AN INVESTIGATION INTO THE INFLUENCE OF VERBS ON ATTENTION AND EVENT PERCEPTION

ISLA JONES

SEPTEMBER 2006

Submitted in partial fulfilment of the MSc in Speech and Language Sciences

Department of Human Communication Science
University College London

Acknowledgements

I would like to thank Maria Black and David Green for all of their advice and expertise. I am indebted to all those who took part in the video production, particularly my Dad, Kirsty, the Harvey family, James Godber, Yvonne and Jeff Mortimer.

I am especially grateful for the unending encouragement and support provided by Yvonne Mortimer, Sarah Harvey and my family.

ABSTRACT

A growing body of evidence suggests that language can influence conceptual representations of objects and events. Most of this evidence comes from cross-linguistic studies and investigations of how events are described. What has been mainly investigated, therefore, is the interaction between cognition and language output. In this study, we explored whether linguistic input, specifically verbs, could affect what we attend to and remember of an event.

Videoed events were seen in two conditions. In the sentence condition, a sentence was seen prior to the video. In the noun condition, a noun sequence was seen prior to the video. We compared responses and reaction times to a property verification task in the two conditions. The property to be verified was located on an area of the actor's body to which we predicted the verb in the sentence condition would focus a viewer's attention. We hypothesized that property verifications to items seen in the sentence condition would be significantly more accurate and faster than property verifications to items seen in the noun condition, as a result of the attention-directing effect of the verb.

Subjects were significantly more accurate in their responses to property verification tasks, when viewing events in the sentence condition. This difference was also seen across items. The six subjects who responded fastest to the property verification task responded significantly faster when the video had been preceded by a sentence than when it had been preceded by a noun sequence. Our results support the claim that verbs can influence what we pay attention to, and therefore what we process and remember of an event. Our results also point to the complex interaction of top-down and bottom-up influences on attention and event processing.

CONTENTS

TABLES AND FIGURES

1. INTRODUCTION

The relationship between language, perception, thought and action has received much attention over the years. The idea that thought is shaped by language is most commonly associated with the writings of Benjamin Lee Whorf (Whorf 1956). Whorf proposed that the categories and distinctions of each language enshrine a way of perceiving, analyzing, and acting in the world. He argued that speakers of different languages would also differ in how they perceive, think and act in objectively similar situations.

In its strongest form, Whorf's hypothesis is no longer considered plausible (Gentner and Goldin-Meadow 2004). However, there has been renewed interest in weaker forms of his position, with a shift from linguistic determinism, to 'linguistic relativity'. Some theorists have argued that in some conceptual domains language can influence conceptual representations, so that some aspects of the representations of events are language-dependent and not universal (Gentner and Boroditsky 2001). There are many studies to support this position, demonstrating that speakers of different languages focus on different aspects of an event, or an object, and this can affect their perception and memory of that event or object.

For example, Lucy and Gaskins (2001) investigated whether differences in the way English and Yucatec Mayans talk about objects and substances affected the way speakers of these languages conceptualise the shapes and materials of objects. Unlike English, where most concrete nouns are count nouns with singular and plural forms, in Yucatec Mayan, nouns tend to be closer to mass nouns in that they refer to substances and require a "unitiser phrase" when counting. For instance, 'two candles' in English would correspond to something like 'two long thin units of wax' in Yucatec Mayan. In this experiment, English speakers and Yucatec Mayans were shown an example object and were asked to choose which of two other objects was more similar to this example. The two choices varied from the example either in shape, or in material.

English speakers preferred the shape match, judging the two objects in the same shape as similar. Yucatec Mayans, on the other hand, preferred the material match judging the two objects of the same material to be similar. Lucy and Gaskins concluded that because aspects of Yucatec grammar refer to objects as if they were substances and pay more attention to materials and substances, they English speakers do. Boroditsky (2003) after reviewing several investigations of linguistic relativity concludes that language influences many aspects of human cognition, including conception of space, time, colour, shapes and events. Studies have also been carried out to investigate what effect linguistic requirements have on event perception and memory. Robinson & Triesch (2004) found that when subjects were asked to verbally encode the colour or name of objects within an event, this verbal encoding affected their memory of the event and their ability to detect changes to a previously viewed event. For example, subjects who were asked to name the objects in an event were significantly more accurate in their detection of object additions, than those subjects who had been asked to name the colours of objects in the event.

However, most of the current evidence for "linguistic relativity" or for the interaction between language and cognition comes from studies of event description, rather than event perception and memory. Slobin (1996) suggested that language has an influence on event perception only during the process of translating a visual and/or cognitive representation into a "message" appropriate for linguistic communication – a stage of processing he termed "thinking for speaking". Our study is an attempt to further explore the effect of language, and verbs in particular, on event processing rather than event description.

1.1 Aim of the study

Apart from studies on the effects of the semantic properties of motion verbs on event memory and categorisation (see Section 1.4 below), to our knowledge, there have been no studies that have investigated more general effects of verb semantics on event perception and memory. This study aims to investigate the attention-directing effects of verbs that, by virtue of the type of action they express, make reference to a part of the body of the actor. Participants in this study viewed either a sentence or noun sequence and were then shown a video clip of an event, after which they had to verify if the sentence or noun sequence correctly represented that event. They were then asked a question about the presence of a property within the event (the property verification task). The participants' responses and time taken to respond were recorded.

We tested the hypothesis that the verb in a preceding sentence would have a different effect on attention, and therefore event processing, than a noun sequence. We predicted that the particular semantic properties of the verbs we used would focus viewers' attention on a part of the actor's body where the property to be verified was located. As a result, videos seen with a preceding sentence would be responded to with significantly higher accuracy rates, and significantly faster response times, than videos seen with a preceding noun sequence. For example, we hypothesized that when a viewer saw the sentence '*Bill was coughing in his chair*' they would respond more quickly to the question 'Did *he have a moustache*?' than if the event had been preceded by the noun sequence '*Bill Chair*'. This is because the verb '*cough*' directs attention specifically to Bill's face in a way that the noun 'Bill' does not.

This hypothesis relies on a number of assumptions: firstly, that what we recall of an event is directly related to what we attend to. Secondly, that what we attend to in an event can be modulated by various factors including language. Thirdly, that the verb in a sentence will have an attentional effect than a sequence of two nouns. Fourthly, that speed of recall reflects the amount of attention given to an object in a scene. Each of these assumptions is discussed below.

1.2 Event Perception and Attention

The claim that what we recall of an event is directly related to what we attend to, relies on a body of research regarding how we perceive events. Chun & Wolfe (2002) state that "at any given time the environment presents far more perceptual information than can be effectively processed". So we use visual attention to select the information that is most relevant to us. But is the visual information that we do not actively attend to not encoded or represented in visual memory? Studies suggest that it is not. Rensink, O'Regan & Clark (1997) tested participants' ability to detect changes to objects, when viewing rapid presentations of an original and modified scene. They inserted a blank field between each presentation, so that the participants' ability to detect changes between the scenes relied on their ability to compare a scene they were currently viewing with the conceptual representation of a scene they had just viewed.

Rensink et al. (1997) found that identification of changes to an object in these conditions was extremely difficult. They concluded that focused attention was necessary for perception of change in an object. When objects do not receive this attention they make a brief sensory impression on the visual memory system, but this is simply overwritten by subsequent stimuli, making it difficult to compare between an object just seen and a newly viewed object. Other studies of object memory support the claim that attention is required on or near an object in order for it to be encoded in the visual memory system (Henderson & Hollingworth 1999). If attention to objects is needed for them to be encoded in visual memory, we can infer that, when a property verification response was correct, the participant had attended to the target property in the event.

To further understand how we perceive events, and why attention is essential in the process, requires some understanding of what attention is and how it functions. Talmy (2000) suggests that attention is a cognitive system with limited resources. It can link up with other cognitive systems and in so doing enhance the processing of the linked up system. According to Chun and

Wolfe (2002), it also plays a critical role in selecting relevant information for processing. Therefore, by selecting areas of importance in an event, the attentional system can focus its limited resources on these areas, thus enhancing the processing and therefore the encoding in memory of those aspects of a scene. If, as we hypothesise, the verbs we have used in our experiment direct attention to the target body areas of the actor in the videoed events, then we should see higher accuracy rates in verification of properties located on these areas, because attention will be directed there and thus facilitate encoding in the viewers' visual memory.

1.3 Influences on Attention in Event Perception

If attention plays a crucial role in event perception, what determines what we attend to? This question is related to our second assumption that what we attend to in an event can be modulated by various factors including language. There are two broad types of processing that affect what we attend to when viewing a visually presented event (Zacks & Tversky 2001). Bottom up processing refers to properties of the scene or event that draw or guide the attention of the viewer. Top-down processing refers to processing directed by a viewer's knowledge and intentions.

Evidence that bottom-up processing occurs has been documented as early as 1967 in a study by Yarbus. He studied subjects' eye movements and points of fixation as they looked at pictures. Studying eye gaze is a uniquely valuable tool when investigating attention because attention is inextricably bound to eye fixations (Henderson and Ferreira 2004). Yarbus found that uniform and empty regions of a scene are typically not fixated; fixations tend to land on objects and clusters of objects.

Corbetta and Shulman (2002) found that properties that were perceptually different from their background would receive more attention than properties that were less contrastive; for example, a poppy in a field of grass would receive more attention than a poppy in a field of tulips. Abrams & Christ (2003) conducted experiments in which subjects had to identify target letters from a range of targets and distracters. They found that speed of identification was significantly faster if the target letters began to move while the subject was viewing the range of targets and distracters, suggesting that onset of motion attracts attention.

Many other features and properties have been associated with bottom-up processing, although which stimulus properties are critical and how they are combined to influence attention has yet to be determined (Henderson and Ferreira 2004). We account for bottom-up processing in our experiment by having the same video event preceded by a sentence and by a noun sequence, so that any difference in recall will be due to the preceding sentence or noun sequence and not to any of the visual features of that particular item.

There is also a body of evidence to support the notion of top-down processing. For example, Yarbus (1967) found that, when he instructed participants to verify the age of people in a scene, they fixated upon the people's faces. On the other hand, when he instructed them to guess at the material wealth of the people, participants fixated on the furniture and clothing. Clearly, viewers' intentions and processing of the task instructions guided their attention in viewing the scenes. Henderson and Ferreira (2004) noted that top-down processing also occurs as a result of knowledge related to the stimulus scene or its event type.

We are basing our experiment upon the evidence that there are top-down factors that determine what a viewer attends to and, therefore, what they perceive and remember. The factor we are investigating is language, more specifically verbs.

1.4 The Influence of Language on Event Perception

What evidence is there that language directs attention in event perception? At a theoretical level, current theories of lexical meaning, especially within a cognitive linguistic approach, claim that linguistic meaning is primarily a device to direct or focus attention on different aspects of cognitive representations (Black, 2003; Black & Chiat, 2003 and in press). The semantic properties of words can focus attention on different components of events; for instance, both the verbs *slap* and *kick* express actions where a part of the body of the actor moves and makes contact with something but different parts of the body are focused – the hand or some hand-held instrument in the case of *slap*, and the leg and foot in the case of *kick*. But, in both cases, it is the body of the actor rather than that of any other participant in the event that the verb meaning directs attention to. With other verbs, verb meaning directs attention to other participants in the event and the change they undergo; for instance, the meaning of a verb such as *cut* directs attention to the specific change undergone by the non-actor participant. Theories that take this approach to meaning, therefore, imply a connection between linguistic meaning and attention.

At the empirical level, a number of studies have been done to investigate whether there are differences cross-linguistically in event perception and memory when differences between languages are exploited. The results have not been conclusive. Boroditsky, Ham, & Ramscar (2002) examined differences in verb syntax between Indonesian and English, and its effect on people's representations of events. One part of the investigation involved participants seeing an action in one of three tenses and later being asked to pick the picture that they had seen. For example, subjects were shown a picture of a man about to kick a ball and, in a later test phase, they were shown three pictures: one of the man about to kick a ball, one of the man kicking a ball, and one of the man having kicked a ball. Subjects were asked to choose which one they had previously seen.

English viewers were better at remembering which of the three tense versions of a picture they had seen as English requires tense to be obligatorily encoded, thus directing attention to whether the event was occurring in the past, present or future. This is not the case in Indonesian where temporal markers are optional and the temporal location of an event is often inferred from context. According to Boroditsky et al (2002), Indonesian speakers did not perform as well as English speakers, because their language does not single out the temporal contrasts associated with linguistic tense as an important aspect of an event.

In contrast, Papafragou, Massey and Gleitman (2002) investigated differences between the way in which English and Greek speakers encode manner and direction of motion, and the possible effects of these differences on processing and memory of motion events. In English, manner of motion is typically encoded within the verb, while information about the direction of motion appears in separate prepositional phrases. In Greek, the verb usually encodes the direction of motion, while the manner information is often omitted, or encoded in separate phrases. Papafragou et al (2002) gave monolingual English speakers and monolingual Greek speakers two tasks. In one task, they were asked to verbally describe a motion scene. In the other task they were asked either to remember that scene, or to match it to another scene.

Papafragou et al (2002) argue that the linguistic relativity hypothesis would predict that English speakers would be more sensitive to manner of motion and categorise events along this dimension, while Greek speakers would be more likely to remember the direction of motion and categorise events accordingly. Results from these experiments showed that, even though the two linguistic groups differed significantly in terms of their linguistic preferences for describing the events, their performance in the non-linguistic tasks of memory and categorisation was identical. Thus, cross-linguistic differences were not found to affect event perception and memory in this experiment.

There is some evidence, however, that linguistic factors in experimental tasks can alter attention, and therefore perception and memory of an event. When Rensink, O'Regan and Clark (1997) gave subjects the name of the object that was going to be changed, the viewer's ability to perceive when the change had occurred while viewing rapid presentations of the event, was greatly increased. When invalid written cues were given, subjects' ability to identify change decreased. These results point to the attention directing effect of language. Subjects were able to focus their attention on the named object, and thus remember features of this object and notice when these changed. When invalid cues were given, subjects focused their attention on the distracter object, at the expense of the target object, and were therefore slower at noticing when a change had occurred in the target object.

There have been some studies investigating the effect of verbs on event perception and memory. Bilman, Swilley and Kryan (2000), (cited in Gennari et al (2002)) found that hearing verbs of motion and manner (e.g., *walk*) or motion and path (e.g., *enter*) when encoding motion events, affected English speakers' subsequent recognition of the events. Participants tended to be less sensitive to changes on the dimension that they had not verbally encoded. This suggests that the type of verb used during encoding of events can influence recognition by either helping retrieval or shaping encoding.

Clearly, there is some evidence that language can have an attention-directing effect that may alter event perception and memory. However, the potential difference between the attention-directing effects of verbs and nouns, which is relevant to our experiment, has not been investigated in any depth. Meyer (2004), reviewing the use of eye fixations in studies of sentence generation, noted differences in eye gaze when speakers were asked to name objects with single nouns, as opposed to when they were asked to produce more complex descriptions involving a whole sentence. These results suggest language affects attention in the process of 'thinking for speaking'. Our experiment aims to discover whether verbs can affect attention and event perception and memory when subjects are not required to describe the event.

1.5 Visual Memory and Event Perception

The final assumption on which our hypothesis and the design of our experiment lies is that speed of recall reflects the amount of attention given to an object in a scene. This claim is based on the traditional association between ease of processing and speed of response. The line of reasoning is as follows: the more attention given to a property in an event, the more detailed its representation in the viewer's visual memory (Henderson and Ferreira 2004).

If the verb in the sentence condition directs attention to a particular part of the actor's body in our videoed events, then the conceptual representation of this body part should be relatively detailed. Therefore we would expect participants to be more accurate in their ability to verify the presence of a property on this area of the actor's body, and to be faster in their responses as a result. If the representation of the target area is less detailed, as we would predict to be the case when a noun sequence precedes the videoed event, participants may make more mistakes in verifying the presence of a property on this target area, and take longer to respond to the verification task.

In sum, there is considerable evidence to suggest that what we recall of an event is directly related to what we attend to, that language can modulate what we attend to, and therefore what we can remember of an event, and that, at least when viewing an event in order to describe it, verbs and nouns have different attentional effects. Consequently, we have a sound rationale for investigating the effect of a verb within a sentence on attention and event processing.

2. METHOD

2.1 Aim

To investigate whether verbs modulate attention, increasing accuracy and speed of response to a property verification task.

2.2 Design of Experiment

Participants were shown a video clip and asked to verify whether a preceding sentence, or a preceding noun sequence, accurately represented the video (the sentence verification task). They were then asked to verify the presence of a property in the video (the property verification task). The video along with its sentence and property verification constituted an experimental item, and each participant saw thirty items in total. Responses and reaction times to the property verification task for all items were recorded via the computer programme on which the experiment ran.

As no suitable video materials were available, new video clips were created for this project. Forty video clips were recorded and edited to depict the scenes expressed by our critical items in the sentence condition and our practice items (see Appendix 2 for a full list of items). The video clips created and used in the experiment can be viewed on the accompanying CD. We used 19 actors for the various events comprising 9 males and 10 females, and 7 children and 12 adults with an age range of 11 months – 52 years.

After relevant filtering of the data, the reaction times to correct property verifications were analysed using a variety of statistical tests. The design is a repeated measures design with a single factor of matching task (sentence vs noun).

2.3 Description of Item Types

Each participant saw a 'list' of items which comprised of 20 critical items and 10 filler items.

Critical Items

The critical items were defined by the following characteristics:

- They required a 'yes' response to sentence verification,
- They required a 'yes' response to property verification,
- The property to be verified was located on the body-area to which the verb in the sentence condition might focus attention to. For instance, when the verb was *sneeze* (a verb that directs attention to the head/face), the property verification was '*Did she have glasses*?'(Item 6). When the verb was *clap* (a verb that directs attention to the hands), the property verification was '*Did he have a watch?*' (Item 16). When the verb was *jump (*a verb that directs attention to the legs), the property verification was '*Did she have bare feet?' (Item 25).*

Across participants there were 30 different critical items. These were composed of video clips of events involving actions related to three body areas of the actor. Ten videos involved actions related to the head/face of the actor, and named by English verbs such as cough, smile, sneeze. Ten depicted actions related to the arm/hand, expressed by verbs such as, hug, clap, elbow, and ten showed actions related to the leg/foot and named by verbs such as hop, jump, and limp. Classification of the actions and their corresponding verbs was based on the body-part classifications motivated by Pulvermuller, Harle, & Hummel (2001) as well as on independent semantic grounds (Black & Chiat, 2003).

Each critical item video was seen by a participant after a sentence (Sentence Condition), e.g., *the girl was kicking the ball,* or after a sequence of two nouns (Noun Condition), e.g., *girl ball.* All the sentences for the sentence condition were 6-7 words in length and had a simple structure with one main verb followed by either a noun phrase or a prepositional phrase. The subject noun phrases always expressed the actor participant in the event depicted by the video clip. The noun sequences always involved two nouns that consisted of the actor and the last noun in the sentence condition, which had a variety of participant and non-participant roles. A complete list of critical items in sentences and noun sequences can be found in Appendix 2 table A.r.

Filler Items

The filler items involved the same video clips used for the critical items but with different preceding sentences/nouns, and property verifications. The sentences/nouns for the filler items differed from those of the critical items with the same video clip in one of three ways: they involved either a change of agent, a change of action or a change of the second event participant or background event feature (the "non-agent").

Response to sentence/noun sequence verification was 'no' for half of the fillers and 'yes' for the other half of fillers given to any one participant. The property verification for all filler items was different from the critical items and the response required was 'no'.

2.4 Rationale for Experimental Design

Across participants the same property was responded to in different contexts. One participant responded to a property verification having seen a sentence before the video clip, a different participant responded to the same property verification having seen two nouns before the video clip. This ensured that the same target property is responded too in different contexts so that aspects of word frequency and property salience did not confound the results, therefore making the role of the verb more explicit in modulating attention and changing accuracy rates and reaction times.

Filler items were used at the beginning of the task to ease participants into the experiment, so that factors such as novelty would not skew results for critical items. Filler items were grouped together so that it appeared to participants that there were more 'no' responses to sentence verification and property verification tasks than was actually the case. Fillers that required a 'no' response to sentence/noun verification were important in ensuring the participant attended to the sentence. This is vital, as it was the effect of the verb on property verification that was of interest.

For the purposes of the item analyses (i.e. between subject analyses) the critical items were rotated so that although each participant only responded to twenty critical items, we had data on all thirty critical items in both conditions. To do this we created 3 lists of items, each comprising 20 critical items and 10 filler items. Each list was also reversed so that any priming or position effects would be eliminated from the data set as a whole. Fillers were distributed in a fixed position between critical items across lists (see Appendix 2 table A.s for the distribution of items). This ensured that the position of the fillers was not a factor in the differing results between participants. There were no more than two consecutive items with property verifications focusing on the same body part, again to avoid priming affects that may skew results.

The properties requiring verification for both critical and filler items were not implied by the verb but were plausible features of the event. This ensured that the sentence condition did not provide any further cues as to property verification, other than to direct attention to the relevant body part. It also ensured that the participant was unable to rely on world knowledge to correctly respond to the property verification task, encouraging close attention to the video clip in both conditions.

2.5 Distribution of Critical and Filler Items.

Critical Items

Critical items were split into three sets of 10 items. (A, B and C). Each set contained videos with actions relating to the mouth/face, arm/hand, and leg/foot of the actor. This ensured that every participant saw items with a range of actions so they were not primed to look at a particular area of the actor's body. Each set has a noun version (e.g. Set An) and a sentence version (e.g. Set As).

Filler Items

The filler items presented to each participant consisted of the ten video clips that the participant did not see in the critical items. Filler items were composed of a mixture of noun sequences and sentences. The filler items were placed in three groups set A, B and C each set had the same video clips as the critical item set with the same name (e.g. Set A had the same video clips as set A in the critical items).

The distribution of items across 3 subjects is shown in Table 2.p below.

	List 1	**List 2**	**List 3**
	S1	S2	S3
Critical Items	**An**	**Bn**	**Cn**
	Bs	**Cs**	**As**
Filler Items	Set C	Set A	Set B

Table 2.p: distribution of items across 3 subjects

Participants 4 – 6 viewed items from the same sets as subjects 1-3 but each list of items was presented in reverse order.

2.6 Experimental Procedure

Participants

Twenty-four monolingual native English speakers participated in the experiment. Eighteen were female and six male, all educated to degree level or above. The participants were unpaid volunteers.

Participants were given the following written instructions.

> *'You will be shown a sentence or two words followed by a video clip. When the clip has stopped you will be asked to press a response key to indicate whether the sentence or words describe/match the video clip. After you have answered this question you will be asked another question about some aspect of the event. There will be 30 video clips and questions within the test and 10 practice clips. Each video clip takes 2 seconds so the test should not take longer than 10 minutes. Sometimes the sentence or words will not match the video clip. It is important that you press the correct key in response. Your reaction time is going to be measured so try to answer each question as accurately but also as quickly as you can. If you have any questions please feel free to ask before we begin.'*

The participant then began the practice items which were identical in format to the test items but involved different videos, preceding sentences or nouns, and property verifications. The experimental task proceeded as outlined below:

1) The participant pressed the space bar to begin the practice items or test items.

2) The participant saw the first sentence/noun sequence which was left on the screen for 1.5 seconds. For instance, 'The boy was licking the lolly' or 'Paul Friend' or 'Sarah jumped over the wall'.

3) The sentence/nouns then disappeared and the video played.

4) The video clip lasted for 2 seconds and, after it finished, the word 'match?' appeared on the screen along with 'yes' 'no' icons.

5) The participant verified whether the video clip matched the sentence/nouns by pressing the designated 'yes' or 'no' key.

6) There was then a 100ms blank interval.

7) The property verification question was then displayed along with 'yes' and 'no' icons. For instance, 'Did he have glasses' for the sentence 'The boy was licking the lolly', or 'Did he have long sleeves' for the noun sequence 'Paul Friend', or 'Did she have bare feet' for the sentence 'Sarah jumped over the wall'.

8) The property verification remained on the screen until the participant had responded.

9) Once the participant had responded, there was an inter-trial interval of 1500ms and then the next sentence or noun sequence appeared.

<u>Data Filtering</u>

Once responses and reaction times had been collected for all critical item property verification tasks, the data was filtered. Individual reaction times to property verifications were eliminated if they met any of the following criteria:

1) The response to the sentence verification or property verification task was incorrect
2) The item across subjects had a less than 75% correct response rate
3) The reaction time was an outlier (i.e. more than 1.5 times the interquartile range) compared with either the subject's data or the item's data.

This ensured that data for items that had been responded to at chance level was not taken into account in the analysis, and that variance was reduced, so that the statistical analysis represented the vast majority of reaction times calculated and was not skewed by abnormal reaction times.

2.6 Analyses of Results

For the subject analysis, responses and reaction times for the two conditions were summed over items for each subject, giving an accuracy percentage and a mean reaction time for that subject's correct responses to the property verification task in the sentence condition, and in the noun condition.

For the item analysis, responses and reaction times were summed over subjects to give an accuracy percentage and mean reaction time for each item correctly responded to in sentence and also in the noun condition.

3. RESULTS

3.1 Treatment of the data

Correct responses were recorded on the computer programme on which the experiment ran, and were totaled with respect to the data set being considered. Reaction times were computed for participants' responses to all critical items in both sentence and noun conditions. Twenty participants supplied reaction times for 20 critical items, and 4 participants supplied reaction times for 19 critical items. Thus the total number of potential responses to be analysed was 476.

Items were then filtered so that any item which had a less than 75% correct response rate, was eliminated from the data set, thus ensuring data was not analysed for items that were responded to correctly at chance. This resulted in the elimination of reaction times for 12 items, thus reducing the number of potential responses to be analysed by 188 (39%).

Error responses to the sentence verification task were then counted for the remaining items, a total of 7 error responses (2.43%) across all 24 subjects. Data for these items were removed. Error responses to the property verification task were then counted and totaled 15 (5.3% of the remaining potential responses to be analysed). These responses were then eliminated from the data.

Finally outliers (i.e. responses that were more than 1.5 times the interquartile range for the given subject or item) were eliminated. This totalled 26 responses (9.8%) of the data to be analysed. Therefore 240 responses to critical items were analysed corresponding to 50.4% of the initial potential responses to be analysed.

3.2 Main Findings

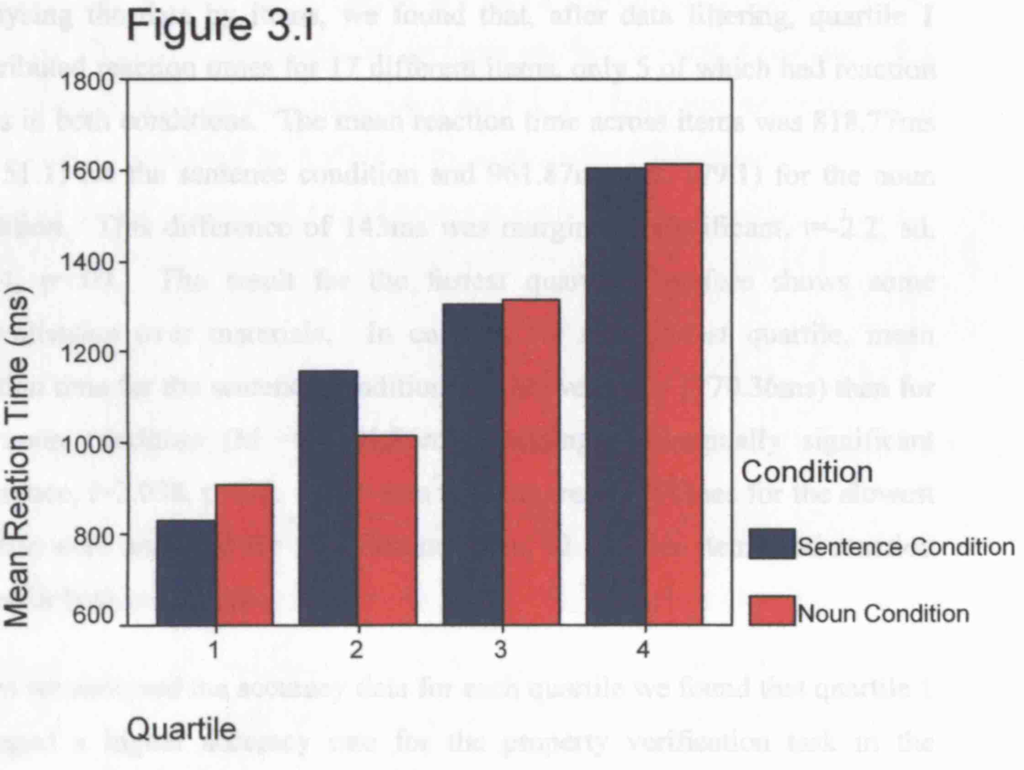
Our hypothesis was that compared to the noun condition, the verb in the sentence condition would have an attention directing effect, producing faster reaction times and more accurate responses to the property verification task. The reason was that the property to be verified in the critical item was located as an area of the actor's body to which the verb in the sentence condition directed attention. Therefore we predicted that participants would be significantly more accurate in their response rate to critical items in sentence condition than in noun condition, and that accuracy for items would be greater in the sentence condition than the noun condition. We also predicted that subjects would demonstrate a significant difference in their reaction time to property verification for sentence and noun conditions, reacting faster in the sentence condition, and that the items would produce faster reaction times in the sentence condition.

There was a significant difference between the accuracy of responses to the property verification task in the sentence and noun conditions when the subject data was analysed. Mean accuracy in percentage of correct responses for the sentence condition was 93.8% (sd, 12.5) and 87.4% (sd, 16.8) for the noun condition. This difference of 6.4% was found to be significant in a paired samples t-test, t=2.3 df, 23 p=.03. In the item analysis a marginally significant difference was found between the items in a sentence condition and noun condition. The mean accuracy of response across items in the sentence condition was 96% (sd,7.8), while in the noun condition the mean accuracy of response to the same items was 91% (sd 10.7). In comparing these means with a paired samples t-test, we found the difference to be marginally significant, t=1.9 df, 17 p=.07.

There was no overall difference in reaction times to the property verification task between conditions when the sample was considered as a whole (subject variability is displayed in Appendix 1, Figure 1 and variability across items is displayed in Appendix 1, figure 2). Mean Reaction time across subjects was 1223.9ms, sd, 335.7 in the sentence condition, and 1213ms, sd, 312.3 in the noun condition (t=.2 df=23, sd, 243.4, p=0.8). Similarly by items, the mean reaction time across items in sentence condition was 1280.4ms with a standard deviation of 296.1 and 1183.2ms in noun condition with a standard deviation of 223.6, t=1.2 df=17, sd, 333 p=.20.

The data was further analysed by averaging reaction times for correct responses in sentence and noun condition and dividing the subjects into four groups based on the speed of correct response averaged over correct response times for the sentence and noun conditions. Our intention was to contrast performance on the fastest quartile with that on the slowest quartile.

Quartile 1 is designated the fastest quartile and quartile 4 the slowest. The average reaction times for these quartiles in both the sentence and noun conditions are displayed in Figure 3.i.



Mean Reaction Time for Subject Quartiles

Figure 3.i shows that subjects in quartile 1 (the six subjects who responded quickest) verify properties fastest in the sentence condition compared to the

noun condition. In contrast there is no difference as a function of condition for those in the slowest quartile. A repeated measures mixed factor ANOVA with the between-subjects factor of quartile (fastest vs slowest) and the within-subjects factor of condition (sentence condition vs noun condition) yielded no overall interaction between quartile and condition, F (1,10) =.183, p=.7. However, t-tests confirmed a significant difference between subjects' reaction time to property verification in the sentence and noun condition for faster subjects (quartile 1) but not for the slowest (quartile 4) (see Table 3q).

Quartile	**Mean Reaction Time (ms) Sentence Condition**	**Mean Reaction Time (ms) Noun Condition**	**T**	**Significance**
1	831.02	909.86	2.96	.032 (p<.05) S
2	**1156.95**	**1014.71**	**2.667**	**.045 (p<.05) S**
3	1304.70	1314.33	-.09	.932 (p>.05) NS
4	**1602.86**	**1613.10**	**-.064**	**.951 (p>.05) NS**

Table 3.q: Mean Reaction Times and T-test results for all quartiles

Analysing the data by items, we found that, after data filtering, quartile 1 contributed reaction times for 17 different items, only 5 of which had reaction times in both conditions. The mean reaction time across items was 818.77ms (sd, 51.1) for the sentence condition and 961.87ms (sd, 179.1) for the noun condition. This difference of 143ms was marginally significant, t=-2.2, sd, 146.4, p=.09. The result for the fastest quartile therefore shows some generalisation over materials. In contrast for the slowest quartile, mean reaction time for the sentence condition was slower (M = 1770.36ms) than for the noun condition (M = 1501.86ms) yielding a marginally significant difference, t=2.038, p=.07. After data filtering, reaction times for the slowest quartile were analysed for 18 different items, 12 of these items had reaction times for both conditions.

When we analysed the accuracy data for each quartile we found that quartile 1 averaged a higher accuracy rate for the property verification task in the sentence condition (those items in which we would expect to see the effect of the verb). All quartiles averaged a higher accuracy rate in the sentence condition than the noun condition (Table 3.r)

Quartile	Accuracy of response (%)			
	Sentence Verification for critical items	Property verification for critical items	Property verification in Sentence Condition (critical items)	Property verification in Noun Condition (Critical Items)
1	97	97	100	92
2	94	95	96	89
3	99	93	93	91
4	98	90	87	78

Table 3.r. Accuracy of response across quartiles.

3.3 Exploring response times for Quartile 1

In order to explore why the fastest participants (quartile 1) were significantly faster in verifying properties in the sentence condition than in the noun condition, and why this was not the case for the remaining 18 subjects, we looked to see whether properties of the items responded to differed across quartiles. Table 3.s. illustrates the potential number of correct responses that participants in each quartile could have made in verifying properties related to the face/head, arm/hand or leg/foot of the actor in the video event.

Quartile	Number of head and face target properties	Number of arm and hand target properties	Number of leg and foot target properties	Total number of critical items potentially to be analysed.
1	36 (52%)	12 (17%)	21 (31%)	69
2	24 (33%)	16 (23%)	31 (44%)	71
3	28 (39%)	16 (22%)	28 (39%)	72
4	24 (33%)	17 (23%)	32 (44%)	73

Table 3.s. Distribution of different property verification types across quartiles.

Clearly, quartile 1 had opportunity to respond to a greater number of head/face related property verifications than subjects in the other 3 quartiles.

We then looked to see whether this imbalance was reflected in the correct responses of the participants in the different quartiles. These results are shown in Table 3.t

Quartile	Percentage of head and face property verifications
1	52%
2	32%
3	35%
4	30%

Table 3.t percentage of correct responses that were responses to head and face property verifications.

Evidently quartile 1 responded to a higher proportion of head and face property verifications than the subjects in the other quartiles. This may be significant in explaining the differing results between quartiles 1 and quartiles 2, 3 and 4, a possibility explored in Section 4.4 of the Discussion.

3.3 Exploring response times for Quartile 2

Inspection of Table 3.q reveals that subjects in the slowest two quartiles behaved similarly. Unexpectedly, subjects in quartile 2 showed a reverse effect to those in quartile 1, averaging significantly faster response times to property verifications for items seen in the noun condition, compared with property verifications to items seen in the sentence condition. On closer inspection, we found that this reverse effect was seen only in 3 of the subjects in quartile 2 who had seen lists 2 & 5 (Lists 2 & 5 are composed of the same items but reversed). The other three subjects in quartile 2 demonstrated similar results to those in quartile 3 and 4, in that there was no real difference between the sentence and noun conditions, although reaction times to the property verification task in the sentence condition were slightly faster than in the noun condition. These results are summarised in Table 3.u.

Lists viewed	Mean reaction time in sentence condition	Mean reaction time in noun condition
1 & 3	1154.8	1058.33
2 & 5	1157.75	963.33

Table 3.u. Mean reaction times for subjects in quartile 2 according to lists seen.

In order to further explore whether these reverse effects were due to peculiarities associated with list 2 and 5, we looked at all the participants who had seen lists 2 & 5. We found that six of the eight participants who had seen these lists produced a reverse effect in their reaction time data. We then looked at all the participants who had produced reverse effects and the lists they had seen. The distribution of lists across subjects who produced reverse effects can be seen in Table 3.v.

Subject	**List**
2	1
4	**1**
5	4
6	**2**
8	2
12	**2**
17	5
19	**5**
20	5
21	**6**
23	6

Table 3.v Distribution of lists across subjects who responded faster to property verifications in the noun condition than the sentence condition.

Therefore 6 of the 11 participants who had produced reverse effects had seen lists 2 & 5 the remaining five participants were distributed across the other lists.

3.4 Summary of Results

Subjects were significantly more accurate in their responses to property verification tasks, when viewing events in the sentence condition. This difference was also seen across items. Items were responded to at significantly higher accuracy rates, when they were presented in the sentence condition, than when they were presented in the noun condition. These results support our hypothesis.

The six subjects who responded fastest to the property verification task, responded significantly faster when the video had been preceded by a sentence, than when it had been preceded by the noun sequence. This significant difference was also seen across the items that the fastest participants responded to. These results also support our hypothesis.

The results for the participant group as a whole did not support the hypothesis regarding speed of response. Reaction times for subjects in quartile 2 elicited a 'reverse effect', although the effect is probably a consequence of list-specific peculiarities.

4. DISCUSSION

4.1 Review of Hypothesis

Our experiment tested the hypothesis that the verb in a preceding sentence would have a different effect on attention, and therefore processing of a videoed event, than a noun sequence. We predicted that this effect would be manifested in higher accuracy rates and faster responses to property verification in the sentence condition. We hypothesised that we would obtain these results because the verb in the sentence condition focused attention on a part of the actor's body where the property to be verified was located.

4.2 Evaluating our Hypothesis: Accuracy Results

The results regarding accuracy of response support our hypothesis, and other evidence, that language can influence attention and therefore event processing. We suggest that subjects were more accurate in their responses in the sentence condition, because the verb focused the viewer's attention on the target area of the actor's body.

Although accuracy rates in the noun condition were significantly lower than those in the sentence condition, they were still very high (Table 3.r). This is evidence that for the majority of viewings the target properties were processed without the attention focusing influence of the verb. If, in the majority of cases, the target body areas of the actor, and therefore the properties located on these areas, were processed, the influence of the verb can be understood as a further 'push' to attend to, and process in more depth, these target areas resulting in higher accuracy rates. This explanation corresponds with Chun and Wolfe's (2002) description of the two critical roles of attention: firstly, to select behaviourally relevant information, and secondly, to modulate or enhance this selected information.

The reasons why our target areas would be processed without the focusing effect of the verb can be understood with reference to what we know of top-down and bottom-up processing. We know, for example, that attention is drawn to the onset of motion (Abrams and Christ 2003). Some of our events were motion events, and the target property was necessarily related to the body part initiating the movement, therefore attention would probably be drawn to the target body area. We also know that attention tends to focus on the people in an event: Yarbus (1967), Weith, Castelhano and Henderson (2003), found eye fixations would accumulate around people in an event. All of the target properties in the items analysed were located on the body of the actor. Therefore a number of bottom-up influences made the processing of our target area likely.

Top-down influences may also account for the processing of our target area without the attention-directing effect of the verb in the sentence condition. The nouns in the noun sequence may also have exerted an attention-directing effect, focusing the viewers' attention on the actor, although not necessarily on the target area of the actor's body. There is evidence that nouns have an attention-directing effect. Rensink, O'Regan and Clark (1997), asked subjects to view rapid presentations of the same scene, and identify when a change in the scene had occurred and which object had changed. Subjects were significantly faster at identifying when the change had occurred, if the object that changed was named prior to viewing. Therefore, when subjects were given the name of the object that was going to change, they were able to direct their attention to this object, and thus identify the point of change more rapidly, than when the target object had not been foregrounded by their attention. Given our results, we may postulate that nouns direct attention to whole entities, which in the context of our experiment was the actor's whole body, while the verbs we used directed attention to specific areas of the actor's body.

Task knowledge may also have been a top-down influence on the deployment of attention when viewing the scenes. The fact that viewers knew they would be asked property verification questions may have influenced their viewing patterns, and the amount of visual information they encoded when watching the event. There is evidence to suggest that prior knowledge of the task requirements affects viewing patterns, as well as the detail and amount of visual information encoded. Robinson and Triesch (2004) gave subjects two tasks: in the first, they were asked to describe objects in natural images reporting a specific property of each object (e.g., its colour or name) when a crosshair appeared above it. In the second, subjects viewed a modified version of each scene, and were asked to detect which of the previously described objects had changed.

The subjects were split into two groups, those that had prior knowledge of the second task (Experiment 1) and those that did not (Experiment 2). The description task affected performance on the memory task only for those subjects who had no prior knowledge of the memory task (Experiment 2). When subjects in Experiment 2 described the colour of objects, their ability to detect colour changes was increased, whereas when they named the objects their ability to detect object additions increased. These results, and the fact that they were not shared with subjects in Experiment 1, are interpreted by Robinson and Triesch to suggest that people automatically encode a surprising amount of visual detail when viewing natural scenes, but the details encoded depend strongly on the task people have been set, and their understanding of that task's overall demands.

Evidently a number of bottom-up and top-down influences may have caused our viewers to process the target properties. It is likely that the deployment of attention, and thus the processing of these properties, relied on a convergence of top-down and bottom-up processing. This explanation concurs with Zacks and Tversky's (2001) argument regarding the requirements of event processing. They claim event processing involves; "the interaction of mechanisms .... that are domain general and domain specific, and that are experience dependent and experience independent."(18).

Our accuracy results suggest that verbs may be one of the top-down factors that influence the selection of visual information to be processed, and the depth of processing aspects of an event receive.

4.3 Evaluating our Hypothesis: Reaction Time Results

The effect of the verb in significantly increasing speed of response to the property verification task, was restricted to those subjects who were fastest and most accurate (Quartile 1). Subjects in quartiles 2 were significantly faster at responding to the property verification task in the noun condition, and were less accurate than subjects in quartile 1 in their responses to the property verification task in the sentence condition. Quartiles 3 and 4 showed no significant difference in their response times to the property verifications in the sentence and noun condition and were less accurate in their responses to the property verification task in the sentence condition than subjects in quartiles 1 and 2 (see tables 3.q and 3.r).

Why should the effect of the verb on speed of property verification be restricted to a cohort of all participants? An explanation may be found in the materials and methodology of the experiment. Although the lists of items seen by participants were carefully balanced before viewing, after data filtering, there were discrepancies between the lists with regard to the total number of responses that could be analysed. Discrepancies were also seen between lists in regard to the proportion of property verifications relating to the actor's head/face, arm/hand and leg/foot that could be analysed. Subjects in quartile 1 were able to provide a higher proportion of property verifications related to the head/face of the actor than the other quartiles. As a result, 50% of the data set for quartile 1 reflected responses to head/face property verifications, whereas between 30-35% of the data sets of quartiles 2 3 and 4 consisted of property verifications related to the actor's head/face (see Table 3.t).

The effect of these list discrepancies on accuracy and reaction time data can be explained with reference to what we know about event processing. As we have already discussed, deployment of attention and the processing of events seem to depend on the interaction of top-down and bottom-up influences. Some researchers have suggested that when top down and bottom up influences converge and reinforce each other, processing is facilitated (Black, 2003; Black and Chiat, in press; Sacchet, 2005). One of the bottom-up factors that may be at play in our experiment is the tendency to attend to people's faces in processing events. Weith, Castelhano and Henderson (2003) found that when viewers studied pictures of a janitor cleaning, and a janitor stealing from an office, a large proportion of fixations fell on the janitor's face. This would suggest a cognitive bias for paying attention to the faces of people in a scene.

Such a cognitive bias would converge and be reinforced by the face-focusing effect of those verbs in our experiment that are head/face related. The effect of the verb should be most evident when it enhances processing of an area already biased to receive attention. Thus, it would not be surprising to find the strongest verb effect on accuracy and speed of responses for subjects in quartile 1, whose data set comprised over 50% head and face property verifications. When the verb in the sentence condition directed attention to an area of the actor's body that was not already biased to receive attention (i.e., verbs focusing on arm/hand or leg/foot), a conflict between bottom and top-down factors may have occurred. Resolution of such a conflict may have incurred processing costs such as the slowing of responses.

List specific features may also account for the anomalous reaction time results recorded for quartile 2. On closer inspection of the data, it was found that not all 6 participants in this quartile produced a reverse effect, responding faster to property verifications in the noun condition rather than the sentence condition. It was only the three participants who had seen lists 2 and 5 (the same lists but in different orders) who had this reverse effect. After data filtering, lists 2 and 5 had noticeably more property verifications related to the actor's leg/foot than was the case with the other lists. However, the exact reason why this should make participants faster in responding to property verifications in the noun condition is unclear.

Our reaction time data does support our hypothesis, although only for the fastest and most accurate responders. Interestingly, examination of the data and items seen by participants, again point to the complex interaction of top-down and bottom-up influences that affect attention deployment and event processing. These interactions will have to be taken into account and be better controlled in future investigations into the possible attention-directing effects of language in visual event processing.

4.4 Alternative Interpretations

It might be possible to argue that the sentence condition encouraged a greater degree of processing over all areas of the event than the noun condition, and therefore the difference in accuracy and reaction time between the conditions was a reflection of overall depth of processing, rather than the attention-directing effect of the verb. According to this interpretation we would expect to see a significant difference in accuracy of property verification response across all items, including filler items, in which the property verification is not located on the target body area. In fact there is very little difference in the accuracy of response in the two conditions when all items are considered (Appendix 1 table A.q). This supports our hypothesis that the difference between the two conditions for the critical items is due to the attention directing effect of the verb. When some of the properties were located on areas of the actor's body that the verb did not focus on, or in other parts of the event, no significant difference was found between the conditions.

It could also be argued that the differences between the quartiles in relation to accuracy and reaction time data, were a result of individual differences regarding processing styles and response strategies. It may be that some individuals are more sensitive to the attention-directing influence of language than others. It would be interesting to further explore this possibility by measuring eye movements and eye fixations, which would allow us to track more precisely individual differences in scanning patterns and fixation durations.

4.5 Evaluation of Materials and Methodology

The high accuracy rates obtained across all items in our experiment showed, in accordance with the current literature on visual event processing, that viewers are able to scan and encode a surprisingly large amount of visual detail in a short amount of time. Evidence from studies that use eye-tracking devices to record where in an event viewers look, and for how long, also suggest viewers are able to shift attention and retain visual information at a remarkable speed (Henderson and Ferreira 2004).

However, there were some shortcomings to our materials and methodology that should be overcome in future research into this area. With regard to the materials we used, it should be noted that a very small number of items could have been responded to relying on general knowledge. For example, Item 30, '*The Scotsman was marching on parade – Was he wearing a kilt?*' This may have encouraged a different form of top-down processing, relying on this knowledge rather than the viewer's conceptual representation of the event. Additionally, some of the items may have attracted attention because of their visual or cognitive prominence, regardless of whether attention had been directed to the relevant body part by the verb; for example, the bruise on Dan's face (Item 5), and the bandage on the actress's hand (Item 18) may have been particularly salient.

These confounding factors associated with the target properties could be overcome by ensuring all items could not be responded to on the basis of experience and knowledge, and by taking baseline measures of property salience to ensure equal variance across lists. Alternatively, participants could be asked to verify the presence of a small dot, or crosshair, on a given area of the event. However, a possible drawback of this method might be that subjects would then search for the dot and results may purely reflect the influence of task knowledge, and not any other top-down or bottom-up influences on event processing.

Some methodological issues could also be resolved if this study was to be replicated. It may be that the length of the video was too long to fully reveal the attention directing effect of the verb in the sentence condition. As discussed earlier, it is clear that the participants were able to gain a detailed representation of the event in the two-second viewing time in both conditions. Perhaps the attention directing effect of the verb would be seen most clearly in eye fixations early in event processing. Therefore, the use of eye tracking devices might be a better measure of the attention directing effect of the verb. As Henderson and Ferreira (2004) point out, "Eye movements provide an unobtrusive, sensitive, real-time behavioural index of ongoing visual and cognitive processing."

The use of a property verification task to measure the effects of language on attention and event processing had some drawbacks. It may be that no significant difference in reaction times to the property verification task were elicited for quartiles 3 and 4 because of strategies used to respond to the question. Perhaps these participants were not relying on their most recent conceptual representation to answer the property verification tasks, but were also considering other factors in responding to these tasks. For example, it may be that these participants attempted to 'check' that that their instinct was correct by considering other events in which the actor had appeared. If they were considering other factors when responding to the property verification task in both conditions, then the effect of the verb on speed of response would be reduced. Again, this would argue for the use of eye-tracking devices in any further investigation into this area, as eye movement patterns would give more direct evidence of the influence of language on attention and event processing.

As discussed earlier, task knowledge may also have affected our results. Participants knew that they would be asked a property verification question and this may have influenced the amount of visual detail they encoded, and the aspects of the scene they focused their attention on. Again, eye tracking devices would be a better measure of the attention-directing effect of the verb because they may not be as affected by task instructions as the property verification measure.

4.6 Conclusion

Our results support our hypothesis that the verbs used in the sentence condition had a different attention-directing effect than the nouns, and that this difference was reflected in accuracy data for property verification tasks, and in reaction time data for the six fastest responders. Language can influence what we pay attention to, and therefore what we process and remember of an event. Furthermore, differences in attention and event processing can be found as a result of changes as small as the inclusion or exclusion of a verb in a preceding sentence, or as a result of the semantic properties of the verbs involved. Our results also point to the complex interaction of top-down and bottom-up influences on attention in event processing. The influence of any one of these factors, including language, is increased or reduced according to its relationship with the other attention-directing factors, internal and external to the event itself.

Word count: 9958

References

Billman, A, Swilley and Krych, M, (2000). Path and manner priming: verb production and event recognition. In: Proceedings of the 22nd Annual Conference of the Cognitive Science Society. Hillsdale, New Jersey: Lawrence Erlbaum Associates

Black, M. (2003) Current theories of event structure : how can they help clinical practice? Paper presented to the International Conference on Verbs: Properties, Processes and Problems. UCL, London, April 2003.

Black, M. & Chiat, S. (in press) Interfaces between cognition, semantics and syntax. In M.J. Ball, M. Perkins, N. Muller, & S. Howard (Eds.) Handbook of Clinical Linguistics. Blackwell Publishing.

Boroditsky, L (2003) Linguistic Relativity Galley Article – 00567 Cambridge Massachusetts, MIT.

Boroditsky, L, Ham, W and Ramscar, M. (2002). What is universal in event perception? Comparing English and Indonesian speakers. Proceedings of the 24th Annual Meeting of the Cognitive Science Society.

Castelhano, M,S, and Henderson, J.M. (2005) Incidental Visual Memory for objects in Scenes. Visual Cognition 12 (6) 1017-1040

Chun, M,M and Wolfe, J,M. (2002). Visual Attention. In Goldstein, E,B (Ed). Blackwell Handbook of Perception. Blackwell Publishing.

Corbetta, M, and Shulman, G.L (2002). Control of goal-directed and stimulus-driven attention in the brain. Nature Reviews Neuroscience 3, 201-215

Findlay, J.M. (2004). Eye scanning and visual search. In J.M. Henderson and F. Ferreira (Eds). The Interface of language, vision, and action: Eye movements and the visual world. New York: Psychology Press

Gennari, S,P, Sloman, S,A, Malt, B,C & Tecumseh Fitch, W. (2002). Motion events in language and cognition. Cognition 83, 49-79

Gentner, D. and Boroditsky, L. (2001). Individuation, relativity, and early word learning. In M. Bowerman & S. Levinson (Eds) Language acquisition and conceptual development. Cambridge: Cambridge University Press.

Gentner, D. and Goldin-Meadow, S (eds) (2003) Language in Mind Massachusetts MIT Press

Henderson, J, M and Ferriera, F. (Eds). (2004). The Interface of Language, Vision and Action: Eye Movements and the Visual World. New York: Psychology Press

Henderson, J. M, and Hollingworth, A. (1999). High-Level scene perception. Annual Review of Psychology 50, 243-271

Lucy, J & Gaskins, S (2001) Grammatical categories and the development of classification preferences: a comparative approach. In Bowerman, M and Levinson S (eds) Language Acquisition and Conceptual Development Cambridge, UK. Cambridge University Press.

Meyer, A,S. (2004). The Use of Eye Tracking in Studies of Sentence Generation. In J.M. Henderson and F. Ferreira (Eds). The Interface of language, vision, and action: Eye movements and the visual world. New York: Psychology Press

Papafragou, A, Massey, C, and Gleitman, L. (2002). Shake, Rattle, 'n' Roll: the representation of motion in language and cognition. Cognition 84, 189-219

Pulvermuller, F, Harle, M & Hummel F. (2001). Walking or Talking?: Behavioral and Neurophysiological Correlates of Action Verb Processing. Brain and Language 78 143-168

Rensink, R, A, O'Regan J, K and Clark J,J. (1997) To see or not to see: The Need for Attention to Perceive Changes in Scenes. Psychological Science 8, 368-373

Robinson, A, and Triesch, J. (2004). Visual Memory for natural scenes: automatic and task dependent components. Journal of Vision. 4, 8, 823a

Sacchett, C.L.M. (2005). An investigation of the relationship between conceptualisation and non-linguistic communication: evidence from drawing production in severe aphasia. Unpublished Ph.D. thesis, University College London

Slobin, D.I (1996) From "thought and language" to "thinking for speaking" in J.J Gumperz & S.L Levinson (eds) Rethinking Linguisitc Relativity Cambridge University Press.

Talmy, L. (2000 Toward a Cognitive Semantics. Volume 1 Concept Structuring Systems. Massachusetts London England: Massachusetts Institute of Technology

Weith, M., Castelhano, M.S, & Henderson, J.M (2003, May) I see what you see: Gaze perception during scene viewing. Presented at the annual meeting of the vision sciences society, Sarasota, Florida.

Whorf, B. (1956). In J.B. Carroll (Ed.), Language, thought and reality: selected writings of Benjamin Lee Whorf. Cambridge, MA: MIT Press

Yarbus, A.L. (1967) Eye Movements and Vision. New York. Plenum

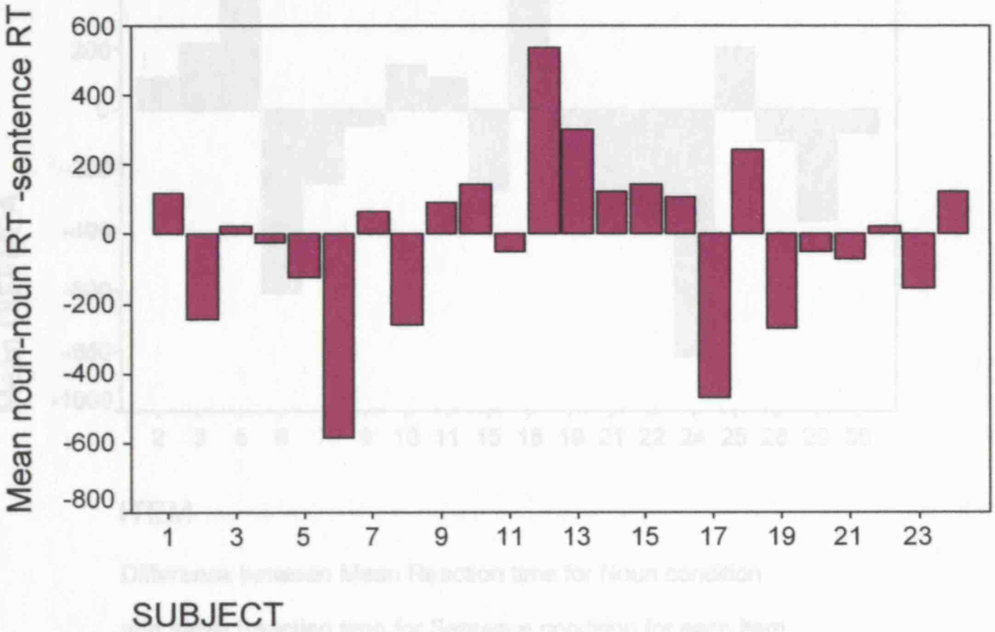
Zaacks, J.M & Tversky, B. (2001) Event Structure in Perception and Conception. Psychological Bulletin. 127 3-21

Appendix 1

A.1 Reaction Time Data

The difference between subjects mean reaction time for items in the noun condition and items in the sentence condition are plotted on the bar chart in Figure 1. Those bars that represent positive figures indicate that for these subjects their mean reaction time in the sentence condition was faster than their mean reaction time for items in the noun condition (as predicted by our hypothesis). Those bars that represent negative numbers indicate that for these subjects their mean reaction time for items in the sentence condition was slower than their mean reaction time for items in the noun condition, contrary to our hypothesis.

Figure 1



Difference between Mean Reaction Time Noun Condition

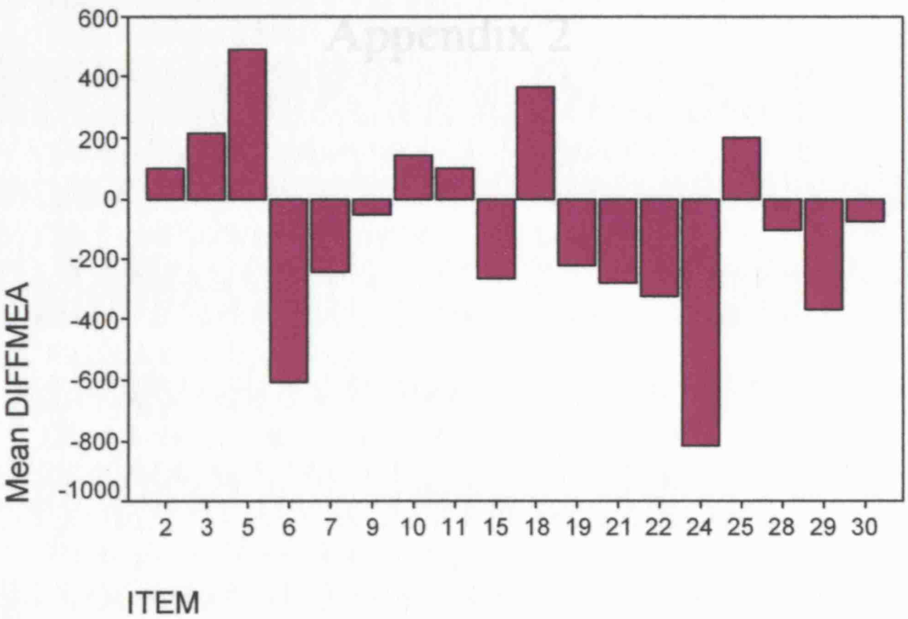
and Mean Reaction Time Sentence Condition for each subject

Results were also analysed for the individual items in the experiment. The differences between Mean Reaction time for each item in Noun condtion and Mean Reaction Time for each item in Sentence condition are displayed in Figure 4. Bars that represent positive numbers indicate for that item reaction times in sentence condition were faster than those in noun condition as predicted by our hypothesis. Bars that represent negative numbers indicate the reverse, contrary to our prediction.

Of the 17 items analysed 7 showed differences that were supportive of our prediction while 10 showed differences between means that were contrary to our prediction.

Not surprisingly, a T-test confirmed no significant difference between the mean reaction times for items recorded in noun and sentence condition. The mean reaction time across items in sentence condition was 1280.43ms, in noun condition 1183.17ms, $t=1.239$ $df=17$ $p>.05$ ($p=.232$).

Figure 2



Difference between Mean Reaction time for Noun condition

and Mean Reaction time for Sentence condition for each item

A.2 Accuracy Data

Number of subjects	Mean accuracy rate to property verifications in sentence condition	Mean accuracy rate to property verifications in the noun condition
24	82%	80%

Table A.q Accuracy data for all subjects across all items (Critical and Filler).

Appendix 2

Item	Critical Item in sentence condition	Critical Item in noun condition	Property verification
1	**The schoolboy was crying in the garden**	**Schoolboy Garden**	**Did he have a mark**
2	Bill was coughing in his chair	Bill Chair	Did he have a moustache
3	**The chef was grinning at the woman**	**Chef Woman**	**Did he have a beard**
4	Sarah was kissing the baby	Sarah Baby	Did she have a nose ring
5	**Dan was smiling at the baby**	**Dan Baby**	**Did he have a bruise**
6	The teacher sneezed by the board	Teacher Board	Did she have glasses
7	**The was licking the lolly**	**Boy Lolly**	**Did he have glasses**
8	The girl was yawning with boredom	Girl Chair	Did she have a lip ring
9	**The baby was sucking a toy**	**Baby Toy**	**Did he have curly hair**
10	Lucy was biting into a sandwich	Lucy Sandwich	Did she have plaits
11	**Sam was hugging his friend**	**Sam Friend**	**Did he have long sleeves**
12	The woman was stroking the cat	Woman Cat	Did she have a ring
13	**The wife was wiping the counter**	**Wife Counter**	**Did she have a ring**
14	Lucy was pinching her brother	Lucy Brother	Did she have a bracelet
15	**The schoolboy was clapping on the sofa**	**Schoolboy Sofa**	**Did he have a watch**
16	Bill was tapping his pipe	Bill Pipe	Did he have a plaster
17	**The warrior was waving a sword**	**Warrior Sword**	**Did he have gloves**
18	The actress was patting her hair	Actress Hair	Did she have a bandage
19	**Paul was elbowing his friend**	**Paul Friend**	**Did he have long sleeves**
20	The robber was grabbing the bag	Robber Bag	Did he have a tattoo
21	**Paul was hoping in the garden**	**Paul Garden**	**Did he have socks**
22	The hiker was wading through the stream	Hiker Stream	Did he have rubber boots
23	**The child was kicking a ball**	**Child Ball**	**Did she have sandals**
24	The soldier was limping across the field	Soldier Field	Did he have boots
25	**Sarah jumped over the wall**	**Sarah Wall**	**Did she have bare feet**
26	Tom was walking up the hill	Tom Hill	Did he have shorts
27	**The athlete was running down the road**	**Athlete Road**	**Did she have a bandage**
28	Sarah was bouncing on the trampoline	Sarah Trampoline	Did she have trousers
29	**The schoolboy was skipping in the garden**	**Schoolboy Garden**	**Did he have shoes**
30	The Scotsman was marching on parade	Scotsman Parade	Did he have a kilt

Table A.r Critical and Filler items in sentence and noun sequence condition

Item Distribution

Critical and filler items were distributed in a fixed position across lists so that results were not skewed by thus ensuring the position of the fillers were not a factor in differing results between participants.

Viewing sequence	Item Type	Viewing Sequence	Item Type
1	**Critical**	**16**	***Filler***
2	Critical	**17**	Critical
3	**Critical**	**18**	**Critical**
4	*Filler*	19	Critical
5	**Critical**	**20**	***Filler***
6	*Filler*	**21**	*Filler*
7	***Filler***	***22***	**Critical**
8	*Filler*	***23***	Critical
9	**Critical**	**24**	**Critical**
10	Critical	**25**	Critical
11	**Critical**	**26**	**Critical**
12	*Filler*	**27**	Critical
13	**Critical**	**28**	***Filler***
14	Critical	**29**	Critical
15	***Filler***	**30**	**Critical**

Table A.s Distribution of critical and filler items for lists 1-3. Lists 4-6 have the same distribution but see items in reverse order.