Exploring Impulsive Activation During Spoken Language Comprehension

by

Cara Kar-Lee Tsang

A thesis submitted in conformity with the requirements
for the degree of Doctor of Philosophy
Department of Psychology
University of Toronto

© Copyright by Cara Kar-Lee Tsang (2012)
Abstract
Exploring Impulsive Activation During Spoken Language Comprehension

Cara Kar-Lee Tsang

Department of Psychology, University of Toronto

Doctor of Philosophy, 2012

A language comprehension mechanism that immediately starts processing language as it is encountered is typically thought of as one that speeds and facilitates spoken language comprehension. However, there exist cases where the earliest parts of a word or phrase encode information that is somewhat at odds with the remainder of the word or phrase in full. Examples of these "potentially misleading" cases include compound words where the initial subpart of the compound belongs to a different syntactic category than the entire compound (e.g., "popcorn", "greyhound"), or noun phrases where the initial element of the phrase signals perceptual properties possessed by the referent of the noun phrase (e.g., some Chinese Cantonese classifier-noun phrases).

Using a visual-world methodology, this dissertation explores the kinds of unintended or "impulsive" activation that are triggered when listeners encounter such cases, as well as how syntactic and contextual cues can constraining this impulsive activation. Experiment 1 examines whether hearing compound subparts (e.g., "pop-" in "popcorn") activates conceptual associates across syntactic categories, and Experiment 2 examines whether this activation is moderated by listeners' expectations about the syntactic structure of the sentences they encounter. Experiments 3 and 4 investigate the processing of compounds whose initial subparts correspond to colour terms (e.g., "greyhound"). Experiment 3 explores whether these colour subparts trigger the activation of phrasal-level descriptions of non-target objects in the visual display, and whether this activation is influenced by the presence/absence of motivation to use colour descriptions when naming screen objects. Experiment 4 further explores whether a perceptual
mismatch between a target object and the colour term in its name increases this impulsive activation. Finally, Experiment 5 investigates whether listeners use the perceptual information encoded in pre-nominal classifiers in Cantonese Chinese to guide their consideration of referential candidates whose perceptual properties do or do not match the classifier semantics.

The findings from this dissertation point to the highly interactive nature of spoken language comprehension, suggesting that the kinds of impulsive activation under current discussion are rampant and automatic, but can also be suppressed to varying degrees by the syntactic, semantic, and contextual cues available to the listener.
Acknowledgements

This work would not have been possible without the guidance and support of my supervisor, Craig Chambers. His patience, generosity of spirit, and gentle humour helped pull me through the many ups and downs of my graduate career, and I could not have asked for a better mentor.

I am also very grateful to my committee members Eyal Reingold and Elizabeth Johnson. Their wisdom and advice throughout the research and writing process have been invaluable, and my work is better for their input. My gratitude also goes to Whitney Tabor and George Cree: they made the defense process as painless as possible, and I am fortunate to have their insightful feedback on this dissertation.

Thank you to Daphna Heller for acting as my academic big sister – her guidance helped me navigate the sometimes-bewildering rules and conventions of academia, and our many interesting conversations shared over many plates of sushi never failed to remind me of why I fell in love with psycholinguistics in the first place.

My thanks also go out to Aashita Pattini for her assistance with data collection and data coding. Her sunny disposition made all those lab hours go by much more quickly.

Finally, I know I could not have made it through this process and remained sane without the help of my family and the friends and colleagues I've met over the past 7 years. My family has been unfailingly supportive, even if they might not fully understand what it was, exactly, that I've been doing all this time. Sean Kersey had the good "fortune" to meet me during the final throes of my PhD, and has offered me endless kindness, consideration, and patience ever since – thank you for easing my way, in so many ways. I have been incredibly lucky to have made so many good friends in graduate school, and Christine Burton, Karen Campbell, Doug and Laura Garrett, Jessica Ellis, Kate Dupuis, Marnie Hirshhorn, Mindaugas Mozuraitis, Signy Sheldon, Stephanie Stalinski, Sarah Tinsley, Marieke van Heugten, and Dominique Vuvan deserve special mention for being always at the ready with a sympathetic
ear, encouraging words, and sometimes, a much-needed glass of wine. Thank you all for travelling this road with me and for making sure I came out the other side alive!

This work was supported by Postgraduate Scholarships from the National Sciences and Engineering Research Council of Canada, and Ontario Graduate Scholarships.
Table of Contents

Abstract ........................................................................................................................................... ii
Acknowledgements ......................................................................................................................... iv
Table of Contents ............................................................................................................................... vi
List of Tables ................................................................................................................................. x
List of Figures ............................................................................................................................... xi
List of Appendices ........................................................................................................................... xiii
Chapter 1: Background and Overview .......................................................................................... 1
Introduction ....................................................................................................................................... 1
Immediacy in Processing the Form of Words ................................................................................ 3
Immediacy in Processing the Meaning of Words ......................................................................... 6
Immediacy in Conceptual Activation ......................................................................................... 8
Immediacy in Referential Interpretation .................................................................................... 11
Interactions Among Form and Meaning in the Processing of Spoken Words .................... 13
Consequences of Immediacy in Processing ............................................................................ 15
Dissertation Overview .................................................................................................................... 17
Chapter 2: Impulsive Conceptual Activation Across Lexical Categories ............................... 20
Experiment 1 .................................................................................................................................... 28
Method ............................................................................................................................................. 28
Participants ..................................................................................................................................... 28
Materials ......................................................................................................................................... 28
Procedure ......................................................................................................................................... 31
Results ............................................................................................................................................. 33
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consideration of Target Object Relative to the Competitor</td>
<td>79</td>
</tr>
<tr>
<td>Consideration of Competitor Object Relative to the Unrelated Distractor</td>
<td>80</td>
</tr>
<tr>
<td>Discussion</td>
<td>80</td>
</tr>
<tr>
<td>Experiment 4</td>
<td>83</td>
</tr>
<tr>
<td>Method</td>
<td>83</td>
</tr>
<tr>
<td>Participants</td>
<td>84</td>
</tr>
<tr>
<td>Materials and Procedure</td>
<td>84</td>
</tr>
<tr>
<td>Results</td>
<td>84</td>
</tr>
<tr>
<td>Consideration of Target Object Relative to the Competitor</td>
<td>88</td>
</tr>
<tr>
<td>Consideration of Competitor Object Relative to the Unrelated Distractor</td>
<td>89</td>
</tr>
<tr>
<td>Discussion</td>
<td>90</td>
</tr>
<tr>
<td>Summary</td>
<td>92</td>
</tr>
<tr>
<td>Chapter 4: Impulsive Processing in Referential Interpretation</td>
<td>96</td>
</tr>
<tr>
<td>Noun Classifiers</td>
<td>98</td>
</tr>
<tr>
<td>Experiment 5</td>
<td>106</td>
</tr>
<tr>
<td>Method</td>
<td>107</td>
</tr>
<tr>
<td>Participants</td>
<td>107</td>
</tr>
<tr>
<td>Materials</td>
<td>108</td>
</tr>
<tr>
<td>Procedure</td>
<td>110</td>
</tr>
<tr>
<td>Results</td>
<td>112</td>
</tr>
<tr>
<td>Consideration of Target Object Relative to the Competitor</td>
<td>113</td>
</tr>
<tr>
<td>Consideration of Competitor Object Relative to the Unrelated Distractor</td>
<td>113</td>
</tr>
<tr>
<td>Summary</td>
<td>115</td>
</tr>
</tbody>
</table>
Classifier Semantics and Referential Mapping of Atypical Cases .................................................. 117

The Uptake of Classifier-Encoded Information ................................................................................. 119

Chapter 5: Conclusions ......................................................................................................................... 120

Can Impulsive Activation be Observed Across Lexical Categories? ................................................. 120

Can Phrasal Descriptions be Activated by Word Subparts? ................................................................. 123

Does Potentially Misleading Information Always Trigger Impulsive Activation? ......................... 125

Why do Cantonese Classifiers Not Entail Misleading Interpretations? .......................................... 126

Can Impulsive Activation be Constrained? ....................................................................................... 129

How Might a Rational Comprehension System be Characterized? ............................................... 131

References ............................................................................................................................................ 134

Appendices ........................................................................................................................................ 141
List of Tables

Table 4.1. Example of common Cantonese sortal classifiers and associated nouns..... 100
List of Figures

Figure 2.1. Example of visual display in Experiments 1 and 2, showing the competitors in the two experimental conditions. The target was the swimsuit. ...................................................... 29

Figure 2.2. Likelihood of fixating screen objects across the four experimental conditions in Experiment 1. Probabilities are depicted on the logit scale. The first vertical line on each graph marks the average end of the Verb region, the second line marks the average end of the Post-Verb region. ................................................................................................................. 34

Figure 2.3. Competitor advantage scores across the four experimental conditions in Experiment 1. ................................................................................................................................. 37

Figure 2.4. Likelihood of fixating screen objects across the four experimental conditions in Experiment 2. Probabilities are depicted on the logit scale. The first vertical line on each graph marks the average end of the Verb region, the second line marks the average end of the Post-Verb region. ................................................................................................................. 46

Figure 3.1. Potential kinds of activation resulting from encountering colour compounds such as "greyhound". ..................................................................................................................... 57

Figure 3.2. Example display of a trial in the pre-test. ......................................................... 69

Figure 3.3. Example of a critical trial display in Experiment 3. The target in this trial was the greyhound. The competitor was the grey spatula in the contrastive- and noncontrastive-competitor conditions, and the stapler in the no-competition condition. In the contrastive-competitor condition, one of the screen objects (e.g., the helicopter) was replaced with an item of the same kind, but different colour, as the competitor (e.g., the red spatula) in order to create a contrast set. ..................................................................................................................... 72

Figure 3.4. Likelihood of fixating screen objects across the three experimental conditions in Experiment 3. Probabilities are depicted on the logit scale. The first vertical line on each graph
marks the average end of the Adjective region, the second vertical line marks the average end of the Post-Adjective region.

Figure 3.5. Likelihood of fixating screen objects across the three experimental conditions in Experiment 4. Probabilities are depicted on the logit scale. The first vertical line on each graph marks the average end of the Adjective region, the second vertical line marks the average end of the Post-Adjective region.

Figure 4.1. Example display from Experiment 5. The target was always a G+S+ item (e.g., the scarf for the classifier "tiu[4]"), and the competitor was one of four objects: (a) a G+S+ item (e.g., the jump rope); (b) a G+S- item (e.g., the key); (c) a G-S+ item (e.g., the flag); or (d) an unrelated item (e.g., the strawberry). The teacup and present in this display were additional unrelated objects.

Figure 4.2. Likelihood of fixating screen objects across the four experimental conditions in Experiment 5. Probabilities are depicted on the logit scale. The vertical line on each graph marks the average end of the classifier in the classifier-noun phrase.
List of Appendices

Appendix 1. Critical trial object arrays in Experiments 1 and 2. Screen objects were presented as single objects in Experiment 1, and three duplicated objects (e.g., three identical ring boxes) in Experiment 2. ................................................................. 141

Appendix 2. Critical trial object arrays in Experiments 3 and 4. Main colour of the target in Experiment 3 corresponded to the colour term in the target name. Colour in brackets indicates the main colour of the target object in Experiment 4. ........................................ 142

Appendix 3. Critical trial object arrays for Experiment 5. ................................................. 143

Appendix 4. Images used in critical trials for Experiment 5. ............................................. 144
Chapter 1
Background and Overview

Introduction

Though seemingly simple from the perspective of our conscious experience, the real time comprehension of language requires the access and integration of many different streams of information within a comparatively limited period of time. For example, in the case of spoken sentences, speech sounds in an unfolding utterance must be identified and mapped onto items in the mental lexicon, words must be combined according to the syntactic rules of the language, and the broader meaning of the sentence must be computed by enriching and integrating the linguistic content with pragmatic and discourse-level information.

Traditionally, the core lexical and syntactic processes involved in language comprehension were considered to operate in a way akin to a sophisticated "catch-up game" (as described by Tanenhaus, 2004). That is, comprehension was thought to proceed in a staggered manner whereby incrementally-encountered information was held in a type of working-memory cache, with only minimal processing occurring before an eventual point (the ends of words, phrases, sentences, etc.) at which "lookup" or "assembly" processes were engaged. This view was plausible in view of the fact that partial information based on the beginning of a word, phrase, or sentence is typically ambiguous or indeterminate as to its identity or meaning. As such, it would be reasonable to delay processing until a later time point when more determinate information was available.

However, contemporary perspectives instead adopt the view that the mechanisms underlying sentence comprehension are characterized by considerable immediacy: a range of processes are initiated from the earliest moments of encountering the input, and operate
continuously as additional information is encountered. This view is informed by a large number of studies exploring a variety of phenomena and drawing on a wide range of experimental techniques (for overviews, see Dahan & Magnuson, 2006; Pickering, 1999; Rayner & Clifton, 2009). However, certain intuitive illustrations of immediacy also exist in daily life. For example, listeners sometimes feel consciously surprised at how a speaker's utterance finishes, based on what had been said in the earlier part of the utterance. Upon hearing "We have nothing to fear but...", a listener might be surprised by a speaker continuing with "... the consequences." The surprise experienced in these instances could occur only if listeners were processing word-by-word at a comparatively deep level, generating expectations about the structure and meaning of a string of upcoming sentence units on the basis of only partial utterance information. A less obvious but no less important type of immediacy characterizes the processing of an individual spoken word, whose form and meaning are often accessed before the end of the word is reached. Although not as apparent to conscious awareness, spoken word recognition arguably provides a more compelling illustration of immediacy due to the time interval involved. For instance, the average spoken monosyllabic English noun has a duration of only approximately 350 to 700 milliseconds (ms), and as such the relevant processes must occur within a much shorter period of time compared to phrasal- or clausal- level processing.

Developing a precise account of what happens "immediately" as an element of language is encountered is one of the central goals in contemporary work on real-time comprehension, and also serves as the major motivation for the studies described in this dissertation. However, a distinguishing feature of the work reported here is a focus on linguistic examples where the early and rapid processing that occurs at the beginning of a word or phrase may be rather unhelpful for the processing of the constituent as a whole. For example, processing the earliest part of the
spoken word "flyswatter" might lead the listener to expect fly to be a verb or the beginning of a modified noun phrase (e.g., "flying squirrel") to a much greater extent than the beginning of an individual noun. The importance of these examples for current theorizing stems from the fact that they provide an opportunity to explore the "downside" of immediacy in processing – namely the access or activation of information that is ultimately irrelevant. However, studies on various components of language comprehension have shown that the impact of certain kinds of misleading information can often be reduced through the use of linguistic or contextual information that is available at the point where the relevant unit is encountered. If these effects are found in all circumstances, we might conclude that language comprehension is driven by a highly "rational" system that optimizes all available information to both favour the activation of appropriate information, and to prevent the activation of irrelevant or unhelpful elements. It is possible however that this type of rational system does not accurately characterize the processes underlying language comprehension. Accordingly, a second goal of this dissertation is to examine how contextual information might suppress or enhance the potentially troublesome consequences of immediacy in these types of cases.

In the remainder of this chapter, I provide an overview of the relevant literature on the processing of word-level units, and outline the series of studies that will be presented in the following chapters.

**Immediacy in Processing the Form of Words**

A word is a pairing of a perceptual form and semantic meaning(s), and immediacy in processing can accordingly be considered in light of both form and meaning. Because the focus of the current thesis involves spoken language, the type of perceptual form considered in most of the work reviewed below is a sound pattern (i.e., a sequence of phonemes).
Immediacy in the identification of spoken word forms is reflected in the incremental evaluation of each unfolding sound segment in the phonetic signal against the form (sound pattern) of stored lexical alternatives. This process is most commonly studied by focusing on a point within a word where the available information is compatible with more than one possibility. For example, as the word "candle" unfolds, there is a point at which the initial sounds (e.g., "can-") are consistent with the sound pattern of lexical entries such as "cantaloupe", "candy", and "candid". As additional sounds are heard, alternatives that do not match these sounds will no longer be viable candidates. In this way, word recognition is often modelled as a competition process within an activation framework, whereby alternatives "compete" for activation dynamically as the signal unfolds. In addition to the degree of match with speech input, other factors determining the relative activation level of a lexical candidate include frequency of use, as well as the number and frequency of use of similar-sounding alternatives ("phonological neighbours").

One of the best-known early studies that provided evidence for the immediate uptake of speech information during spoken word recognition is a study by Grosjean (1980). In this "gating" experiment, participants heard repeated presentations of recorded spoken target words that had been truncated prior to their endpoints. Starting with an initial presentation of only the first 30 ms of the spoken target, participants heard successively longer samples of the target until they could accurately identify the word. Participants wrote down their guesses about the identity of the word after each presentation. Grosjean found that listeners did not need to hear the entire word before they correctly guessed the identity of the target word; they merely needed to have heard enough information to distinguish it from its phonological neighbours. Prior to this "uniqueness point", the collective guesses of the full set of participants were understood to reflect
the consideration of lexical alternatives that were consistent with the available speech information. As participants heard subsequently longer portions of the word stimulus, the number of lexical alternatives reflected in the collective responses diminished accordingly.

One criticism of Grosjean's (1980) findings, however, is that a gating task may not accurately reflect typical language processing. In normal conversation, listeners are rarely required to identify isolated words from truncated speech, and even more rarely do they hear consecutively longer segments of the same truncated word. More direct evidence for Grosjean's proposal about the immediate and continuous uptake of information during speech recognition comes from studies using more naturalistic tasks such as those using eye monitoring methodologies. For example, Allopenna, Magnuson, and Tanenhaus (1998) report a study in which participants viewed computer displays containing multiple clip-art objects. On critical trials, the display contained a target object (e.g., a beaker), an object whose name started with the same speech sounds as the target's name (e.g., a beetle; a phonological "cohort" competitor), an object whose name rhymed with the target's name but did not begin with the same speech sounds (e.g., a speaker; a rhyme competitor), and a phonologically unrelated object (e.g., a stroller; the unrelated distractor). As their eye movements were recorded, participants carried out a spoken instruction to pick up (i.e., click on) the target object and move it to a specified location on the grid.

Allopenna et al. found that listeners started launching eye movements to the target object approximately 200 ms after the onset of the target noun in the spoken instruction, demonstrating the rapid uptake of speech information. Further, in addition to measures related to the target object, eye movements to competitor objects provided important insights into the activation level of lexical alternatives. To illustrate, as the first part of the target word unfolded, listeners made
more eye movements to the cohort competitor (e.g., the beetle) than to the unrelated distractor, with visual consideration of the cohort competitor being comparable to that of the target object. This is due to the temporary ambiguity whereby the initial speech sounds in the target and the cohort competitor names constituted equally good candidates for the unfolding signal. Once disambiguating information was encountered in the speech stream (e.g., the /k/ in "beaker"), the proportion of fixations to the cohort competitor decreased, while fixations to the target object continued to rise. Given that the mean target noun duration was 375 ms, and taking into account that an eye movement requires about 200 ms to plan and execute in this type of paradigm (Matin, Shao, & Boff, 1993), these results provide strong evidence that phonetic information is taken up and matched against lexical entries (thereby eliminating incorrect candidates) as soon as it becomes available. Further, from a methodological perspective, these findings show that listeners’ eye movements are closely time-locked to speech comprehension, and that the "visual-world paradigm" is one that is capable of capturing the immediate and incremental use of information contained in speech.

Interestingly, Allopenna et al. found that the rhyme competitor (e.g., the speaker) also received referential consideration, even though the input from the start of the target word (e.g., /b/) was inconsistent with that lexical alternative. This effect demonstrates that later subparts of an unfolding signal can still trigger lexical activation despite the fact that the listener had earlier information indicating a mismatch between the onset sounds of the candidate and target. The later information was therefore used to independently propose lexical alternatives to some degree, regardless of the candidates generated from earlier information.

**Immediacy in Processing the Meaning of Words**
The meaning of words is often considered on two different levels: conceptual meaning and referential meaning. The conceptual meaning of a word reflects the abstract knowledge contained as part of our mental lexicon, built up through a lifetime of experience. For example, the concept of CAT contains information about a cat’s generic physical appearance, its behaviours, the sounds it can emit, the types of food it will eat, and so forth. Although some additional shades of meaning can become integrated into the concept (e.g., learning that referring to someone as "cat-like" connotes a sense of sleekness or slyness), conceptual representations are largely stable over time and over different contexts. Knowing that some specific cats are hairless or encountering a cat with a missing leg, for instance, does not change our general conceptualization of cats as furry four-legged animals.

The referential "meaning" of a word, however, reflects how the word is being used by a speaker in an expression that denotes a specific object, property, or event/state in the domain of discourse, and is often more fluid or more complex than the conceptual meaning. As alluded to in the preceding example, the term "cat" could be used by a speaker to refer to a three-legged hairless cat, despite the mismatch between the exemplar and the stored abstract representation of a cat. Pronouns provide another particularly useful illustration of this difference. For example, the word "she" has a fairly limited conceptual meaning: it indicates only a single female third-person entity. However, when used by speakers in different contexts, the meaning of "she" can take on very different semantic values, and might refer to a complete stranger, a dear friend, or a bitter enemy. Other cases of reference illustrate how the use of a particular word form depends on the situational context at hand. For instance, a speaker who states "There's a lecture today about the prevalence of big zoo animals in captivity" probably intends "big" to refer to something on the scale of a lion or an elephant. A speaker who states "There's a lecture today
about the prevalence of big hair in the 60's", on the other hand, is probably not referring to a hairstyle that is equivalent in size to a lion or elephant. Still other cases show that the referential and conceptual meanings of a word can in some cases appear to be inconsistent with each other, and yet still allow the speaker's intended meaning to be understood with relative ease. Take, for example, a phrase like "I'm parked over by the school". Although the conceptual meaning of "I" (the speaker) is inconsistent with the action expressed in the sentence (i.e., a person cannot typically be parked), few listeners would have trouble resolving the referential meaning of "I" to be MY VEHICLE.

**Immediacy in Conceptual Activation.** The processing of conceptual meaning is typically measured in two ways: by examining the activation of concepts associated with the word in question (e.g., the activation of NURSE when one encounters "doctor"), and through examining the specific case of lexically ambiguous words such as "bug" or "bark" where the presence of alternative meanings is of interest. Conceptual associates and alternative meanings are stored and stable forms of information associated with words, and are useful for examining conceptual activation because they allow one to disentangle conceptual effects from those driven by reference resolution. Because these effects are observed via the consideration of concepts that are not actually required for resolving reference, it can be inferred that the activation of conceptual associates or alternative meanings stems from the listener accessing the conceptual meaning of a word, prior to any referential mapping processes that may take place.

In the case of ambiguous words, a classic finding comes from lexical decision studies and suggests that all alternate meanings of homophones receive activation in the early moments of processing, even when only one of those meanings is contextually supported by information encountered beforehand. For example, Swinney (1979) presented homophones like "bug" in
auditory sentence contexts that biased interpretation towards one meaning of the homophone (e.g., an insect). Despite the bias, lexical decisions to visually-presented probe words appearing as the homophone was heard showed speeded responses relative to control words both for probes associated with the supported meaning (e.g., ANT) and for probes associated to the non-supported meaning of the homophone (e.g., SPY). Converging evidence for this finding comes from Tanenhaus, Leiman, and Seidenberg (1979), who found activation for all meanings of a homophone even in the presence of strong syntactic cues that made only one of those meanings appropriate. Examining the activation of simple semantic associates reveals similar effects. For example, reading a word prime (e.g., "doctor") resulted in faster lexical decisions to a semantically-related target word (e.g., "nurse") compared to an unrelated control (e.g., "chair", Meyer & Schvaneveldt, 1971). These findings support the idea that accessing a word also results in the automatic and incidental activation of associated conceptual information, which includes any alternative lexical meanings the word might have. Accordingly, these results cast doubt on the idea the processing occurs via a fully "rational" system whereby already-encountered information (e.g., semantic and syntactic biases) acts to suppress activation of inappropriate alternatives.

However, it may be argued that lexical decision tasks, which interrupt the flow of input and require comprehenders to make explicit judgments about language, do not reflect a naturalistic way of processing. Consequently, it may be the case that these paradigms fail to detect the moderating effects of context that accompany normal language comprehension behaviour. More recent research has used naturalistic methods to explore the extent to which this kind of activation is truly insensitive to context during the initial moments of processing. For instance, Mirman, Magnuson, Strauss, and Dixon (2008) examined the activation of alternate
meanings of balanced homophones, and found that this activation was decreased when only one meaning of the homophone was contextually supported by the global, non-linguistic context. In this case, the global context was construed as whether the alternate meaning of the homophone target could be easily depicted as a concrete object within the visual-world paradigm. Huettig and Altmann (2004) report a related finding such that consideration of the subordinate meaning of a homophone was increased if that meaning had been biased by a prior sentential context.

Unlike the classic work examining homophone processing using lexical decision tasks, however, findings from the kinds of visual world experiments described above demonstrate that the influence of context can be observed from the earliest moments of processing (e.g., as the target word is still unfolding), and not only after the sound pattern of the ambiguous word has been matched to its corresponding forms in the mental lexicon. Critically however, in both Mirman et al. (2008) and Huettig and Altmann (2004), although context did reduce the consideration of the unintended meaning of a homophone to some extent, activation of that homophone was nonetheless still observed. Thus, although the results from these visual world studies indicate that top-down contextual knowledge does interact with bottom-up effects to shape semantic access as the input unfolds, these findings are still broadly consistent with many other findings suggesting that it is very difficult to fully suppress alternative meanings with context, unless that meaning is also comparatively infrequent to begin with. The more typical result with regard to homophone comprehension, at least during the early moments of processing, is to see attenuation without complete suppression of the unintended homophone meaning.

Effects of conceptual activation as measured by the real-time consideration of semantic associates can also be seen using lexically unambiguous stimuli. For example, using a visual world paradigm, Huettig and Altmann (2005) showed that the activation of conceptual associates
occurs alongside the activation of a target word. Participants viewed a computer display containing four objects: a target (e.g., a piano), a categorically-related competitor (e.g., a trumpet), and two distractors that were phonologically and semantically unrelated to the target (e.g., a goat and a hammer). As participants heard the target word unfolding, they made more transient fixations to the conceptually-related competitor than to the distractors. Crucially, the proportion of eye movements to both the target and the conceptual associate competitor began to increase approximately 200 ms after the onset of the target word. In other words, the activation of conceptual associates occurred during the earliest moments of spoken word recognition, and was not, for example, something that occurred only after processing the full word form for the target word. Further, it is quite striking that the activation of conceptual associates was so easily detectable in this type of paradigm, given that the linking of a word to a specific token in the visual display is seemingly much more related to referential meaning. One might, for example, therefore imagine that these "internal" conceptual processes would not be evident in the canvassing of external objects in the visual world. The ability of these conceptual-level processes to be reflected in word-to-world mappings highlights their pervasive and automatic nature, and also mirrors myriad classic findings using techniques such as lexical decision tasks.

**Immediacy in Referential Interpretation.** Effects of immediacy can also be seen during the real-time mapping of referential expressions to their corresponding referents. These effects are distinguished from those discussed above in that reference to entities is typically understood as a function of phrases rather than individual words. Immediacy in reference is therefore most often investigated by observing how various phrasal elements (e.g., articles and modifiers, in combination with nouns) are successively used to generate and refine implicit hypotheses about the identity of a referent for the entire phrase. For example, using instructions such as the
"Touch the starred yellow square", Eberhard et al. (1995) showed that listeners integrated semantic information from each successive word heard during the unfolding sentence to narrow their consideration of relevant referential objects. In addition to the intended target, displays in this study contained competitor objects whose features were manipulated to allow for comparatively early versus later individuation ("disambiguation") of the target object. In the early-disambiguation case, the target was the only object that fulfilled the property denoted by the first descriptor (e.g., the only starred object). In the mid-disambiguation condition, the target was the only object that fulfilled both the first and second descriptors (e.g., the only object that was both starred and yellow). In the late-disambiguation condition, the target could only be identified after hearing all three descriptors (e.g., there were two starred yellow objects in the scene, but only one of them was a square). Eberhard et al. found that listeners were faster to fixate the target object in the early-disambiguation than in the mid-disambiguation condition. Fixations to the target were slowest in the late-disambiguation condition.

These results demonstrate that listeners can interpret referring phrases immediately and incrementally, and rapidly combine each phrasal element with subsequent information to narrow the set of referential alternatives under consideration to a unique candidate (consistent with the use of the definite article "the"). The results also highlight how referential interpretation is not simply a by-product of activating conceptual information from individual words but instead involves more complex form of combinatorial processing. For example, the referential alternatives considered upon hearing "yellow" in "the starred yellow square" were not simply the set of yellow objects in the display. Rather, consideration was only given to those objects that were both starred and yellow, as only those objects would be compatible with the information at this point.
Interactions Among Form and Meaning in the Processing of Spoken Words.

The "visual world" studies described in the above sections have been used successfully to measure: (i) conceptual activation effects found in classic experimental tasks such as lexical decision, (ii) the immediate uptake of speech information to map the signal against possible word forms, and (iii) the incremental interpretation of words in referring expressions. Recently, researchers have begun to use this technique to explore interactions among these components of the interpretation process. More specifically, this research examines conceptual activation and referential processing at a more fine-grained level in the temporal domain, namely within spoken words as they unfold in time. Thus the question is whether, when, and how parts of words might drive semantic processes as the speech signal is encountered.

Using an experimental design similar to Huettig and Altmann (2005), Yee and Sedivy (2006) corroborated Huettig and Altmann's basic finding that the activation of semantic associates occurs during the comprehension of a target word. Furthermore, the authors revealed an interesting chain of events whereby subparts of target words activate alternatives in the mental lexicon, resulting in activation of concepts associated with these alternatives. On critical trials, participants viewed a computer display containing four objects. These included a target (e.g., some logs) that shared initial speech sounds with specific lexical alternatives that were not displayed (e.g., a lock). In addition to the target, there was a competitor that was semantically associated to one of those undisplayed lexical alternatives (e.g., a key, an associate of "lock"), as well as two distractors that were phonologically and semantically unrelated to the target (e.g., a deer and an apple). As participants heard the target word unfolding, they made more transient fixations to the competitor than to the distractors. Encountering the target word "logs" apparently activated its phonological neighbour LOCK, and this activation spread to the neighbour's
semantic associate KEY. Crucially, the proportion of eye movements to the semantic competitor began to increase approximately 200 ms after the onset of the target word, again demonstrating that the effect is not one that only occurs after lexical access is complete. Interestingly, this kind of phonosemantic activation is found in children as young as 5 years of age (Huang & Snedeker, 2010). However, it is relevant to note that the activation of semantic associates in these experiments could still occur under a strictly rational system of processing, because at the point that associate activation was observed, there was nothing yet in the signal that would have ruled out "lock" as the actual target word being heard, for example.

These findings carry important implications for morphologically-complex lexical items such as compound words, whose processing has been most extensively examined using written stimuli. The most straightforward kinds of compounds are composed of two or more clearly identifiable words, such as "blackboard" or "flyswatter". The primary question in this area of research has revolved around the extent to which the sub-components of the compound are processed individually (i.e., "decomposed") or whether the processing of a compound word is essentially the same as the processing of a morphologically simple word (e.g., Libben, 1994; Seidenberg & Gonnerman, 2000; Taft & Forster, 1975). The evidence presented above, however, suggests that in the case of spoken language, the activation of subconstituents at the level of form and meaning (at least for a moment in time) may be a general side-effect of word recognition processes, and is not specific to compounds. Nevertheless, compounds can provide a particularly useful test case to explore this kind of incidental activation and how this activation might be guided by various types of constraints. This is because, despite the presence of some subtle subphonemic cues that might help indicate the subpart’s status as an embedded element within a longer compound, the initial subpart of a compound nonetheless yields a fleeting but broadly
"complete" match with a lexical form that is different than the one corresponding to the full word that is eventually encountered. The bulk of the work in this dissertation will explore the scope of the kinds of incidental activation observed in Yee and Sedivy (2006), and the extent to which activation occurs even when it would ideally be suppressed under a rational processing system.

**Consequences of Immediacy in Processing**

While there are certain advantages to processing immediacy, such a system clearly comes at some cost. If the activation of form and various levels of meaning occur from the earliest moments of processing, the listener will in many cases not yet have encountered enough information to constrain activation to only those lexical forms, concepts or referents intended by the speaker. If the early information is ultimately irrelevant in some way, it may in turn hinder or delay processing. Take for example, the case of compound words briefly described in the previous section. The nature of some compound words is such that the first embedded word may not even belong to the same syntactic category as the entire compound, as in the case of compounds such as "flyswatter" or "blackboard". In the former case, where the initial element of a compound noun also has a verb meaning, hearing the first part of the compound (e.g., "fly-") may initially lead the listener to activate concepts associated with that verb meaning (e.g., AIRPLANE, KITE, etc.) which are unrelated to the intended referent concept. In the latter scenario, where the initial element of the compound overlaps with the form of an adjective (e.g., "black-"), incrementally processing that element could plausibly result in a processing "boggle" if the object in question does not have the property described by the adjective (e.g., in the case of a blackboard with a green surface).

Of course, we must remember that words do not occur in a vacuum. Based on what has already occurred in a sentence and within the discourse and broader context, listeners create
expectations about the form and meaning of utterance elements they are likely to encounter. Indeed, research has shown that contextual information can streamline processing at the level of reference (e.g., Chambers, Tanenhaus, Eberhard, Filip, & Carlson, 2002) and, as described above, can modulate the activation of alternative meanings (e.g., Tanenhaus, Leiman, & Seidenberg, 1979). An important question, however, is whether and how these effects might extend to the specific cases of interest here, where the initial portion of an expression provides temporarily misleading information. Taking the cue from recent studies of syntactic processing, it may be the case that misleading incidental activation is not entirely constrained by available informational cues. Tabor, Galantucci, and Richardson (2004) demonstrated that readers momentarily posit syntactic relationships within a sentence that are globally inconsistent with syntactic information encountered earlier. For instance, readers were slower to process reduced-relative sentences with like "The coach smiled at the player tossed a Frisbee" than one like "The coach smiled at the player throw a Frisbee", even though they are syntactically equivalent. This delay is attributed to the local coherence of "... the player tossed...", which could be parsed as the beginning of an active clause in which the player is the agent of the tossing action. This is contrasted with "...the player thrown...", which is much more likely to be only (and correctly) parsed as the beginning of a reduced relative clause because the past tense form of the verb "throw" cannot be interpreted as a direct action of an agent (in this case "the player"). Importantly, regardless of the presence of local coherence at "... the player tossed...", syntactic information provided earlier in the sentence (i.e., "The coach smiled at...") indicated that "the player" must be the patient, and this should in principle have ruled out the possibility that "the player tossed" was the beginning of an active clause. Despite this earlier syntactic information, however, effects of local coherence were still observed. (See Konieczny, Müller, Hachmann,
Schwarzkopf & Walter, 2009 and Kukona, Fang, Aicher, Chen, & Magnuson, 2011 for related findings.) If this phenomenon is comparatively general and captures the nature of the incidental activation of information within words and phrases, we might expect to find converging evidence when examining processing that is triggered when listeners encounter the kinds of compound nouns described above, whose initial elements could be interpreted as belonging to a different syntactic category than the entire compound itself.

**Dissertation Overview**

The series of experiments in this dissertation will investigate the nature of incidental activation during spoken language comprehension. The specific focus involves what I will refer to as "impulsive activation". Impulsive activation in this context is used to refer to cases where the information being activated is ultimately irrelevant to processing in some way. More specifically, I define the term as follows:

"activation triggered by the earliest elements of an unfolding unit that is ultimately detrimental for the comprehension of the unit as a whole"

This term helps differentiate the cases of interest from circumstances in which bottom-up activation is neutral in terms of its consequences. Impulsive activation, in contrast, involves only those instances where information that is incidentally and unintentionally accessed might hinder the identification of current or upcoming elements. Three such cases are considered, each focusing on a different type of processing where impulsive activation can occur at a comparatively fine temporal grain within an unfolding word or phrase. The first case examines the comprehension of compound nouns like "flyswatter" or "swimsuit", where the first part of the compound can match a word belonging to a different syntactic category (i.e., a verb) than that of the whole compound. These cases will allow for testing the extent to which encountering the
"embedded" verb triggers impulsive activation of verb-related semantic and syntactic knowledge. The second case will make use of compound words with a colour adjective embedded in the initial part of the word (e.g., "greyhound", "blackboard"). The question in this case is whether the recognition of the target noun form (e.g., "greyhound") impulsively triggers the parallel consideration of an alternative that bears the relevant colour but does not share the sound pattern (e.g., a spatula, which is grey), or which both bears the relevant colour and whose likely description would also partially overlap with the target name (e.g., again a grey spatula, but now accompanied by a red spatula, thereby motivating the complex description "grey spatula"). Finally, the third case will explore how pre-nominal elements within a noun phrase might impulsively lead listeners to make incorrect referential expectations in cases where there is a conflict between the semantic information conveyed by the pre-nominal element and the semantic features of the noun object. While there are some instances of this in English (e.g., the preposition "on" in phrase "on the bus" is used when an entity is inside of, rather than on top of, the vehicle), more reliable and straightforward cases of this usage are seen in classifier-noun phrases in languages such as Cantonese Chinese. For example, it is perfectly grammatical to say "yat tiu gam yue " (i.e., "one [classifier] goldfish") in Cantonese, even though the classifier "tiu" normally denotes something that is long, narrow, and flexible, yet goldfish are not thought of as canonically long or narrow. The degree to which a match/mismatch between the classifier semantics and noun object properties will be varied and the resulting effects on referential mapping will be examined.

Together, the series of experiments described in this dissertation provide different avenues to explore the extent to which impulsive activation occurs. Further, the tests of whether
impulsive activation can be moderated by contextual expectations will assess the degree to which this kind of activation occurs within the confines of a strictly rational system of comprehension.
Chapter 2  

Impulsive Conceptual Activation Across Lexical Categories

A large body of research has indicated that, when processing a word, activation of its associated concepts and meanings occurs in parallel. This is the case even when using naturalistic methodologies and where linguistic or pragmatic contextual information biases only one of several meanings (e.g., Mirman et al., 2008). As well, recent visual world studies that reveal comprehension processes as spoken words unfold have shown that the initial subparts of words also trigger the activation of concepts associated with lexical competitors that are phonologically similar to those subparts. For instance, the initial sounds in "logs" activate the lexical form "lock" in addition to "logs", with the former triggering the activation of a semantic associate such as KEY (Yee & Sedivy, 2006). However, the research examining this "phonosemantic activation" effect has focused only on noun-noun relationships, where hearing an unfolding noun triggers fixations to conceptually related objects that would also be denoted by a noun. It remains unclear whether this kind of activation occurs when the subpart in question could be identified as a lexical item belonging to a different syntactic category than the word as a whole, and it might be reasonable to expect stronger activation only among elements of the same conceptual kind or within the same syntactic category. Further, although the studies of noun-based effects have been informative, a consideration of other lexical classes may provide potentially richer insights into the nature of the information that is incidentally activated via lexical competitors. For example, studies of sentence processing have shown that encountering a verb rapidly evokes its associated syntactic frame(s), which reflects how noun-phrase arguments are structurally realized in relation to the verb (see, e.g., Tanenhaus, Carlson, & Trueswell, 1989). To illustrate, a transitive verb such as "kick" in English evokes a subject-verb-object
syntactic frame, where the subject (in a canonical active-voice sentence) is the agent of the action, and the object a patient (e.g., "John kicked Mary"). On the other hand, a ditransitive verb like "give" or "tell" creates an expectation for two objects, one direct and one indirect (e.g., "John gave the chair to Mary"). As well, hearing a ditransitive often evokes a patient that is somehow transferred by an agent to a recipient. Finally, intransitive verbs like "elapse" do not require an object at all (e.g., "Time elapsed"), and realize their sole argument in subject position. Thus, although verbs will evoke associated object concepts, they also evoke specific linguistic and semantic ways in which these concepts are related to the verb. A second question is therefore whether a verb embedded within a more complex lexical item simply triggers the activation of any relevant conceptual associates, or whether this activation is influenced by the argument structure frame that forms part of the syntactic knowledge associated with those verbs. For example, is it the case that hearing a potentially transitive verb like "swim" in "swimsuit" will evoke the concept FISH? Or would this concept be unlikely to be strongly activated given that the word "fish" would be unlikely to follow the verb "swim" due to the verb's argument structure preferences (fish swim, but one doesn't swim a fish)? The question therefore addresses whether associate activation is at some level unconstrained by expectations about syntactic argument structure, or whether it is influenced by patterns governing the sequencing of elements within sentences.

With regard to the first question (activation of concepts across lexical classes), some evidence from the single word processing literature suggests that reading nouns can lead to increased activation of related verbs. For instance, McRae, Hare, Elman, and Ferretti (2005) found that the naming of a verb (e.g., "swim") was speeded if the verb was preceded by a related location noun prime (e.g., "ocean"). A similar facilitative effect was observed for verbs
preceded by relevant instrument nouns (e.g., "spoon" – "stir"), typical agents (e.g., "nun" – "praying"), and typical patients (e.g., "dice" – "rolled"). The reverse is also true - there is evidence that reading a verb prime results in faster decisions about its associated agents, patients, and instruments (Ferretti, McRae, & Hatherell, 2001). However, it is possible that the presence of a sentence context and the broader behavioural task used in a visual-world style experiment might entail different results. For example, one might expect that the depiction of isolated objects in the visual display, along with sentence frame information (e.g., "Click on the..."), could lead listeners to develop strong expectations for nouns and their corresponding object concepts. If so, the action concepts associated with verbs would be comparatively unactivated. The "embedded" verb in a noun such as "popcorn" would therefore be less likely to trigger fixations to verb-related associates. This question then addresses the generality and strength of the phonosemantic activation effect observed in past studies.

With regard to the second question (whether activation from an "embedded" verb is influenced by that verb's argument structure frame), we can begin by considering evidence that verbs immediately increase consideration of their corresponding concepts in cases where these concepts are eventually realized as post-verbal expressions. Altmann and Kamide (1999) found that listeners can make use of the selectional semantic restrictions encoded in verbs to anticipate the upcoming referent. Participants heard sentences such as "The boy will move/eat the cake", and their eye movements were monitored as they viewed a corresponding depicted scene. The verb included in the instruction was manipulated such that in one condition (e.g., "The boy will move the cake"), there was little restrictive semantic information conveyed by the verb. That is, any object in the concurrent visual scene could theoretically be moved. In another condition, however, the verb in the instruction (e.g., "The boy will eat the cake") set semantic constraints
upon what the sentence-final noun could be (e.g., something edible). In the latter case, participants made eye movements to the object denoted by the sentence-final noun even before it had been heard in the speech stream. However, this anticipatory effect was not seen when non-restrictive verbs (e.g., "move") were used. These findings suggest that listeners can use their knowledge about verb semantics immediately to compute which noun associates are and are not likely to follow an encountered verb.

More specific effects of argument structure were demonstrated by Boland (2005), who reported that listeners in a visual-world task were more likely to look at a recipient after hearing a ditransitive verb than at that same object/person after a transitive or intransitive verb. For example, listeners heard sentences such as "One window was broken, so the handyman mentioned it right away to / fixed it right away for] the owners", and viewed displays containing depictions of a window, a handyman, the building owners, and some tools. Listeners who heard the "mentioned it right away" version of the sentence made more eye movements to the owners than those who had heard the "fixed it right away" version. This is because the verb "mention" is ditransitive, and introduces the syntactic argument of a recipient for the information being mentioned. A verb like "fixed", on the other hand, does not introduce a recipient, and therefore does not drive looks to the depicted owners in the same way.

While these findings are informative, they do not address the question of whether a subpart perceived as belonging to one grammatical category can trigger analogous effects when the "complete" word belongs to another grammatical category. The most relevant findings for this question may be syntactic-level effects observed within subparts of whole sentences. As mentioned earlier, research by Tabor et al. (2004) has shown that local combinatorial relationships between verbs and their arguments are incidentally computed as the sentence is encountered,
despite global evidence that should in principle block the computation of these relationships. For example, when reading the sentence "The coach smiled at the player tossed the Frisbee", readers are slower to read "the Frisbee" than when reading "The coach smiled at the player thrown the Frisbee." This effect is thought to be driven by the reader momentarily parsing "the player tossed" as if it were a legitimate NP+V sequence forming the beginning of an active clause, despite the fact that the information already encountered in the sentence (e.g., "The coach smiled at ... ") makes this parse impossible. This difference across conditions was not observed when participants read versions of the sentences containing unreduced relative clauses, where the local NP+V relationship that could have been momentarily parsed as the beginning of an active clause was interrupted (e.g., "The coach smiled at the player WHO WAS tossed/thrown the Frisbee").

The difference between the two versions of the reduced cases then does not appear to be the result of simple form-based associations, but rather of knowledge about the syntactic constructions in which the relevant forms can participate. Thus, although real-time processing seems to be characterized in part by impulsive local processes that do not seem particularly helpful for ongoing global processing, they are principled in certain ways.

In the current series of experiments, a similar question is explored, but with two salient differences. First, the focus is narrowed to effects that occur within the context of an unfolding single word. Second, one of the elements involved in the impulsive and local combinatory link that might be computed (namely, the argument for the compound subpart that could be interpreted as a verb) is never in fact linguistically expressed in any form. For instance, the first element (i.e., "pop") in the compound "popcorn" might be locally coherent with the noun phrase "the balloon". However, this noun phrase is never actually heard, and it is the remaining element of the compound (i.e., "-corn") that is encountered. As such, this item (e.g., a balloon) merely
has the status of a conceptual associate for the verb action. If effects of verb argument structure influence associate activation in this case, this would provide a particularly compelling demonstration of the idea that the impulsive processing occurring within the span of a word can involve more than shallow semantic-level associations.

In both of the experiments that follow, a visual-world methodology was used. Participants heard pre-recorded spoken instructions to click on target pictures in the visual displays, and their eye movements were monitored as they carried out the task. On critical trials, target names were always compound words like "popcorn" or "swimsuit". These compounds all contained initial embedded elements whose form corresponded to the form of an isolated verb (e.g., "pop" and "swim" respectively, in the examples above). In some cases, the initial element of the compound did reflect its verb meaning (e.g., "driveway"), but in other cases, the initial element would commonly be understood as taking a noun meaning (e.g., "flyswatter"). For the sake of simplicity, these compounds will be referred to as "verb-noun compounds" and their initial subparts as "embedded verbs" throughout this dissertation. However, the use of these descriptors should not be taken to imply that the initial elements of all of these compounds would actually be considered verbs in linguistic analyses of the compounds' structure.

The verb-noun compounds and associated object pairs in these experiments were selected such that the embedded verb portion of the compounds had one of two possible syntactic relationships with their paired semantic associates, which were depicted in the visual display but never mentioned in the corresponding linguistic materials. One kind of compound had embedded verbs that would in principle allow their respective associates to either occur after them as direct objects (e.g., a balloon, for the embedded verb "pop" in "popcorn"; it is possible to say "pop the balloon"), or before them as sentential subjects (e.g., "the balloon popped"). The other kind of
compound had embedded verbs that could not take their respective semantic associates as direct objects. Rather, the typical structural realization of these entities would be in subject position (e.g., for the embedded verb "swim" in "swimsuit"; it is not syntactically sound to say "X swims the fish"; however "the fish swims" is possible). For current purposes, the relevant difference is therefore whether the associate can or cannot occur in the structural position following the embedded verb.

If the transient activation effect found in Yee and Sedivy (2006) extends to associations that cross the boundaries of different types of concepts (objects and actions), relatively increased visual consideration of the semantic associate for the embedded verb should be observed as a listener processes the initial part of the compound target word. Further, if the argument structure of the embedded verb indeed plays a role in spoken-word recognition, it might be the case that incidental activation may be stronger in instances where the embedded verb's associate could occur as a direct object. For example, upon hearing "pop" in "popcorn", participants might briefly consider the depicted associate (i.e., a balloon) because its name could occur after "pop" as a sentential object. If argument structure information plays a moderating role, however, the same effect should be less likely to be observed for compounds of the "swimsuit" type. This is because the name of the depicted associate (i.e., a fish) could only occur before "swim" as a subject, and not after it as a direct object. On the other hand, if an embedded verb's argument structure is not considered, there should be no differences resulting from verb type.

These questions were explored in two separate experiments. In the first experiment, recorded instructions such as "Click on the popcorn/swimsuit" were used in order to assess whether effects of transient associate activation can be detected in a syntactic context that did not highlight the possibility of verbal descriptions of target objects. Because every spoken
instruction (on critical and filler trials) ended in a noun, and because the visual display available before and throughout the instruction depicted the isolated objects that these nouns denoted (and not, for example, a more complex scene where the accompanying language might denote potential or past actions), the experimental design of Experiment 1 created a global bias towards expecting nouns to be used when naming potential target objects. Further, participants had no reason to expect that they would be asked to manipulate screen objects in any other way except to click on them with the computer mouse. Together, these factors ensured that verb-denoted actions (other than "Click on...") were not expected in general, which in turn means that listeners would be relatively unlikely to evoke action concepts upon hearing the target word in the instruction. Importantly, given the overall noun bias, listeners would be especially unlikely to evoke action concepts that relate most strongly to an initial fragment of the noun rather than to the meaning of the entire noun itself.

The second experiment was designed to examine whether the visual consideration of semantic associates might increase in situations where there was an increased possibility of explicitly referring to them (i.e., listeners might consider the balloon as "popcorn" unfolds because of some implicit expectation that this object might be referred to as a "popping balloon"). In the second experiment, the critical trials involved sentences like "Mary thought about popcorn/swimsuits", and participants were instructed to click on the last thing mentioned in the sentence. The use of this syntactic frame was intended to increase listeners' expectations that the depicted semantic associate would actually be the denoted target object, given the increased number of ways in which the sentence could be completed (e.g., "Mary thought about... popping a balloon"). In addition, the experiment included filler items of the type "Mary thought about sleeping cows", or "Mary thought about delivering pizzas". In these fillers, verb
meanings are integrated into actual descriptions rather than existing merely as an incidental by-product of lexical activation from the initial parts of compound nouns. The inclusion of these fillers should also plausibly increase listeners’ expectations to hear a gerundive adjective or present-tense participle phrases in the target description. If a marked increase in associate fixations is found, this might suggest that the source of the effect is tied at least in part to expectations for modifier-noun descriptions, rather than simply the noun-driven transient activation of semantic associates.

Experiment 1

Experiment 1: Method

Participants. Thirty-eight native English-speaking students at the University of Toronto at Mississauga participated in this study, and received either course credit or monetary compensation for their participation. Prior to the experiment, each participant completed a language background questionnaire detailing their age of acquisition of English, as well as their oral and written fluency in English. All participants in the final sample reported learning English from birth and indicated full fluency. Data from one additional participant was excluded from the analyses because North American English was not learned from birth.

Materials. The visual displays used in the study consisted of a 3 by 3 grid with four objects displayed in the four corner squares (see Figure 2.1). The target item named in the corresponding recording was always present on the screen, and the names of targets on critical trials were always verb-noun compound words (e.g., "popcorn", "swimsuit", etc.). On critical trials, displays also contained either a competitor object that was a semantic associate of the embedded verb in the target compound (the "SA"; e.g., a balloon for the target "popcorn", given the embedded verb "pop"), or a corresponding control object whose shape was similar to the SA
but was not semantically related to either the embedded verb in the target name or the target object itself (e.g., a hand mirror for the target "popcorn"). Further, the syntactic relationship between the embedded verb in the target names and SAs were categorized into two conditions, based on the verb's argument structure frame. In the "Both" condition, the SA competitor commonly occurs in natural language as either the subject or object of the embedded verb in the target name (e.g., a balloon for the target "popcorn"). In the "Subject Only" condition, however, the SA competitor typically occurs only as the subject of the embedded verb in the target (e.g., a fish for the target "swimsuit") (see Appendix 1 for a full list of critical target and competitor items). The manipulation varying the type of competitor was crossed with the manipulation of the argument frame for the embedded verb, creating four conditions in total.

![Figure 2.1. Example of visual display in Experiments 1 and 2, showing the competitors in the two experimental conditions. The target was the swimsuit.](image)

The critical targets and SA competitors were carefully selected to ensure that the portion of the critical target compound that is temporarily compatible with a verb form would not
normally be used in a linguistic description of the depicted SA. Take, for example, the target "park bench", which was paired with the SA "car." Because the embedded verb "park" in the target "park bench" could in principle be used in a referential phrase to denote a "parked car", the car depicted in the visual display was one that implied a moving vehicle. Likewise, the SA "hand" for the target "punchbowl" was depicted with the fingers extended and palm facing outwards, rather than with the fingers closed, to eliminate the possibility that participants might expect the SA to be referred to as a "punching hand/fist".

One concern is that any differences observed in the two Both and Subject Only conditions might stem from differences in the strength of the conceptual association between the embedded verbs and SA's in these conditions, rather than factors related to argument structure. To control for this possibility, conceptual relatedness measures for the embedded verb and the name of the semantic associate were calculated using the pairwise comparison tool on the Latent Semantic Analysis (LSA) website (http://lsa.colorado.edu; topic space: General Reading up to 1st year college). The cosine measures provided by this tool reflect the strength of the conceptual relationship between a pair of words based on their occurrence in the same contexts in written corpora. Using these measures, it is therefore possible to explicitly compare the strength of the conceptual associations between embedded verbs and SA names in the Both and Subject Only conditions. An unpaired t-test found no significant difference in the conceptual relatedness scores for verb-SA pairs in the Both ($M = 0.17$, $SD = 0.11$) versus the Subject Only ($M = 0.12$, $SD = 0.12$) conditions ($t(18) = 0.80$, $p = .43$). This result indicates that the two kinds of SA's were conceptually related to their respective embedded verbs to a similar degree. In other words, it is not the case that the conceptual tie between words such as "pop" and "balloon" is overall stronger than that between words like "swim" and "fish". As a result, it is unlikely that any
differences in the pattern of results could be attributed to differences in the strength of conceptual associations between the verbs and SA’s in each of the argument structure categories.

In addition to the target and SA/control objects, two distractors were also present in the display on critical trials, and these objects were semantically, phonologically, and visually unrelated to the target, SA, and control objects. The assignment of SA and control objects was cycled across different versions of the experiment, so that each target appeared once with their respective SA and once with their respective control object, creating two stimulus lists.

Instructions to click on target objects were recorded by the author (a native-like English speaker), and took the form of "Click on the [target name]". For critical trials, target names did not share initial speech sounds with the names of any other object on the screen. The placement of the various object types on critical trials (target, SA, control object, distractors) within the grid display was counterbalanced. There were 20 critical trials and 46 filler trials in the experiment. Two additional practice trials were used at the start of the experiment. Filler trials were designed to prevent participants from expecting compound nouns as target names, and from discerning the relationship between target compounds and competitor objects on critical trials. Some fillers included target objects that were initially homophonous with another object in the display (e.g., fans [sports] vs. fan [desk], flour vs. flowers), as well as objects described using adjectives that required a subjective judgement by the participant (e.g., "fast boat" [speedboat vs. rowboat], "nice dress" [new dress vs. old torn dress]). Based on participant reports after the experiment, these types of trials were found to be highly salient, and no participant reported noticing that a certain proportion of target names were verb-noun compounds, or the nature of the relationship between targets and SA competitors on these trials.

**Experiment 1: Procedure**
Participants were tested individually. After completing the questionnaire, participants were seated in front of a computer monitor and fitted with the head-mounted optics for an eye tracking system (Eyelink II, SR Research, Mississauga, Ontario). Participants were advised that a grid containing various objects would appear on the screen, after which they would hear a tone signalling that the recorded instruction was about to occur. The instruction asked them to click on one of the objects in the display, and they were to click on the named object as quickly and accurately as possible. If the correct target was selected, a chime indicating a correct response was heard. Clicking an incorrect target produced no result until the correct object was selected (however, participants almost always chose the correct target on the first attempt on critical trials). The trial ended once the correct target had been selected.

Following initial eye tracker setup and calibration, participants were presented with two practice trials to familiarize them with the experimental task. Once participants indicated understanding of the procedure, they continued with the remainder of the trials. A calibration point appeared on a blank screen before each trial, and participants were instructed to fixate and click on the point to allow the eye tracking system to compensate for possible drift correction. The trial began automatically following this procedure, with re-calibration performed as necessary throughout the experiment. Each display was visible for 1500 ms prior to the onset of the tone signalling the start of the trial. The recorded instruction started playing 500 ms after this tone, and participants could click on screen objects immediately after the onset of the target word, which occurred an average of approximately 666 ms after the start of the instruction. At the end of the experiment, the experimenter asked participants whether they had noticed any unusual patterns in the stimuli, or whether they had relied on any special strategies to perform the task. No participant showed any awareness of the purpose of the study.
Experiment 1: Results

Participants' eye movements were coded using DataViewer (SR Research, Mississauga, Ontario). The temporal region depicted in Figure 2.2 begins at the average onset of the target noun during the speech stream, and ends 2000 ms after average target onset, by which point all processing related to target identification should have been completed.

In order to facilitate statistical analyses using general linear approaches, a quasi-logit transformation was applied to the raw fixation probabilities (values of 0 and 1 are first replaced with corresponding decimal near-equivalents, e.g., 0.0000001 and 0.9999999, to avoid undefined values). Figures of the fixation probabilities are therefore depicted on a logit scale. The logit scale preserves relative differences between fixations to alternative objects, such that a comparatively higher value corresponds to a greater likelihood of fixation. The zero point on the Y axis corresponds to a 50% probability of fixating a given object.

Figure 2.2 shows that across all conditions, looks to the target object rise as the target noun unfolds, reflecting listeners' incremental comprehension of the target noun. Additionally, the proportion of fixations to the distractor object are similar across conditions, reflecting their lack of phonological or semantic relationship with any part of the target word, and suggesting that listeners did not consider them to be likely referential candidates. Further examination of the figure shows that fixations to the SA’s in the Both condition (top right panel) appear to be elevated relative to fixations to the distractor objects, and that this elevation is not evident for the control objects in this condition (top left panel). The proportion of fixations to the SA’s and control objects in the Subject Only condition (bottom right and left panels, respectively) seem more similar to the proportion of fixations to distractors.
Figure 2.2. Likelihood of fixating screen objects across the four experimental conditions in Experiment 1. Probabilities are depicted on the logit scale. The first vertical line on each graph marks the average end of the Verb region, the second line marks the average end of the Post-Verb region.

In order to test the reliability of these differences, two temporal regions of interest were defined for statistical analyses. The first region, the Verb region, captures the processing driven by the interpretation of the embedded verbs in the target names as they unfold in time. Because the average duration of embedded verbs in the Both and Subject Only conditions was 382 ms and 409 ms, respectively, the Verb region spanned from 200 ms to 580 ms after target noun onset in the Both condition, and from 200 ms to 600 ms after target noun onset in the Subject Only
condition. (The left and right boundaries of the Verb region are adjusted by a 200 ms margin, reflecting the typical time interval required for planning and launching eye movements in this paradigm. Further, the probability of fixations was calculated in 20 ms time bins, and so all speech landmarks are rounded to the nearest 20 ms.) The second region of interest, the Post-Verb region, captures any residual effects from the initial moments of processing, and allows for an examination of how information from the embedded verb might shape the continued processing of the target compound as the "noun" portion is heard. The duration of this region is the same as the first, but begins immediately after the first region ends. In the Both condition, the Post-Verb region spans from 581 ms to 960 ms after target noun onset, and spans from 601 ms to 1000 ms after target noun onset in the Subject Only condition.

Two measures were used in order to test for differences in the likelihood of fixating SA competitors versus shape-matched controls in the experimental conditions. The first measure, the target advantage score, was calculated by subtracting participants' average likelihood of fixating the competitor from their likelihood of fixating the target object within the relevant time window, averaged across trials. This provides a method of assessing how quickly listeners were able to differentiate the target from the competitor and identify it as the intended referent (see, e.g., Arnold, Eisenband, Brown-Schmidt, & Trueswell, 2000; Heller, Grodner, & Tanenhaus, 2008). The second measure, the competitor advantage score, was calculated by subtracting participants' average likelihood of fixating an unrelated distractor from their likelihood of fixating the competitor object, averaged across trials. This second measure provides a method of assessing how much referential consideration a competitor object received, as compared to an unrelated object also present in the display (see, e.g., Dahan & Tanenhaus, 2005; Yee, Huffstetler, & Thompson-Schill, 2011).
Consideration of Target Object Relative to the Competitor. A 2 (Competitor Type) x 2 (Argument Structure) repeated-measures Analysis of Variance (ANOVA) was conducted using the target advantage measure. Separate analyses were conducted for the Verb and Post-Verb regions. Recall that a higher target advantage score corresponds to a greater degree of differentiation of the target object from the competitor. In the Verb region, there was a main effect of competitor type, such that the target advantage scores were lower when SA’s were present ($M = 2.72$) than when control objects were present ($M = 5.42$; $F(1, 36) = 4.70, p < .05$). There was no significant effect of argument structure nor was there a significant interaction (both $p's > .20$). In the Post-Verb region, there was no significant main effect of competitor type or argument structure, nor was the interaction significant (all $p's > .46$). Together, the results from both time intervals indicate that as the embedded verb unfolded, participants were slower to differentiate the target from the SA competitors than from the shape-matched controls. However, by the time information about the rest of the compound noun was encountered, participants differentiated the targets from the SA competitors and control objects to a similar degree.

Consideration of Competitor Object Relative to the Unrelated Distractor. The competitor advantage scores were submitted to a 2 (Competitor Type) x 2 (Argument Structure) ANOVA, analogous to the target advantage analyses. In the Verb region, neither the main effect of competitor type, the main effect of argument structure, nor the interaction reached significance (all $p's > .15$). However, recall that Figure 2.2 showed what appeared to be an increased rate of fixations to SA competitors in the Both argument structure condition. Although the interaction between condition and argument structure did not reach significance ($F(1, 36) = 2.14, p = .15$), there does appear to be a difference between the proportion of fixations to SA’s in the Both condition and that to other competitor objects. Figure 2.3 directly depicts the competitor
advantage scores across the two different argument structures in the Verb region. (Recall that competitor advantage scores reflect the degree to which competitors were distinguished from unrelated distractors.) SA competitors in the Both argument structure condition appeared to have been differentiated from distractors to a greater degree than control competitors in that same argument structure. This difference, however, was not apparent for competitor differentiation in the Subject Only argument structure condition.

Figure 2.3. Competitor advantage scores across the four experimental conditions in Experiment 1.

Because these numerical differences lie in the same direction as would be predicted by a priori hypotheses about the effect of argument structure in the Verb region, and because this was the region where verb-driven effects were considered most likely to surface, following through with planned pairwise comparisons seems warranted. These planned contrasts confirmed
that there was greater consideration of the SA competitors relative to distractors when embedded verbs had a Both argument structure \((M = 2.28)\) rather than a Subject Only structure \((M = 0.85, F(1, 36) = 4.35, p < .05)\). In the Both structure condition, there was greater consideration of the competitors relative to distractors when that competitor was an SA \((M = 2.28)\) versus when it was a control object \((M = -0.29, F(1, 36) = 5.51, p < .05)\). However, there was no such difference found between the SA and control competitors in the Subject Only condition \((p = .78)\).

These results suggest that during the initial moments of embedded verb processing, there is some tendency for listeners to give more consideration to those SA's that can occur with their respective embedded verbs as direct objects. This difference can be considered as somewhat tentative, however, as the interaction term in the 2 x 2 ANOVA did not reach significance.

In the Post-Verb region, there was a main effect of competitor type such that competitor advantage scores were higher for the SA objects \((M = .77)\) than for the control objects \((M = -0.33; F(1, 36) = .46, p < .05)\). There was a marginal effect of argument structure, such that competitor items tended to receive more fixations relative to distractors in the Both argument structure condition \((M = 0.55)\) than in the Subject Only structure condition \((M = -0.12; F(1, 36) = 2.91, p = .10)\). This result is somewhat surprising because it suggests that control objects, in addition to target objects, received more consideration in the Both condition. As such, it would seem to be best understood as reflecting a simple bias to consider objects with particular visual contours, rather than a true semantically-driven effect of argument structure. However, this result was at the outer limit for a conventionally-defined marginal effect \((F(1, 36) = 2.91, p = .10)\).

**Experiment 1: Discussion**
The results from this study are broadly in line with previous findings that concepts that are semantically associated with the name of the target object are partially activated during real-time spoken-word interpretation. Further, like Yee and Sedivy (2006), subparts of the target name were found to increase consideration of semantic associates for non-depicted lexical competitors. One novel contribution from the current results is the finding that word subparts can apparently drive semantic activation even across word classes during on-line language comprehension (i.e., from verb concepts to noun concepts: POP triggers BALLOON). Second, these semantic activation effects are found even when the "triggering" subpart is embedded within a lexical item that belongs to a different syntactic category (i.e., "pop" is contained within the noun "popcorn").

The finding that embedded verbs can activate semantically associated nouns might not initially seem surprising in light of prior evidence that (complete) nouns can prime related verbs in lexical decision tasks (e.g., Ferretti et al., 2001). However, it is relevant to note that a lexical decision task likely amplifies sensitivity to the types of semantic relationships of interest, in that any boost in the activation of words related to the prime is facilitative for performing the task at best and irrelevant (but not detrimental) at worst. For example, if the target is semantically related to the previously encountered prime, it enjoys an increase in activation and is verified as being a real word more quickly. Conversely, if the target is semantically unrelated to the previously encountered prime, its base activation level is not suppressed; it merely remains at baseline. So, a participant in a lexical decision task has little to lose by maximizing their (implicit) attention to the semantic relationships between verbs and nouns. In the current study, on the other hand, attending to verb-noun relationships can in fact hinder processing. For one, critical verbs were embedded within more complex noun compounds, and attending to the verb
at all is in some sense detrimental to the explicit experimental task (i.e., to identify and select the referent of a noun). In addition, the task demands create a strong linguistic and visual bias towards noun concepts. Participants in the study were explicitly told that they would hear instructions to click on screen objects. At no point during the experiment did participants hear instructions that involved actions other than clicking, nor did they hear sentences that might have highlighted the semantic or syntactic relationships between the embedded verbs and semantic associates. That participants considered the embedded verb at all might be compatible with the idea that lexical activation is a largely automatic, bottom-up process that is at least momentarily impervious to constraints of syntactic or pragmatic expectations.

In addition to the observed general activation of semantic associates, some weak effects of argument structure were also found in this experiment. The results suggest that during the initial moments of embedded verb processing, listeners do not activate semantic associates of different types of verbs to a similar degree. Rather, associates that could occur as direct objects of an encountered verb tend to receive greater consideration than those that could only typically occur as sentential subjects. The most plausible explanation for this difference between argument structure types is that, because the embedded verbs in the Both condition could be syntactically linked with an upcoming (i.e., as yet unheard) noun, listeners may briefly consider the semantic associates that could fulfil this function in anticipation of their potential occurrence. Embedded verbs in the Subject Only condition, on the other hand, would typically occur only after their requisite nouns, and because these nouns had not yet been encountered by the time the embedded verb was heard, their semantic associates were not considered in this way.

The results from Experiment 2 can be thought of as refining the "grain size" at which impulsive combinatorial processes take place. Recall that the "merely local" effects described by
Tabor et al. (2004) refer to cases where readers or listeners posit incorrect syntactic relationships due to the local coherence between two sentence elements, despite that fact that the input encountered earlier within the sentence makes that parse syntactically impossible. By this logic, in the current study, listeners may have momentarily posited a verb phrase upon hearing an embedded verb, and accordingly briefly considered those noun concepts that could have been uttered following the embedded verb in speech (i.e., the SA’s in the Both condition).

On the other hand, because these effects were only statistically significant in the planned contrasts and not in the overall ANOVA, and because they were not maintained in the Post-Verb region, firm conclusions about the role of argument structure cannot be drawn. The mild and somewhat unreliable effects of argument structure in the current study suggest that these impulsive but syntactically-guided combinatorial links may be best detected when the "linked" elements are actually realized, as in the Tabor et al. (2004) study (i.e., as opposed to in the present case, where an anticipated element only exists as a potentially discourse-relevant object that is never actually named). Another possibility is that the pragmatic constraints in this task were strong enough to reduce the effects of local coherence. Because all the instructions in this experiment took the form of "Click on the [adjective-noun/noun]", participants likely had strong expectations that target objects would be described by a noun phrase. If the latter explanation is correct, reducing this pragmatic expectation might reveal that semantic associates whose names could follow encountered embedded verbs would receive greater consideration than those whose names typically occurred before the embedded verbs.

Experiment 2 was designed to test the influence of pragmatic expectations with regard to the forms used to name the target. This was done by using a different sentence construction in the spoken stimuli, and by including filler items that both reduced the expectation that the main
verb would be followed by simple noun phrases, and also boosted the relevance of verb semantics to the task. In this experiment, all sentences began with "Mary thought about...", and participants were asked to click on the last thing named in the sentence. Instructions on filler trials referred to target objects using noun phrases containing gerundive adjectives (e.g., "Mary thought about sleeping cows") or a present-participle construction (e.g., "Mary thought about delivering pizzas"). Although targets on critical trials were still named using unmodified compound nouns (e.g., "Mary thought about popcorn"), the nature of the filler instructions may increase listeners' attention in general to the embedded verbs in the target compounds. This in turn might provide a stronger opportunity to observe argument structure effects such that semantic associates of embedded verbs that commonly occur with both subjects and direct objects would receive more consideration than those of verbs that commonly occur only with sentential subjects. Further, unlike a command such as "Click on the ...", which picks out a particular token of a noun concept, the semantics of so-called "psych verbs" such as "think about" put less emphasis on the physical form of a noun concept, and allow for references to objects whose linguistically-described states may differ from their visual depiction in a corresponding display (e.g., it is acceptable to pair "Mary thought about popping balloons" with a visual depiction of an intact balloon, while the same is not true for "Click on the popping balloon"). This in turn may make the actual depiction of the objects in the visual scene less relevant for the task.

Aside from the syntactic constructions of the filler items, and the fact that using psych-verbs allows for more flexibility with regard to how referent objects might be depicted, there were two other important differences between the spoken instructions used in the first and second experiments. First, in order to make the use of the gerundive adjective and present
participle verb phrase constructions possible, the names of all of the target objects had to be pluralized. This was done to avoid the use of a determiner in the filler items (e.g., "Mary thought about a sleeping cow"), which would have alerted listeners that a noun phrase was being heard before listeners had heard the adjective or participle in question, thereby eliminating the indeterminacy that the fillers were intended to create. To minimize differences with Experiment 1, however, the images in the visual displays were kept the same (i.e., as arrays of singular objects). Because the use of a psych verb like "thought about" in the spoken instructions is compatible with an indefinite or generic interpretation of a noun phrase, the mismatch between the plurality of the target noun in the sentence and the singular depicted objects on the computer screen is not expected to disrupt conceptual activation in any substantial way.

Second, unlike the "Click on the ..." instructions in Experiment 1, the auditory stimuli in Experiment 2 did not direct participants to perform an action. Instead, participants were asked at the beginning of the experiment to click on the last object named in the sentences they heard. The task in Experiment 2 then perhaps more closely approximates the kind of tasks used in visual world studies by Altmann and Kamide (1999), Huettig and Altmann (2004), and Yee and Sedivy (2006), where participants were only asked to listen to spoken sentences or words and view concurrent visual scenes without the action required by the participant being mentioned in the accompanying speech. Crucially, the results from those studies still reflect the types of semantic and syntactic effects of interest in the current study. Therefore, despite this difference between Experiments 1 and 2, the capacity to detect effects of semantic activation should be comparable.

**Experiment 2**

**Experiment 2: Method**
**Participants.** Thirty-eight native English-speaking students at the University of Toronto at Mississauga participated in this study, and received either course credit or monetary compensation for their participation. Prior to the experiment, each participant completed a language background questionnaire detailing their age of acquisition of English, as well as their oral and written fluency in English. All participants in the final sample reported learning English from birth and indicated full fluency. Data from four additional participants were excluded from the analyses either due to technical difficulties during testing, or because North American English was not learned from birth.

**Materials.** The experimental design, visual arrays, and target objects used in this experiment were identical to those in Experiment 1. Spoken instructions were pre-recorded by the author (a native-like English speaker), and took the form of "Mary thought about [target name, pluralized]." As in Experiment 1, there were 20 critical trials and 46 filler trials in this experiment. In 16 of these fillers trials, the spoken instruction now used a gerundive adjective to refer to the target object (e.g., "Mary thought about sleeping cows"). In 17 of the filler trials, a present-participle verb phrase was used (e.g., "Mary thought about delivering pizzas"). The remaining 11 filler items involved simple noun phrases (e.g., "Mary thought about beds"), so that in total, there were 33 instructions (including those on critical trials) that referred to target objects using simple noun phrases, and 33 instructions that referred to targets using phrases that involved verbal elements (either a deverbal [gerundive] adjective or a participle). Two additional practice trials (one of which involved a present-participle verb phrase) were included at the start of the experiment to ensure participants adequately understood the task.

**Procedure.** Participants were tested individually, using the same apparatus and procedures as in Experiment 1. At the end of the experiment, the experimenter asked participants
whether they had noticed any unusual patterns in the stimuli, or whether they relied on any special strategies to perform the task. No participant showed any awareness of the purpose of this experiment.

**Experiment 2: Results**

Participants' eye movements were coded using DataViewer (SR Research, Mississauga, Ontario). The temporal region depicted in Figure 2.4 begins at the average onset of the target noun during the speech stream, and ends 2000 ms later, by which point all processing related to target identification should have been completed.

In order to explore these differences statistically, two regions of interest were defined, following the same criteria as in Experiment 1. As before, the probability of fixations was calculated in 20 ms time bins, and so all speech landmarks are rounded to the nearest 20 ms. In the current experiment, the average duration of embedded verbs in the Both and Subject Only conditions was 358 ms and 400 ms, respectively. The Verb region thus spanned from 200 ms to 560 ms after target noun onset in the Both condition, and from 200 ms to 600 ms after target noun onset in the Subject Only condition. The Post-Verb region for the Both condition spanned from 561 ms to 920 ms after target noun onset, and from 601 ms to 1000 ms after target noun onset for the Subject Only condition. The left and right boundaries of the Verb region were once again adjusted rightward by a 200 ms margin to allow for the planning and launching of eye movements. As in Experiment 1, two measures were calculated to examine activation patterns as the compound noun unfolded.
Consideration of Target Object Relative to the Competitor. A 2 (Competitor Type) x 2 (Argument Structure) ANOVA was conducted using the target advantage measure in the Verb and Post-Verb regions. In the Verb region, no significant main effect of argument structure or competitor type was found, and the interaction between the two factors was also not significant (all p's > .18). In the Post-Verb region, a significant effect of argument structure was found, such that target objects were differentiated more slowly in the Both condition ($M = 17.95$) than in the Subject Only condition ($M = 21.13$; $F(1, 36) = 7.85$, $p < .01$). There was also a marginal effect of
competitor type in this region, suggesting that participants were slower to differentiate targets from SA's ($M = 18.52$) than from control objects ($M = 20.56$; $F(1, 36) = 2.80, p = .10$). The interaction between argument structure and condition was not significant ($p = .90$).

**Consideration of Competitor Object Relative to the Unrelated Distractor.** In the Verb region, a significant effect of competitor type was found, such that listeners fixated SA competitors ($M = 0.54$) more than control competitors ($M = -0.66$; $F(1, 36) = 8.05, p < .01$). Neither the main effect of argument structure nor the interaction was significant (both $p$'s $>.68$). Planned contrasts were performed to facilitate comparison with Experiment 1, and revealed that SA competitors ($M = 1.81$) received marginally greater consideration than control objects in the Both condition ($M = -0.09$; $F(1, 36) = 2.91, p = .10$). However, no significant difference was found between looks to the SA versus control competitors in the Subject Only condition ($p = .14$).

In the Post-Verb region, there was a significant effect of competitor type ($F(1, 36) = 6.00, p < .01$). However this was moderated by a significant interaction between argument structure and competitor type ($F(1, 36) = 4.48, p < .05$). Pairwise contrasts revealed that this interaction was driven by increased fixations to the SA competitors in the Both argument structure ($M = 2.47$) relative to the Subject Only argument structure ($M = 0.29$; $F(1, 36) = 11.30, p < .01$), a difference that did not exist for the control objects across the two argument structure conditions ($p = .71$). The main effect of argument structure was not significant in this region ($p = .16$).

The results from the competitor advantage scores indicated that during the earliest moments of embedded verb comprehension, semantic associates of the verb received greater consideration in general. There was a slight trend for listeners to make more eye movements to
SA's in the Both condition as the embedded verb unfolded. This trend was more strongly manifested in the moments after initial verb processing, where semantic associates that could occur as both subjects and direct objects of their respective verbs received greater consideration than associates that could only occur as subjects of their verbs. This post-verbal activation might reflect a kind of spillover effect, whereby listeners reconcile initial verb-driven processes with the new information (i.e., the rest of the noun compound) being encountered.

**Experiment 2: Discussion**

Similar to Experiment 1, the findings from the current study show that activation of semantic associates can occur across syntactic categories, and can be driven by subparts of words. While the effect of semantic association was observed in the target advantage measure in Experiment 1, in Experiment 2 this effect was reflected in the competitor advantage measure, whereby competitor objects that were semantically associated with the embedded verb received greater consideration than control objects. This difference between experiments might have arisen from the presence of filler instructions that used gerundive adjective and present participle verb phrases, which highlighted the relevance of verb semantics. Since the competitor objects on critical trials were now potentially more relevant to the semantics of the embedded verb, the competitor measure may have been more sensitive to effects of semantic association.

Analysis of the Verb region in Experiment 1 revealed some evidence that only those associates that could follow their respective verbs as direct sentential objects (vs. those that typically occurred as sentential subjects) received greater consideration. This interaction between argument structure and semantic association is once again observed in Experiment 2; however the effect was only marginally significant, and, as in Experiment 1, was only apparent in planned contrasts and was not related to a significant interaction in the main analysis. However, in
contrast to Experiment 1, in the Post-Verb region in Experiment 2, the interaction did reach full statistical significance. There was fully reliable evidence in this region that semantic associates in the Both argument structure condition received greater consideration than those in the Subject Only argument structure condition. Critically, this difference was not observed between the control items in the two respective conditions.

The across-experiment differences in the effect of argument structure can be seen as a result of the reduced certainty participants had with respect to the syntactic form that the spoken instructions would take. In Experiment 1, every instruction followed the structure of "Click on the [adjective-noun phrase / simple noun]". In Experiment 2, however, the main verb could be followed by simple noun phrases, modified noun phrases, present participles, or gerundive adjective noun phrases. In light of this greater syntactic uncertainty, it is reasonable to posit that listeners' comprehension processes were relatively delayed, and that stronger effects of argument structure would not be seen until further downstream. Taken together, the findings from Experiment 2 indicate that the manipulations increased the relevance of embedded verb semantics (i.e., as reflected by the competitor measure being more sensitive to semantic association) and also increased the likelihood that the potentially "local" argument structure (i.e., between the embedded verb and the names of semantic associates in the Both argument structure condition) would actually be realized during critical trials. For example, the inclusion of gerundive adjectives and present participle verb phrases in the fillers introduced the possibility that embedded verbs in target compounds might have actually formed the beginning of a verbal phrase in which the semantic associate would eventually be named (e.g., "Mary thought about popping balloons").
More generally, the findings from Experiment 2 suggest that in a context where verb elements are more likely to occur, effects that are evocative of local coherence effects can be observed, even when the initial element that "triggers" the coherence relationship exists as only a subpart of a longer word. The finding that semantic associates are selectively activated based on the syntactic relationship they could have with the embedded verbs shows that word subparts can also trigger the activation of their associated syntactic frames. It appears that when listeners were provided with a task where the constituent following the main verb was always a noun phrase (e.g., Experiment 1), the argument structure of embedded verbs did not seem to be activated to any great degree. However, when this expectation is reduced, effects arising from the different argument structures associated with the embedded verbs can be revealed. It is interesting that these effects can be detected even though the noun that would theoretically follow the embedded verb was never actually realized, but rather existed only as a potential referent. That is, these effects presumably arise because listeners give greater consideration to candidates whose names could follow the embedded verb, even though those candidates are never actually named. This indicates that activation of verb argument structures occurred quickly and automatically even when the immediately subsequent 500 ms of the speech stream indicated that the linguistic unit being processed was something other than a verb.

Note that the inclusion of gerundive adjective and present participle verb phrases in the fillers should have "boosted" listeners' expectations of both kinds of semantic associates being referred to with verb-relevant phrases, if no separate influence of argument structure were present. The gerundive adjective phrases increased the likelihood that semantic associates in the Subject Only argument structure condition would be referred to (e.g., the speaker could have said "Mary thought about swimming fish"), while the present participle verb phrases did the same for
associates in the Both condition (e.g., the speaker could have said "Mary thought about popping balloons"). If associate consideration was driven by listeners' expectations that the associate name merely could follow the embedded verb, then activation of both kinds of semantic associates should have increased equally. However, activation was greater for associates in the Both condition, suggesting that there is something special about the syntactic relationship between these embedded verbs and their respective semantic associates. Recall that LSA measures indicated that the semantic relatedness between embedded verbs and associate concepts was similar for the Both and Subject Only conditions, and so the observed activation difference between the two kinds of associates cannot be due to inherently stronger semantic relationships in the Both condition.

As a caveat, it is important to note that, to this point, the major difference between that the Both and Subject Only conditions has been described in terms of the order in which arguments (unnamed semantic associates) would occur relative to the embedded verb when these verbs and arguments are expressed in canonical sentence frames. Semantic associates in the Both condition could be direct objects of the embedded verb, whereas associates in the Subject Only condition could not. However, another way to characterize the difference between the two classes of verbs is not about linear ordering, but simply the overall number of possible argument "slots". If potential verb arguments receive activation upon hearing a verb, it is not unreasonable to believe that a noun concept that is associated with two possible argument slots for a particular verb would receive more activation than one that is associated with only one slot. Although the current study was not designed to distinguish between these two possibilities, it should be noted that both explanations suggest that argument structure information is playing a role at some level.
Summary

The findings from the studies described in this chapter serve as evidence that embedded verbs within compound nouns are incidentally computed as "real" verbs during incremental comprehension, which in turn can activate concepts semantically related to the verb, as well as argument structure information associated with the verb. One point of interest across both experiments is that, contrary to expectations, semantic associates in the Both argument structure condition were not consistently or strongly differentiated from their shape-matched controls. For example, in the Post-Verb region of Experiment 1, while there was a marginal effect of argument structure such that competitors in the Both argument structure condition were differentiated from distractors to a greater degree, this was the case even for competitors that were merely shape-matched control objects semantically unrelated to the embedded verb. Semantic associates in the Both argument structure in Experiment 2 attracted only marginally more attention than the control objects within the Verb region, and it was not until the Post-Verb region that the predicted interaction between semantic association and embedded verb argument structure was reliably observed.

At first glance, this is surprising: why should control objects in the Both argument structure condition attract any more attention than control objects in the Subject only argument structure condition? After all, these control objects were selected to be semantically unrelated to the embedded verb, and as such, verb-related differences (e.g., such as argument structure information) should not result in corresponding differences in the consideration of control objects across conditions. In order to address this issue, it is useful to turn to studies that have examined the role of visual information during visual world tasks. For instance, Dahan and Tanenhaus (2005) found that accessing a target concept (e.g., a snake) also results in the
activation of the canonical shape contours associated with that object (e.g., a coil). This shape information in turn drives listeners' eye movements to co-present screen objects that also possess the canonical shape of the target concept (e.g., a rope). Yee et al. (2011) report a related finding, such that listeners make eye movements to screen objects (e.g., a pizza) that typically possess the same shape properties as the target referent (e.g., a coin), even in cases where the specific depiction of the competitor does not match the shape contours of the target object (e.g., where the pizza is depicted as a slice instead of a whole).

These findings suggest that, within a visual world experiment such as one used in Experiments 1 and 2, visual features of screen objects are quickly assessed against a listeners' knowledge about the shape properties of referents. Further, screen objects that share visual contours corresponding to canonical shapes of target referents receive increased visual attention. The control objects in Experiments 1 and 2 were initially matched with their corresponding semantic associate objects on the basis of rough visual contours in an attempt to ensure that any differences observed between the semantic associate conditions and control conditions could not be due to some inherent preference for the semantic associate's shape. However, it appears that this shape-match may have inadvertently resulted in listeners deploying greater attention to the control objects because they shared visual features with semantically associated concepts. This shape-based activation likely accounts (at least in part) for the absence of a consistent and strong interaction between semantic association and argument structure throughout Experiments 1 and 2.

In line with findings demonstrating that processing a word can result in semantic activation across syntactic categories (e.g., Ferretti et al., 2001; McRae et al., 2005) results from Experiment 1 indicate that this spreading activation is triggered even by subparts of words as
they unfold in real time, and even in the presence of cues signalling that the verb subparts are merely part of a more complex target noun. The findings from Experiment 2 suggest that when the strength of these cues are decreased, listeners' knowledge of the syntactic information associated with embedded word subparts is reflected in their consideration of referential candidates, such that candidates whose names could grammatically follow the embedded verb as a direct object received more attention. These findings might reflect an anticipatory version of the "merely local" syntactic processes discussed in Tabor et al. (2004), in which adjacent sentence elements are momentarily computed to be part of some immediate syntactic unit, even though this unit is incompatible with the global interpretation of the sentence. Although the various structures of the instructions in Experiment 2 create a situation in which embedded verbs in either of the syntactic argument conditions might be linked with the semantic associate objects (i.e., because one can say "Mary thought about popping balloons" or "Mary thought about swimming fish"), it was still observed that associates with names that could follow the embedded verb as a direct object received more attention than those with names that could not. It may be the case then, that the manipulations in Experiment 2 increased listeners' sensitivity in general to potentially-coherent elements in a way that was simply additive to the effect of argument structure information.

The overall pattern of findings in the current series of experiments points to the highly interactive nature of language comprehension processes. Both semantic concepts and syntactic information associated with word subparts were shown to be activated to various degrees as listeners encounter those subparts, even though this activation might in principle have been suppressed by listeners' expectations. It is especially noteworthy that hearing even these relatively fleeting word subparts can result in the access of cross-category lexical items and
associated syntactic frames. The depth at which this impulsive processing is seen to occur suggests that even phrasal-level structures might be activated by word subparts. That is, since hearing a "verb" subpart seemed to selectively increase the consideration of candidate objects whose names might follow a verbal phrase that began with the word subpart, it may be the case that phrasal structures are momentarily posited as word subparts are encountered. On the other hand, it was also observed that this kind of impulsive syntactic activation was only observed in cases where listeners did not have strong expectations about how utterances would unfold. Chapter 3 will further explore the question of whether phrasal-level candidates are also impulsively activated during spoken word comprehension, with specific focus on how listener expectations (as mediated by visual context) might constrain this process.
Chapter 3

Impulsive Phrasal-Level Competition During Spoken Word Recognition

It is evident from the work described in prior chapters of this dissertation that lexical and semantic activation both occur immediately as the speech signal is processed (e.g., Allopenna et al., 1998, Yee & Sedivy, 2006), and that hearing the initial constituent of a compound (e.g., "flyswatter") results in the incidental activation of distinct lexical items whose forms overlap with that of the constituent (e.g., "fly", which corresponds to a possible verb form), which in turn activates associated semantic concepts (e.g., AIRPLANE). More specifically, the experiments in Chapter 2 provided a demonstration that these effects occur even when the initial constituent belongs to a different grammatical category than the one to which the entire compound belongs, and that the range of activated information can also include syntactic knowledge associated with the initial constituent, such as verb-encoded argument structure preferences. In other words, the embedded verb portion of a verb-noun compound appeared to trigger the range of linguistic processes that are characteristic of "real" verbs, despite the fact that the containing constituent is in fact a noun.

Recall that the critical starting point in the chain of events is that ultimately non-selected lexical alternatives sharing the same form (initial speech sounds) as the target word are temporarily activated to some degree (e.g., "logs" activates "lock", "popcorn" activates "pop"). Recall also that the range of activated forms is thought to dynamically compete for activation as speech sounds continue to unfold in time. Importantly, most primary research on spoken word recognition examining the dynamics of activation processes has characterized the competitors in this process as independent lexical units. However, compound nouns may provide a unique
opportunity to address whether this competition process also involves *phrasal* units, composed of multiple words.

![Diagram showing potential kinds of activation resulting from encountering colour compounds such as "greyhound".](https://example.com/diagram)

*Figure 3.1. Potential kinds of activation resulting from encountering colour compounds such as "greyhound".*

Consider the case of compound nouns composed of an adjective and a noun, for example, where the first constituent of the compound is a colour term (e.g., "greyhound"; herein referred to as "colour compounds"). As a starting point, we would already expect that hearing the word "greyhound" might result in transient activation of the entries for "gravy", "grateful", and "gradient" (amongst others) in the mental lexicon (see Figure 3.1). However, on the assumption that embedded adjectives trigger transient processes that are similar to the processing of "real" adjectives (in the way that embedded verbs were observed to temporarily act like "real" verbs), it is possible that the alternative forms being considered also include phrasal units such as "grey
Although phrasal combinations of words are unlikely stored as such in the mental lexicon, the availability of entities in the contextual environment might boost the likelihood of certain multi-word descriptions. Compounds of this type can therefore provide an interesting test case to consider how the mapping of speech sounds to lexical entries involves not only the degree of "match" between the unfolding signal and stored lexical alternatives, but also against more complex phrasal expressions. Does competition arise only from lexically-stored alternatives? Or is the system truly interactive, such that these phrasal-level descriptions that are computed dynamically based on contextually-relevant entities can also compete for attention? If it is the latter, what factors might influence the extent to which this competition is observed?

In the case of adjectives that modify nouns, it is important to consider past research examining how these elements/expressions/units are interpreted in real time. Previous work has shown that the interpretation of these expressions involves the immediate integration of lexical knowledge and referential context. In fact, research in this area has proven to be one of the clearest examples of how the visual world paradigm can reflect the interactive processes that take place as listeners use context to streamline real time interpretation. For example, as mentioned in Chapter 1, work by Eberhard et al. (1999) showed that listeners interpret multi-adjective noun phrases such as "Touch the starred yellow square" incrementally, and that the speed of identifying an intended referent depends on whether the other (non-target) objects in the referential domain also possess the properties denoted by the adjectives (e.g., being yellow, having a starred pattern, etc.). Given this, it is likely that the interpretation of colour compounds and the kinds of competition that might be observed will be differentially affected depending on the presence or absence of discourse-relevant objects and their colour properties.
Of course, in cases where there is nothing else in the context that could be referred to using the relevant colour term, competition with the names for other display objects would not occur, and the intended object should be identified quickly as the compound noun begins to unfold. For instance, in a visual display containing a greyhound, a red spatula, and a beige chair, the greyhound is the only object that could likely be referred to with a description beginning "the grey...", on both the basis of a single noun or a potential phrasal-level expression. However, in situations where the relevant colour terms might be used in descriptions for other objects, these compounds could plausibly trigger the processes that characterize the incremental and immediate interpretation of stand-alone colour adjectives. For instance, Sedivy, Tanenhaus, Chambers, and Carlson (1999) presented listeners with visual displays of objects and monitored listeners’ eye movements as they carried out spoken instructions like "Touch the blue pen". Display objects were manipulated to create an early and a late disambiguation scenario. In the early disambiguation case (e.g., a display containing a blue pen, a yellow rubber duck, a red notebook, and a pink comb), listeners had enough information as soon as they had heard the colour adjective to accurately identify the intended target object. In the late disambiguation case (e.g., a display containing a blue pen, a blue bowl, a yellow duck, and a red notebook), listeners needed to wait until the onset of the noun in the instruction to accurately select the intended target.

Sedivy et al. (1999) found that listeners were faster to make eye movements to the target object in the early disambiguation case than in the late disambiguation case, suggesting that listeners were integrating the semantics of the colour adjective with the properties of the discourse-relevant objects to guide their interpretation of the spoken instruction. A second experiment in Sedivy et al. (1999) revealed further evidence that unfolding speech input is continuously evaluated against listeners’ expectations about how other discourse-relevant objects
will be referred to. In this experiment, participants viewed visual displays which contained a contrast set of two objects that differed only in colour (e.g., a yellow comb and a pink comb), another object that shared the colour of one of the objects in the contrast set (e.g., a yellow bowl), and an unrelated distractor object (e.g., a silver knife). On critical trials, the target object was either the object within the contrast set that shared the same colour as the other display object (e.g., the yellow comb), or the display object that had the same colour as one of the contrast set objects (e.g., the yellow bowl). Participants were asked to touch display objects according to spoken instructions that contained a colour adjective to describe the target object (e.g., "Touch the yellow comb/bowl"). In the former case, the use of a colour term in the target object description was motivated by the visual context, because there were two objects present that could be referred to by the same noun (e.g., the yellow and pink combs), and so further specification would be required to differentiate the target. In the latter case, however, there is no contextual motivation to use a colour descriptor in the sentence. Because there was only one bowl in the display, for example, the speaker could have simply said "Touch the bowl" without an additional need for modifiers.

In sum then, Sedivy et al.'s (1999) most important finding was that listeners were faster to fixate target objects that were part of a contrast set than those that were not. That is, upon hearing a description consisting of an adjective-noun sequence, listeners were able to compute the intended target more quickly when there was contextual motivation to use that adjective in a referring expression. With regard to the processing of colour compounds (e.g., "greyhound"), these findings might indicate that phrasal-level interpretations of "grey..." (i.e., where "grey" is classified as a stand-alone adjective belonging to an unfolding modified noun phrase) will compete most strongly for consideration in cases where there is a contrast set present such that
one of the items within the set could be differentiated from the other using the relevant colour term (e.g., a contrast set consisting of a grey spatula and a red spatula).

If the incremental interpretation of a colour compound noun is misled by a context that motivates the potential use of the relevant colour term as part of an adjective noun phrase, how might interpretation be affected by a context containing an object that could be described with the relevant colour term, but where there is no motivation to actually use that colour term to refer to the object? Imagine, for instance, hearing the sentence "Click on the greyhound" in a situation where a greyhound and a grey spatula are both present. While the spatula does in fact possesses a property that matches the semantics of the unfolding compound constituent "grey-" and thus could be described as "the grey spatula", a speaker has no strong reason to use a modified noun phrase to refer to the chair. Because there is only one chair in the current referential domain, a speaker should choose simply to call that object "the spatula" (in accordance with the Gricean maxim of quantity, for example, which states that speakers should not provide more information than is necessary for achieving the goal of the communicative utterance; Grice, 1975).

On the other hand, work by Pechmann (1989), Sedivy (2003), and others shows that speakers sometimes produce colour adjectives even when there is little in the situation to warrant their use. For example, Sedivy (2003) showed that when asked to instruct conversational partners to move objects, speakers spontaneously produced colour adjectives almost 40% of the time, even though the target could be uniquely identified using an unmodified noun. In light of this finding, it may be the case that the mere presence of a colour-matching object in the visual context, regardless of how necessary it would be to use a colour term to refer to that object, will be enough to increase attention to that object as the colour portion of the compound word is heard.
In addition to factors related to context, more basic semantic factors may also influence the competitiveness of phrasal-level alternatives upon hearing a compound such as "greyhound". The issue here is that in many cases, the colour term in the compound does not actually correspond to the colour of the entity it denotes (e.g., despite general properties of the conceptual kind, or the historical basis of the word). For example, greyhounds can have fur of various colours, blackboards can have black, green, or grey surfaces, and redheads can have hair that ranges from strawberry blonde to orangey-red in colour. Unlike the effects reported in prior work by Eberhard et al. (1999) and Sedivy (1999; 2003), which have only examined colour adjectives in contexts where objects unambiguously bear the relevant colour, the initial constituent in a colour compound may be "misleading" in that it does not accurately describe the target object itself. How might the complete versus partial match between the colour term in an object’s name and the actual colour of the object affect impulsive activation? For example, are the semantics of the embedded element activated and recruited during compound comprehension, resulting in a delay in identifying the target or in greater competition from phrasal-level alternatives like "grey spatula"? Or will the semantics of colour compounds be interpreted akin to "frozen" expressions, such that the degree of correspondence between the semantics of object name subparts and object properties are essentially ignored?

The purpose of the current series of visual-world studies was twofold. The first goal was to investigate the effects that different visual contexts have on the likelihood of considering an adjective-noun sequence instead of a compound noun as a colour compound term is heard in real-time. The second goal was to explore whether a mismatch between the perceptual properties of the intended referent and the semantics of the colour term in that referent’s name has any effect on incremental mapping, and whether this effect interacts with the effect of visual context.
Two experiments were designed to this end. In Experiment 3, listeners were asked to click on target objects whose names were colour compounds (e.g., a greyhound, a blackboard). These targets always possessed the colour that corresponded to the colour term in their names (e.g., the surface of the blackboard was black, the greyhound was grey), but were cycled through three different visual contexts. Across all the experimental conditions, all non-target screen object on critical trials would be named with (unmodified) nouns that did not overlap with the initial speech sounds of the target compound. In the no-competition condition, no other screen object possessed a colour that corresponded to the colour term in the target’s name. Target recognition is expected to occur quickly and this condition will serve as a baseline for comparison to the other conditions. In the contrastive-competitor condition, the target object was presented along with a contrast set of two differently-coloured but otherwise identical objects. One of the objects in the contrast set shared the same colour as the colour term in the target’s name, and served as the competitor object in this condition. For example, the target greyhound was paired with a grey spatula and a red spatula. It is expected that target recognition will be the slowest in this context if listeners’ sensitivity to the presence of the contrast set guides their expectations about how screen objects will be named. As the target colour compound (e.g., "greyhound") unfolds, the competitor (e.g., the grey spatula) would likely attract a large amount of consideration, because its physical attributes constitute a good match for the colour term being heard, and there is contextually-based motivation to refer to it using a colour descriptor.

Finally, in the noncontrastive-competitor condition, the target object (e.g., a greyhound) was presented along with a competitor object that possessed the colour corresponding to that in the target object’s name (e.g., a grey spatula). However, this object was not part of a contrast set, and as such did not need to be differentiated from another screen object on the basis of colour.
(e.g., the two other screen objects were a yellow taxi, and a blue helicopter). The extent to which this competitor object is considered as the target compound unfolds during this colour-only context will be the most interesting to examine. Although the colour of the competitor object allows for that competitor to be named using a phrase that begins with the colour term in the target compound, there is no motivation to actually use a modified noun phrase to refer to that object. In this context, simply using the competitor's unmodified name ("spatula") would be sufficient to distinguish it from all other screen objects. This final condition is the critical one, where the degree to which the competitor receives attention will reveal whether phrasal descriptions for objects are available candidates for the unfolding speech even though the referential situation does not require their use.

Experiment 4 was designed to assess how a mismatch between the colour of the target object and the colour term in that object's name might affect the process of mapping the speech sounds in the target name to alternative candidates. The same three referential contexts described above were used, however the target objects in this experiment possessed a colour that was in contrast to the colour term in their names (e.g., a brown greyhound, a green blackboard). In the no-competition condition, it will be interesting to see whether this perceptual mismatch results in slower target recognition, even though the visual context contains no potential for ambiguity. In the contrastive-competitor and noncontrastive-competitor conditions, the question of interest will be whether the mismatch between the target's colour and the colour term in its name results in any increase in consideration of the competitors. If target recognition is delayed and competitor consideration increased in these contexts relative to Experiment 3, it would suggest that the semantics of the initial embedded element in target compounds are processed and integrated into referential mapping processes to some extent, despite the fact that at the level of linguistic form
alone, the target can be easily identified without the use of this information. Further, any mismatch between target object properties and the semantic features activated by target name subparts may result in greater levels of competition from phrasal-level descriptions, perhaps even in cases where the use of these descriptions is not particularly motivated by the referential context.

Before conducting these experiments, however, it is important to address an important consideration that arises with regard to the comprehension of spoken colour compounds. The specific issue involves well-known prosodic differences between compounds and genuine multi-word phrases. For example, the compound term "the White House" has stronger prosodic stress on "White" compared to "House" whereas the analogous modified noun phrase "the white house" does not. As well, the durations of the initial subparts of multi-syllabic words (e.g., "cap" in "the captain") have been observed to differ from their individual-word counterparts (e.g., "cap" in "the cap tucked"; Salverda, Dahan, & McQueen, 2003). It is possible that phonetic cues such as stress and vowel duration might be available to a listener as either a genuine adjective or the adjective portion of a colour compound is heard, thereby disambiguating the type of expression on the basis of sound-level information. However, it is in fact unclear whether these differences are manifested at the initial compound segment, or whether it is the relative difference in the stress, prosodic, or rhythmic patterns between the two compound elements that conveys this segmental information. If it is the latter, phonetic cues about the type of expression would be available mainly when the second constituent is heard. As a result, the processing of the initial constituent in a colour compound would be more likely to proceed without the benefit of potentially-disambiguating prosodic cues.
To assess this possibility, a pre-test was carried out prior to the main experiments. The goal was to examine whether the subtle phonetic information present in the initial element of colour compounds can in fact lead participants to interpret that colour term as part of a compound word versus an adjective noun phrase.

**Experiment 3**

**Pre-test: Method**

**Materials and Procedure.** In this study, 10 native speakers of English heard recordings of sentences containing truncated words or noun phrases, and were asked to decide which of two alternatives was being said. The stimuli, which contained the 12 target words used in Experiments 3 and 4 as well as control and filler items, were recorded by the author, a native-like English speaker. The target words were all compound nouns in which the first element embedded in the compound was also a colour adjective (e.g., "greyhound", "blackboard", "greenhouse"; see Appendix 2 for a full list).

For critical trials, each colour compound was realized as the last word of a sentence beginning with "Click on the…". The recordings initially consisted of the preceding sentence and the entire colour compound (e.g., "Click on the greyhound") as a whole, but were then clipped immediately before the point at which the initial sound of the second compound constituent could be perceived (e.g., "Click on the grey-"). In other words, sentences on critical trials were truncated such that they contained only the phonetic information up until the end of the colour term in the colour compound. These recordings were paired with visual displays containing two written alternatives. One alternative was the colour compound in full (e.g., "greyhound") and the other alternative was the name of the corresponding competitor object in Experiments 1 and 2 (e.g., "grey spatula").
It is possible that listeners have an underlying preference in general to interpret truncated phrases or morphologically complex words as belonging to an adjective + noun sequence rather than a compound word. The inverse might also be true: when confronted with truncated speech, listeners might have a bias to interpret the shortened element as part of a compound word. A bias of this type could confound the interpretation of results, for example, if participants showed high rates of selecting a compound rather than a modified noun upon hearing any truncated compound. Specifically, it would be unclear whether effects stem from listeners using prosodic cues present in the initial constituent of the compound, or from biases towards selecting a single word rather than a longer modified expression.

In order to test for general response biases, control trials were created in which spoken tokens of modified noun phrases (e.g., "dog treat"; "eye test") were recorded as the last words of sentences starting with "Click on the…". These recordings were then truncated such that they contained only phonetic information up until the end of the first word in the noun phrase (e.g., "Click on the dog-"; "Click on the eye-"). The visual displays paired with these recordings included the actual token that had been said and a compound word whose initial element matched the first word of the noun phrase (e.g., "dog treat" vs. "doghouse"; "eye test" vs. "eyelid").

If listeners have a bias towards interpreting truncated elements as belonging to single words when possible, they should tend to make incorrect choices in the control trials (i.e., decide incorrectly that the "dog-" truncated from "dog treat" is part of "doghouse" rather than "dog treat"). Conversely, if listeners have a bias to perceive the stimulus as an adjective-noun sequence, they should overwhelmingly choose the correct alternative in the control condition,
since both the prosodic information and the underlying bias would lead listeners to expect a noun-phrase continuation of the truncated sentence.

Two types of filler trials were included in the study to prevent participants from discerning the compound versus modified noun manipulation. One type of filler items consisted of truncated sentences that presented enough phonetic information for listeners to accurately pick out the correct token amongst the given alternatives. These trials included either both single-word alternatives (e.g., "Click on the deal-", presented with "dealer" vs. "decoy"), or both noun-phrase alternatives (e.g., "Click on the pine-", presented with "pine tree" vs. "pie slice"). The other type of fillers was created such that listeners would not be able to reliably pick out the correct token based on the edited recording. These fillers included truncated words that shared initial speech sounds with their corresponding alternatives (e.g., "Click on the can-", created from "Click on the candy", which was paired with a visual display showing "candy" vs. "candle" as the alternatives). Also included were truncated utterances including modified noun phrases, such as "Click on the big-", created from editing the utterance "Click on the big cat". These recordings were presented with alternatives that shared the same first word (e.g., "big cat" vs. "big kite"). There should be minimal prosodic differences in how the alternatives in these kinds of filler trials would be pronounced, and so these items served to reduce listeners' explicit attention to prosody throughout the experiment.

Each visual display consisted of the alternative choices printed on the left and right side of the screen (see Figure 3.2). The locations of the correct versus incorrect alternatives were counterbalanced, and trials were randomized. There were 12 critical trials, 10 control trials, 28 filler trials, and 4 practice trials at the beginning of the study to familiarize participants with the procedure. Participants were tested individually, and the study was conducted using a laptop
computer with the recorded sentences played over headphones. Written instructions were displayed on the laptop screen at the start of the study. These instructions explained that participants would hear part of a recorded sentence, and then would see two alternative words or phrases displayed on the computer screen. Participants were informed that the recordings were truncated, and that their task was to decide which of the alternatives was the last word or words being said. If they had not heard enough information to know which of the alternatives was the correct choice, they were to make their best guess.

![Example display of a trial in the pre-test.](image)

*Figure 3.2. Example display of a trial in the pre-test.*

At the beginning of each trial, a blank screen with a fixation cross was displayed for 500 ms, after which point the spoken recording was heard. The written alternatives were displayed 250 ms after the recording ended, and stayed on the screen until participants indicated which alternative they thought was being named in the recording by pushing keyboard buttons that were labelled to correspond to the symbols depicted under each alternative. Participants were
asked to look at the fixation cross on the screen whenever they saw it appear, and to try to complete the task as quickly and as accurately as possible.

**Pre-test: Results and Discussion**

Over the course of the study, each participant saw 12 critical trials, and so could have made up to 12 correct responses. If there were prosodic cues present in the first element of colour compounds that reliably cued listeners to expect that the colour term was part of a longer compound word rather than part of a colour adjective noun phrase, one should observe the number of correct responses to be higher than chance (i.e., more than 6 out of 12 per participant). If, on the other hand, there were no prosodic cues present in the truncated compounds that led listeners to expect a compound word over an adjective noun phrase, then the proportion of correct answers should be at chance. To test this hypothesis, the number of correct answers per participant on critical trials were submitted to a one-sample *t*-test against chance. It was found that the number of correct responses was not significantly different than chance ($M = 6.2$, $t(9) = .31$, $p = .76$).

If listeners have an underlying preference to interpret truncated speech as belonging to a modified noun phrase rather than a compound, this preference might wash out a separate influence of prosodic cues on listeners’ judgments when hearing the truncated compounds. Recall that if there is an overarching preference to interpret truncated elements as being parts of noun phrases, it should be observed that participants overwhelmingly choose the correct option in control trials in which of the auditory stimulus was a truncated modifier + noun sequence (e.g., "Click on the dog"*(treat)*), and in which the incorrect alternative was a compound (e.g., "doghouse"). In order to assess this possibility, the number of correct answers per participant on the control trials were also submitted to a one-sample *t*-test against chance. It was found that the
proportion of correct answers was not significantly different than chance ($M = 4.7$, $t(9) = .41$, $p = .69$).

Taken together, the findings from the pre-test study render it unlikely that there is sufficient prosodic information in the initial element of a colour compound to allow listeners to deduce that the colour term is part of a compound word rather than an adjective-noun phrase. This means that sound-level information is unlikely to mask or overwhelm the manipulations of referential context and match/mismatch between the semantics of the colour term and the physical properties of the target object used in Experiments 3 and 4.

**Experiment 3: Method**

**Participants.** Twenty-four native English-speaking students at the University of Toronto Mississauga participated in this study, and received compensation of either course credit or $5 (CAD) per each 30 minutes of participation. Participants in this study did not take part in the pre-test. Prior to the experiment, participants completed a language background questionnaire detailing their age of acquisition of English, as well as their oral and written fluency in English. All participants in the final sample reported that they had learned English from birth, and indicated full fluency.

**Materials.** The visual displays used in this study consisted of a 3 by 3 grid with four objects displayed in the top center, middle left and right, and bottom center squares (see Figure 3.3). Spoken instructions to click on screen objects were recorded by the author (a native-like English speaker). The target item named in the instruction was always present on the screen, and the names of targets on critical trials were always colour-adjective compound words (e.g., "greyhound", "blackboard", etc.). The main colour of target objects in critical trials always matched the colour term in the target name.
Figure 3.3. Example of a critical trial display in Experiment 3. The target in this trial was the greyhound. The competitor was the grey spatula in the contrastive- and noncontrastive-competitor conditions, and the stapler in the no-competition condition. In the contrastive-competitor condition, one of the screen objects (e.g., the helicopter) was replaced with an item of the same kind, but different colour, as the competitor (e.g., the red spatula) in order to create a contrast set.

Critical trials fell in to one of three experimental conditions: the no-competition condition, the contrastive-competitor condition, or the noncontrastive-competitor condition. In the no-competition condition, targets (e.g., a greyhound) were presented with "competitor" items whose colour did not match the colour term in the target’s name, and whose name did not have the same initial speech sounds as the target name (e.g., a pink stapler). The two other screen objects (i.e., the distractors) presented in this condition were similarly unrelated to the target name (e.g., an orange helicopter, and a yellow taxi). The competitors in the no-competition condition then were functionally identical to the unrelated distractors, and were not expected to receive greater consideration relative to those distractors. Nonetheless, these objects were
identified as being "competitors" because they were the objects against which competitors from other conditions (i.e., which might plausibly be expected to receive increased consideration) were compared.

In the contrastive-competitor condition, the target on critical trials (e.g., a greyhound) was paired with a competitor object whose main colour corresponded to the colour term in the target name (e.g., a grey spatula). Further, one of the distractor objects in this condition belonged to the same category as the competitor, but had a different colour (e.g., a red spatula). The other distractor object neither had a colour that matched the colour term in the target name, nor had a name that shared the same initial speech sounds as the target name (e.g., a yellow taxi).

Finally, in the noncontrastive-competitor condition, targets (e.g., a greyhound) were paired with competitor object whose main colour corresponded to the colour term in the target name (e.g., a grey spatula). However, unlike the contrastive-competitor condition, the competitor object was the only exemplar of its kind – both distractor objects in this condition were items whose colour did not match the colour term in the target name, and whose names did not share initial speech sounds with the target name (e.g., an orange helicopter, a yellow taxi). The assignment of the three different kinds of visual context was cycled across different versions of the experiment, so that every target appeared once in each of the no-competition, contrastive-competitor, and noncontrastive-competitor conditions, creating three stimulus lists.

Instructions to click on target objects took the form of “Click on the [target name]”. The placement of the various object types on critical trials (target, competitor object, distractors) within the grid display was counterbalanced and randomized. There were 12 critical trials and 50 filler trials in the experiment, with 2 additional practice trials at the start of the experiment to ensure that participants understood the task. Due to the fact that some target objects were named
using their subordinate level name rather than their basic level name (e.g., "greyhound" instead of "dog"), targets on filler trials were sometimes named using subordinate level terms (e.g., "La-Z-Boy chair" to refer to a recliner; "evergreen" to refer to a tree). Other filler trials required listeners to make subjective judgements about screen objects (e.g., "Click on the expensive watch", where two watches are depicted). Finally, filler trials also included sets of objects that had to be contrasted based on dimensions other than colour (e.g., a large vs. small mug; a new vs. old car). On some of these trials, one of the objects in the contrast set was named as the target, while on other trials, neither of the objects within the contrast set were named. These various kinds of filler trials were included to prevent listeners from discerning the colour focus of the experiment, and also to eliminate any bias stemming from critical trials in the contrastive competitor conditions to expect that a contrastive object in the visual scene would not be named.

**Experiment 3: Procedure**

Participants were tested individually. After completing the language background questionnaire, participants were seated in front of a computer monitor and fitted with the head-mounted optics for an eye tracking system (Eyelink II, SR Research, Mississauga, Ontario). Participants were advised that a grid containing various objects would appear on the screen, after which they would hear a tone signalling that the recorded instruction was about to occur. The instruction would ask them to click on one of the objects in the display, and they were to click on the named object as quickly and accurately as possible. If the correct target was selected, a chime indicating a correct response was heard. Clicking an incorrect target produced no result until the correct object was selected. The trial ended once the correct target had been selected.

Following initial eye tracker setup and calibration, participants were presented with two practice trials to familiarize them with the experimental design. Once participants indicated
understanding of the procedure, they continued with the remainder of the trials. A calibration point appeared on a blank screen before each trial, and participants were instructed to fixate and click on the point to allow the eye tracking system to compensate for possible drift correction. The trial began automatically following this procedure, with re-calibration performed as necessary throughout the experiment. Each display was visible for 1500 ms prior to the onset of the tone signalling the start of the trial. The recorded instruction started playing 500 ms after this tone, and participants could click on screen objects immediately after the onset of the target word, which occurred 1000 ms after the start of the instruction.

**Experiment 3: Results**

Participants' eye movements were coded using DataViewer (SR Research, Mississauga, Ontario). Figure 3.4 shows the likelihood of fixating screen objects in the three experimental conditions as the target compound unfolds, averaged across trials and participants. A quasi-logit transformation was applied to the raw fixation probabilities to allow for statistical analyses using general linear approaches (see Chapter 2 for further details). Figures of the fixation probabilities are therefore depicted on a logit scale. The time frame depicted in Figure 3.4 begins at the average onset of the target compound during the speech stream, and ends 1200 ms after average target onset, by which point all processing related to target identification should have been completed. Recall that the zero point on the Y axis corresponds to a 50% probability of fixating a given object, and that a higher value on the logit scale corresponds to a higher probability of fixation.

Figure 3.4 shows that across all conditions, fixations to the target object rose as the target noun unfolded, reflecting listeners' incremental comprehension of the target noun. Additionally, the distractor object received a similarly low proportion of fixations across all three conditions,
demonstrating their lack of phonological or semantic relationship with any part of the target word and suggesting that listeners did not consider them likely candidates. It is also evident that there were clear differences in the proportion of fixations to targets and competitors across the three conditions.

Figure 3.4. Likelihood of fixating screen objects across the three experimental conditions in Experiment 3. Probabilities are depicted on the logit scale. The first vertical line on each graph marks the average end of the Adjective region, the second vertical line marks the average end of the Post-Adjective region.

The top right panel depicts the probability of fixating screen objects in the \textit{no-competition condition}. Recall that in this condition, non-target objects in the display did not have names that
shared the same initial phonemes as the target compound, nor did they possess the colour
denoted by the embedded adjective in the target compound name. As the target name was heard,
fixations to the target object rose quickly, even as the embedded colour term within the
compound was still unfolding. As expected, because no other screen object had names that
overlapped with the initial speech sounds of the target compound, and because no other screen
object's colour matched that denoted by the colour term embedded in the compound, little
consideration was given to non-target screen objects. As such, the competitor in this condition
was akin to an unrelated distractor, and serves as the baseline from which to compare competitor
consideration in the contrastive- and noncontrastive-competitor conditions.

Inspection of the graph for the contrastive-competitor condition (the top left panel) shows
a very different fixation profile. Here, the probability of fixating target objects was relatively
slow to rise, which was a result of the greatly increased attention to the competitor object. Recall
that in this condition, although the unmodified name of the competitor did not overlap with the
target compound, a referring expression that might be used to describe the competitor would
begin with the colour adjective embedded in the target compound, because the colour of the
competitor object corresponded to that colour term. Further, the presence of another screen
object that could be differentiated from the competitor on the basis of colour created a scenario
in which there would be motivation to use a colour adjective if one were to name the competitor
object. These factors combined appeared to have led participants to strongly consider the
competitor object to be as likely as the target object to be the intended referent.

Finally, the bottom panel of Figure 3.4 shows the probability of fixating screen objects in
the noncontrastive-competitor condition, in which the colour of the competitor object
corresponds to the colour term in the target compound, but where there were no other screen
objects present that would motivate the use of that colour term in referring to the competitor object. This means that, in principle, the names of the competitors in this condition should not have matched the speech sounds in the unfolding target compound, because there was no need to use an adjectival expression to refer to the competitor. Because of this, one might expect that the competitor in this condition would have received minimal attention as the target compound unfolded. Interestingly, however, the competitor in this condition did appear to receive increased consideration, suggesting that as the colour term in the target compound unfolded, listeners accessed the adjectival description of the competitor, despite the lack of contextual motivation to actually use an adjective in describing the competitor object.

In order to test whether the differences between conditions were statistically reliable, two time regions were defined for analysis. The first region, the Adjective region, captures the processing that would be driven by the unfolding colour term in the target compound. Because the average duration of the embedded colour terms was 482 ms, the Adjective region spans from 200 ms to 680 ms after the onset of the target compound. (The left and right boundaries of the Adjective region are adjusted by a 200 ms margin to account for the time required for planning and launching eye movements in this paradigm, and all speech landmarks are rounded to the nearest 20 ms because the probability of fixations was calculated in 20 ms time bins.) The second region of interest is the Post-Adjective region. Fixations in this region reflect any residual effects from the initial moments of colour-term processing, and allow for an examination how information from the colour term might guide the continued processing of the target compound as the "noun" portion is heard. The duration of this region is the same as the first, but spans from 681 ms after target onset (i.e., immediately after the end of the first region) to 1160 ms after target onset.
As in the experiments described in Chapter 2, two measures were used in order to test for differences in the levels of competitor consideration in the different experimental conditions. The first measure was the target advantage score, which was calculated by subtracting participants' average likelihood of fixating the competitor from their likelihood of fixating the target object within the relevant time window, averaged across trials. The second measure was the competitor advantage score, and was calculated by subtracting participants' average likelihood of fixating the unrelated distractor from their likelihood of fixating the competitor object, averaged across trials.

**Consideration of Target Object Relative to the Competitor.** Target advantage scores were submitted to a one-way repeated measures analysis of variance (ANOVA), with the three different display conditions as a within-subjects factor. Separate analyses were conducted for the Adjective and Post-Adjective regions. A higher target advantage score corresponds to a greater degree of differentiation of the target object from the competitor. In the Adjective region, the omnibus ANOVA was significant \( F(2, 23) = 18.34, p < .001 \), indicating a significant effect of condition. Planned comparisons revealed that the target was differentiated from the competitor to a greater degree in the no-competition condition \( M = 14.44 \) than in both the contrastive-competitor condition \( M = -1.76; F(2, 23) = 36.55, p < .001 \) and the noncontrastive-competitor condition \( M = 5.52; F(2, 23) = 11.07, p < .01 \). The target was also differentiated from the competitor to a greater degree in the noncontrastive-competitor condition \( M = 5.52 \) than in the contrastive-competitor condition \( M = -1.76; F(2, 23) = 7.39, p < .01 \).

In the Post-Adjective region, the omnibus ANOVA was once again significant \( F(2, 23) = 26.59, p < .001 \), indicating reliable differences across conditions. Planned comparisons revealed that the differences found in the earlier region were maintained. The target was
differentiated from the competitor to a greater degree in the no-competition condition \((M = 29.05)\) than in both the contrastive-competitor condition \((M = 4.18; F(2, 23) = 50.24, p < .001)\) and the noncontrastive-competitor condition \((M = 11.41; F(2, 23) = 25.27, p < .001)\). The target was also differentiated from the competitor to a greater degree in the noncontrastive-competitor \((M = 11.41)\) than in the contrastive-competitor condition \((M = 4.18; F(2, 23) = 4.25, p = .05)\).

In summary, the target advantage measure shows a clear effect of condition in both regions of target compound processing. As the colour compound unfolds, the target object is most quickly identified when it is the only screen object that could be referred to with a name or an expression that matches the initial speech sounds of the target compound. Target identification is slowed greatly when the visual scene contains a competitor object that could be referred to using an expression beginning with the colour term embedded in the target compound, and the visual context motivates the use of that colour term to refer to the competitor. However, target identification in the presence of a colour-relevant competitor object is also reliably slowed even if there is no motivation to use that term to refer to the competitor, albeit to a lesser extent.

**Consideration of Competitor Object Relative to the Distractor.** Competitor advantage scores for each region were submitted to separate one-way repeated measures ANOVAs, with the three different display conditions as a within-subjects factor. Recall that a higher competitor advantage score corresponds to a greater degree of differentiation of the competitor object from the distractor. In the Adjective region, the omnibus ANOVA revealed a significant effect of condition \((F(2, 23) = 16.01, p < .001)\). Planned comparisons showed that the competitor was differentiated from the distractor to a greater degree in the contrastive-competitor condition \((M = 10.22)\) than in both the no-competition condition \((M = -0.84; F(2, 23) = 31.86, p < .001)\) and the
noncontrastive-competitor condition ($M = 4.04; F(2, 23) = 9.95, p < .01$). The competitor was also differentiated from the distractor to a greater degree in the noncontrastive-competitor ($M = 4.04$) versus the no-competition condition ($M = -0.84; F(2, 23) = 6.21, p < .05$).

In the Post-Adjective region, the omnibus effect of condition was once again significant ($F(2, 23) = 20.48, p < .001$). Planned contrasts showed that, as in the earlier region, the competitor was differentiated from the distractor to a greater degree in the contrastive-competitor condition ($M = 11.00$) than in the no-competition condition ($M = -0.29; F(2, 23) = 37.94, p < .001$). The competitor was also differentiated from the distractor to a greater degree in the noncontrastive-competitor condition ($M = 8.12$) than in the no-competition condition ($M = -0.29; F(2, 23) = 21.02, p < .001$). However, there was no longer a difference in the degree of competitor differentiation between the contrastive- and noncontrastive-competitor conditions ($F(2, 23) = 2.48, p = .12$).

Results from the analyses of the competitor advantage measure suggest that, as the colour term in the target compound unfolded, competitors that possessed a colour corresponding to this constituent (and which could therefore be referred to with an expression that matches the initial speech sounds of the compound) received greater consideration than those that did not. Further, consideration of a competitor was greatest when that competitor possessed the compound-relevant colour and was within a context that motivated the use of that colour term to refer to it. As the "noun" portion of the target compound was heard, increased consideration of competitors that possessed a compound-relevant colour was maintained, however the additional "boost" in consideration conferred by an adjective-motivating context seemed to be attenuated.

**Experiment 3: Discussion**
The findings from Experiment 3 clearly illustrate that as listeners heard a colour compound and attempted to map it onto a referent object, potential phrasal-level descriptions of other referentially-relevant objects were also available. Competition was, perhaps not surprisingly, found in cases where the situation motivated the use of the colour term in the colour compound to refer to another object. These results are reminiscent of findings from prior work showing that the interpretation of modified noun phrases reflects the integration of knowledge about the semantic properties of referents as well as expectations about how discourse-relevant objects might be named (e.g., Eberhard et al., 1999; Sedivy et al., 1999). In the case of the current study, however, it was demonstrated that adjectival interpretations were also strongly triggered when the display contained an object (the target) whose simple (unmodified) name contained the same sounds as a stand-alone adjective. This pattern provides some basic evidence that simple lexical terms corresponding to objects are not necessarily more accessible in the early moments of processing than more complex descriptions motivated by the context.

Perhaps even more surprisingly, the lack of a supporting context for using colour terms in referential expressions did not seem to eliminate the accessibility of phrasal-level expressions. That is, even when there was no need to use a complex expression to refer to an object (e.g., when there is no reason to use "the grey spatula" to refer to a grey spatula because there is only one spatula in the visual context), these expressions were still available to listeners as they heard adjective-noun compounds unfold. Although competition arising from the activation of adjective-noun referring expressions was weaker in non-motivated contexts than in motivated contexts, these potential descriptions of display objects were clearly still posited as an unfolding colour compound was being interpreted.
Given the increase in consideration of competitor objects that possessed colour features corresponding to the colour term in the target name, one can conclude that the semantics of unfolding colour terms in colour compounds were accessed, and that listeners evaluate the perceptual properties of visually-available objects against those semantics. In the current experiment, the target object also possesses the perceptual property denoted by the colour term in its name. However, as mentioned in the introductory section of this chapter, the nature of objects named by colour compounds is such that they are not always the same colour as that denoted by the colour term embedded in their names (e.g., a brown greyhound, a green blackboard). How might such cases of mismatch affect the consideration of potential phrasal-level expressions observed in this study?

Experiment 4 was carried out to address this question, and uses an experimental design that is similar to Experiment 3. Target objects in Experiment 4, however, did not possess the colour denoted by the colour term embedded in their names. If an unfolding (embedded) colour term in a compound is impulsively perceived as an adjective, it would presumably be more rapidly mapped on to objects that bear the relevant colour. A mismatch between target object properties and the semantics of elements in the target name should then increase the consideration of phrasal-level expressions as the compound unfolds, because the properties of the competitor object would actually constitute a better match for the semantics of the adjectival element in the colour compound than the target object. If this were the case, as listeners interpret an unfolding colour compound, one should see even greater consideration of colour-matching competitors than that observed in Experiment 3. A difference observed between the findings from Experiment 3 and Experiment 4 would further serve to confirm that the effects observed in Experiment 3 really do arise from the embedded colour term being impulsively processed as an
adjective, rather than from some experimental artefact, such as the relative frequencies of colour-  compounds versus colour adjective-noun phrases, or subordinate names (e.g., "greyhound") versus basic-level names (e.g., "dog"), for example.

Experiment 4

Experiment 4: Method

Participants. Twenty-four native English-speaking students at the University of Toronto Mississauga participated in this study, and received compensation of either course credit or $5 (CAD) per each 30 minutes of participation. No participant in Experiment 4 took part in either Experiment 3 or the pre-test. Prior to the experiment, participants completed a language background questionnaire detailing their age of acquisition of English, as well as their oral and written fluency in English. All participants in the final sample reported that they had learned English from birth, and indicated full fluency.

Materials and Procedure. The current study employed the same spoken recordings, visual displays, and testing procedure as in Experiment 3, with the exception that the target items in Experiment 4 possessed a colour other than the one named by the embedded colour term in their names (see Appendix 2 for a full list). The same three visual display conditions as in Experiment 3 were used, and each critical target item was once again cycled across the three different conditions, creating three stimulus lists.

Experiment 4: Results

Figure 3.5 shows the likelihood of fixating screen objects in the three experimental conditions as the target compound unfolded, averaged across trials and participants. The same quasi-logit transformation used in Experiment 3 was applied to the raw fixation probabilities, and figures of the fixation probabilities are once again depicted on a logit scale. Each graph depicts
the time frame starting from target onset until 1200 ms after target onset, by which point all target-identification processes should have been resolved. The vertical lines on the graphs again correspond to speech landmarks within the unfolding compound, and define the same two regions of analyses as in Experiment 3 (i.e., the Adjective and Post-Adjective regions). Recall that the same recordings from Experiment 3 were used in the current study, and so the average duration of the embedded colour terms and the duration and timing of the regions of analysis remain the same. The first vertical line on each graph marks the average end of the embedded colour term in the target compound (680 ms after target onset), and the second vertical line marks the end of the Post-Adjective region (1160 ms after target onset), with both regions adjusted by 200 ms to account for the time required for planning and launching eye movements.). As in Experiment 3, speech landmarks are rounded to the nearest 20 ms.

Figure 3.5 shows that across all conditions, looks to the target object rose as the target compound unfolded, reflecting listeners’ incremental comprehension of the target noun. The distractor object, which was unrelated to the target compound both phonologically and semantically, once again received a low proportion of fixations across all three conditions. It is clear that the probability of fixating competitor objects in the no-competition condition (top left panel) was very different from that in the contrastive-competitor and noncontrastive-competitor conditions (top right and bottom panels, respectively). Predictably, the competitor object in the no-competition condition attracted little consideration. Recall that the competitor in this condition was identical to a distractor object in that its name did not match the speech sounds in the target compound, and because it did not bear the colour described by the colour term in the target compound. Looks to target objects increased relatively quickly, but appear somewhat delayed when compared to the no-competition condition in Experiment 3.
Figure 3.5. Likelihood of fixating screen objects across the three experimental conditions in Experiment 4. Probabilities are depicted on the logit scale. The first vertical line on each graph marks the average end of the Adjective region, the second vertical line marks the average end of the Post-Adjective region.

The bottom panel of Figure 3.5 shows the probability of fixating screen objects in the noncontrastive-competitor condition, in which the colour of the competitor object again corresponded to the colour term in the target compound, but the visual context did not motivate the use of that colour term in referring to the competitor. Unlike in Experiment 3, where competition was somewhat moderated, the noncontrastive competitor in the current study appeared to attract as much consideration as the contrastive competitor. Further, this competitor
also seemed to be considered as a better candidate than the target object as the colour term in the target compound was heard. The semantics of the embedded colour term appeared to have been activated and assessed against the properties of screen objects, driving eye movements toward objects bearing the colour described by the colour term. In order to test whether the differences between conditions were statistically reliable, the two measures used in Experiment 3 were calculated for each analysis region in the current study. The target advantage score was calculated by subtracting participants' average likelihood of fixating the competitor from their likelihood of fixating the target object within the relevant time window, averaged across trials. The competitor advantage score was calculated by subtracting participants' average likelihood of fixating the unrelated distractor from their likelihood of fixating the competitor object, averaged across trials. As in Experiment 3, each of these scores were submitted to a one-way repeated measures Analysis of Variance (ANOVA), with the three different display conditions as a within-subjects factor. Separate analyses were conducted for the Adjective and Post-Adjective regions.

**Consideration of Target Object Relative to the Competitor.** In the Adjective region, the ANOVA using the target advantage scores revealed a marginally significant effect of condition \((F(2, 23) = 2.694, p = .06)\). Planned comparisons indicated that targets were differentiated from competitors to a greater degree in the no-competition condition \((M = 5.77)\) than in the contrastive-competitor condition \((M = -0.61; F(2, 23) = 4.80, p < .05)\). Targets were also differentiated from competitors to a greater degree in the no-competition condition \((M = 5.77)\) than the noncontrastive-competition condition \((M = -.04)\), however this difference was on the cusp of the conventional threshold for statistical significance \((F(2, 23) = 3.98, p = .05)\).
Unlike in Experiment 3, there was no significant difference in target differentiation between the contrastive- and noncontrastive-competitor conditions ($F(2, 23) = .39, p = .85$).

In the Post-Adjective region, the ANOVA using target advantage scores revealed a significant effect of condition ($F(2, 23) = 14.51, p < .001$). Planned comparisons showed that targets were differentiated from competitors to a greater degree in the no-competition condition ($M = 19.09$) than in both the contrastive-competitor condition ($M = -0.62; F(2, 23) = 25.45, p < .01$) and the noncontrastive-competitor condition ($M = 2.85; F(2, 23) = 17.28, p < .001$). No significant difference was observed in target differentiation between the contrastive- and non-contrastive competitor conditions ($F(2, 23) = .79, p = .38$).

As in Experiment 3, analysis of the target advantage measure showed effects of condition in both regions of target compound processing. As the colour compound unfolded, the target object was most quickly identified when no screen objects had names with overlapping speech sounds with the target name, or possessed a colour that corresponded to the first part of the target compound. Target identification was again slowed greatly when the visual scene contained a competitor object that possessed a colour corresponding to the colour-term in the target’s name and the visual context motivated the use of that colour term in a phrasal description to refer to the competitor. Target identification in the presence of a colour-relevant competitor object was also slowed even when there was no motivation to expect the use of a colour term in referring to the competitor, however, unlike in Experiment 3, the delay observed in the current study was of the same magnitude as when there was contextual motivation to use colour terms in naming competitor objects. These results suggest that a mismatch between target object colour and the semantics of the colour term in the target’s name hindered target identification such that any
Consideration of Competitor Object Relative to the Distractor. The ANOVA conducted on the competitor advantage measure yielded a significant effect of condition in the Adjective region \((F(2, 23) = 16.75, p < .001)\). Planned comparisons revealed that competitors were differentiated from distractors to a greater degree in both the contrastive-competitor \((M = 12.48; F(2, 23) = 33.47, p < .001)\) and noncontrastive-competitor \((M = 6.60; F(2, 23) = 9.13, p < .01)\) conditions as compared to the no-competition condition \((M = 0.17)\). Competitors were also differentiated from distractors to a greater degree in the contrastive-competitor condition \((M = 12.48)\) than the noncontrastive-competitor condition \((M = 6.60; F(2, 23) = 7.64, p < .01)\).

In the Post-Adjective region, the omnibus ANOVA again revealed a significant effect of condition \((F(2, 23) = 13.19, p < .001)\). As in the earlier region, competitors were differentiated from distractors to a greater extent in both the contrastive-competitor \((M = 12.48; F(2, 23) = 16.60, p < .001)\) and noncontrastive-competitor \((M = 10.90; F(2, 23) = 22.51, p < .001)\) conditions than in the no-competition condition \((M = 1.26)\). However, there was no significant difference in competitor differentiation between the contrastive- and non-contrastive competitor conditions \((F(2, 23) = .45, p = .51)\).

Results from the analyses of the competitor advantage measure in Experiment 4 mirrored those found in Experiment 3. As the colour term in the target compound unfolded, competitors that possessed a colour corresponding to the colour term in the target name received greater consideration than those that did not. During these initial moments of compound processing, consideration of a competitor was greatest when that competitor possessed the compound-relevant colour and was within a visual context that motivated the use of that colour term in
referring to the competitor. Similar to Experiment 3, as the "noun" portion of the target compound was heard, the increase in consideration of competitors with a compound-relevant colour was maintained, but the additional "boost" in consideration conferred by context was again attenuated. It would appear then that the presence/absence of referential entities that can be named using colour-adjective expressions still guides the consideration of competitor objects, but that this process is not affected by a mismatch between target object colour and the semantics of the colour term in the target's name.

**Experiment 4: Discussion**

Experiment 4 was designed to assess how instances of mismatch between target properties and the semantics denoted by the colour term in the target’s name might influence the consideration of potential phrasal descriptions of non-target objects. It was found in Experiment 3 that competition from phrasal expressions was strongest when there was contextual motivation to refer to non-target objects with the colour term embedded in the target compound (e.g., when a grey spatula and a red spatula were present while hearing the target "greyhound"). This result was maintained in the current study – the "incongruent" colour of the target object seemed to be irrelevant in this scenario. This finding is perhaps intuitive: because prior work (e.g., Sedivy et al. 1999) has shown that listeners are very sensitive to the presence of contrast sets in the referential context, this experimental condition can be considered as a ceiling for competition effects.

In Experiment 3, a moderating influence of referential context was observed, such that competition from potential phrasal expressions was decreased in cases where the use of the relevant colour terms in phrasal expressions was not warranted by the objects in the visual display. Recall that although a colour-relevant competitor object (e.g., a grey spatula) still
received some consideration as listeners heard a colour compound (e.g., "greyhound"), the fact that the competitor object did not exist as part of a contrast set (i.e., was the only object of its category) seemed to diminish the amount of consideration it received. Interestingly, this moderating effect was not found in the current study. All colour-relevant competitor objects received a high proportion of consideration, even when there was no contextual motivation to use a colour term when referring to that object. Given this, it is reasonable to conclude that the semantics of the colour term in the unfolding compound are impulsively computed and linked to referential entities that best instantiate this property. In the case where the target is not among these entities, target identification is hindered as listeners momentarily give preference to those objects whose perceptual properties do match the semantics denoted by unfolding the colour term in the compound, and this preference overrides any effects of contextual motivation.

While the results from Experiment 4 indicate that "incongruent" target perceptual properties interact with expectations about how referential objects are likely to be named, these findings alone cannot be used to determine whether the identification of an incongruent target is slowed even when the referential domain contains no potential for competition. In order to do this, one must compare the likelihood of target consideration in the no-competition condition between Experiments 3 and 4. To this aim, independent samples t-tests were performed using the target advantage scores from the no-competition condition of each experiment. Separate t-tests were performed for each region of analysis. Results showed that in the Adjective region, targets were differentiated from competitor objects to a lesser degree in Experiment 4 ($M = 5.77$) than in Experiment 3 ($M = 14.44$; $t(46) = 3.45, p < .01$). This difference was maintained in the Post-Adjective region, with targets again being differentiated from competitors to a lesser degree in Experiment 4 ($M = 19.09$) than in Experiment 3 ($M = 29.05$; $t(46) = 3.99, p < .001$).
This cross-experimental comparison shows that a mismatch between target object properties and the semantics of elements embedded in target compounds can hinder the mapping of a compound noun to a candidate, even in situations where there is nothing else in the situation whose name should in principle provide a better fit with the embedded adjective in the target compound. Because every screen object belonged to a different conceptual category (i.e., there were no screen objects that were the same kind of object), listeners should expect that the forms used to name screen objects should all be uniquely identifying, regardless of the visual properties of the target object. Despite this, mapping the target name onto its referent was still delayed in cases where the target's visual properties conflicted with the semantics of the unfolding embedded adjective. This finding speaks to the automaticity with which comprehenders try to integrate semantic knowledge from unfolding elements with the perceptual properties of discourse-relevant objects, and also demonstrates that eye movements in this paradigm are clearly not driven by task-specific strategic expectations for how objects will be named once the display appears.

Summary

The research described in Chapter 2 demonstrated that verbs embedded within compound nouns temporarily act like "real" verbs during online comprehension. The evidence from the current series of studies extends this phenomenon to embedded colour terms, which appear to momentarily have the same effects as "real" colour adjectives as listeners encounter an unfolding colour compound. As has been observed with real colour adjectives (e.g., Eberhard et al., 1995; Sedivy et al., 1999), listeners encountering these colour terms map them against potential phrasal descriptions of contextually-present objects, and this mapping can be modulated by specific expectations about how different referential objects would be named. One novel contribution
from this set of studies is that, at the level of linguistic form, embedded colour terms are temporarily understood to form part of a phrasal description in parallel with a single noun description, even when the context does not warrant the use of the former type (cf. the results in the noncontrastive-competitor conditions). This is the case even though listeners in these studies arguably have sufficient time (upwards of 2 seconds) to consider the display objects (and indeed sometimes report doing so as a strategy to help them complete the task), and realize that the use of colour adjective noun phrases would be unnecessary. The structure of filler trials also ensured that adjective noun phrases were not used superfluously, but rather only in instances where an adjective was required to distinguish a target object from another screen object of the same kind, further reinforcing that adjective use in this task could be reliably predicted based on visual context, if that is a strategy that comprehenders happened to adopt.

In light of these findings, one might ask if the observed effects are merely due to low-level perceptual processes. In other words, do listeners look at the grey spatula in the noncontrastive-competitor condition, for instance, merely because it is a grey object, and not because they expect the word "grey" to appear in an expression describing the object? Evidence from Sedivy et al. (1999) suggests that this is unlikely to be the case. When listeners in that experiment heard instructions such as "Touch the tall glass" for example, looks to a tall pitcher did not increase, even though it was technically the tallest object in the visual display. Rather, listeners interpreted the adjective with reference to a contrast set of objects (e.g., a tall glass and a short glass), presumably because there was no contextual motivation to use an adjective to refer to the pitcher, which was the only object of its kind in the display. Along this reasoning, it seems likely that the effects observed in the current series of studies arise from listeners' expectations about how objects would be named, rather than a simple property-matching process.
While colour is a property that seems much more intrinsic to an object than height or size, which is typically interpreted relatively, there exist many cases where the use of a colour term in speech also entails an interpretation of colour as a relative construct. For example, "white wine" refers to a liquid that is not really white at all. The term only has meaning because it is meant to distinguish that liquid from its darker counterpart "red wine" (which is also not typically canonically red). Other uses of colour terms in referring expressions are figurative in nature, and invoke understandings of colour semantics in an abstract way (or at least in a way that is not tied directly to perception). For example, blackmailers are not physically black, and bluegrass is perceptually much closer to green than it is to blue. Finally some phrases that involve colour-terms may have originated from colour-relevant concepts, but modern use has rendered their current meanings so far from their origin that it is unlikely listeners would recruit colour semantics when understanding them (e.g., "blue-blood", "red herring", "yellow-bellied"). Therefore, it is not unreasonable to think that colour compounds might be interpreted as "frozen" expressions, and that colour semantics would not be activated during their comprehension. However, the current series of studies shows that the interpretation of colour terms embedded within compound nouns are in fact guided by contextual and semantic factors, in a way that is similar to the interpretation of actual colour adjectives.

Up until this point, the questions examined in this dissertation have focused on how "misleading" elements within individual words trigger the activation of semantic associates, syntactic arguments, and phrasal-level expressions. These activated elements can be considered incidental, because they do not ultimately help identify the intended referent. Further, the intended referent is eventually revealed to be fully incompatible with the initial misleading element in that the embedded verbs and embedded colour terms used in these studies in the end
only exist as part of a morphologically-complex noun. One of the hallmarks in research on processing immediacy, however, involves exploring the factors that influence the integration of *multiple* words within a noun phrase. Are there cases where the first element of a noun phrase is misleading for referential mapping?

There are some examples in English where the initial element in a multi-word expression is somewhat misleading with regard to the eventual referent. For example, the sentence *"She is on the bus"* is initially misleading in that the phrase *"She is on..."* typically describes someone being on top of something (consider the scenario that the syntactically-equivalent sentence *"She is on the car"* evokes, for instance). Instead, this phrase actually describes a state of events in which a person is *inside* a vehicle, and not on top of it. While such cases of mismatch do exist in English, they are more difficult to find *within* a multi-word description for an object, which defines the scope of this thesis. More frequent and obvious examples of this can be found in other languages, where certain grammaticalized elements must occur before nouns in noun phrases, but where the semantics of those elements might be inconsistent with the properties of the respective referent objects. Chapter 4 of this dissertation will explore the consequences of immediacy and impulsive processing in these instances.
Chapter 4

Impulsive Processing in Referential Interpretation

The studies described in this dissertation so far have explored the nature of processing immediacy using spoken stimuli, which allows for the examination of effects that occur within relatively short time frames. The specific focus of this work has involved cases where the initial element of the utterance is in some way incompatible with the ultimate concept or referent that the utterance denotes. The results to this point have indicated that encountering an embedded verb in a verb-noun compound (e.g., "popcorn") activated concepts semantically associated with the verb (e.g., a balloon), even though the activation of these associates was ultimately not helpful in identifying the target referent (Experiments 1 & 2). It was also found that encountering an embedded colour term in an adjective-noun compound (e.g., "greyhound") resulted in the consideration of referentially-instantiated alternatives that could be named by an expression that temporarily matched the signal of the embedded colour term (e.g., a grey spatula). This consideration was even greater in cases where the target referent of the compound had a colour that did not fully correspond to the colour term embedded in its name (e.g., brown greyhound; Experiments 3 & 4). Again, however, the fact that these processes were engaged was not directly helpful for interpreting the compound noun, and contextual information that could in principle suppress the consideration of non-target elements was not always effectively used. The results are therefore perhaps best understood as an indication that the core mechanisms underlying processing immediacy may operate in a way that is at least partially orthogonal to the notion of processing efficiency.

These effects have demonstrated the consequences of processing immediacy for "misleading" initial elements of single (morphologically-complex) words. The final study in this
dissertation complements these findings by examining how elements of a noun phrase are incrementally integrated and mapped onto candidates during referential interpretation. This more direct focus on reference (i.e., leaving behind issues of associate activation or indeterminacy in the form against which unfolding speech is matched) requires a shift away from examining noun targets (including compounds) to evaluating these time-course effects within phrasal-level constructions. The incremental comprehension of an unfolding sentence ultimately requires the integration of information that is encoded across multiple words, and the kinds of potentially misleading information seen within lexical items can also be observed within a phrase. Recall, for instance, the example of the semantic difference between "He is on the car" versus the syntactically similar "He is on the bus". In the former case, "He is on..." refers to someone being on top of a vehicle, but in the latter, the same construction refers to someone being inside a vehicle. "He is on the radio" refers neither to someone being inside nor on top of a radio, but rather to someone whose voice can be heard via a radio. In the last two cases, accessing the canonical meaning of "on" (i.e., being on top of something) would be "misleading", and might plausibly result in delaying or disturbing comprehension processes as listeners try to integrate the rest of the phrase with the semantics of the initial elements. In order to explore the effects of this kind of potentially misleading initial information during the mapping of noun phrases to referents, it is necessary to switch to another language where the presence of misleading information in these constructions occurs within the span of a noun phrase's elements and is not particularly marked or unusual. One kind of linguistic construction that fulfills these criteria is the classifier-noun construction occurring in certain noun phrases in a language such as Cantonese. Noun classifiers are linguistic elements that signal agreement-like grammatical relationships with their associated nouns but also encode semantic properties that are typically
possessed by the objects named by these nouns. As will be explained below, classifiers occurring in perfectly well-formed and natural noun phrases can be partially "misleading" in terms of this semantic information. The specific goal of this chapter is to discover how the information encoded in these expressions influences the process of incrementally mapping noun phrases to real-world referents.

**Noun Classifiers**

Noun classifiers are words or morphemes used in some languages to signal a category or class to which an associated noun belongs. To a certain degree, the use of classifiers mirrors the way in which articles in some Romance languages (e.g., French, Spanish) mark the grammatical gender of nouns. However, in the case of classifier languages, the corresponding noun classes are not semantically arbitrary. In standard linguistic analyses, a noun classifier "denotes some salient perceived or imputed characteristic of the entity to which an associated noun refers (or may refer)" (Allan, 1977, p. 285). Many of the world's languages use classifiers (see Allan, 1977; Gao & Malt, 2009; Saalbach & Imai, 2007; and Zhang & Schmitt, 1998 for discussion), and although the rules for classifier use vary across languages, all of these languages have constructions in which classifier use is compulsory. Cantonese Chinese is an example of a language where classifiers are obligatory whenever a numerical noun phrase or definite noun phrase is used. For example, it is ungrammatical to say the translation equivalent of "one scarf" in Cantonese Chinese; one must say "one + [classifier] + scarf", where the relevant classifier in this case (i.e., "tiu[4]") can be roughly understood as meaning something that is long, narrow, and flexible. (The number inside the square parentheses indicates the tone of the word; see Chao, 1947 for further description.) Cantonese noun classifiers can be divided into two of the major categories described in standard frameworks (see, e.g., Allan, 1977). Mensural classifiers are those used
when referring to masses or quantities (somewhat similar to English expressions like "a bushel of" or "a carton of"), whereas sortal classifiers are used when referring to individuated entities and reflect properties that are more stable and specific to their associated noun class (examples of this classifier type and associated nouns are shown in Table 4.1). In some cases, sortal classifiers have a broadly taxonomic character and are used with nouns denoting similar kinds of concepts within a coherent semantic field (e.g., "bow[6]" is used with nouns for machines). In other cases, the corresponding noun concepts are relatively unrelated in terms of semantic category but reflect common perceptual or action-based features. For example, the Cantonese classifier "ba[2]" occurs with nouns defining objects that must be grasped during use (e.g., knife, hair brush, umbrella, etc.) and "tiu[4]" (noted in the example above) occurs with nouns defining objects that are long, relatively narrow, and flexible (e.g., scarf, snake, hair, trousers, and rope). The properties of these "shape" classifiers will be the focus of the current chapter.

Crucially, the canonical perceptual information signalled by a shape classifier is not always fully reflected in the noun that follows it (Cantonese shape classifiers always occur before their associated noun). This is because, within the set of nouns that occur with a given shape classifier, there are certain nouns that fail to reflect a straightforward match with the perceptual features associated with the classifier (see Lakoff, 1986; Erbaugh, 2002; Gao & Malt, 2009; Matthews & Yip, 1994 for additional discussion). This shape mismatch is somewhat analogous to the kind of mismatch between an object's colour and the colour semantics encoded by subparts of a compound name, as explored in Experiment 4 (e.g., a brown greyhound). However, in the case of Cantonese classifier-noun pairings, the mismatch occurs between separate lexical entities within a noun phrase, rather than within a single unfolding word. Further, the classifier is not a frozen element that always coincides with the noun in question.
For example, in plural descriptions (e.g., the equivalent of "the scarves"), the shape classifier is no longer present.

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Meaning</th>
<th>Associated Noun Objects</th>
</tr>
</thead>
<tbody>
<tr>
<td>TIU[4]</td>
<td>Long, narrow, flexible</td>
<td>Snake, Scarf, Key, Problem (e.g. mathematical), Life (e.g., a human life)</td>
</tr>
<tr>
<td>ZI[1]</td>
<td>Rigid, narrow, long, small</td>
<td>Twig, Pen, Candle (all kinds), Tube of Toothpaste, Song (slang)</td>
</tr>
<tr>
<td>JEUNG[1]</td>
<td>Sheet-like, flat, broad, flexible</td>
<td>Paper (sheet), Bedsheet, Greeting Card, Table, Chair</td>
</tr>
<tr>
<td>GAU[6]</td>
<td>Lump-like</td>
<td>Rock, Cloud, Meat (if not sliced or diced), Cake (piece or slice, not whole)</td>
</tr>
<tr>
<td>JEK[3]</td>
<td>General animal classifier, also small individual objects</td>
<td>Dog, Elephant, Eye (singular), Ring (i.e., jewellery), Banana</td>
</tr>
</tbody>
</table>

Table 4.1. Example of common Cantonese sortal classifiers and associated nouns.

To illustrate some exceptional pairings involving shape classifiers, the classifier "tiu[4]" not only occurs with the nouns for scarf, snake, etc., but is also used with the nouns for a river, a human or animal life, and a mathematical problem or equation. In these cases, it is clearly
necessary to adopt a more abstract or figurative sense of winding through time or space. Further, even when consideration is limited to nouns denoting concrete physical objects, there are instances that reflect an imperfect fit with the corresponding shape classifier. One such example is that the noun for key requires the classifier "tīu[4]". Although a key can be thought of as being narrow, it is not prototypically long or flexible. Finally, it is important to note that changing the form or properties of an object does not necessarily change the associated classifier used in its description (unlike the case with adjectives). For example, "tīu[4]" is still used with the noun for scarf even if the scarf being referred to is tightly wadded up into a ball and no longer appears long and narrow. This variation in the degree of match between the semantics encoded by a classifier and the physical properties encoded by the corresponding noun concept makes classifier-noun phrases an ideal case for exploring immediacy in processing in cases where the initial element of a referring expression is somewhat misleading for referent identification. In addition, recent work has suggested that shape information can drive the temporary consideration of competitor objects in the visual world paradigm (e.g., Dahan & Tanenhaus, 2005; Huettig & McQueen, 2007; Yee et al., 2011). It is therefore reasonable to expect that any semantically-driven effects triggered by shape classifiers should be detectable in studies drawing on this technique.

Before continuing, however, it is important to clarify other ways in which shape classifiers may exert an influence during processing. As noted above, classifiers are in many ways similar to articles marking grammatical gender in languages such as French to the extent that the grammatical well-formedness of an expression involves selecting the "correct" article or classifier, regardless of semantic characteristics (i.e., just as a key is not long and flexible, the "sentry" [a feminine noun in French] is not prototypically feminine.) Prior work has
demonstrated that the grammatical information expressed by a gendered article in French can suppress the activation of referential alternatives whose names could not grammatically occur with the article. Dahan, Swingley, Tanenhaus, and Magnuson (2000) used a spoken language eye tracking methodology to examine French-speaking participants' interpretation of definite noun phrases. Listeners heard sentences such as "Cliquez sur le bouton" ("Click on the button") while viewing a visual display containing a button (Fr. "bouton", the target object) and a bottle (Fr. "bouteille"), among other objects. Because "bouton" and "bouteille" begin with the same speech sounds, listeners would normally be equally likely to fixate either of these objects until the final disambiguating portion of the noun is heard (e.g., Allopenna et al., 1998). However, because "bouton" is a grammatically masculine word and "bouteille" is grammatically feminine, the gender cue provided by the preceding masculine article (i.e., "le") was found to suppress consideration of the bottle upon encountering the target noun.

Given these findings, it is possible that encountering a noun classifier might increase the consideration of referential alternatives with names that can grammatically follow the classifier. The more interesting question for the focus of this dissertation, however, is whether and how the perceptual information encoded in shape classifiers might either streamline or mislead the process of mapping language to real-world referents as noun phrases, depending on the character of these referents. Recall the results of Experiments 3 and 4 of this dissertation, in which candidates whose properties matched the colour semantics of subparts of unfolding words received greater consideration. If this type of impulsive semantic effect is sustained across separate lexical elements, a shape classifier might strongly trigger mappings to referential candidates bearing the relevant shape contours. The strongest case of this would presumably be reflected by listeners being "misled" into considering these candidates even in cases where their
names do not constitute grammatically legitimate continuations of the classifier-noun phrase. A more moderate effect of classifier semantics might involve a "boost" in the consideration of shape-matching candidates, but only when those candidates have names that could grammatically follow the classifier in question. Alternatively, if classifier semantics exert very little or no influence on referential mapping, it may be observed that any candidate whose name could follow the classifier receives increased consideration, regardless of how well their physical properties correspond to their classifier's semantics.

One relevant issue to consider is overt understanding of a classifier: how do listeners consciously think about classifiers when directed to do so? It is clear from past work that, when asked specifically to articulate the reasons for why certain classifiers occur with certain nouns (e.g., Gao & Malt, 2009), Chinese speakers report awareness of the semantics of noun classifiers. However, it is unclear whether they actually consider the degree of fit between a noun object's properties and the semantics of its corresponding classifier when considering classifier-noun pairings. Tsang (2007) explored this question, using a rating task to determine whether native Cantonese speakers use shape information when explicitly evaluating the appropriateness of a given classifier for a set of objects. Participants were presented with sets of three pictured objects and were asked to rank each object in terms of how well they fit with a particular classifier. Each set included one object whose basic-level name was grammatically legitimate to follow the classifier, and which also possessed physical properties that constituted a good match with the classifier's canonical shape semantics. Another object in each set was also grammatically legitimate, but its shape features did not fully match the canonical shape semantics of the classifier. A third object in each set was not grammatically legitimate (i.e., it could not grammatically occur with the classifier), but did match the classifier's canonical semantics. In
addition to making the forced-choice rankings, participants were also asked to provide justifications for why they ranked the noun objects they way they did.

The findings showed that participants gave similarly high rankings to objects whose names could grammatically follow the given classifier, regardless of their shapes. In other words, shape information did not reliably influence participants' judgments of the legitimacy of specific classifier-noun pairings, even though the task forced participants to explicitly rank two grammatically viable objects (within each set of three objects) that differed in terms of whether they exhibited the canonical shape features for that classifier. Further, participants' justifications for their rankings almost never referred to shape features and instead mentioned criteria such as frequency (e.g., "Object ranked #1 is said most often"), familiarity (e.g., "That's how I hear it being said", "[X] sounds more natural / awkward"), grammatical convention (e.g., "Only [X] and [Y] are used with [CL]").

It appears then that Cantonese speakers do not appear to use semantic information in assessing the appropriateness of classifier-noun pairings, even in an explicit ranking task. However, these types of off-line tasks involving conscious reflection might underestimate the influence of perceptually-based semantic information for linguistic processing. To illustrate, although an expression like "The road goes through the desert" is not overtly understood as describing a motion event, previous work has suggested that the interpretation of these "fictive motion" sentences may trigger the latent computation of motion semantics (Richardson & Matlock, 2007). In light of this, it is interesting to consider how the shape information encoded in classifiers might drive online referential processing, thereby misleading listeners in some circumstances.
Experiment 5 tested how the consideration of various referential candidates is affected by the degree of match between candidate physical properties and classifier semantics, and drew on a methodology similar to those used in previous chapters. As participants followed spoken instructions to click on screen objects, their attention to various types of referential alternatives was monitored. Because an object’s fit with a given classifier can be evaluated in two ways – by whether its name can grammatically follow the classifier, and by whether it possesses the perceptual properties denoted by the classifier – this results in four potential kinds of referential candidates that might conceivably receive attention as a classifier-noun construction is heard. One kind of candidate was compatible with the classifier in the spoken instruction both on the basis of grammatical compatibility and of semantic features (the "G+S+") competitors; e.g., a jump rope for the classifier "tiu[4]"). In other words, the name of this candidate could grammatically follow the classifier in the target noun phrase, and its shape features fully matched the classifier’s canonical semantics. Another type of candidate had a name that could grammatically follow the classifier heard in their respective instructions, but did not possess shape features typical for that classifier (the "G+S-") competitors; e.g., a goldfish, for the classifier "tiu[4]"). A third type of candidate had a name that could not grammatically follow the classifier, but did possess shape features that matched the classifier’s semantics (the "G-S+") competitors; e.g., an unfurling roll of toilet paper, for the classifier "tiu[4]"). This is the critical competitor type for exploring the extent to which classifier semantics could temporarily induce listeners to consider candidates that are in fact grammatically incompatible. Finally, the last kind of candidate was not compatible with the classifier in the target noun phrase on the basis of grammatical compatibility or in terms of whether it had fully-matching shape features (the "G-S-") competitors; e.g., a strawberry for the classifier "tiu[4]"). The differences in the consideration
of the different candidate types will reflect how the shape semantics of noun classifiers guide real-time referential mappings.

**Experiment 5**

In this experiment, native Cantonese speakers' eye movements were monitored during the presentation of spoken instructions to manipulate depicted objects. On each trial, participants viewed a grid containing four clip-art objects, presented on a computer monitor. For all critical trials, the target object referred to in the instruction was always grammatically compatible with the classifier in the corresponding noun phrase, and possessed the canonical semantic properties for the classifier (e.g., a scarf, for the classifier "tiu[4]" [long, narrow, flexible]). Each target was paired once with each of the four competitor types described above (e.g., the G+S+, G+S-, G-S+, and G-S- competitors), creating four experimental conditions (see Figure 4.1). In addition to the target and the competitor items, each display also included two distractor items that were not related to the classifier in the spoken instruction in terms of grammatical compatibility or having appropriate shape features.

If listeners use the semantic information encoded within shape classifiers to generate referential predictions, they should consider grammatically-sanctioned competitor objects whose shape properties reflect a canonical match with the classifier occurring in the unfolding expression (i.e., those in the G+S+ conditions) to a greater degree than those whose shape properties do not match the classifier semantics (i.e., G+S- objects). Most critically, if the semantic information is salient, increased attention to the G-S+ competitors should also be observed, despite the fact that these objects do not have names that could grammatically follow the classifier. However, if classifier semantics do not guide referential mapping in this way, then increased consideration might only be observed for competitors whose names can grammatically
follow the classifier, regardless of their shape properties (i.e., both the G+S+ and G+S- objects). Finally, if neither the semantic nor grammatical information provided by noun classifiers is used to boost or damp consideration of referents during real time interpretation, there should be no influence associated with the various competitor types.

Figure 4.1. Example display from Experiment 5. The target was always a G+S+ item (e.g., the scarf for the classifier "tiu[4]"), and the competitor was one of four objects: (a) a G+S+ item (e.g., the jump rope); (b) a G+S- item (e.g., the key); (c) a G-S+ item (e.g., the flag); or (d) an unrelated item (e.g., the strawberry). The teacup and present in this display were additional unrelated objects.

Experiment 5: Method

Participants. Participants were native Cantonese-speaking students at the University of Toronto at Mississauga and received either course credit or $10. Data from 36 participants were included in the final analysis with data from an additional eight participants excluded for failing
to meet criteria for fluency in Cantonese. The criteria for determining verbal Cantonese proficiency are described in the following section.

**Materials.** The classifiers "tiu[4]", "zi[1]", and "jeung[1]" were chosen for use in the critical trials of this study because their semantic meanings are the most clearly defined in terms of shape features. As mentioned above "tiu[4]" typically occurs with nouns for objects that are long, narrow, and flexible, like snakes and ropes. "Zi[1]" typically occurs with nouns for relatively small, long, rigid, narrow objects (e.g., drinking straws, chopsticks, pens), and "jeung[1]" typically occurs with nouns for broad, flat, flexible sheet-like objects (e.g., bedsheets, paper, quilts). Four trials were assigned to each of the three classifiers, yielding 12 critical trials in total (see Appendices 3 and 4 for a full list of critical items and images, respectively). The target object on critical trials always constituted a full match with the classifier in the corresponding noun phrase, in terms of both grammar and shape semantics. Each target was presented with a competitor object and two unrelated distractors (see Figure 4.1). As described above, the competitor was systematically varied in terms of whether its corresponding noun was grammatically compatible with the classifier in the target noun phrase (compatible/incompatible) and also in terms of whether it reflected the canonical shape semantics for the classifier (match/mismatch). The assignment of competitor type (G+S+, G-S+, G+S-, G-S-) was cycled across different versions of the experiment such that every object array was paired with each type of competitor once, creating four stimulus lists, with each participant encountering a given target object only once. The two unrelated items in each object array remained constant across conditions (see Appendix 3). The object arrays were also designed to ensure any differences in consideration of the competitor across conditions could not simply result from basic (non-shape) semantic associations between target and competitor nouns (see, e.g., Huettig, Quinlan,
McDonald & Altmann, 2006). To confirm this, the semantic relatedness of all target-competitor noun pairs was calculated on the basis of Latent Semantic Analysis (LSA) cosine measures, using the pairwise comparison tool on the LSA website (lsa.colorado.edu; topic space: General Reading up to 1st year college). Recall that LSA measures provide an index of conceptual relatedness that is not simply a product of the co-occurrence of words in sentences, and so the relatedness measures are assumed to hold across languages to a reasonable degree. (To my knowledge, there is no comparable resource for Cantonese.) No reliable differences in LSA cosine measures for target-competitor pairs were found across conditions ($p > .69$).

Two spoken Cantonese instructions were used for each trial. The instructions were pre-recorded by the author, a Cantonese-English bilingual with native-like Cantonese pronunciation as judged by native Cantonese-speaking pilot participants. These instructions had the following structure: "Look at the cross. Take [object X], and put it in the square above / below / next to [reference object A]. Then take [object Y], and put it in the square above / below / next to [reference object B]." On experimental trials, the critical instruction was always uttered first, followed by a filler instruction referring to a new object. Recall that due to the nature of Cantonese syntax, the classifier (CL) for each noun immediately precedes the noun phrase in these sentences. For example, if the instruction was to take the scarf and put it in the square above the cloud, the Cantonese structure would be: "Take-[CL]-scarf, put-at-[CL]-cloud-above-[possessive]-square." The target noun did not have overlapping onset sounds with nouns that would be used to denote any other screen object. The placement of the various object types (target, competitor, distractors) within the grid display was counterbalanced on critical trials. The eventual destinations of the target objects were also counterbalanced.
An additional 24 filler trials were included to prevent participants from discerning the purpose behind the study. For example, trials were included in which two unmentioned objects shared the same classifier to neutralize any suspicion that objects sharing a classifier with other scene objects might be more likely to be mentioned. To further divert attention away from the fact that targets often shared a classifier and shape characteristics with displayed alternatives, several filler trials included similarly-shaped objects (e.g., a mop and a broom, a crayon and a rocket ship) that did not share a classifier. Also included were fillers whose targets were differentiated from otherwise-identical distractors by using descriptions containing color or size adjectives, as these types of distinctions are very salient in comparison to the subtle contrasts under investigation. Finally, sentences in filler trials often contained sortal classifiers specifying either animacy (e.g., "jek[3]"), very loosely defined perceptual categories (e.g., "gau[6]"; lump-like, no size specified), or taxonomic class (e.g., "bou[6]"; machine-like), as well as a general classifier that has little to no consistent semantic meaning (e.g., "go[3]"). This ensured that participants would not strategically attend to perceptual features and consequently could not anticipate a particular type of classifier or reference object.

Experiment 5: Procedure

Participants were tested individually. Prior to beginning the experiment, participants completed a language background questionnaire detailing their age of acquisition of Cantonese and English and their level of oral proficiency in Cantonese. All participants included in the final sample reported that they had been speaking Cantonese since birth, and indicated that they had good proficiency at producing and comprehending spoken Cantonese. After completing the language background questionnaire, each participant was seated in front a computer screen on which the experimental stimuli were displayed and was then fitted with an EyeLink II head-
mounted eye tracking system (SR Research Ltd.). Each trial consisted of an image of a 5 x 5 grid with a fixation cross in the center square. The inner 3 x 3 grid contained objects that could be clicked and dragged in the display and the identity of these objects varied from trial to trial. At the beginning of each trial, these objects were presented in the four squares immediately above, below, and to the left and right of the central fixation square, and participants manipulated these objects as directed by the spoken instructions. On critical trials, one of the depicted objects was the target referent, another was the competitor object, and the remaining two were unrelated distractors. In addition to these objects, there were four reference objects in the grid (a sun, a moon, a star, and a cloud, not depicted in Figure 4.1). These reference objects remained constant in their locations at the furthermost corners of the grid throughout all the trials and were distinguished from the moveable items by a darker background square. Participants were told that they would hear instructions telling them to use a computer mouse to pick up specific objects on the screen and move them to another location on the grid, and were asked to carry out the spoken instructions as quickly and accurately as possible.

Following calibration, two practice trials were presented to familiarize participants with the task. After the participant carried out the first instruction in a trial, the experimenter pressed a key to initiate the second instruction. The experimenter also manually advanced the display at the end of each trial once the participant had successfully completed both spoken instructions. A fixation point appeared in the middle of the screen between trials to allow drift corrections to be made if necessary. Calibration checks were performed after every 13 trials.

Upon completion of the eye tracking portion of the experiment, participants were shown pictures of 36 common objects, and asked to name them out loud in Cantonese using a numerical noun phrase. This served as a secondary validation of participants' self-reported proficiency at
Cantonese. Participants who displayed a very non-native-like accent or who produced more than five incorrect classifier-noun pairings were excluded from the analyses (eight participants total).

**Experiment 5: Results**

DataViewer analysis software (SR Research, Ltd.) was used to calculate real-time fixations to depicted objects starting from onset of the classifier in the critical instruction and ending 1200 ms later, reflecting the approximate average endpoint of the noun phrase (1121 ms: average classifier duration = 371 ms, average noun duration = 750 ms). Figure 4.2 shows the average probability of fixating the various scene object types over time in the four experimental conditions. Because the displays contained two distractor objects, fixations to a single randomly-selected distractor are depicted. As in the previous experiments in this dissertation, fixation probabilities were first submitted to a logit transformation to facilitate statistical analysis and are depicted on a logit scale. The depicted time interval begins at 200 ms following classifier onset in view of the approximate time lag for eye movements to reflect information in the speech signal. The vertical line that bisects each graph marks the end of the classifier region (classifier offset plus 200 ms).

The data show that target fixations increased as the noun phrase unfolded in time, reflecting the incrementality that characterizes referential interpretation in spoken language. The average proportion of fixations to the unrelated distractor object varied little across conditions, reflecting the fact that the distractors were neither phonologically nor semantically related to either the classifier or noun in the spoken instruction. Target advantage scores and competitor advantage scores were once again calculated for each region of interest, using the same method as earlier in this dissertation.
**Consideration of Target Object Relative to the Competitor.** Inspection of the graphs suggests that targets were more easily differentiated from competitors in conditions where the name of the competitor could not grammatically follow the classifier in the spoken instructions. This is reflected in the larger, more consistent gap between target and competitor fixations in the two right panels compared to the two left panels. This gap is most apparent after noun onset. To conduct a statistical evaluation of the overall results, the mean likelihood of fixating the target vs. competitor object was calculated for each participant. Separate analyses were performed for (i) the classifier region and (ii) three successive 100 ms intervals within the noun, beginning at noun onset. The latter analyses capture the earliest moments of noun processing and any residual processing effects related to the classifier. These values were then submitted to a mixed-model analysis of variance (ANOVA), with grammatical match (match/mismatch between classifier and competitor name) and shape match (match/mismatch between competitor properties and canonical classifier meaning) as within-factors. Stimulus list was included as a between-participants factor (Pollatsek & Well, 1995). No significant effects of either grammatical match or shape match were observed in the classifier region or during the first 100 ms of noun processing (all \( p > .18 \)). Within the second 100 ms of noun processing, there was a significant main effect of grammatical match (\( F(1, 32) = 6.61, p < .05 \)) but no significant effects of shape match (\( p > .54 \)). In the third 100 ms of noun processing, the analysis yielded a marginally significant effect of grammatical match (\( F(1, 32) = 3.54, p = .07 \)). No significant effect of shape match was observed in this region (\( p > .69 \)), and there were no significant Grammatical Match x Shape Match interactions in any of the analysis regions (all \( p > .62 \)).

**Consideration of Competitor Object Relative to the Unrelated Distractor.** Inspection of Figure 4.2 suggests that, in the classifier region, there was a slight tendency to consider the
distractor object more than the competitor in the classifier region in the G-S+ and G-S- conditions (right panels). This may reflect some unintended greater visual interest associated with the distractor. Importantly, however, this pattern was reversed in the G+S+ and G+S- conditions (left panels), where it was the competitor that received more consideration. A tendency to fixate the competitor over the distractor was also apparent at later points in the noun region in the G+ conditions (right panels) compared to the G- conditions (left panels).

Figure 4.2. Likelihood of fixating screen objects across the four experimental conditions in Experiment 5. Probabilities are depicted on the logit scale. The vertical line on each graph marks the average end of the classifier in the classifier-noun phrase.
To provide a statistical analysis of these patterns, the mean likelihood of fixating the competitor object relative to the distractor was calculated within the same time intervals used in the preceding analyses and using the same ANOVA model. In the classifier region, the results yielded a significant effect of grammatical match ($F(1, 32) = 6.22, p < .05$), reflecting greater consideration of the competitor relative to the distractor when the competitor's name was grammatically compatible with the classifier. No significant effect of shape match was found in the classifier region, nor was the interaction significant ($p's > .39$). A marginal effect of grammatical match was found in the first 100 ms of noun processing ($F(1, 32) = 3.30, p = .08$). As in the previous region, there was no effect of shape match, nor was the interaction significant ($p's > .27$). In the second 100 ms of noun processing (667 ms to 766 ms), there was a grammatical match ($F(1, 32) = 6.78, p < .05$), but there was again no effect of shape and no interaction ($p's > .38$). Finally, in the third 100 ms of noun processing, no significant differences were observed across conditions ($p's > .20$).

**Summary**

The goal in the current series of studies was to explore whether and how the perceptual information encoded in Cantonese "shape classifiers" – a type of sortal noun classifier – influences the real-time mapping of spoken noun phrases to referent objects. Although these (and other) classifiers behave like markers of abstract grammatical gender from a morphosyntactic perspective, the pairing of shape classifiers to noun classes is not semantically arbitrary, and instead reflects canonical shape features of the corresponding noun class. On the other hand, these semantic relationships are not fully reliable, and there are "exception" cases where noun objects do not possess all of the canonical perceptual properties denoted by their associated classifiers. Consequently, a question of particular interest is whether the shape information
encoded in the classifier might lead listeners to momentarily consider shape-appropriate referential alternatives, perhaps even when those alternatives are not grammatically sanctioned to occur within the noun phrase.

The findings from Experiment 5 indicate that with regard to the real-time interpretation of referential classifier-noun phrases, the contribution of shape classifiers is most strongly reflected in their agreement-like grammatical properties and not their shape semantics. When the name of the competitor object in the display could grammatically follow the classifier in the spoken instruction, listeners (i) were less effective at differentiating the target object from the competitor as the noun unfolded, and (ii) were more likely to consider a competitor relative to an unrelated distractor both during the classifier and as the noun phrase continued. This lingering effect of grammatical match is intriguing because the sounds in the noun onset were in principle sufficient to distinguish the target referent from all alternatives (names of display objects did not have overlapping onset sounds). Consequently, the elevated consideration of grammatically-compatible competitors during noun processing suggests these agreement-like cues are comparatively robust. This outcome is broadly compatible with recent results showing prolonged effects of lexical-level factors on real-time comprehension. For example, in a study of Dutch listeners, Dahan and Gaskell (2007) found that a high-frequency competitor ("koffie" [coffee]), by virtue of its increased activation level, continues to "compete" with a low-frequency target ("koffer" [suitcase]) beyond the point where it no longer matches the unfolding phonological input.

However, the statistical analyses did not detect any effects related to the match/mismatch between the competitor object and the canonical semantics of the classifier in the spoken instruction. The fact that the analyses did not reveal even a subtle influence of shape semantics is
striking given that information about the shape features of objects was directly available in the visual display. In contrast, the information underlying listeners' increased looks to grammatically compatible alternatives (i.e., contingencies between classifiers and the names for referents) cannot be gleaned from perceptual features of objects. It appears that when target referents possess physical properties that constitute a good match for their respective classifiers' semantics, the consideration of referential alternatives is not guided by the semantic information encoded by the classifier, but rather by its encoded grammatical information.

**Classifier Semantics and Referential Mapping of Atypical Cases.**

The findings from this study must, however, be regarded in light of related results from a similar eye-tracking study in Tsang (2007). In this visual world task, listeners followed the same kind of spoken instruction as in Experiment 5, except the target nouns referred to objects whose names could grammatically follow the classifier in the instruction, but whose perceptual properties failed to constitute a good match with the classifier semantics. In other words, the experiment in Tsang (2007) used G+S- objects as targets. In this case, an effect of shape match was observed, such that any objects with perceptual properties denoted by the encountered classifier received increased consideration, even when their names could not grammatically follow the classifier (i.e., both G+S+ and G-S+ competitors received greater consideration). This suggests that, under some circumstances, classifier semantics do guide referential mapping, and can even momentarily "mislead" the comprehender into considering alternatives that should be ruled out on the basis of grammatical information.

Why then, might an effect of shape match/mismatch be detected when processing G+S-classifier-noun parings, but not G+S+ classifier-noun pairings? One reason might be that, because the majority of objects within a classifier category do in fact possess the properties
denoted by that classifier, the classifier-noun pairings used as target phrases in Tsang (2007) are relatively rare. It may be the case that encountering these relatively "atypical" targets results in a kind of "boggle" or delay in processing. In support of this idea, the findings from Tsang (2007) also revealed that an effect of grammatical match (which would be expected, given Dahan & Tanenhaus' (2005) findings and the findings from Experiment 5) was absent. Listeners did not seem to consider referential alternatives more when their names could grammatically follow the encountered classifier. In addition, the design of the visual-world experiment in Tsang (2007) was such that one of the referential competitors presented with the target referent possessed properties that were indisputably a better match for the canonical semantics of the relevant shape classifier. These factors combined might have had the result of increasing the role of classifier-encoded shape information during referential mapping.

Further, a mismatch-driven disruption in target identification processes might allow more subtle effects of classifier semantics to be detected. For instance, although the statistical analyses in Experiment 5 did not reveal any effects of shape match, certain features of the fixation profiles suggest there may have been some weak influence of shape classifier semantics on real-time comprehension. For example, the gap between target and competitor fixations appears to have been narrowed in the G+S+ condition (top left panel of Figure 4.2) compared to the G+S- condition (bottom left panel of Figure 4.2), suggesting greater consideration of the shape-matching competitor object. Similarly, competitor objects appeared to receive more sustained attention relative to distractor objects in the G+S+ compared to the G+S- condition. Therefore, it is possible that this hint of shape influence arose from a latent sensitivity to classifier semantics, which can be inflated in the cases of more atypical classifier-noun pairings as used in Tsang (2007).
The Uptake of Classifier-Encoded Information.

The results from Experiment 5 did not reveal any sensitivity to the semantics of shape classifiers during online comprehension. Recall that, in previous studies of real-time interpretation in English, an effect of shape information was detected during the interpretation of nouns (e.g., Dahan & Tanenhaus, 2004, Yee et al., 2011), which obviously convey many semantic features other than shape. The semantic content of shape classifiers, in contrast, conveys little apart from information about shape features. As a result, one would expect shape information to be highly salient and therefore be a potentially misleading source of information during incremental referential interpretation (a point also consistent with the claim that language is grounded in perceptual meanings, e.g., Zwaan, Stanfield, & Yaxley, 2002). Experiment 5 showed, however, information encoded in shape classifiers did not increase attention to candidates with the appropriate shape properties, and that a mismatch between classifier semantic and referent properties did not delay or otherwise hinder referential mapping.

These findings suggest that the semantics of classifiers are not processed in the same way as those of embedded verbs or adjectives within compound nouns, despite the possibility that they also provide putatively "conflicting" information. Instead, shape classifiers appear to serve as grammatical cues only, with the semantic information encoded in them largely untapped by listeners during referential processing. This difference between the processes triggered by embedded verbs and adjectives versus shape classifiers will be addressed further in the final chapter of this dissertation.
Chapter 5

Conclusions

A wealth of research in the domain of spoken language processing has demonstrated that the comprehension of unfolding speech is an immediate process, in which new information is quickly integrated into old, guiding expectations about upcoming elements. At the lexical and phrasal level, the strategy of interpreting unfolding speech sounds as soon as they occur is typically thought to contribute to the overall efficiency of interpretive processes. However, there exist cases where the initial element of a noun phrase or morphologically-complex word is "misleading" in the sense that it triggers the consideration of linguistic or referential alternatives that seem more semantically or syntactically inappropriate than standard competitors in an activation-based framework. These cases provide a useful tool to explore the information sources and mechanisms that underlie incremental interpretation. As discussed in the Introduction (Chapter 1), if comprehension occurs via a strictly "rational" system whereby available top-down information serve to eliminate or suppress bottom-up activation that is inconsistent with already-encountered input, this should prevent the impulsive activation of elements that are irrelevant to the processing of the intended target word or referent. The current dissertation examined two issues that help reveal whether impulsive activation indeed occurs under this rational framework, and specifically sought to explore 1) the scope of this activation, and 2) the ability for various contextual constraints to suppress this unintended activation.

Can Impulsive Activation be Observed Across Lexical Categories?

One way to examine the scope of impulsive activation is to determine whether such activation can occur across lexical categories (e.g., from nouns to verbs). The first series of experiments (Chapter 2) used compounds with "embedded" verbs (e.g., "popcorn", in which the
form of the first element of the compound also corresponds to a verb) to explore whether encountering those verb subparts (e.g., "pop" in "popcorn") triggered the activation of verb-relevant noun concepts (e.g., BALLOON), even when the subpart (i.e., a verb form) belonged to a different syntactic category than the entire word (i.e., a noun). This situation is relevant for examining how activation spreads across different conceptual/grammatical categories during naturalistic language comprehension, and also the extent to which the activation of verb-related meanings is reduced if a noun was expected. Experiment 1 found that semantic associates of embedded verbs were indeed activated, even though there was a strong expectation (from the sentence frame used consistently on every trial) that instructions would end in simple noun phrases (e.g., "Click on the popcorn"). Experiment 2 provided a comparison case where these expectations were changed. This was achieved by using instruction sentences that took the form of "Mary thought about [target compound]". The use of a "psych" verb such as "thought about" in target instructions reduced the expectation that the target object would be depicted in a way that specifically corresponded to the description of that object (e.g., it is acceptable to pair a depiction of intact balloons with the phrase "Mary thought about popping balloons", but not with "Click on the popping balloons"). Further, this change in the syntactic structure of target instructions also allowed for the inclusion of filler items that used gerundive adjective phrases (e.g., "Mary thought about sleeping cows") or present participle verb phrases (e.g., "Mary thought about delivering pizzas") in their instructions. Although the target instructions always ended with an unmodified noun (e.g., "Mary thought about popcorn"), the gerundive adjective and present participle filler sentences emphasized the possibility that an unfolding element might actually be a verb-derived word preceding the noun, in which the verb-related meaning is relevant for target identification. These fillers also reduced the expectation that simple noun
phrases would always be used to name targets. Finally, these modifications combined resulted in reduced certainty about where the target noun actually began.

Under these conditions, the activation of concepts that were semantically associated to the target noun was observed as in the prior experiment; however these effects were delayed relative to Experiment 1. This result presumably arose from the greater variation in the syntactic forms of the spoken instructions used in Experiment 2 and the reduced certainty about when the onset of the target noun was, which may have delayed listeners' identification of a depicted object corresponding to the target compound in the sentence. Additionally, unlike in Experiment 1, there was some evidence that the argument structures of unfolding embedded verbs were accessed, in turn influencing semantic associate activation. That is, semantic associates whose names could occur as either the object or subject of an embedded verb (e.g., BALLOON, for the compound "popcorn") received greater consideration than those whose names could only occur in subject position (e.g., FISH, for the compound "swimsuit"). Apparently, when target identification was delayed or made less certain to some degree as it was in Experiment 2, knowledge about embedded verb argument structure is more clearly reflected in candidate consideration. However, this activation is damped in cases where target identification occurs with greater speed or certainty (e.g., Experiment 1). These findings indicate that the kinds of "impulsive" processing that occur during comprehension may be more widespread than previously thought. In addition to the noun-to-noun associate activation effects seen when examining partially overlapping word forms (e.g., "logs" \(\rightarrow\) "lock" \(\rightarrow\) KEY; Yee & Sedivy, 2006), Experiments 1 and 2 showed conceptual activation across syntactic categories, from verbs to nouns. Experiment 2 further demonstrated that under certain conditions, this verb-driven activation of associated noun concepts can be constrained based on knowledge about verb
argument structure. Taken together, these results point to the possibility that latent conceptual and syntactic knowledge are activated at various points during spoken language comprehension, and that the course of this activation can be shaped or suppressed by contextually-driven pragmatic expectations.

**Can Phrasal Descriptions be Impulsively Activated by Word Subparts?**

Another avenue to explore the scope of impulsive activation is to examine compounds whose initial elements match an adjective form. Given that embedded verbs in the first series of studies were observed to act momentarily like "real" verbs, the next question was whether embedded colour terms in compound nouns such as "greyhound" might similarly be briefly considered as actual adjectives, and trigger the kinds of processing effects observed when an unfolding adjective occurs within a phrasal expression. If so, it would be possible to conclude that, in addition to lexical and semantic activation, the activation of phrasal-level expressions can also be triggered by encountering embedded word subparts. In other words, the forms that listeners consider in the early moments of so-called spoken word recognition might also include multi-word units that are linguistically relevant given the contextual situation. This result would provide a novel source of evidence supporting the idea that there are dense interactions among referential, lexical, and semantic processes during real-time interpretation.

Experiment 3 was carried out to explore this possibility, and found that encountering the adjective portion of an unfolding colour compound (e.g., "greyhound") led to consideration of competitors that could be named using modified noun phrase containing the embedded colour term (e.g., a grey spatula). This consideration was greatest when the competitor was part of a contrast set (e.g., a grey spatula and a red spatula). Quite strikingly, however, some consideration was given to the competitor even when it was the only object of its kind in the visual scene. In
this case, there was little pragmatic reason to expect the competitor to be named using a modified noun phrase rather than simply an unmodified noun (e.g., "the spatula"). Nonetheless, consideration of the competitor was still observed as the colour compound unfolded. Further, in Experiment 4, when the target object's colour did not correspond to the colour term in its name (e.g., a brown greyhound), the competition effects described above increased such that the competitor received as much consideration when the contrast set was absent as when it was present. This latter result suggests that perceptual features of candidates were immediately and automatically assessed against the semantics of the embedded adjective even when the process did not require this sensitivity (i.e., at the level of linguistic form, the task of matching the unfolding speech to the sound pattern in the word "greyhound" should in principle not depend on the referent's greyness). As a consequence, a mismatch between the colour of the target object and the colour evoked by the first element of its name slowed processing by increasing the consideration of alternatives that provided a better perceptual match. The findings from this series of experiments further broaden the scope of the types of processes that occur as a result of immediacy during comprehension. In addition to cross-category associate activation (Experiments 1 & 2), hearing an unfolding compound can also trigger the access of potential phrasal level descriptions of discourse-relevant objects. Additionally, the semantics of elements "embedded" at the beginning of a compound were impulsively activated and assessed against the properties of candidate objects, and the strength of this activation was modulated depending on the visual properties of the target object. The findings from Experiments 3 and 4 further demonstrate that i) activation resulting from immediacy in comprehension occurs across various linguistic and semantic levels, and ii) these streams of activation dynamically interact with listeners' pragmatic expectations that are built out of the context of the communicative task at
Does Potentially Misleading Information Always Trigger Impulsive Activation?

The first two series of experiments (Chapters 2 & 3) used indeterminacy in the match between the unfolding speech signal and different linguistic forms (i.e., whether the speech sounds correspond to the first constituent of a compound noun vs. an individual verb or adjective) as a means to enrich our understanding of immediacy in (i) the activation of conceptual associates and (ii) the activation of lexical forms. As described in the introduction, a third prominent approach to studying immediacy in spoken language interpretation concerns incremental referential interpretation and involves cases in which there is no such temporary indeterminacy in the names of potential candidates. Rather, this question is typically studied by examining the uptake of information from multiple elements within a noun phrase. Experiment 5 of this dissertation (Chapter 4) thus focused on instances where potentially-misleading information came not from any part of the signal getting "mistaken" for something else at the level of linguistic form, but rather from a mismatch between the semantics of one noun phrase element and those of a subsequent element. In order to explore this, it was necessary to move away from working in English in favour of a language in which these cases of mismatch are more common and useful for our purposes. Noun phrases in Cantonese Chinese containing classifiers satisfied these requirements. Classifiers serve a grammatical agreement-like function, but also carry their own semantics, typically signalling some semantic, functional, or perceptual category to which their associated noun concepts belong. However, because their function is grammatical, the pairing of classifiers to nouns is ultimately governed by grammatical rules, and cases of semantic mismatch between classifier meaning and noun object properties can exist (e.g., the classifier "tiu[4]" grammatically precedes the Cantonese noun for "goldfish", even hand.
though "tıu[4]" denotes a category of long, flexible, narrow things, and a goldfish does not fit canonically into that category).

Of particular use for the purposes of this dissertation are Cantonese shape classifiers, which encode perceptually-based semantics typical of the noun objects with which they occur. These classifiers were used to compare processing in cases where there was a match or mismatch between classifier semantics and noun object properties, and whether a mismatch between these elements might have resulted in "misled" referential processing. The results from Experiment 5 showed that the effect of the classifier was primarily to boost the referential consideration of candidates with names that could grammatically follow the encountered classifier. The degree of consideration was not affected by the goodness of fit between a candidate's shape properties and the shape semantics of the classifier. It appears that listeners are either not sensitive to the semantic relationship between classifiers and their associated noun objects' properties, or that the grammatical constraints encoded by classifiers are strong enough to override any effects that might arise from this semantic relationship.

**Why Do Cantonese Classifiers Not Entail Misleading Interpretations?**

Taken together, the results from the first two empirical studies (Experiments 1-2, and 3-4) showed that as listeners encountered an unfolding expression whose initial part is in some way at odds with the expression as a whole, semantic, syntactic, and lexical features of the "conflicting" part were indeed immediately, but incidentally, activated, and this activation could in some cases delay or hinder successful processing of the target word. Further, the findings demonstrated that a listener's expectations (such as expectations about the syntactic structure of an utterance, or about how a speaker might refer to co-present objects) were not sufficient to eliminate consideration of conceptual or referential alternatives whose names, descriptions, or
perceptual properties were momentarily consistent with part of an unfolding word. However, the final study (Experiment 5) indicated that semantic information encoded within pre-nominal elements such as shape classifiers did not result in the same activation of semantically-appropriate alternatives. Rather, it seems that the classifiers’ effect on referential attention was such that only grammatically-sanctioned alternatives were considered as classifier-noun phrases unfold, regardless of how well the perceptual properties of those alternatives matched the perceptual semantics of the respective classifier.

How can these findings be reconciled? One important consideration is that the source of the misleading information was different in Experiment 5 than in Experiments 1-4. In the first four experiments, due to the dynamic nature of speech, the sounds in the unfolding target compound (e.g., "flyswatter", "greyhound") were momentarily compatible with the sound pattern of not only the intended noun but also other lexical or phrasal alternatives ("fly", "flying", "grey spatula", "grey dog", etc.). It is this overlap at the level of linguistic form that drove the increased eye movements to competitors that were semantically (Experiments 1 & 2) or perceptually (Experiments 3 & 4) related to the initial word subpart. However, in Experiment 5, the misleading information came not from the fact that the unfolding signal momentarily matched some other word or phrasal description, but rather from the semantics of the classifier clashing with the perceptual properties of a subsequent referent. In Experiments 1 through 4, the initial element of the compound could also stand on its own as a verb or adjective, and its meaning can be (and often is) fully independent of the noun that followed them in the compound. Shape classifiers, on the other hand, do not have any meaning other than denoting a shape category corresponding to their associated noun concepts. Even in cases where they are used independently instead of within a noun phrase, they function as anaphors (i.e., referencing an
absent noun). Therefore, although both classifiers and the first constituent of a compound noun share the potential to be a "misleading" initial element, the analogy between the two no longer stands when it comes to their status as rich and independent linguistic units. The reduced "autonomy" of the classifier as compared to elements embedded at the start of compounds might correspond to a decrease in the importance of classifier semantics during comprehension, in turn reducing the kinds of impulsive effects they trigger upon being heard.

An alternative explanation for the lack of incidental semantic activation in Experiment 5 arises from the nature and function of classifiers in Cantonese Chinese. Unlike verbs (Experiments 1 & 2) and colour terms (Experiments 3 & 4), classifiers serve more as closed-class function units than semantically-rich content words or adjectives. Despite being aware of classifier semantics when directly to elaborate on them (Gao & Malt, 2009), native speakers do not appear to consider these semantics when evaluating the relationship between classifiers and nouns. As such, it is possible that classifiers have been sufficiently grammaticalized within the language such that they function as agreement markers, making their semantic effects weaker during the kinds of implicit comprehension tasks such as that employed in Experiment 5. This grammaticalization would by definition involve some degree of "semantic bleaching", with the result that native listeners end up largely insensitive to the degree of perceptual match between shape classifier semantics and noun object properties, unless the task exaggerates or emphasizes this goodness-of-fit in an explicit way (e.g., Gao & Malt, 2009) or involves target referents whose properties consistently mismatch the semantics of their associated classifiers (Tsang, 2007).

A classifier's status as a closed-class grammatical element (vs. an open class word) also results in classifiers preceding nouns with much greater frequency in speech than do the
embedded adjectives or verbs used in Experiments 1 through 4 (for example, Erbaugh, 2002, found that classifiers are produced in over 60% of utterances that include nouns). As a result, the processing of classifier-noun pairings might be highly automatized. On this view, listeners’ expectations about potential upcoming nouns might be so strongly "trained" that any influence of classifier semantics would not be apparent under typical circumstances. Although the correspondence between the semantics of a classifier and the properties of the associated noun object may not always be consistent, the grammatical rules governing the pairing of a noun to a classifier and noun are invariant: regardless of classifier semantics, a pairing is either grammatical or ungrammatical. Both the semantic-bleaching account and frequency-based explanation are consistent with the finding that the grammatical information encoded by classifiers had stronger effects in guiding the consideration of referential alternatives during comprehension.

**Can Impulsive Activation be Constrained?**

The evidence presented in this dissertation suggests that syntactic and pragmatic cues, whether encountered in the prior input or available in the immediate situation, do not necessarily eliminate the activation of semantic alternatives that are incompatible with those cues. Although pragmatic expectations can decrease the degree to which semantic alternatives are activated (e.g., Experiment 4), this activation was not seen to be completely suppressed. This is consistent with certain claims including Kuperberg's (2007) proposal that comprehension involves two simultaneous, separate, and competing streams of processing: one driven by the syntactic expectations built up combinatorially over the course of the sentence, and the other (possibly faster) stream driven by the stored semantic information associated with the individual elements of the sentence (see also Bever et al., 1998).
The idea that semantic and contextual cues merely constrain the activation of alternatives to varying degrees, rather than fully suppress that activation, is also reflected in findings from research on lexical ambiguity resolution and from work on sentence comprehension. In Swinney's (1979) experiment examining how comprehenders process lexically ambiguous words, semantically biasing one meaning of a homophone (e.g., biasing the insect meaning of "bug") increased the activation of the biased meaning during a lexical decision task, but did not eliminate the activation of the alternate meaning (e.g., the spy-device meaning). Using a visual world paradigm, Huettig and Altmann (2004) and Mirman et al. (2008) also found that contextual cues can increase the consideration of the subordinate meaning of a homophone. These studies also showed that while the dominant (i.e., unbiased) meaning of the homophone was somewhat suppressed, it was nonetheless still active despite the incongruent context. Research from the sentence processing domain has also demonstrated that bottom-up processing is not eliminated by potentially constraining top-down knowledge. Tabor et al. (2004) found increased reading times for sentences which contained locally-coherent elements even when the syntactic cues contained in the preceding elements should have ruled out the locally-coherent parse. Konieczny et al. (2009) and Kukona et al. (2011) found similar effects for spoken language comprehension using a visual-world paradigm. The findings from the current dissertation demonstrate that the scope of this kind of impulsive activation extends to word-level processes, and show that while syntactic and pragmatic cues can exert probabilistic constraints on the degree to which conceptual associates and potential phrasal-level descriptions are activated, they do not act as absolute filters that necessarily eliminate the impulsive activation altogether.
How Might a Rational Comprehension System be Characterized?

One important theme throughout the current studies is the idea that processing immediacy and its consequences in terms of "activation" may not be necessarily beneficial for the end goal of comprehending language in an optimally efficient manner – in most cases, incidental activation was seen to delay processing in some way. One could therefore argue that these processing "boggles" make for a poor trade off for speed, and that a process that delayed aspects of interpretation until slightly further downstream the utterance (e.g., 1000 ms after word onset, at which point the entire compound would have been heard) would eliminate much of the incidental activation observed in these studies. On the other hand, it is important to remember that these effects are relatively short-lived, and comprehension is typically resolved without much difficulty. For example, the kinds of inadvertent activation seen in the current studies are somewhat analogous to cross-linguistic lexical activation seen in bilinguals, in which hearing an unfolding word in the active (i.e., relevant) language results in the temporary activation of a lexical item sharing similar speech sounds in the inactive (i.e., irrelevant) language (e.g., Chambers & Cooke 2009; Spivey & Marian, 1999). Despite this additional lexical competition that occurs subconsciously in the bilingual mind, these listeners comprehend the linguistic input with little difficulty.

Further, there may be situations in which incidental activation could be considered a useful feature of language comprehension. Take the case of metonymy, for example, in which a subpart, property, or associate of a concept is used to refer to the whole. A movie critic who utters "Despite the vast innovation in special effects, Hollywood has really gone to the dogs in the last decade," is using the word "Hollywood" to refer to the general film industry, rather than the municipality itself. In this case, impulsively activating the concepts of SHOW BUSINESS
and ACTORS upon encountering "Hollywood" arguably helps the listener make sense of the utterance more quickly. The common use of figurative or metaphoric language results in another area in which maintaining alternate meanings and conceptual associates in activation can increase processing efficiency. In fact, it is the very fact that these forms of language activate both the direct, literal meaning and an alternate interpretation that makes these linguistic devices useful. For instance, in order to fully understand metaphors about anger, such as "she jumped down his throat" or "he blew his stack", listeners must not only access the literal meaning of the utterance, but also activate more general conceptual associations between the idea of anger and animal behaviour and heated liquids under pressure, respectively (e.g., Gibbs et al., 1997). Full comprehension of sarcastic statements (e.g., "Gee, you sure are a ray of sunshine today", as uttered to a depressed person) also requires listeners to simultaneously access two different meanings – in this case, both the literal and opposite meaning of the utterance.

In some cases then, the latent activation of non-stated forms and meanings is actually required as part of successful comprehension. Further, beyond these non-literal forms, speakers display a wide and often unpredictable range of communicative behaviours. Given the possibility for disfluencies, self-repairs, and ungrammatical utterances by children or non-native speakers, amongst others, it may be in fact inaccurate to construe an efficient and "rational" system as one that would actually suppress the kinds of incidental activation observed in the studies described in this dissertation. Rather, is possible that efficiency in spoken language comprehension might be better reflected by a system that considers a very broad range of alternatives, and only seems "overzealous" when examined using the relatively predictable, grammatical, and fluent kinds of language used in laboratory studies. Under this view, it may be more productive to re-frame the conception of a "rational" comprehension system as one that allows for impulsive activation
because a) it may be required to understand the linguistic cases described above, and b) in cases where impulsive activation is ultimately unhelpful, it nonetheless does not appear to disrupt comprehension in a particularly long-lasting and unrecoverable way.

In sum, the findings from the work presented in this dissertation underscore the wide – and sometimes unexpected – range of knowledge that is accessed during spoken language comprehension. From conceptual associates of fleeting word subparts, to locally-coherent syntactic structures, to phrasal-level descriptions of potentially-relevant referential candidates, these streams of information are activated and modulated dynamically as linguistic input is encountered. Far from being a "catch up game", the results from this dissertation point to a language comprehension mechanism that immediately taps into various levels of representation as the input unfolds.
References


Appendix 1

Critical trial object arrays in Experiments 1 and 2. Screen objects were presented as single objects in Experiment 1, and three duplicated objects (e.g., three identical ring boxes) in Experiment 2.

<table>
<thead>
<tr>
<th>Target</th>
<th>SA Competitor</th>
<th>Shape-Matched Control</th>
<th>Unrelated Distractor 1</th>
<th>Unrelated Distractor 2</th>
<th>Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ring box</td>
<td>Bell</td>
<td>Hat</td>
<td>Camera</td>
<td>Coffee maker</td>
<td>Both</td>
</tr>
<tr>
<td>Flyswatter</td>
<td>Air plane</td>
<td>Pickaxe</td>
<td>Strawberry</td>
<td>Teddy bear</td>
<td>Both</td>
</tr>
<tr>
<td>Waterfall</td>
<td>Plant</td>
<td>Lamp</td>
<td>Cat</td>
<td>Racquet</td>
<td>Both</td>
</tr>
<tr>
<td>Rolling pin</td>
<td>Tire</td>
<td>Plate</td>
<td>Pig</td>
<td>Gift box</td>
<td>Both</td>
</tr>
<tr>
<td>Driveway</td>
<td>Tractor</td>
<td>Roller skates</td>
<td>Laptop</td>
<td>Milk</td>
<td>Both</td>
</tr>
<tr>
<td>Park bench</td>
<td>Car</td>
<td>Stapler</td>
<td>Thermos</td>
<td>Match</td>
<td>Both</td>
</tr>
<tr>
<td>Popcorn</td>
<td>Balloon</td>
<td>Mirror</td>
<td>Belt</td>
<td>Dollar bill</td>
<td>Both</td>
</tr>
<tr>
<td>Clothespin</td>
<td>Door</td>
<td>Easel</td>
<td>Map</td>
<td>Perfume bottle</td>
<td>Both</td>
</tr>
<tr>
<td>Pushpin</td>
<td>Stroller</td>
<td>Chair</td>
<td>Feather</td>
<td>Boot</td>
<td>Both</td>
</tr>
<tr>
<td>Pitchfork</td>
<td>Baseball</td>
<td>Orange</td>
<td>Sweater</td>
<td>Television</td>
<td>Both</td>
</tr>
<tr>
<td>Revolver</td>
<td>Earth</td>
<td>Beach ball</td>
<td>Book</td>
<td>Ladder</td>
<td>Subject Only</td>
</tr>
<tr>
<td>Passport</td>
<td>Football player</td>
<td>Sprinter</td>
<td>Cherry</td>
<td>Drum</td>
<td>Subject Only</td>
</tr>
<tr>
<td>Hopscotch</td>
<td>Rabbit</td>
<td>Gopher</td>
<td>Garbage can</td>
<td>Coin</td>
<td>Subject Only</td>
</tr>
<tr>
<td>Diving board</td>
<td>Submarine</td>
<td>Whale</td>
<td>Hairbrush</td>
<td>Remote control</td>
<td>Subject Only</td>
</tr>
<tr>
<td>Steamroller</td>
<td>Kettle</td>
<td>Purse</td>
<td>Soda</td>
<td>Flag</td>
<td>Subject Only</td>
</tr>
<tr>
<td>Chopsticks</td>
<td>Axe</td>
<td>Hockey stick</td>
<td>Hanger</td>
<td>Packing tape</td>
<td>Subject Only</td>
</tr>
<tr>
<td>Swimsuit</td>
<td>Fish</td>
<td>Kite</td>
<td>Bus</td>
<td>Cheque</td>
<td>Subject Only</td>
</tr>
<tr>
<td>Punchbowl</td>
<td>Hand</td>
<td>Baseball glove</td>
<td>Rose</td>
<td>Pail</td>
<td>Subject Only</td>
</tr>
<tr>
<td>Jump rope</td>
<td>Kangaroo</td>
<td>Ground hog</td>
<td>Train</td>
<td>Towel</td>
<td>Subject Only</td>
</tr>
<tr>
<td>Staircase</td>
<td>Eyes</td>
<td>Sunglasses</td>
<td>Book bag</td>
<td>Radio</td>
<td>Subject Only</td>
</tr>
</tbody>
</table>
Appendix 2

Critical trial object arrays in Experiments 3 and 4. Main colour of the target in Experiment 3 corresponded to the colour term in the target name. Colour in brackets indicates the main colour of the target object in Experiment 4.

<table>
<thead>
<tr>
<th>Target</th>
<th>Competitor</th>
<th>Screen Object</th>
<th>Distractor</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Contrastive &amp; Noncontrastive Condition</td>
<td>Contrastive Condition &amp; Noncontrastive Condition</td>
<td></td>
</tr>
<tr>
<td>Blackboard (green)</td>
<td>Black chair</td>
<td>Tan chair</td>
<td>Headphones</td>
</tr>
<tr>
<td>Greenhouse (clear/brown)</td>
<td>Green bowl</td>
<td>Brown bowl</td>
<td>Truck</td>
</tr>
<tr>
<td>Blueberry (purple)</td>
<td>Blue folder</td>
<td>Green folder</td>
<td>Mug</td>
</tr>
<tr>
<td>Blackberry [phone] (pink)</td>
<td>Black pants</td>
<td>Purple pants</td>
<td>Necklace</td>
</tr>
<tr>
<td>Greyhound [dog] (brown)</td>
<td>Grey spatula</td>
<td>Red spatula</td>
<td>Helicopter</td>
</tr>
<tr>
<td>Whiteboard (grey)</td>
<td>White mouse</td>
<td>Pink mouse</td>
<td>Jump rope</td>
</tr>
<tr>
<td>Redhead (orange hair)</td>
<td>Red pepper</td>
<td>Green pepper</td>
<td>Comb</td>
</tr>
<tr>
<td>Blacksmith (tan clothes)</td>
<td>Black teapot</td>
<td>Orange teapot</td>
<td>Strawberry</td>
</tr>
<tr>
<td>Bluetooth (silver)</td>
<td>Blue scissors</td>
<td>Red scissors</td>
<td>Carrot</td>
</tr>
<tr>
<td>Blue whale (grey)</td>
<td>Blue shirt</td>
<td>Orange shirt</td>
<td>Ring</td>
</tr>
<tr>
<td>Yellow Pages (grey/white)</td>
<td>Yellow duck</td>
<td>Red duck</td>
<td>Grocery cart</td>
</tr>
<tr>
<td>Blue Jay [baseball player] (white uniform)</td>
<td>Blue car</td>
<td>Green car</td>
<td>Racquet</td>
</tr>
</tbody>
</table>
## Appendix 3

Critical trial object arrays for Experiment 5

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Target</th>
<th>Type</th>
<th>Unrelated</th>
<th>G-S-</th>
<th>G-S-</th>
<th>G-S+</th>
<th>G-S+</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ziu</td>
<td>Snake</td>
<td>295</td>
<td>Toilet paper</td>
<td>Kite</td>
<td>Popside</td>
<td>Lollipop</td>
<td>Belt</td>
</tr>
<tr>
<td></td>
<td>Tie</td>
<td>296</td>
<td>Kite</td>
<td>Kite</td>
<td>Popside</td>
<td>Lollipop</td>
<td>Belt</td>
</tr>
<tr>
<td></td>
<td>Rope</td>
<td>297</td>
<td>Stocking</td>
<td>Kite</td>
<td>Popside</td>
<td>Lollipop</td>
<td>Belt</td>
</tr>
<tr>
<td></td>
<td>Necklace</td>
<td>298</td>
<td>Kite</td>
<td>Kite</td>
<td>Popside</td>
<td>Lollipop</td>
<td>Belt</td>
</tr>
<tr>
<td></td>
<td>Jump rope</td>
<td>299</td>
<td>Kite</td>
<td>Kite</td>
<td>Popside</td>
<td>Lollipop</td>
<td>Belt</td>
</tr>
<tr>
<td></td>
<td>Extension cord</td>
<td>300</td>
<td>Kite</td>
<td>Kite</td>
<td>Popside</td>
<td>Lollipop</td>
<td>Belt</td>
</tr>
<tr>
<td></td>
<td>Pencil</td>
<td>301</td>
<td>Kite</td>
<td>Kite</td>
<td>Popside</td>
<td>Lollipop</td>
<td>Belt</td>
</tr>
<tr>
<td></td>
<td>Paintbrush</td>
<td>302</td>
<td>Kite</td>
<td>Kite</td>
<td>Popside</td>
<td>Lollipop</td>
<td>Belt</td>
</tr>
<tr>
<td></td>
<td>Welcome mat</td>
<td>303</td>
<td>Kite</td>
<td>Kite</td>
<td>Popside</td>
<td>Lollipop</td>
<td>Belt</td>
</tr>
<tr>
<td></td>
<td>Playing card</td>
<td>304</td>
<td>Kite</td>
<td>Kite</td>
<td>Popside</td>
<td>Lollipop</td>
<td>Belt</td>
</tr>
<tr>
<td></td>
<td>Bed sheet</td>
<td>305</td>
<td>Kite</td>
<td>Kite</td>
<td>Popside</td>
<td>Lollipop</td>
<td>Belt</td>
</tr>
<tr>
<td></td>
<td>Curtain</td>
<td>306</td>
<td>Kite</td>
<td>Kite</td>
<td>Popside</td>
<td>Lollipop</td>
<td>Belt</td>
</tr>
<tr>
<td></td>
<td>Jew</td>
<td>307</td>
<td>Kite</td>
<td>Kite</td>
<td>Popside</td>
<td>Lollipop</td>
<td>Belt</td>
</tr>
</tbody>
</table>

Note: The table lists the critical trial object arrays for Experiment 5, with each array consisting of a combination of related and unrelated items.
Appendix 4

Images used in critical trials for Experiment 5.

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Target</th>
<th>G+S+ Competitor</th>
<th>G+S- Competitor</th>
<th>G-S- Competitor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tiu[4]</td>
<td>(long, narrow, flexible)</td>
<td><img src="image1.png" alt="Image" /></td>
<td><img src="image2.png" alt="Image" /></td>
<td><img src="image3.png" alt="Image" /></td>
</tr>
<tr>
<td>Zi[1]</td>
<td>(rigid, narrow, long, small)</td>
<td><img src="image4.png" alt="Image" /></td>
<td><img src="image5.png" alt="Image" /></td>
<td><img src="image6.png" alt="Image" /></td>
</tr>
<tr>
<td>JEUNG[1]</td>
<td>(sheet-like, flat, broad, flexible)</td>
<td><img src="image7.png" alt="Image" /></td>
<td><img src="image8.png" alt="Image" /></td>
<td><img src="image9.png" alt="Image" /></td>
</tr>
</tbody>
</table>