion type had an effect on the proportion of trials (computed by subject) where participants did not launch a saccade to the target object after verb-onset, a repeated-measures 2 X 3 ANOVA ($N = 38$) was conducted. Neither verb type ($F(1, 74) = 1.81, p = .19$) nor motion type ($F(2, 74) < 1, p = .88$) had a significant main effect, nor was there a significant interaction effect between the two independent variables ($F(2, 74) < 1, p = .52$). This indicates that the apparent motion of the agents in the scenes and the verb class did not affect whether participants did not fixate the target object after verb-onset; in other words, the number of trials with no post-verb fixations was evenly distributed across the six conditions. This rules out the idea of object habituation discussed in the manipulation check analyses presented above, where having fixated the target object more frequently in the Towards condition did not lead to a disproportionate lack of post-verbal fixations.

The third source of missing data derived from trials in which participants happened to be fixating the target object at verb-onset. This occurred in 41 (6.4%) of the trials. These trials had to be excluded from any analyses examining the effect of verb type on subsequent eye movement behaviour, because of the inability to discern whether participants continued to fixate the object because they heard the more semantically-restrictive verb that could have involved that object or not. A 2 X 3 repeated-measures ANOVA ($N = 38$) was conducted to examine the effect of verb type and motion type on the proportion of these trials (by subjects). The results indicated that there was a marginally significant main effect of verb type, $F(1, 74) = 3.84, p = .06$, contrary to our hypothesis. However, there was no main effect of motion type, $F(2, 74) < 1, p = .64$, nor a significant interaction effect, $F(2, 74) < 1, p = .98$, as expected. These results confirm the idea that the agent's apparent direction of motion did not affect whether participants were fixating the target object at the time the verb was spoken. However, they do not support the notion that the verb class did not affect whether participants were looking at the object at the time the verb was spoken, given that this was a retroactive (and therefore impossible) effect. Finally, these results cannot confirm that these trials were evenly distributed across the two verb conditions.

To further explore what may have led participants to be fixating the target object at verb-onset, apart from the stimulus variables, a chi-square analysis was conducted. Here, the effect of whether or not the target object was fixated before verb-onset on whether it was being fixated at verb-onset was examined. The analysis revealed that there was no significant effect, $\chi^2(1, N = 608) = .05, p > .05$. This does not support the prediction that having previously attended to an object might elicit further fixations at some point in the future.

In summary, the pattern of missing data attributable to two of the three sources (corrupt data trials and trials where the participant never fixated the target object after the verb) was randomly distributed across the six conditions. However, trials where the participant was already fixating the target object at verb-onset were not evenly distributed across the two verb conditions. This indicates that the results of the following analyses can mostly be interpreted meaningfully, although some caution is warranted.

Anticipatory Eye Movements

In order to determine whether any anticipatory eye movements may have been made, SOTs compared to two other time points in the sentences: the noun-onset and the noun-offset. In addition, we calculated the proportion of trials (trials spoilt due to corrupt data were not counted) in which a saccade was launched towards the target object before the onset of the noun. No such evidence of anticipatory eye movements was found. On average, eye movements were initiated 489 ms after the noun-onset, and 211 ms after the offset of the noun. In addition, saccades were launched towards the target object before the onset of the noun in only 7.5% of the causative trials and 9.3% of the perception trials.

Finally, the difference in time between the offset of the noun and SOT was computed and subjected to a 2 (verb type) X 3 (motion type) repeated-measures ANOVA ($N = 32$), in order to determine whether these first saccades were affected by verb type and motion type. Results indicated that only motion type had a significant main effect, $F(2, 60) = 21.63, p < .0001$, as predicted. A modified Bonferroni/Dunn test, with corrected alpha levels of .03 for each of the three pairwise comparisons, was conducted to explore the significant main effect of motion type. This indicated that the Towards condition ($M = 146$, $SD = 364$) had a lower mean difference than the Away ($M = 492$, $SD = 655$) and Neutral ($M = 718$, $SD = 655$) conditions, $p1 < .01$, $p2 < .01$, as predicted which also significantly differed from each other, $p < .01$, which was not predicted. The effect of verb type was not significant, $F(1, 60) = 2.21$, $p = .15$, contrary to our hypothesis, nor was the interaction, $F(2, 60) = 1.27$, $p = .29$. A paired one-tailed t-test between the Causative-Towards ($M = 161$, $SD = 493$) and Perception-Towards ($M = 180$, $SD = 595$) conditions did reveal a significant difference, $t(30) = -2.69$, $p < .01$, as hypothesized.

In order to determine whether participants were able to anticipate the object that the sentence referred by initiating a saccade towards it *before* the agent in the scene reached it (in the Towards condition, as the agent never interacted with the object in the Away and Neutral conditions), the difference between SOT and the time at which the agent touched the object was computed. It was found that participants launched a saccade towards the target object 1122 ms before the agent made contact with the object, indicating that participants were not simply following the motion of the agent towards the object but were in fact using some combination of the motion and linguistic contextual factors to anticipate the object of interest.

A similar computation was done for the Away trials, in order to determine whether participants initiated a saccade towards the target object before the agent left the scene. The difference between SOT and the point at which the agent exited the scene was computed. It was found that participants launched a saccade towards the target object 707 ms before the agent left the scene, indicating again that participants were not simply following the agent and decided to look at the target object only once he or she had left.

Target Object Saliency

The first analysis examined the correlation between target object saliency ratings, pre-verb fixation durations, post-verb fixation durations and total fixation durations. We hypothesized that the saliency ratings would correlate positively and significantly with the amount of time spent looking at the target objects. A Pearson's correlation ($N = 567$) indicated that target object saliency ratings did correlate significantly with the time spent looking at the target object before verb-onset ($r = .09, p = .04$), but correlated only marginally with the total amount of time ($r = .08, p = .06$), and not at all with the time spent looking after verb-onset ($r = .04, p = .32$). Because this correlation was not significant, target object saliency was not included as a covariate in the main analysis described below.

Second, the relationship between target object saliency and SOTs was computed. We hypothesized that there would be a significant positive correlation, which a Pearson's correlation ($N = 476$) did not confirm ($r = -.02, p = .58$). For that reason, target object saliency was not included as a covariate in the main analysis described below.

Third, the relationship between target object saliency ratings and whether or not the target object was being fixated at verb-onset was examined. We expected that there would be a significant positive correlation, such that the higher the saliency rating, the more likely the target object would be fixated at verb-onset. A one-tailed point biserial correlation ($N = 608$) was computed, which indicated that there was no significant correlation ($r = 0, p = .50$), contrary to our hypothesis.

Target Event Saliency

The same set of analyses described above was conducted with target event saliency instead of target object saliency. First, the relationship between target event saliency ratings and the three fixation durations was examined and was expected to yield significant positive correlations. This hypothesis was supported: a Pearson's correlation ($N = 567$) indicated that the amount of time spent looking at the target object before verb-onset correlated significantly with target event saliency ($r = .15, p < .001$), as did the time spent fixating after ($r = .10, p = .01$), as well as the total time spent fixating ($r = .16, p = .0001$).

Next, the relationship between target event saliency and SOTs was computed. We hypothesized that there would be a significant positive correlation, which a Pearson's correlation ($N = 476$) did not confirm ($r = -.08, p = .10$). Because of this, target event saliency was not included as a covariate in the main analysis described below.

Finally, the relationship between target event saliency ratings and whether or not the target object was being fixated at verb-onset was examined. We expected that the more predictive a scene was in terms of the target event, the more likely the target object (implicated in the target event) would be fixated at verb-onset. A one-tailed point biserial correlation ($N = 608$) was computed, which indicated that there was a significant correlation ($r = .08, p = .02$), such that the higher the target event saliency rating, the more likely that the target object would be fixated upon at verb-onset.

Main Analysis: First Post-Verb Saccades to Target Object

The first analysis examined the effect of verb type and motion type on post-verbal eye movement behaviour, namely saccade onset time (analysis by subjects). The data from only 31 participants was included, due to missing data from seven of them (from the sources described above). We hypothesized that both verb type and motion type would have a main effect on SOT. However, this was not confirmed, as shown in Figure 10 (see Appendix N for the ANOVA tables relevant to the analyses for this experiment). The results indicated that only motion type had a significant effect on SOT, *F*(2, 60) = 17.63, *p* < .0001. Verb type did not have a significant main effect, *F*(1, 60 = 2.17, *p* = .15, nor was the interaction significant, *F*(2, 60) < 1, *p* = .52. A modified Bonferroni/Dunn test, with adjusted alpha levels of .03 for each of the three pairwise comparisons, was conducted to explore the significant main effect of motion type. This indicated that all three groups significantly differed from each other, such that the Away group (*M* = 1211, *SD* = 458) had a lower mean than the Neutral group (*M* = 1432, *SD* = 666), *p* = .02 the Towards group (*M* = 907, *SD* = 297) had a lower mean than the Away group, *p* = .001, and the Towards group had a lower mean than the Neutral group, *p* < .0001.

In order to test the hypotheses outlined in the Analyses section, four planned comparisons were conducted. The first hypothesis was that the Away and Towards conditions would differ significantly, such that the Towards condition would have a lower mean SOT than the Away condition. A 2 X 2 (motion type X verb type) ANOVA was conducted to that effect (*N* = 33), and revealed that motion type did indeed have a significant main effect, *F*(1, 32) = 38.03, *p* < .0001), with the Towards condition (*M* = 917, *SD* = 293) having a lower mean SOT than the Away condition (*M* = 1234, *SD* = 467), as predicted. The second hypothesis was that the Neutral and Towards conditions would differ significantly, such that the Towards condition would have a lower mean SOT than the Neutral condition. Again, this hypothesis was confirmed with a 2 X 2

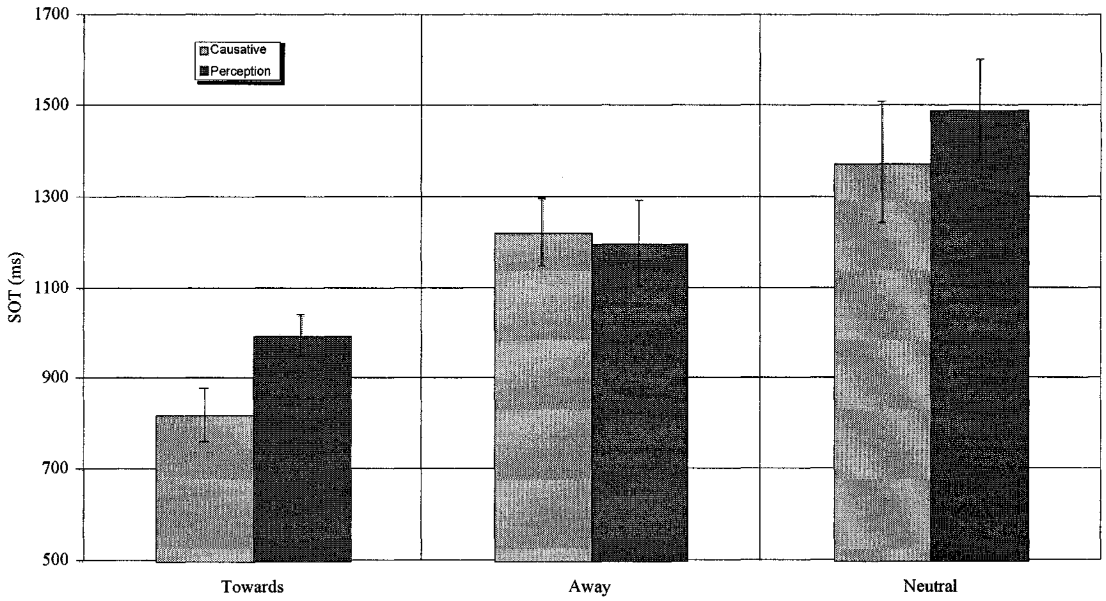


Figure 10. Mean SOTs (± *SE*) as a function of verb type and motion type, computed by subject.

repeated-measures ANOVA (*N* = 34), *F*(1, 33) = 32.60, *p* < .0001: the Towards condition (*M* = 945, *SD* = 341) had a lower mean SOT than the Neutral condition (*M* = 1461, *SD* = 652). Third, to compare the two verb types without the confounding effects of the motion context, the Causative-Neutral and Perception-Neutral groups were compared. In the absence of any apparent motion in the scenes, we expected that the Causative condition would lead to lower SOTs than in the Perception condition. A one-tailed paired t-test failed to lend support to this notion (*t*(34) = -.80, *p* = .21), although the trend was in the predicted direction, with the Causative-Neutral condition (*M* = 1386, *SD* = 705) having a lower mean than the Perception-Neutral condition (*M* = 1493, *SD* = 607). Finally, to compare the two verb types in the Towards condition, we compared the Causative-Towards and Perception-Towards conditions. We expected that the visual context (agent appearing to move towards the target object) would aid in the semantic interpretation of the verb, such that SOTs would be lower in the Causative-Towards than in the Perception-Towards condition. A one-tailed paired t-test supported this notion, *t*(36) = -2.71, *p* < .01, with the Causative-Towards condition (*M* = 860, *SD* = 343) having a lower mean than the Perception-Towards condition (*M* = 1033, *SD* = 258).

The second analysis was identical to that described above, except that it was conducted by items and not by subjects. The same hypotheses held for this analysis, which produced the same pattern of results as that computed by subjects. A 2 X 3 repeated-measures ANOVA revealed a significant effect for motion type, *F*(2, 32) = 13.62, *p* < .0001, but not for verb type, *F*(1, 32) = 1.16, *p* = .30. The interaction was not significant, *F*(2, 32) = 1.18, *p* = .32. A modified Bonferroni/Dunn test, with corrected alpha levels of .03 for each of the three pairwise comparisons, was conducted to explore the significant main effect of motion type. All three groups differed significantly from each other. The Towards condition had a significantly lower mean than the Away and Neutral conditions, while the Away condition also had a shorter mean SOT than the Neutral condition: Away vs. Towards, $p = .01$; Neutral vs. Towards, $p < .0001$, Away vs. Neutral, $p = .02$.

Planned comparisons for this data set also remained the same, as did their respective hypotheses: (1) Away < Towards; (2) Neutral < Towards; (3) Causative-Neutral < Perception-Neutral; and (4) Causative-Towards < Perception-Towards. The first hypothesis was supported by a 2 X 2 repeated-measures ANOVA; the main effect of motion type was significant, $F(1, 16) = 10.88, p < .01$), with the Towards condition having a shorter mean SOT than the Away condition, as expected. The second hypothesis was also confirmed; the main effect of motion type in a second 2 X 2 repeated-measures ANOVA was significant ($F(1, 16) = 23.42, p < .001$), such that the Towards condition also had a lower mean than the Neutral condition. The third hypothesis was not supported; a one-tailed paired t-test indicated that the Causative-Neutral condition did not have a significantly lower mean than the Perception-Neutral condition: $t(16) = -.04, p = .49$. Finally, the fourth hypothesis was supported; the Causative-Towards condition had a significantly lower mean than the Perception-Towards condition, $t(16) = -3.75, p < .001$.

As mentioned above, although SOT did correlate significantly with target event saliency, these analyses were not repeated as ANCOVAs with target event saliency as the covariate because the correlation was not sufficiently high to warrant this analysis.

Analysis of Early Post-Verb Cumulative Saccades to the Target Object

The effects of verb type, motion type and sentence point on the cumulative proportion of saccades to the target object after verb-onset was analyzed using a 2 X 3 X 3 ANOVAs (by subjects). We expected that there would be a significant interaction between sentence point and verb type, such that the difference between the two verb types would increase as the sentence unfolded, as well as a significant interaction between sentence point and motion type, such that the difference between the Towards and Away/Neutral conditions would increase as the sentence unfolded. The results (which are plotted in Figures 11 and 12) partially confirmed our hypotheses, such that the interaction between sentence point and motion type was significant, $F(4, 148) = 6.00, p < .001$, while the interaction between sentence point and verb type was not, $F(2, 148) < 1, p = .90$. However, the three-way interaction of sentence point, verb type motion type was also significant ($F(4, 148) = 3.06, p = .02$), which was not expected.

To further explore this three-way interaction, a test of simple interaction effects was conducted, which indicated that the interaction effect of sentence point and motion type differed at the two levels of verb type. Specifically, the interaction effect of sentence point and motion type was significant at the Causative level of verb type, $F(4, 148) = 9.12, p < .0001$, while it was not at the Perception level of verb type, $F(4, 148) < 1, p = .61$. Furthermore, three tests of the simple effects of motion type at Causative across all levels of sentence point were conducted. These indicated that the difference between Away and Neutral was not significant, $F(1, 74) < 1, p = .81$, although the difference between Away ($M = .070, SD = .142$) and Towards ($M = .168, SD = .256$) was, $F(1, 74) = 13.29, p < .001$, as was the difference between Neutral ($M = .076, SD = .152$), $F(1, 74)$

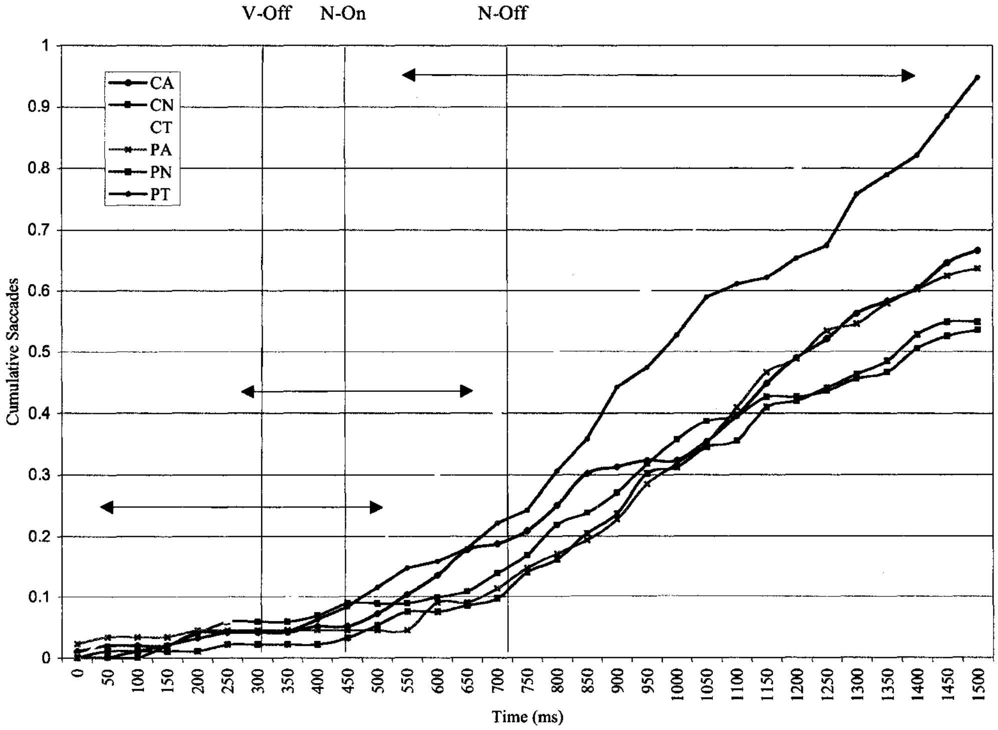


Figure 11. A plot of the mean cumulative average number of fixations (by subject) to the target object after verb-onset. Each line refers to a single condition, and each point to one 50-ms bin. The origin of the X-axis refers to the verb-onset, and the three vertical lines mark the temporal boundaries of the verb and noun (average onset and offset). The double-headed horizontal arrows on each boundary indicate the range of onsets and offsets at the points in time relative to the verb-onset. This was done because in order to take into account the variable lengths of each of the sentence segments (i.e. verbs and noun phrases). The point at which each of the coloured lines (referring to cumulative fixations for each condition) intersects with the three critical sentence points (verb-offset, noun-onset and noun-offset) were computed and compared in an ANOVA.

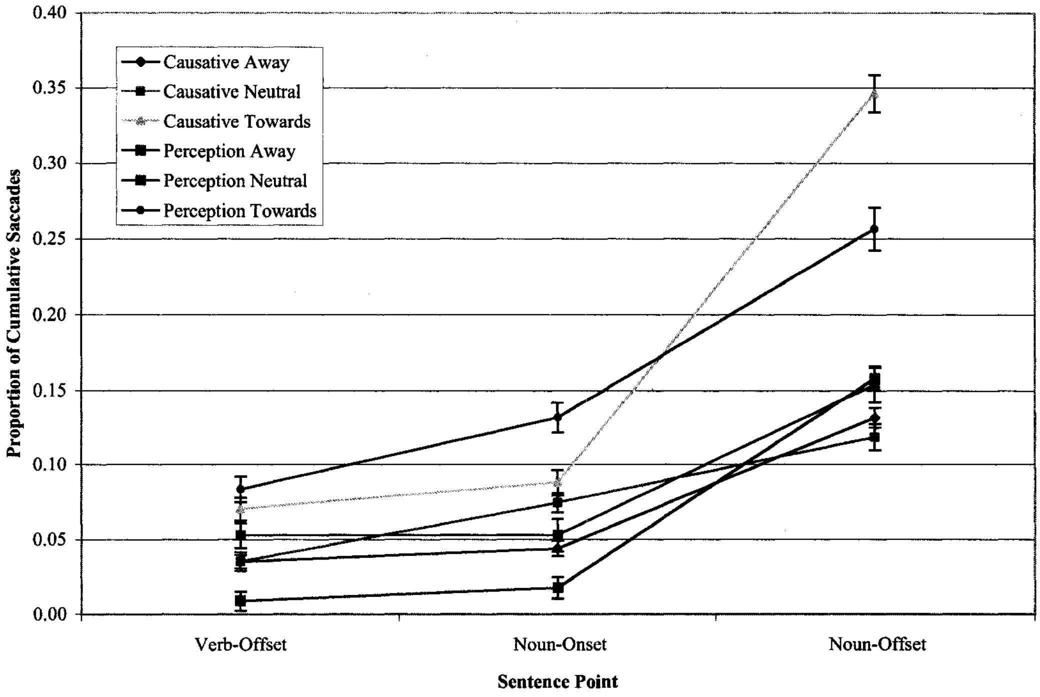


Figure 12. Mean number of cumulative saccades towards target object (± *SE*) at each of the three critical sentence points, verb-offset, noun-onset and noun-offset.

= 6.82, p = .01, and Towards, such that the Towards group had a higher mean than the other two groups.

To explore the two significant simple effects of motion type at Causative, a series of simple comparisons were conducted. First, Away and Towards at V-Off and at Causative were compared in a one-tailed paired t-test; this difference was not significant, p = .12. Second, Away and Towards at N-On and at Causative were compared; this difference was marginally significant, p = .07. Third, Away and Towards at N-Off were compared; this difference was significant, p < .0001. Fourth, Neutral and Towards at V-Off and at Causative were compared; this difference was not significant, p = .14. Fifth, Neutral and Towards at N-On and at Causative were compared; this difference was also not significant, p = .36. Sixth, Neutral and Towards at N-Off were compared; this difference was significant, p < .0001.

Planned comparisons tested the hypotheses that the Away group would exhibit significantly higher means than the Neutral and Towards groups across all sentence points. These hypotheses were not supported: there was no significant difference between the Away group (M = .078, SD = .194) and the Towards group (M = .163, SD = .087), F(1, 74) < 1, p = .92, although the trend was in the right direction. In addition, the difference between the Neutral (M = .069, SD = .165) and Towards groups was not significant, F(1, 74) < 1, p = .58, although again, the trend was in the right direction. Next, the hypothesis that the Causative-Neutral group would have a higher mean than the Perception-Neutral group across all sentence points was not supported. Results indicated that the Causative-Neutral condition (M = .076, SD = .152) was not significantly different from the Causative-Perception condition (M = .061, SD = .178, F(1, 74) < 1, p = .62, although the trend was in the predicted direction. Finally, the hypothesis that the Causative-Towards condition (*M* = .168, *SD* = .256) would have a lower mean than the Perception-Towards condition (*M* = .157, *SD* = .266) across all sentence points was not supported, *F*(1, 227) < 1, *p* = .81.

Discussion

The results of this experiment supported some, but not all, of our hypotheses. Firstly, with regards to anticipatory eye movements, the saccade onset times reported here are not consistent with the notion that participants initiated saccades towards the target object before the noun-onset or the noun-offset. In fact, on average, saccades were launched 211 ms after the end of the noun's utterance. However, this figure is lower than that obtained in Experiment 1, where the first saccades were launched 486 ms after the offset of the noun. Although the reasons for this are not clear, it may be linked to the finding that less than 10% of the trials contained a pre-verb look to the target object. Although participants might not have directly fixated the object, they may have encoded its location and therefore when its relevance became more apparent later in the trial saccades may have been launched more quickly to further examine it. In addition, these figures differed by motion context, such that all three conditions differed from each other, where saccades were launched soonest in the Towards condition (145 ms after the offset of the noun), followed by the Away (492 ms after) condition, then the Neutral (717 ms after) condition. This did not differ by verb type, however, with the saccades being launched 405 ms after the offset of the noun in the Causative condition and 499 ms after in the Perception condition.

As mentioned above, saccades were launched towards the target object before the onset of the noun in only 7.5% of the causative trials and 9.3% of the perception trials, compared to 6.3 % and 9.7% in Experiment 1; approximately equivalent to those in Experiment 1. In addition, these figures are much lower than those obtained in the second Altmann and Kamide (1999) experiment: 32% of the semantically restrictive trials and 18% of the semantically non-restrictive trials. This is likely because the agents in the movies attracted more fixations, due to the fact that they were in a state of motion, and participants spent less time examining other regions of the scenes. In other words, the less complex the scene (as in Experiment 1 or the Altmann and Kamide, 1999, study), the more likely a given object is to be fixated early in scene viewing.

Another type of anticipatory eye movement that was examined was the difference in time between SOT and the time the agent made contact with the target object in the Towards condition. Here, participants launched a saccade just over a second before this contact was achieved, indicating that participants were not simply following the path of the agent but rather made an anticipatory eye movement towards the object. Similarly, in the Away condition, it was found that on average, participants launched saccades 707 ms before the agents left the scene, also indicating that participants disengaged their attention from the agents and shifted it towards the objects prior to their disappearance.

There was some support for a relationship between target object saliency and the amount of time spent fixating the target object. The correlation was significant for time spent fixating before verb-onset and marginally significant for the total time, but not for time spent fixating after verb-onset. The fact that fixation times before verb-onset increased as the saliency of the object increased, whereas fixation times after verb-onset did not, as in Experiment 1, points to the possibility that saliency has more of an effect early in scene viewing. These objects may attract longer viewing times at first, but once they have been sufficiently inspected, they may no longer do so. However, given that the correlations were so low (approximately .09 or less), target object saliency may not have been a particularly important factor in determining fixation lengths.

The correlation between target object saliency and SOT was not significant, contrary to our hypothesis, and as found in Experiment 1. Again, this indicates that target object saliency did not play a role in determining the speed with which participants initiated saccades towards these objects after hearing the utterance of the verb. Target object saliency may not have been an important factor in determining eye movement patterns after the utterance of the verb, and other scene factors, such as their complexity, may have lead to this lack of a relationship.

Finally, target object saliency did not affect whether the target object was being fixated at verb-onset, as in Experiment 1. Although this is contrary to our hypothesis (a highly salient object should have a higher probability of being fixated at any given time), it is consistent with the idea that these objects may have been sufficiently attended to prior to that point.

With regards to target event saliency, a somewhat different pattern of findings was obtained, relative to target object saliency. Target event saliency correlated with all three measures of fixation durations, which contrasts with the results of Experiment 1, where target event saliency correlated with only time spent fixating the target object after verb-onset. This could indicate that the target event was anticipated by participants early in the trial, especially as the event's predictability increased, but once the event and its participants had been named, fixations towards the object involved may have also increased with increasing event saliency. In other words, the dynamic component of the films may have primed participants to anticipate that an event would eventually take place and activate possible event structures more readily than in the still pictures, where the component of motion was missing.

Contrary to our hypothesis, target event saliency did not significantly and positively correlate with SOT, as in Experiment 1. This indicates that the predictiveness of the scene's target event did not have an effect on the speed with which participants initiated a saccade towards the target object. Why this was so is not clear; scene factors, such as their layout and complexity, may have contributed to this finding, but it may also be that the dynamic element of the films was an additional factor.

Finally, target event saliency did somewhat affect whether or not participants were fixating the target object at verb-onset (the correlation was marginally significant, while in Experiment 1 it was significant at $p < .05$), as predicted. Again, this indicates that the more predictive a scene was of the event described by the sentence (and, by extension, its likely participants or role fillers), the more likely a participant would be to fixate the target object at the point at which the verb was uttered. Interestingly, this may indicate that participants were anticipating which object would be involved in the sentence described, and were fixating it just prior to the utterance of the verb so as to confirm whether or not it would be involved as an event participant. This corresponds to our hypothesis that the information conveyed by the sentence is used to confirm the gist of possible events extracted from the scene.

The main analyses (as well as the analysis of early post-verb eye movement behaviour) constituted the principal means of investigating the major hypothesis of this research, namely whether verb-specific information constrains referential domains. In addition, they investigated the previously unexplored effect of actual agent motion on eye movement behaviour. Overall, the results indicate that verb type mostly did not have an effect on how quickly saccades were launched towards the target object after verb-onset, This lack of an effect was further corroborated by the finding that there was no difference between the Causative-Neutral and Perception-Neutral groups, where it was expected that the neutrality of the agent's positioning could not mediate the effect of verb type. Even with this neutrality, however, the more semantically restrictive causative verbs did not cause participants to launch saccades towards the target objects more quickly. These results were similar to those reported in Experiment 1. On the other hand, a significant difference was found between the Causative-Towards and Perception-Towards groups, indicating that when the agent moved towards the target object, this information served to interact with the selectional restrictions of the verb, yielding faster saccades towards the target object. However, this was not the case at the early moments of sentence interpretation (between the onset of the verb and the offset of the noun).

Again, these findings fail to support the notion that verb-specific information, in the form of thematic roles, can guide eye movements, as has been supported by the bulk of the findings reported in the literature (e.g., Tanenhaus et al., 2000; Altmann & Kamide, 1999; Kamide et al., 2003). As discussed in Experiment 1, one reason may be that not enough participants were tested; the power for verb type was less than .3 (as opposed to motion type, which was approximately 1.0) for both analyses. Another possibility is that the low-level properties of the scenes themselves may have prevented any linguistic influences from affecting eye movements, such as the fact that the scenes were dynamic depictions of true scenes, not *ersatz* still scenes. It may be that the linguistic variables only have an effect on the visual system (as measured by eye movements) when the visual input is relatively impoverished and non-complex, both in terms of its content and layout, and the inclusion of dynamic factors. Here, the scenes were fairly complex, and included actual depictions of people moving about.

Although there are no eye-tracking studies within the visual world paradigm that have utilized motion pictures of events, making it difficult to interpret these results, what may in fact be occurring is that when the visual context is complex and features moving stimuli, the resources of the visual attentional system may be fully allocated to processing these stimuli, preventing the integration of other contextual information (such as linguistic representations) within the attentional system. Thus, eye movements are directed solely in response to these visual stimuli, such that the effects of visual context processed within the attentional system responsible for controlling eye movements are dissociated from the effects of subtle linguistic differences.

On the other hand, motion type did have a significant effect on the speed with which saccades were initiated towards the target object after the verb was uttered. In addition, the scenes in which the agent moved towards the target object led participants to look towards the target object more quickly than when the agent moved away, or when he or she remained in a neutral position with respect to the object. Furthermore, there was no difference in saccade onset time between the Away and Neutral conditions. These findings confirmed our hypotheses, and suggest that the direction of motion of a human agent within a scene can serve to orient eye movements when that person moves towards scene regions of interest, at least when that movement is consistent with the semantics of the sentence being uttered. However, when the person's path of motion is not consistent with the sentence's meaning (i.e., the agent's direction does not match the upcoming event), it may be that a the central processes involved in integrating the linguistic input and visual context require extra time to process the inconsistency between the two inputs.

Because other studies within the visual world paradigm have shown that eye movements are locked to the incremental process of sentence comprehension at the earliest moments of interpretation, a separate analysis was conducted that examined the cumulative number of saccades that had been initiated towards the target object at three critical sentence points: verb-offset, noun-onset and noun-offset. Although the hypothesis that sentence point would interact with verb type (such that the cumulative number of saccades towards the target object would increase more with time in the Causative condition than the Perception condition) was not supported, a main effect for sentence point was detected. This indicates, as would be expected, that the number of saccades initiated towards the target object increased as the sentence unfolded. Contrary to the results obtained in Experiment 1, no main effect of verb type was found. This indicates, for the small proportion of saccades that had been launched by the end of the noun's utterance (a mean of approximately 0.20 saccades across all trials), eye movements were *not* constrained by the verb's semantic restrictions. Corroborating evidence was obtained in the planned comparison, which indicated that the Causative-Neutral and Perception-Neutral groups did not differ significantly; in other words, even in the absence of any direct path of motion of the agent, eye movements were not differentially affected by the linguistic context. This is consistent with the findings of the main analyses reported above, that eye movements are not guided by subtle linguistic differences within a dynamic visual context, and is inconsistent with findings within the literature that eye movement behaviour can be guided at the earliest moments of sentence comprehension. These findings do not support the notion that thematic information contained within the verb can affect eye movements when the visual context is dynamic.

However, the three-way interaction between sentence point, verb type and motion type was significant, such that the interaction effect of sentence point and motion type was different at the two levels of verb type, which was not expected. Specifically, the interaction was significant at the Causative level while it was not at the Perception level such that the mean cumulative number of saccades across all sentence points at the Causative level differed between the Towards condition and the Neutral and Away conditions. In addition, the Towards and Away groups and the Towards and Neutral groups only differed at N-Off. In other words, the Towards condition led to a higher number of cumulative saccades than the Away and Neutral conditions only at the offset of the noun and only in the Causative condition. Consistent with this finding is that planned comparisons did not reveal any differences between the motion conditions, as well as between the Causative-Towards and Perception-Towards conditions.

These results can be interpreted to mean that at the earliest stages of post-verb sentence comprehension and at the beginning of the agent's motion (the disambiguating point), motion context does not affect whether participants launch a saccade to the object referent of the verb's direct object when that verb is a causative verb, but it does when it is a perception verb. This provides some support for the very early interaction of visual and linguistic contextual information. However, overall, when counting all trials, motion context had a stronger influence over the direction of eye movements than the linguistic context.

In summary, the results of Experiments 2 show that there was weak support for the hypothesis that verb-specific information can guide eye movements within a dynamic visual context, even at the very earliest moments of verb phrase interpretation. Much stronger support was shown for the effect of the apparent motion of the agent on eye movements. It appears that when realistic, moving depictions of everyday events are employed, the visual context takes precedence over the linguistic context in constraining the domain of subsequent reference.

General Discussion

The purpose of the work presented here was to examine how linguistic and visual processing interact, and to try to determine the locus of this interaction – whether it is at the central cognitive system, within the visual and linguistic modules, or some point in between. By examining the pattern and timing of eye movements across static and dynamic scenes, we attempted to better understand how two of the most important cognitive faculties, language and vision, interact with each other to govern our experience of the world. The purpose of the Normative Study was to obtain data regarding the saliency of the target object, as well as the semantics of the scene in the form of present and future events. The purpose of Experiment 1 was to replicate previous studies that used still pictures with spoken sentences, but with methodological improvements in the form of naturalistic, true scene depictions. The purpose of Experiment 2 was to replicate Experiment 1 except with the added dimension of dynamic motion in order to determine how motion affected the pattern of eye movements.

Summary of Results

The main purpose of the Normative Study was to gather basic information regarding the relative saliency of the target objects and events, which we hypothesized would have an effect on dependent variables used in the two experiments (e.g., how quickly saccades were initiated towards target objects after the verb). In addition, we were interested in knowing whether these frequencies differed by the direction of apparent motion. The results indicated that the saliency rating, as a whole, did not correlate strongly with post-verbal SOTs, and for that reason, target object saliency was not used as a covariate in the main analyses of the two experiments.

In addition, it was found that the direction of apparent motion of the agents in the scenes did affect how frequently the target object was listed by participants. In particular, results indicated that the target object was listed more frequently in the Towards condition than in the Away and Neutral conditions, which did not differ significantly. The direction of apparent motion did not affect how frequently the two target events were listed across these conditions, both those regarding future events and those regarding current events, or the gist of the scene.

The purpose of the two eye-tracking experiments was to replicate and extend previous findings within the visual world paradigm that have shown that eye movements are driven by linguistic information using more naturalistic visual contexts. In addition, the purpose of Experiment 2 was to determine whether these findings would continue to hold when the visual context was dynamic. Finally, another purpose was to examine the effect of different motion contexts. For the two experiments, it was found that motion context had a significant main effect on how quickly eye movements were launched towards the target object after the verb was uttered, with the Towards condition leading to shorter SOTs than the other two motion conditions. This mirrors the results found in the Normative Study, where the target object was listed more frequently when the agent appeared to be facing or moving towards it. Overall, then, it appears that when the information contained within a verb (the event it describes and its participant entities) is consistent with the visual context, saccades are initiated more quickly than when it is not.

The effect of verb type was not found to be consistently significant: in the still picture experiment, where the analysis was conducted by items, it was significant, while in the movie experiment, it was not significant in both the analyses by subjects and by items. Furthermore, at the earliest moments of post-verb sentence interpretation, for the small proportion of saccades that had been launched by the end of the noun's utterance, eye movements were constrained by the verb's semantic restrictions in Experiment 1. However, this did not hold true for Experiment 2.

Implications of the Results

With regards to the Normative Study, it was found that the direction of apparent motion of the agents in the scenes did affect how frequently the target object was listed by participants, such that the target object was listed more frequently in the Towards condition than in the Away and Neutral conditions. This finding is consistent with other work that has found that a person's gaze direction affects where others look (Henderson & Ferreira, 2004). The direction of apparent motion did not affect how frequently the two target events were listed across these conditions, both those regarding future events and those regarding current events, or the gist of the scene. However, the frequencies were higher for the version "b" propositions (which referred to events occurring presently) than for the version "a" propositions (which referred to likely future events). These higher means imply that the gist of the scene was extracted more easily than the predictiveness of the scene.

For the two experiments, the only consistent finding was that motion context had a significant main effect on how quickly eye movements were launched towards the target object after the verb was uttered, with the Towards condition leading to shorter SOTs than the other two motion conditions. This mirrors the results found in the Normative Study, where the target object was listed more frequently when the agent appeared to be facing or moving towards it. Overall, then, it appears that when the information contained within a verb (the event it describes and its participant entities) is consistent with the visual context, saccades are initiated more quickly than when it is not.

With regards to verb type, the results indicated that it was only significant for the analysis by items in Experiment 1 but not at all significant for Experiment 2. In addition, for the small proportion of saccades initiated prior to the offset of the noun, verb type did have a significant effect in Experiment 1 while it did not in Experiment 2. This fails to support the notion that verb-specific information, in the form of thematic roles, can guide eye movements. However, verb class did affect eye movements at the earliest moments of sentence comprehension (immediately following the verb and before the offset of the noun), but only when the visual context was static. This is somewhat in contrast to the bulk of the findings reported in the literature (e.g., Tanenhaus et al., 2000; Altmann & Kamide, 1999; Kamide et al., 2003), where all first saccades to the target object were affected by the verb restrictions, not simply the small proportion of pre-noun-offset saccades. This may be due to the nature of the scenes presented to participants; as noted in the Introduction, object arrays (as have been employed in other studies) have no discernable gist, while the scenes used in the present studies had a clear meaning (as evidenced by the Normative Study). Therefore, participants may have predicted upcoming events early in scene viewing, and may have relied on the features of the visual context (namely, motion context) instead to determine subsequent regions of interest. In the absence of such a discernable gist, the visual system can only use the linguistic input to guide eye movements, rather than semantic information that is gleaned from a true scene. In addition, the scenes were more complex than those used in other studies, especially in the case of Experiment 2, where the context was dynamic. This indicates that when a naturalistic scene is used (especially a dynamic one) the visual context takes control of eye movements more so than the linguistic context, and although some interaction likely occurs (as evidenced by the significant main effect of verb type found under certain circumstances), the visual factors take precedence over the linguistic ones.

How does this fit with the overarching question of this thesis, which pertains to the modularity (or independence) of the language and vision systems? Are the results observed consistent with Fodor's modular properties? With regards to mandatoriness, it is clear that the interaction, as evidenced by the fact that the target object was not always fixated after the verb was uttered, is not mandatory. However, it is difficult to directly detect whether the representations activated by the linguistic and visual input do or do not always become integrated using the present methodology. It appears that under certain circumstances they do, as participants did fixate the target object some of the time after hearing the verb, although if they automatically do, this may not always be demonstrated in the form of eye movement patterns. As such, it appears that this integration is not necessarily mandatory and likely occurs at the central level.

With regards to speed, it appears, again, with the present methodology, that the integration of concepts activated by the visual context and verb-specific information does not occur particularly quickly. On the whole, saccades towards the target object were initiated approximately 1200 ms after the utterance of the verb. This suggests that whatever integration that occurs between linguistic and visual representations *that is evidenced by eye movements* does not occur very rapidly and as such, indicates that the interaction likely occurs at the central level, as would be expected by the modularity hypothesis.

As stated in the introduction, if a process is not fast and not mandatory, it is probably a central process. Therefore, we can tentatively conclude that the interaction of linguistic and visual information occurs at a central level and that the two modules dedicated to these processes are independent of each other. In addition, we propose that the conceptual information activated by the visual environment, as well as by the linguistic input (specifically in the form of verb-specific thematic information), occurs within conceptual short-term memory. This notion is consistent with Potter's (1993, 1999) CSTM theory.

Potential Confounds and Future Work

With regards to the Normative Study, it may be that the scene layouts (including the presence or absence of various objects, low saliency of the target objects, the activities that the agent appeared to be involved in, and the actual positioning of the objects and agents themselves) may not have successfully set up the conditions necessary to make the scenes predictive of the target events for version "a." For example, in the "milk" scene, although the cup is visible, it is so minute as to be unrecognizable, and the boy appears to be playing with something rather than about to have a glass of milk. Similarly, the semantics of the scenes may not have correlated highly with the verbs used in the sentences, in the sense that although they were likely to occur given the scene layout, they appear not to be the *most* likely. For example, in the "ball" scene, the target event was DROP [MAN, BALL] ("the man will drop the ball"). However, the most commonly generated propositional structure involving the ball was KICK [MAN, BALL] – most participants expected that he would kick the ball in the near future, not unreasonable that it was on the ground a few steps away from him, in a soccer field (as opposed to being in his hands, making it more likely to be dropped). Future studies might address these issues by first constructing the scenes to be as predictive as possible of the events (and specifically the verbs that describe them) of interest, and second, by conducting pilot normative studies prior to pairing them with the sentences in order to determine whether the scenes are actually conveying the message they are intended to.

Similar arguments can be made for version "b" of the study, in that the scene layouts may not have corresponded well to the target events described by the first patch clauses of the sentences in Experiments 1 and 2. For example, in the "picture" scene, the target event structure is UNPACKING [WOMAN, OFFICE], but firstly, this is an infrequent event to take place in an office (more likely events include working, reading, writing, working on the computer, etc.), and secondly, the agent does not appear to be unpacking an office at all, but rather simply entering/leaving the office, or looking at a book (depending on the motion condition). In addition, there were other layout problems associated with some specific scenes. These are presented in Table 5.

In addition, target object and event saliency did not have much effect on eye movement behaviour. Aside from the properties of the scenes used, it may be that our measure of saliency, in the form of listing frequencies, was not an accurate or sensitive measure of saliency. Alternatively, because the scenes used in the present studies were highly realistic, we would expect that eye movement patterns may have been guided by expectations regarding typical objects and their locations – what Henderson and Ferreira (2004) term scene schema knowledge. Thus, typicality rather than saliency may have had a more powerful effect on eye movement behaviour. However, the lack of any research that investigates the effect of both target object and event saliency on eye movement

Table 5

Layout Problems Associated with Some Specific Scenes

Scene	Problem
Cube	Cubes are not rollable due to their non-spherical shape, making it highly unlikely that this event would be generated.
Plate	The face of the agent is obscured, making it difficult to see where his gaze may be directed.
Shoes neutral	Agent is facing away from shoes, and is not in an obviously "neutral" position with respect to it, although he is not obviously walking away as in the away condition.
Vase	All three scenes look similar, in that the agent is facing away from the vase, because in the original film she backs into it (thereby knocking it over).

behaviour does not provide a meaningful context within which to discuss these results. Nevertheless, they may point to more fruitful avenues of study in the future, as this aspect of the visual world paradigm has never been examined before. In order to understand the factors that influence eye movements, future studies should consider these variables, and seek to define and measure them carefully.

With regards to the two experiments, there were a number of limitations and potential confounds. First, the amount of time in which participants were allowed to view scenes before linguistic input varied largely across items in the movie experiment. The problem is that the more time participants have to view the scenes, the more information they extract not only regarding scene gist but object identification and location as well. This means that participants may have looked quicker towards the target object once having heard the verb, especially with longer exposure times. This problem was not present in Experiment 1, where the exposure time was the same in all items. Future studies of this nature should seek to equalize exposure time across all trials.

One possible confound is that participants may have developed a strategy for examining the scenes. Because no distractor trials were used, participants may have detected that not only were there three different motion conditions, but also that the object named in the sentence always appeared opposite to the agent. However, even if participants did develop a strategy, there is no reason why the two verb conditions did not demonstrate a difference, as this should have aided participants in executing their strategy.

Another confound may be that participants may have followed the motion of the agent within the movies most of the time because their motion was inconsistent -

sometimes towards the target object, sometimes away, and sometimes neither. This inconsistency of motion may have disproportionally increased the saliency of the agents, and the attentional system may have allocated more resources to them, therefore increasing the number and length of fixations they received. However, participants did disengage from fixating the agents before they left the scenes in the Away condition, and visual conditions were identical for the two verb conditions, meaning that there still should have been some difference between the two based on their semantic differences.

Similarly, because of the complex nature of the scenes (large number of objects and textures, as well as the motion of the agent in Experiment 2), attention may have remained fixed on the agents because of an inability to meaningfully choose appropriate saccadic targets in the absence of very specific linguistic information (the utterance of the noun itself). Eye movement behaviour centering on the agents was not systematically examined in these experiments; therefore, future studies should attempt to quantify the number and durations of fixations directed towards the human agent.

Yet another possible confound is that attention towards the target object may have shifted in the appropriate direction after the verb's onset without the concomitant shift in eye movements. As Fischer and Breitmeyer (1987) have shown, attention can be shifted without a corresponding shift in fixation location. In addition, because attention is more distributed, such that more information can be gleaned in parallel than serially, when approximately only four fixations are made per second, much information can be extracted in parallel without foveation. Thus, effects of the verb's selectional restrictions may have in fact have occurred, but were not detected with the present methodology. In fact, this covert shifting of attention may have delayed SOTs in that the saccades were not initiated until the noun was uttered. In the case of the movie experiment, the direction of the agent's motion confirmed the event anticipated by participants.

Future work within the visual world paradigm can take several paths. First, related to the last point, we plan to replicate these studies with the scenes projected onto a large screen so as to mimic real-life scene perception, where the entire visual field is filled with the scene of interest. By increasing the visual angle, the eye will be forced to move about the scene to fully inspect it rather than to shift attention without moving the eye as can occur when the visual angle is small. Because some proportion of fixations will be to the target object before the verb, if the object concept has already been activated, upon hearing the verb saccades will be initiated back to the object more quickly.

Another potentially fruitful experimental manipulation would be to freeze the movies at the disambiguating point to determine whether participants look at the target object, because at that point they can no longer follow the motion of the agent. This may eliminate any confounding factors actual motion has after the point that the verb's conceptual information becomes activated.

In order to understand whether implicit arguments are part of a verb's representation, we could use sentences containing instrument verbs (denominal verbs such as *hammer* and *paint*) with implicit arguments but with omitted objects or instruments. Future studies could employ intermediately constraining verbs, such as contact verbs (e.g., *move*, *place*), or other verb types such as possession verbs (*own*, *have*).

Another area of investigation could involve manipulating sentence constructions. Because of the difference in the thematic information conveyed by the two different verb classes, where causative verbs focus more on the event's patient while perception verbs focus more attention on the experiencer, they may differentially affect eye movement patterns. However, these differences constitute the very reason they were employed in these experiments; because causative and perception/psychological verbs belong to coherent semantic-syntactic classes, they are ideally suited to this methodology. To disentangle this effect of the focus being on the experiencer *vs.* the patient, future studies could employ passive sentences, where the subject of the verb is uttered at the end of the sentence.

In summary, this set of studies indicates that when realistic scenes are employed within the visual world paradigm, the linguistic information contained within the verb does not serve to constrain domains of visual reference. However, they are an improvement over the verbs used in the Almann & Kamide (1999) and Kamide et al. (2003) studies, where verbs were divided into semantically restrictive and nonrestrictive classes without any clear semantic/syntactic basis, in the form of Levin's (1993) verb classes. When verb-semantic information and visual scenes are manipulated obeying stronger linguistic and visual constraints, the effects found in the literature do not hold. This is contradictory to the work conducted by other researchers within this domain (e.g., Tanenhaus, Altmann, Kamide and their colleagues), but is consistent with the notion that the language and visual systems process information independently (i.e., they are modular) and is not inconsistent with the idea that conceptual information activated by the language and visual systems can and do interact at the conceptual level (CSTM, likely a central process).

References

Allopenna, P.D., Magnuson, J.S., & Tanenhaus, M.K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language, 38,* 419-439.

Altmann, G., & Steedman, M. (1988). Interaction with context during human sentence processing. *Cognition, 30,* 191-238.

Altmann, G.T.M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition, 73,* 247-264.

Altmann, G.T.M., & Kamide, Y. (2004). Now you see it, now you don't: Mediating the mapping between language and visual world. In J. Henderson & F. Ferreira (Eds.), *The interface of language, vision, and action: Eye movements and the visual world.* New York: Psychology Press, 347-386.

Buswell, G. T. (1935). *How People Look at Pictures.* Chicago: University of Chicago Press.

Carlson, G.N., & Tanenhaus, M.K. (1988). Thematic roles and language comprehension. In W. Wilkins, (Ed.), *Syntax and Semantics, Vol. 21: Thematic Relations,* pp. 263-89. San Diego, CA: Academic Press.

Chambers, C.G., Tanenhaus, M.K., Eberhard, K.M., Filip, H., & Carlson, G.N. (2002). Circumscribing referential domains during real-time language comprehension. *Journal of Memory and Language, 47,* 30-49.

Clifton, C., Jr., Frazier, L., & Connine, C. (1984). Lexical expectations in sentence comprehension. *Journal of verbal learning and verbal behavior, 23,* 696-708.

Cooper, R.M. (1974). The control of eye fixation by the meaning of spoken language. *Cognitive Psychology, 6,* 84-107.

Dowty, D.R. (1991). Thematic proto-roles and argument selection. *Language, 67,* 547-619.

Eberhard, K.M., Spivey-Knowlton, M.J., Sedivy, J.C., & Tanenhaus, M.K. (1995). Eye movements as a window into real-time spoken language comprehension in natural contexts. *Journal of Psycholinguistic Research, 24,* 409-436.

Fischer, B., & Breitmeyer, B. (1987). Mechanisms of visual attention revealed by saccadic eye movements. *Neuropsychologia, 25,* 73-83.

Fodor, Jerry A. (1975). *The Language of Thought.* Cambridge, MA: Harvard University Press.

Fodor, J.A. (1983). *The Modularity of Mind: An Essay on Faculty Psychology.* Cambridge, MA: The MIT Press.

Henderson, J.M., Weeks, P.A. Jr. & Hollingworth, A. (1999). Effects of semantic consistency on eye movements during scene viewing. *Journal of Experimental Psychology: Human Perception and Performance, 25,* 210-228.

Intraub, H. (1999). Understanding and remembering briefly glimpsed pictures: Implications for visual scanning and memory. In V. Coltheart (Ed.), *Fleeting Memories: Cognition of Brief Visual Stimuli*, pp. 47-70. Cambridge, MA: MIT Press.

Jackendoff, R.S. (1983). *Semantics and cognition.* Cambridge, MA: MIT Press.

Jackendoff, R.S. (1987a). Language processing. In *Consciousness and the Computational Mind,* pp. 91-120. Cambridge, MA: MIT Press.

Jackendoff, R.S. (1987b). On beyond zebra: The relation of linguistic and visual information. *Cognition, 26,* 89-114.

Jackendoff, R.S. (1993). The combinatorial structure of thought: The family of causative concepts. In E. Reuland & W. Abraham (Eds.), *Knowledge and Language, Volume II, Lexical and Conceptual Structure,* pp. 31-49. Dordrecht, Netherlands: Kluwer Academic Publishers.

Kamide, Y., Altmann, G.T.M., & Haywood, S.L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language, 49,* 133-156.

Kintsch, W. *The representation of meaning in memory.* Hillsdale, NJ: Erlbaum, 1974.

Kowler, E. (1999). What movements of the eye tell us about the mind. In E. Lepore and Z. Pylyshyn (Eds.), *What is cognitive science?*, pp. 248–262. Malden, MA: Blackwell Publishers.

Levin, B. (1993). Introduction: The theoretical perspective. In *English Verb Classes and Alternations: A Preliminary Investigation,* pp. 1-23. Chicago: University of Chicago Press.

McRae, K., Ferretti, T.R., & Amyote, L. (1997). Thematic roles as verb-specific concepts. *Language and Cognitive Processes, 12,* 137-176.

Marslen-Wilson, W. & Tyler, L.K. (1989). Against modularity. In J.L. Garfield (Ed.), *Modularity in Knowledge Representation and Natural-Language Understanding,* pp. 37-62. Cambridge, MA: MIT Press.

O'Conner, K.J. & Potter, M.C. (2002). Constrained formation of object representations. *Psychological Science, 13,* 106-111.

Potter, M.C. (1975). Meaning in visual search. *Science, 187,* 965-966.

Potter, M.C. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology: Human Learning and Memory, 2,* 50-522.

Potter, M.C. (1993). Very short-term conceptual memory. *Memory & Cognition, 21,* 156-161.

Potter, M.C. (1999). Understanding sentences and scenes: The role of conceptual short-term memory. In V. Coltheart (Ed.), *Fleeting Memories: Cognition of Brief Visual Stimuli*, pp. 11-46. Cambridge, MA: MIT Press.

Potter, M.C., & Levy, E.I. (1969). Recognition memory for a rapid sequence of pictures. *Journal of Experimental Psychology, 81,* 10-15.

Rappaport Hovav, M. & Levin, B. (1998). Building verb meanings. In M. Butt & W. Geuder (Eds.), *The Projection of Arguments: Lexical and Compositional Factors,* pp. 97-134. Stanford, CA: CSLI.

Sedivy, J.C., Tanenhaus, M.K., Chambers, C.G., & Carlson, G.N. (1999). Achieving incremental semantic interpretation through contextual representation. *Cognition, 71,* 109-147.

Spivey, M.J., Tanenhaus, M.K., Eberhard, K.M., & Sedivy, J.C. (2002). Eye movements and spoken language comprehension: Effects of visual context on syntactic ambiguity resolution. *Cognitive Psychology, 45,* 447-481.

Spivey, M.J., Tyler, M.J., Eberhard, K.M., & Tanenhaus, M.K. (2001). Linguistically mediated visual search. *Psychological Science, 12,* 282-286.

Tanenhaus, M.K., Carlson, G.N., & Trueswell, J.T. (1989). The role of thematic structures in interpretation and parsing. *Language and Cognitive Processes, 4,* SI 211-234.

Tanenhaus, M.K., Magnuson, J.S., Dahan, D., & Chambers, C. (2000). Eye movements and lexical access in spoken-language comprehension: Evaluating a linking hypothesis between fixations and linguistic processing. *Journal of Psycholinguistic Research, 29,* 557-580.

Tanenhaus, M.K., & Spivey-Knowlton, M.J. (1996). Eye-tracking. *Language and Cognitive Processes, 11,* 583-588.

Tanenhaus, M.K., Spivey-Knowlton, M.J., Eberhard, K.M, & Sedivy, J.C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268,* 1632-1634.

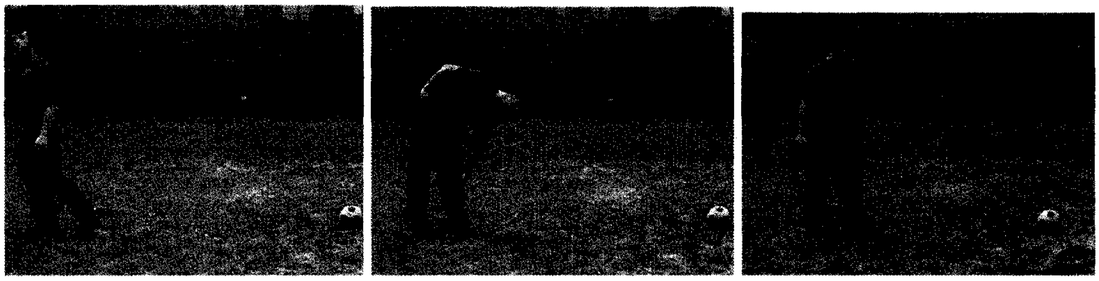
Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature, 381,* 520-522.

VanRullen, R., & Thorpe, S.J. (2001). The time course of visual processing: From early perception to decision-making. *Journal of Cognitive Neuroscience, 13,* 454-461.

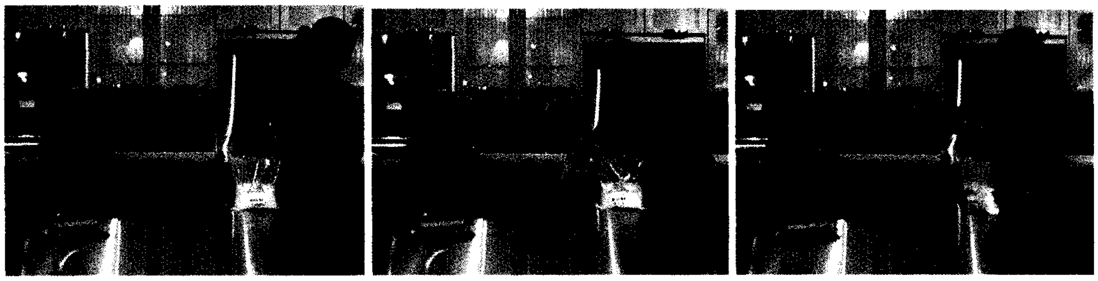
Yarbus, A. L. (1967). *Eye movements and vision.* New York: Plenum Press.

Appendix A

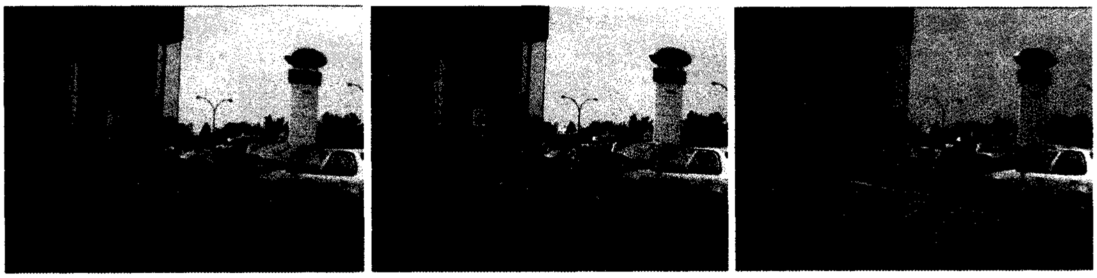
Below are the seventeen scene triplets (Away, Neutral and Towards) and the corresponding sentence pairs used in the experiment. The verb before the forwardslash (/) is the more selectionally restrictive causative verb, while the second verb is the perception verb used in each sentence pair.



1. After his warm up, the athlete will drop/inspect the ball that he uses for drills.

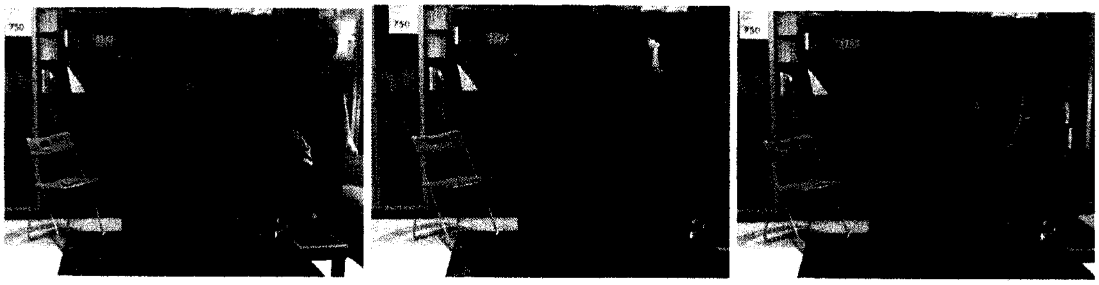


2. In order to bake some muffins, the woman will melt/check the butter that is required for the dough.



3. On her way to the station, the driver will crash/check the car that she just bought.

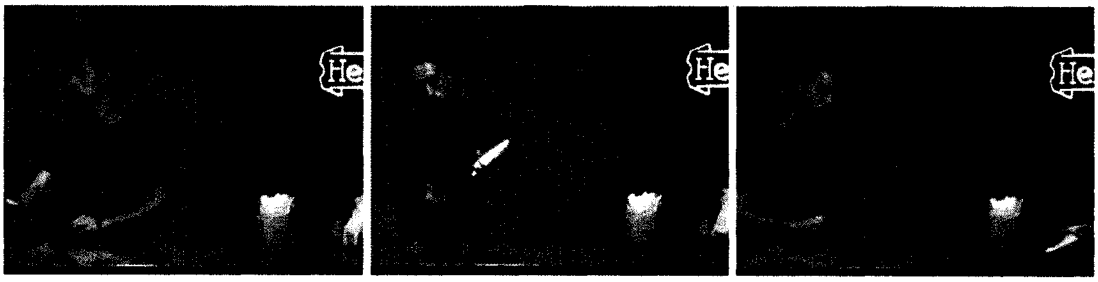
4. Before going to work, the driver will start/check the car that is in front of her house.



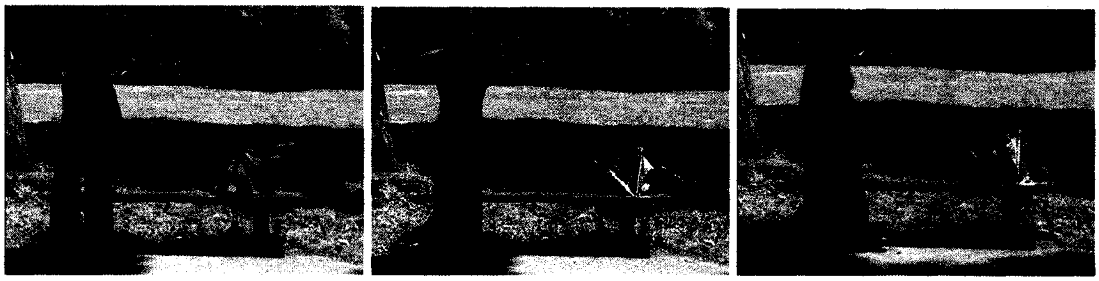
5. While dusting the furniture, the maid will fold/see the chair that is in the living room.

6. While playing with his toys, the infant will roll/notice the cube that is on the floor.

7. Before making the dessert, the cook will crack/examine the egg that is in the bowl.



8. While preparing the drink, the bartender will crush/notice the ice that he has to put in the glass.



9. While standing in the park, the girl will fly/see the kite that is on the bench.

10. While playing with the lid, the child will spill/spot the milk that is on the table.

11. Before preparing the cake, the cook will heat/inspect the oven that is in the kitchen.

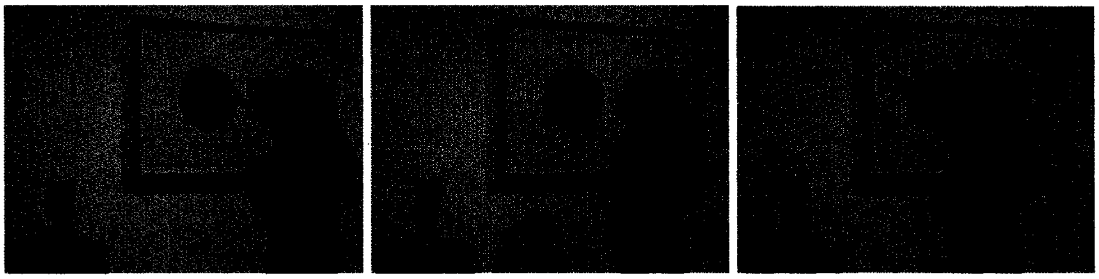
12. After talking on the phone, the secretary will rip/examine the paper that is on the desk.

13. While unpacking her office, the student will hang/study the picture that she bought at the auction.

14. Before ending his shift, the busboy will dry/spot the plate that is on the counter.

15. While packing his clothes, the man will wrinkle/see the shirt that he will use at the meeting.

16. After getting ready for work, the businessman will shine/examine the shoes that he got from his wife.



17. During her visit to the gallery, the girl will break/spot the vase that is on display.

Appendix B

CONSENT FORM TO PARTICIPATE IN RESEARCH

This is to state that I agree to participate in a program of research being conducted by Julia Di Nardo and Stephanie Houston of the Psycholinguistics and Cognition Lab at Concordia University, under the supervision of Dr. Roberto G. de Almeida. This research will contribute to J. Di Nardo's Master's Thesis.

A. PURPOSE

I have been informed that the purpose of the research is to investigate how the brain processes visual information and its interaction with language.

B. PROCEDURES

I have been informed that the experiment involves the following procedures: I will look at still pictures of scenes on a computer screen. A fixation cross ("+") will be present at the center of the screen, and I must keep my eyes fixated on the cross, until it is removed. When the fixation cross is no longer on the screen, I am free to move my eyes. My task will be to view the pictures that will appear on the screen, and then list the objects that I see and make up a sentence related to those pictures.

The completion of this experiment will take about 15 minutes. I have been informed that my name will not be associated with my data in the experiment. I understand that my participation in the experiment, and the information and the data I provide, will be kept strictly confidential. I understand that if the results are published, individual data will be reported but I will be only indicated as "Subject" and then a number.

C. CONDITIONS OF PARTICIPATION

- I understand that I am free to decline to participate in the experiment without negative consequences.
- I understand that I am free to withdraw my consent and discontinue my participation at anytime without negative consequences.
- I understand that my participation in this study is confidential (i.e., the researcher will know, but will not disclose my identity).
- I understand that the data from this study may be published.
- I understand the purpose of the study and I know that I have been made fully aware of the procedures.
-

D. QUESTIONS AND COMMENTS

- If you have any questions or comments with regards to this experiment, please do not hesitate to contact Dr. de Almeida at 848-2424 x2232 or Dr. Michael von Grünau at 848-2424 x 2190.
- If you have any other concerns, you may contact Andrea Rodney at the Office of Research Services at 848-2424 x4887.

I HAVE CAREFULLY STUDIED THE ABOVE AND UNDERSTAND THIS AGREEMENT. I FREELY CONSENT AND AGREE TO PARTICIPATE IN THIS STUDY.

NAME (please print) **SIGNATURE**

_______________________________ _______________________________

WITNESS SIGNATURE **DATE**

_______________________________ _______________________________

Appendix C

NORMATIVE STUDY

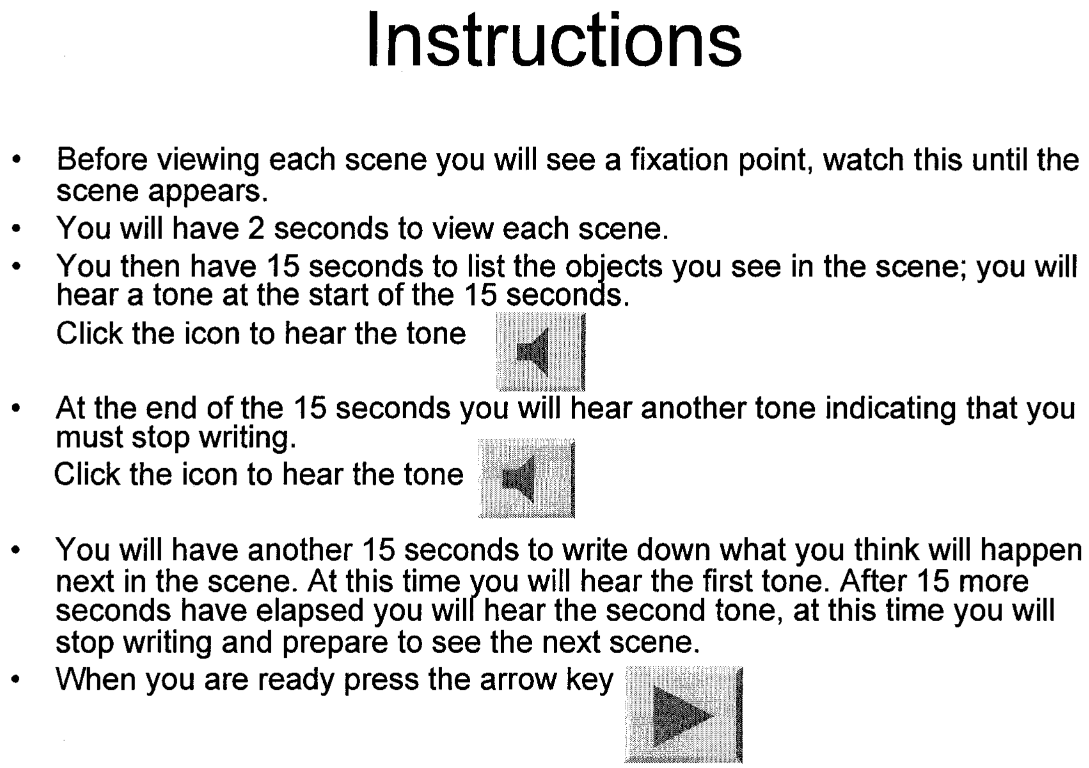
PARTICIPANT INSTRUCTIONS

Thank you for choosing to participate in this study! You will be viewing a number of scenes and we will ask you some questions about what you've seen.

The general format of the experiment is as follows:

1. Before viewing each scene you will see a **fixation cross** in the middle of the screen; keep your eyes focused on this until it disappears and the scene appears.
2. You will have **2 seconds** to view each scene. This is a very short amount of time so pay attention!
3. You then have **15 seconds** to **list the objects** you see in the scene; you will hear a beep at the start of the 15 seconds and another beep at the end of the 15 seconds indicating that you must stop writing. You will have a chance to hear what these beeps sound like at the beginning of the experiment.
4. You will then have another **15 seconds** to **write down a sentence** about what you think will happen next OR what you think is happening in the scene. The experimenter will tell you which question you will be answering during the experiment (it will be the same question throughout). Again, there will be two beeps indicating when you must begin and stop writing.
5. After this, the fixation cross will reappear and the entire process will be repeated.
6. The first three trials are **practice trials** to give you a feel for how the experiment works. After these trials, the experiment will pause and the experimenter will look over your answers to make sure you understand the task. You will also have the opportunity to ask any questions you might have.

Appendix D



Instructions

- Before viewing each scene you will see a fixation point, watch this until the scene appears.
- You will have 2 seconds to view each scene.
- You then have 15 seconds to list the objects you see in the scene; you will hear a tone at the start of the 15 seconds.
  Click the icon to hear the tone 🔈
- At the end of the 15 seconds you will hear another tone indicating that you must stop writing.
  Click the icon to hear the tone 🔈
- You will have another 15 seconds to write down what you think will happen next in the scene. At this time you will hear the first tone. After 15 more seconds have elapsed you will hear the second tone, at this time you will stop writing and prepare to see the next scene.
- When you are ready press the arrow key ▶

Appendix E

Version “a”

1) List the objects in the scene:

____

____

____

____

____

____

2) Write down what you think will happen next in the scene:

____

____

____

Version “b”

1) List the objects in the scene:

____

____

____

____

____

____

2) Write down what you think is happening in the scene:

____

____

____

Appendix F

Table F.1

Analysis of Variance for the Effect of Motion Type on Target Object Listing Frequency

Source	*df*	*F*	*p*
Motion Type (MT)	2	6.32	< .01
Error	48		
Total	50		

Table F.2

Analysis of Variance for the Effect of Motion Type on Target Event Listing Frequency (A)

Source	*df*	*F*	*p*
Motion Type (MT)	2	1.15	.33
Error	48		
Total	50		

Table F.3

Analysis of Variance for the Effect of Motion Type on Target Event Listing Frequency (B)

Source	*df*	*F*	*p*
Motion Type (MT)	2	.10	.90
Error	42		
Total	44		

Appendix G

CONSENT FORM TO PARTICIPATE IN RESEARCH

This is to state that I agree to participate in a program of research being conducted by Julia Di Nardo of the Psycholinguistics and Cognition Lab at Concordia University, under the supervision of Dr. Roberto G. de Almeida. This research will contribute to J. Di Nardo's Master's Thesis.

A. PURPOSE

I have been informed that the purpose of the research is to investigate how the brain processes visual information and its interaction with language.

B. PROCEDURES

I have been informed that the experiment involves the following procedures and that the task is done in the dark: An eye-tracker machine will be placed over my head by the researcher, and the machine will record my eye movements as I look at events unfolding on a computer screen. I will also hear sentences through earphones. A fixation cross ("+") will be present at the center of the screen, and I must keep my eyes fixated on the cross, until it is removed. When the fixation cross is no longer on the screen, I am free to move my eyes. My task will be to view the sentence-related pictures that will appear on the screen, and to listen to the concurrent sentences. In order to initiate each trial, I will be required to press the spacebar on the keyboard.

The completion of this experiment will take about 30 minutes. I have been informed that my name will not be associated with my data in the experiment. I understand that my participation in the experiment, and the information and the data I provide, will be kept strictly confidential. I understand that if the results are published, individual data will be reported but I will be only indicated as "Subject" and then a number.

C. CONDITIONS OF PARTICIPATION

- I understand that I am free to decline to participate in the experiment without negative consequences.
- I understand that I am free to withdraw my consent and discontinue my participation at anytime without negative consequences.
- I understand that my participation in this study is confidential (i.e., the researcher will know, but will not disclose my identity).
- I understand that the data from this study may be published.
- I understand the purpose of the study and I know that I have been made fully aware of the procedures.

D. QUESTIONS AND COMMENTS

- If you have any questions or comments with regards to this experiment, please do not hesitate to contact Dr. de Almeida at 848-2424 x2232 or Dr. Michael von Grünau at 848-2424 x 2190.
- If you have any other concerns, you may contact Andrea Rodney at the Office of Research Services at 848-2424 x4887.

I HAVE CAREFULLY STUDIED THE ABOVE AND UNDERSTAND THIS AGREEMENT. I FREELY CONSENT AND AGREE TO PARTICIPATE IN THIS STUDY.

NAME (please print) ______________________ **SIGNATURE** ______________________

WITNESS SIGNATURE ______________________ **DATE** ______________________

Appendix H

INSTRUCTIONS

In this experiment, you will **see** a series of **pictures** displayed on the screen. At the same time, you will **hear sentences** that refer to the pictures on the screen. Your task will be to simply view the pictures and listen to the sentences.

During this experiment, we will also be recording your eye movements. This will be done through the use of a head-mounted **eye-tracking** device that will sit on your head as you look at the screen. This equipment does not pose any risks, although it may be slightly uncomfortable. Before the experiment begins, please inform the experimenter if you are uncomfortable so that it can be adjusted.

There are a few details to understand before starting. Please read the sequence of tasks carefully, and make sure you understand what you should do in each part of the experimental trials.

1. First, to initiate each trial, you have to **press the spacebar**.

2. Each trial will begin with the presentation of a **fixation cross (+)** displayed in the middle of the screen. You should focus on this cross until it disappears.

3. When the **fixation cross disappears** and the **picture appears**, you will be free to move your eyes and scan the scene.

4. When the trial is over you will see a black screen and then an instruction to **press the spacebar**. This will initiate the next trial.

5. At the end of the experiment, which lasts about 5 minutes, there will be a short **quiz** that will test your memory for what you have heard and seen. Therefore it is important that you **pay attention** to both the visual display and the sentence presented over the earphones.

6. If you have any questions or concerns, do not hesitate to speak to the experimenter.

Have fun!

N.B. Please **do not talk** during the experiment since it can disrupt the eye-tracker.

Appendix I

Thank you for choosing to participate in this experiment.

You will be presented with a series of pictures accompanied by spoken sentences relating to the scene on-screen. You are asked to simply look at the pictures and listen to the sentences. Prior to each trial, there will be a fixation cross (+) in the middle of the screen that you must fixate on. This cross will be red on a black background. The picture will then appear, with the cross still in the centre of the screen. Keep looking at the cross. Once the + disappears, you may look wherever you like on the screen. To move on to the next trial, just press the spacebar. It is important to remember to pay attention to both the pictures and the spoken sentences. After the experiment is finished, you will be given a short memory task to ensure that you have been paying attention.

If at any time you experience discomfort, you may choose to discontinue the experiment.

Now sit back, relax, and enjoy!

Press the spacebar to continue.

Appendix J

Table J.1

Analysis of Variance for the Effect of Verb Type and Motion Type on SOT

Source	*df*	*F*	*p*
Verb Type (VT)	1	.37	.55
Motion Type (MT)	2	8.95	< .001
VT X MT	2	.09	.92
Error	138		
Total	143		

Table J.2

Analysis of Variance for the Effect of Sentence Point, Verb Type and Motion Type on the Cumulative Number of Saccades Initiated Towards the Target Object

Source	*df*	*F*	*p*
Sentence Point (SP)	2	53.16	< .0001
Verb Type (VT)	1	4.22	< .05
Motion Type (MT)	2	2.25	.11
SP X VT	2	1.42	.25
SP X MT	4	2.58	.04
VT X MT	2	.12	.89
SP X VT X MT	4	.14	.96
Error	540		
Total	557		

Appendix K

CONSENT FORM TO PARTICIPATE IN RESEARCH

This is to state that I agree to participate in a program of research being conducted by Julia Di Nardo and Zinnia Madon of the Psycholinguistics and Cognition Lab at Concordia University, under the supervision of Dr. Roberto G. de Almeida. This research will contribute to J. Di Nardo's Master's Thesis and Z. Madon's honours research project for PSYC 430.

A. PURPOSE

I have been informed that the purpose of the research is to investigate how the brain processes visual information and its interaction with language.

B. PROCEDURES

I have been informed that the experiment involves the following procedures and that the task is done in the dark: An eye-tracker machine will be placed over my head by the researcher, and the machine will record my eye movements as I look at events unfolding on a computer screen. I will also hear sentences through earphones. A fixation cross ("+") will be present at the center of the screen, and I must keep my eyes fixated on the cross, until it is removed. When the fixation cross is no longer on the screen, I am free to move my eyes. My task will be to view the sentence-related events that will occur on the screen, and listen to the concurrent sentences. In order to initiate each trial, I will be required to press the spacebar on the keyboard.

The completion of this experiment will take about 25 minutes. I have been informed that my name will not be associated with my data in the experiment. I understand that my participation in the experiment, and the information and the data I provide, will be kept strictly confidential. I understand that if the results are published, individual data will be reported but I will be only indicated as "Subject" and then a number.

C. CONDITIONS OF PARTICIPATION

- I understand that I am free to decline to participate in the experiment without negative consequences.
- I understand that I am free to withdraw my consent and discontinue my participation at anytime without negative consequences.
- I understand that my participation in this study is confidential (i.e., the researcher will know, but will not disclose my identity).
- I understand that the data from this study may be published.
- I understand the purpose of the study and I know that I have been made fully aware of the procedures.

D. QUESTIONS AND COMMENTS

- If you have any questions or comments with regards to this experiment, please do not hesitate to contact Dr. de Almeida at 848-2424 x2232 or Dr. Michael von Grünau at 848-2424 x 2190.
- If you have any other concerns, you may contact Andrea Rodney at the Office of Research Services at 848-2424 x4887.

I HAVE CAREFULLY STUDIED THE ABOVE AND UNDERSTAND THIS AGREEMENT. I FREELY CONSENT AND AGREE TO PARTICIPATE IN THIS STUDY.

NAME (please print) **SIGNATURE**

______________________________ ______________________________

WITNESS SIGNATURE **DATE**

______________________________ ______________________________

Appendix L

INSTRUCTIONS

In this experiment, you will see a series of short movies displayed on the screen. At the same time, you will hear a sentence that refers to the event occurring on the screen. Your task will be to view the sentence-related events.

During this experiment, we will also be recording your eye movements. This will be done through the use of a head-mounted eye-tracking machine that will sit on your head as you watch the movies. This equipment does not pose any risks, although it may be slightly uncomfortable. Before the experiment begins, please inform the experimenter if you are uncomfortable so that it can be adjusted.

There are a few details to understand before starting. Please read the sequence of tasks carefully, and make sure you understand what you should do in each part of the experimental trials.

1. First, to initiate each trial, you have to press the spacebar.
2. Each trial will begin with the presentation of a fixation cross (+) displayed in the middle of the screen. You should focus on this cross until it disappears.
3. When the movie begins and the fixation cross disappears, you are free to move your eyes and scan the scene.
4. It is important that you pay attention to both the visual display and the sentence presented over the earphones.
5. When the trial is over you will see a black screen and then an instruction to press the spacebar. This will initiate the next trial.
6. If you have any questions or concerns, do not hesitate to speak to the experimenter.

Have fun!

Appendix M

Thank you for choosing to participate in this experiment.

You will be presented with a series of short movie clips accompanied by spoken sentences relating to the event on-screen. You are asked to simply watch the movies and listen to the sentences. Prior to each movie, there will be a fixation cross (+) in the middle of the screen that we ask you to fixate on. Once this + disappears, you may look wherever you like on the screen. To move on to the next movie, just press the spacebar. It is important to remember to pay attention to both the movie and the spoken sentence. After the movies are finished, you will be given a short recall task to ensure that you have paid attention to them.

If at any time you experience discomfort, you may choose to discontinue the experiment.

Now sit back, relax, and enjoy!

Appendix N

Table N.1

Analysis of Variance for the Effect of Verb Type and Motion Type on SOT

Source	*df*	*F*	*p*
Verb Type (VT)	1	2.17	.15
Motion Type (MT)	2	17.63	< .0001
VT X MT	2	.66	.52
Error	180		
Total	185		

Table N.2

Analysis of Variance for the Effect of Sentence Point, Verb Type and Motion Type on the Cumulative Number of Saccades Initiated Towards the Target Object

Source	*df*	*F*	*p*
Sentence Point (SP)	2	58.23	< .0001
Verb Type (VT)	1	.03	.86
Motion Type (MT)	2	8.76	< .001
SP X VT	2	.11	.90
SP X MT	4	6.00	< .001
VT X MT	2	.19	.83
SP X VT X MT	4	3.06	.02
Error	666		
Total	683		