Chapter 6: Reaching Sentence and Reference Meaning

Paul E. Engelhardt[1] & Fernanda Ferreira[2]
[1]University of East Anglia, [2]University of California, Davis

1.0 Introduction

The focus of this chapter is how people construct the meaning of sentences and establish reference to objects in the visual world. We will be concentrating primarily on linguistic processing with spoken input, and how attention to objects can be used to study the mechanisms and time course of language comprehension. Like several other chapters in this volume, we focus our review on studies that simultaneously present a spoken utterance with a visual scene. It is assumed that eye movements to particular objects reflect different interpretations of ambiguous or partial input, and the speed in which interpretations are computed (see Spivey & Huette, Chapter 1 for an overview). Because there is a lot of methodological overlap between the different chapters in this volume, we have focused this chapter on a relatively focused set of studies which we cover in depth.

This chapter consists of three main sections. The first focuses on how the Visual World Paradigm has been used to investigate the activation of semantic (or conceptual) representations. A series of papers by Huettig and colleagues have demonstrated that eye movements can be directed by partial semantic overlap between a spoken word and objects in the environment (e.g. Huettig & Altmann, 2005, 2007, 2011; Huettig & McQueen, 2007; for a review see Huettig, Rommers, & Meyer, 2011). This literature is important because it establishes how conceptual activation takes place following lexical access. The second section focuses on lexical-semantic processing of sentences. A widely held assumption is that sentence comprehension occurs incrementally, that is, interpretations are made as each word is processed and integrated with the previous context. The content of this section will show that the comprehension system is an active predictor of upcoming words, as reflected by anticipatory eye movements. For example, given the combination of a subject and a verb (e.g. *the baby drank....*), participants will to some extent be able to anticipate the likely object in the sentence given a number of visual alternatives (Altmann & Kamide, 1999; Altmann & Mirkovic, 2009). We will also consider the mapping of language and event interpretations (Scheepers, Keller, & Lapata, 2008).

The third section focuses on reference, which occurs when a linguistic expression identifies or selects a particular object in the external world (Brown, 1958). Here we will focus on definite noun phrases (e.g. *the large book*) and the quantity of information that is necessary to identify a particular object amongst an array of competitors. Thus, we will be paying close attention to the predictions of the Cooperative Principle and the Maxim of Quantity (Grice, 1975, 1989), which apply to both production and comprehension. The focus of this section is a surprising asymmetry between production tendencies and comprehension efficiency. In the conclusions, we present several limitations and future directions of the Visual World Paradigm, and its role in the study of semantic processing in sentence comprehension and reference.

2.0 Semantic-Conceptual Representations

The structure and representation of conceptual knowledge was an early focus of research in cognitive psychology (Anderson, 1983; Barsalou, 1982; Collins & Quilian, 1972; McClelland & Rumelhart, 1985; Tulving, 1972). For example, Collins and Loftus (1975) proposed that concepts were organized in a network, and they tested the distance between concepts by measuring the reaction time taken to verify statements (e.g. *Is a canary a bird?* or *Is a robin a bird?*). These networks were assumed to operate via spreading activation. Later work based on neuropsychological data and computational models have led to a Conceptual-Structural Model of stored knowledge, which assumes that concepts are represented in a distributed system as patterns of activation between multiple semantic properties (Taylor, Devereux, & Tyler, 2011; Taylor, Moss, & Tyler, 2007; Tyler & Moss, 2001). The neuropsychological case studies have revealed category-specific semantic deficits, in which patients are unable to comprehend or produce words from particular categories (e.g. living vs. non-living things, animals, fruits, etc). However, in most cases, brain damage does not selectively impair a specific type of knowledge or semantic category (Caramazza & Shelton, 1998; Mahon & Caramazza, 2009; Warrington & McCarthy, 1983; Warrington & Shallice, 1984). Neuroimaging studies of healthy participants have been inconsistent in showing distinct (as opposed to overlapping) brain regions for the different semantic categories observed in brain-damaged patients (Martin, 2001; Martin, Wiggs, Ungerleider, & Haxby, 1996; Martin & Simmons, 2008; Moore & Price, 1999; Perani et al., 1995). Connectionist models postulate that stored conceptual knowledge is distributed and based on similarity or to what extent the semantic properties of different concepts overlap (Masson, 1995; McRae, de Sa, & Seidenberg, 1997). At present, no model can account for all of the data, but the Conceptual-Structure Model assumes that conceptual representations are distributed with more similar concepts represented closer together via the sharing of more links (i.e., more shared features between concepts lead to more and closer connections).

In the next three sections, we review studies that investigated language processing using the Visual World Paradigm (for an overview of this methodology, see Chapter 3 "Attention in Vision and Language"). Before moving on, there is one point that needs to be addressed. In this chapter, we attempt to focus on studies that investigated language processing by examining eye movements to particular objects in view. In most studies, however, visual object arrays or scenes are presented before language input, and so it is possible that conceptual activation from visual objects (before the auditory input) affects how the linguistic input is structured and interpreted. In most studies, we believe that it is impossible to fully dissociate the two types of effects from one another, and indeed, most studies discuss findings in terms of the interaction of vision and language (e.g. Henderson & Ferreira, 2004; Knoeferle & Crocker, 2006).[1]

2.1 Activation Pre-Lexical Access

Across many studies, the Visual World Paradigm has been instrumental in showing how people activate different types of information when hearing words. Upon hearing a spoken utterance, listeners must first access the words that they are hearing and then put the words together in order to determine the meaning of the sentence. Therefore, we begin with a brief overview of research that has looked at activation processes during word recognition (Dahan, Manguson, & Tanenhaus, 2001; Eberhard, Spivey-Knowlton, Sedivy, & Tanenhaus, 1995; Magnuson, Dixon, Tanenhaus, & Aslin, 2008). Studies of word recognition have revealed that words with initial phonological overlap (e.g. *dollar* and *dolphin*) compete with each other during lexical access (Dahan & Tanenhaus, 2004; Marslen-Wilson, 1987, 1990;

[1] One study that presented the language before the visual array was conducted by Moores, Laiti, and Chelazzi (2003). They found that during visual search an associated object (lock - key) was target of first saccade 23% of the time, whereas unrelated objects were targeted only 17%.

Marslen-Wilson & Zwitserlood, 1989). Allopenna, Magnuson, and Tanenhaus (1998) tested rhyme competitors (e.g. *dollar* and *collar*) along with phonological onset competitors. Their results showed that as participants heard the word “dollar” there was competition between the phonological competitor (dolphin) early in the word, but towards the end of the word, looks to “collar” began to increase. However, looks to the rhyme competitor never surpassed the phonological competitor. Thus, the competition for visual attention occurred early for phonological onset competitors and late for rhyme competitors, which indicates that linguistically activated phonological representations affected the probability that particular objects in the array would be fixated. The differential activation between the two types of competitors reveals how spoken word processing is temporally tied to unfolding language. Objects that share the same phonological features are more likely to be looked at.

2.2 Activation Post-Lexical Access

In this section, we review the evidence that suggests that conceptual overlap can also mediate the mapping between language and visual attention. Again, the Visual World Paradigm has been a useful tool for understanding the activation of different types of semantic information as words are accessed. This is tested by examining shifts in visual attention depending on the overlapping properties between words in speech and objects in view. The underlying assumption is that eye movements reveal the online activation of semantic information (Cooper, 1974). Dahan and Tanenhaus (2004) looked at contextually constraining verbs in Dutch (e.g. *Never before* ***climbed*** *a goat so high.*—English translation). The visual displays contained the target (goat), a semantic competitor (spider), a phonological competitor (bone), and an unrelated distractor (island). (*Goat* and *bone* have significant phonological overlap, i.e., *bok* and *bot* in Dutch, respectively.) The semantic competitor was a plausible subject of the verb (i.e. spiders climb) but was phonologically unrelated (*spin* in Dutch). Results showed a small semantic competitor effect. There were more looks to the spider than to the unrelated distractor. Moreover, there was no evidence of competition between the cohort competitors, and looks to the target began to increase before the word referring to the target had ended. Based on these results, Dahan and Tanenhaus argued that the mapping from linguistic input to meaning is continuous.

In the next paragraphs, we focus on language-mediated eye movements and the extent to which they are driven by the activation of different types of stored knowledge. As mentioned previously, eye movements reflect the online activation of semantic information. Huettig and Altmann (2005) showed that participants would fixate a semantically related item that was not associatively related to the target.[2] For example, if participants heard the word "piano" they had an increased tendency to fixate an object such as a "trumpet". These words are semantically related because both are members of the same category of musical instruments (Mahon & Caramazza, 2009; Martin et al., 1996). Furthermore, fixation probabilities were correlated with conceptual similarity as measured by feature norms (Cree & McRae, 2003). Thus, language-mediated eye movements are driven by the degree of semantic similarity between spoken words and visual objects in the environment. These types of effects have also been observed in the absence of semantic overlap. Dahan and Tanenhaus (2005) showed that language-mediated eye movements could be directed to objects that have similar perceptual features but are semantically unrelated. For example, participants were more likely to look at a rope upon hearing the word *snake*, because both objects share the same global shape. Huettig and Altmann (2004) investigated perceptual similarity effects using objects that shared the same prototypical color. They found an increased tendency to look towards a *strawberry* upon hearing the word *lips.* This is because both are prototypically red in color (for similar effects in language production, see Huettig & Hartsuiker, 2008, and in computational modelling, see McRae, 2004).

An important question that arises from these findings is whether the fixations are driven by stored knowledge or perceptual information from the environment. As discussed in the introduction, visual objects are typically presented in advance of the linguistic input, and therefore, participants have the opportunity to explore, identify, and potentially name the objects in the array. Huettig and Altmann (2011) addressed the stored knowledge vs. perceptual information question by examining color. Some objects are associated with particular colors; for example, frogs are most often green. Others objects (typically artifacts) are not associated with particular colors; for

[2] Association is typically tested by presenting participants with a word and having them write down the word or words that first come to mind.

example, cars can be any color, which means that color is not a diagnostic property of a car. In the first two experiments, Huettig and Altmann presented participants with a target word, such as *spinach*. The first experiment examined objects that were associated with the prototypical color of the target word, but presented in black and white (e.g. line drawing of a frog). The second experiment examined objects that were associated with the prototypical color, but presented in an atypical color (e.g. yellow frog). Results showed a marginal and relatively late occurring effect in the experiment with atypical colors, which suggests that participants accessed the prototypical color of the target word (*spinach* is green*),* and then matched that to the prototypical color information associated with the frog. Stored knowledge, therefore, seems to have a small and relatively late effect on language-mediated shifts of overt attention. The final experiment examined target words such as *pea,* and two types of competitor objects. The first were objects not associated with a diagnostic color (e.g. a jumper), but which overlapped with the prototypical color of the target, in this case, a green jumper. The second type were conceptual competitors, such as a mushroom (mushrooms and spinach are both vegetables). Results showed significantly more fixations to both the color and conceptual competitors compared to an unrelated distractor. The looks to the color competitor demonstrate that language-mediated eye movements can be directed based on surface color overlap.

In summary, these studies show how meaning information is activated during and after lexical access. Semantic activation from a word spreads to the features associated with that particular word/concept, and it also activates other concepts within the same semantic category (Martin et al., 1996; Moss, McMormick, & Tyler, 1997; Yee & Sedivy, 2006). The spreading activation also results in changes in visual attention, as certain objects in the environment are more likely to be fixated compared to others. These changes in visual attention appear to occur automatically when a visual feature overlaps with the activated information in memory. The tendency to shift attention occurs more quickly for semantically related objects (i.e. objects in the same category) than it does for objects that share similar surface color (Huettig & Altmann, 2011). In the next section, we focus on sentence comprehension; however, a similar mechanism based on the activation of overlapping representations, seems to underlie people's tendency to actively make predictions about upcoming input during sentence processing.

3.0 Sentence Comprehension

The processes underlying syntactic analysis, prediction, and event representations have been effectively studied by monitoring eye movements in the presence of visual scenes. The focus of this section is on thematic role assignment and semantic activation in the process of comprehending sentences (for readers interested in discourse processing, see Chapter 6 “Discourse Level Processing”). In this section, we review three main findings. The first is that a verb’s selectional restrictions can lead to anticipatory eye movements. The second is that the combination of subject and verb together can drive anticipatory eye movements. The third is that the tense of an auxiliary verb can drive eye movements to objects reflecting past or future events. There are other examples in the literature; however, the ones we have just identified are central to the parsing of sentences in visually-situated language comprehension.[3] Again, it is important to keep in mind that eye movements are likely affected by the combination of language input and objects represented in the visual scene. In addition, the relationship between events described in language and the objects in view is also influenced by real-world and episodic information in contained in long-term memory (Hagoort, Hald, Bastiaansen, & Peterson, 2004; Tulving, 1972). Finally, we will briefly discuss the mechanism of prediction in sentence comprehension (i.e. the mechanism that drives language-mediated eye movements and permits a rapid updating of event representations that both an utterance and scene refer to).

3.1 Lexical-Semantic Sentence Processing

In a now classic study, Altmann and Kamide (1999) showed that upon hearing a sentence such as *the boy will eat the cake,* listeners tended to anticipate the object of the verb. That is, they had a tendency to fixate the edible object in the display (e.g. the cake) while they were listening to the verb and before mention of the object. In comparison, when a sentence contained a less restrictive verb such as *the boy will move the cake,* there were no anticipatory looks during the verb because all objects were equally probable given “move”. This pattern of results suggests people actively predict upcoming words based on the selectional restrictions of a verb in

[3] For a comprehensive review of the Visual World Paradigm and the range of topics investigated, see Huettig et al., 2011.

combination with a limited visual world. More specifically, people tend to fixate the object that satisfies the appropriate thematic role of a verb (e.g., *eating* requires something edible).

This type of anticipatory effect was extended to determine whether the combination of a subject and a verb together could be used to predict the object (Kamide, Altmann, & Haywood, 2003). Again, participants were presented with a semi-realistic scene containing several objects. Utterances consisted of, for example, *the man will ride...* and *the girl will ride...* The prediction was that if people can make anticipatory eye movements based on the combination of a subject and verb, then participants would be more likely to fixate a motorcycle after hearing “man”, and more likely to fixate a carousel after hearing “girl”. These predictions were confirmed. Therefore, the mechanism of prediction is not strictly associated with the processing of thematic roles of single verbs, but instead, can reflect the compositional nature of incremental sentence comprehension. Similar sorts of effects have been shown in the comprehension of events. For example, Altmann and Kamide (2007) presented participants with sentences in different tenses (i.e. *the man will drink...* vs. *the man has drunk...*). They found that participants were more likely to look towards a full beer mug with the future tense and an empty wine glass with the past tense. Thus, the tense of the verb triggers eye movements to past events and vice versa for future events (for similar findings see Kamide, 2008; Kamide, Scheepers, & Altmann, 2003; Kukona, Altmann, & Kamide, 2014; Knoeferle & Crocker, 2006, 2007; Knoeferle, Crocker, Scheepers, & Pickering, 2005). These results also indicate that eye movements reveal syntactic processing, but also, that people actively integrate unfolding language with available visual information. The consistency of these studies demonstrates that people update and can predict events on a word-by-word basis.

3.2 Understanding Events

In this section, we review evidence that focuses on event representations and how listeners construct events from metonymic sentences. The process is often referred to as enriched composition, because a noun must be “coerced” into an event (Brennan & Pylkkanen, 2008; Husband, Kelly, & Zhu, 2011; Traxler, Pickering, & McElree, 2002). A sentence such as *the artist started the picture,* incurs processing cost because the complement noun (the picture) must be coerced into an event rather than being an entity. This coercion requires some additional semantic elaboration. A sentence containing the verb phrase “start the picture” can have several interpretations, for example, start **painting** the picture, start **analyzing** the picture, or start **framing** the picture. One possibility is that these different interpretations compete with one another as the sentence is processed.[4]

Scheepers, Keller, and Lapata (2008) tested this “competition” hypothesis by examining eye movements in a visual world study. Visual arrays contained an artist, a painting, and two instruments. One instrument (e.g. paint and paint brushes) went with the dominant meaning, that is, started **painting** the picture. The other instrument (e.g. magnifying glass) went with the less dominant meaning, that is, started **analyzing** the picture. If different interpretations compete with one another, then there should be slower processing and more difficulty establishing the intended interpretation. However, Scheepers et al. found that the instruments associated with the dominant interpretation were anticipated, and there were few looks to the instruments associated with the less dominant meaning. Thus, the evidence failed to show competition between the different interpretations, which suggests that the different interpretations were computed/accessed serially, and so, Scheepers et al. argued for a serial coercion process. We turn next to the mechanism underlying prediction.

3.3 Mechanism of Prediction

Altmann and Mirkovic (2009) hypothesized that the convergence of representations from a linguistic utterance and visual scene flow into a unitary system, which also draws on information from long-term memory. It is the combination of these information sources that permits the rapid prediction of likely linguistic continuations. However, the anticipatory eye movements that were described in previous sections are not exclusively due to linguistic prediction. Instead, there is some motivation on the part of the listener to understand the language input and the events that are described in the language in relation to the external environment. The ability to make predictions is based on the fact that certain words are more likely to go together, and more likely to fulfill different roles within a

[4] A study using a speed-accuracy tradeoff paradigm revealed that these sentences were processed less accurately than sentences that did not involve coercion (McElree, Pylkkanen, Pickering, & Traxler, 2006).

sentence. Linguistic restrictions are often based on the semantic properties of an object, for example, agents must be animate. Further restrictions come from the visual scene and the combination of these information sources leads to predictions about what is most likely to be talked about next.

In summary, the evidence reviewed in the sentence comprehension section suggests three things. The first is that comprehension is expectation based. The second is that interpretations are made as each word in a sentence is accessed and integrated with the previous context, suggesting that the comprehension processes are incremental on a word-by-word basis. Finally, as a sentence unfolds, certain conceptual representations become activated, it is assumed that this activation spreads to other representations and this is how thematic roles and/or events get predicted.

4.0 Establishing Reference

One of the primary functions of language is to direct attention to objects in the world. Referring expressions are linguistic descriptions that single out or identify a particular object in the environment (Olson, 1970; Pechmann, 1989). In this section, we will be focusing on how listeners identify specific objects using modified noun phrases (e.g. *the red book*). In particular, we will focus on the number of modifiers (or quantity of information) in relation to contextual alternatives. The choice of what information and how much information to include in a referring expression depends on what will make an object unique for an addressee (Brenan & Clark, 1996). The language philosopher Paul Grice assumed that people cooperate when interacting in conversation, and he formulated a Cooperative Principle based on this assumption (Grice, 1975). In addition to the Cooperative Principle, Grice also formulated a Maxim of Quantity, which requires speakers to provide enough information but no more than is necessary for an object to be identified.[5] We begin with a short review covering the production of referential expressions that contain a pre-nominal modifier because the predictions of the Cooperative Principle and the Maxim of Quantity were originally formulated in terms of production.

4.1 Production of Pre-Nominal Adjectives

[5] The Maxim of Manner, which requires speakers to avoid ambiguity, is also relevant to the Maxim of Quantity and predictions regarding reference.

Referring expressions produced by adult speakers almost always contain enough information to uniquely identify objects.[6] However, speakers do tend to include extra modifiers. We call such expressions over-descriptions (Belke, 2006; Deutsch & Pechmann, 1982; Engelhardt, 2013; Engelhardt, Bailey, & Ferreira, 2006; Koolen, Goudbeek, & Krahmer, 2013; Nadig & Sedivy, 2002; Pechmann, 1989). Deutsch and Pechmann (1982) showed participants arrays of objects, and asked them to select and then name one of the objects in the array. Their results showed that speakers produced over-descriptions on almost one-quarter of the trials. For example, if *the book* would have been sufficient for unique identification, participants had a tendency to produce expressions, such as *the red book*. Pechmann (1989) found an even higher rate of over-descriptions. In his study, approximately 60% of noun phrases had an unnecessary color modifier. More recently, Belke (2006) showed that when participants were placed under time pressure to begin speaking, they were even more likely to produce over-described utterances. In that study, the object arrays consisted of three objects that varied in size and color. In the time-pressure condition, when a size modifier was required, almost all utterances had an unnecessary color modifier, and when a color modifier was required, approximately half had an unnecessary size modifier. Therefore, across several studies, it has been consistently demonstrated that people will include extra information when producing referential expressions, which is inconsistent with the second part of the Maxim of Quantity. These studies also tend to show that color modifiers are more likely to be included as over-descriptions than are size modifiers, and this is particularly true of artifacts. A great deal of recent work has focussed on modelling the rates and types of modifiers produced by human participants (for reviews, see Dale & Reiter, 1995; Gatt, Krahmer, van Deemter, & van Gompel, 2014; Krahmer & van Deemter, 2012; Reiter & Dale, 2000; van Deemter, Gatt, van Gompel, & Krahmer, 2012).

4.2 Comprehension of Pre-Nominal Adjectives

In a seminal study, Sedivy, Tanenhaus, Chambers, and Carlson (1999) investigated the role of multiple referents on the comprehension of pre-nominal adjectives. They found that if a participant hears an instruction, such as *pick up the*

[6] Some types of referring expressions, such as pronouns, are necessarily ambiguous; however, the focus of this section is modified noun phrases (i.e. adjective(s) + noun).

tall glass, in the presence of a tall and a short glass, the probability of fixating the tall glass will start to increase even before the onset of the noun. This occurs even though the display also contained a competitor object that was taller than, for example, the tall glass. Thus, the comprehension system can predict the intended referent based on the modifier*,* as participants were able to anticipate the referent at a point where the linguistic input was still ambiguous. In later work, Sedivy (2003) looked at color modifiers. She found that when participants heard an instruction such as *pick up the yellow book,* no anticipatory looks were made to the yellow book based on the presence of a color contrast (i.e. *a red book*). However, in another experiment that examined instructions such as *pick up the yellow banana,* and included visual contexts containing both a yellow and a blue banana, Sedivy did observe anticipatory eye movements, as is typically found with size and material modifiers (Sedivy, 2006). Therefore, a highly salient and atypical color contrast, such as a yellow versus a blue banana, resulted in participants making use of the visual context to a greater extent than they otherwise might (see also Huettig & Altmann, 2011). Based on these data, it seems that participants can generate a prediction about the quantity of information needed to distinguish contrasting objects in the environment in advance of linguistic input (see also, Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). In these cases, participants are extremely efficient at establishing reference, and in certain cases, can anticipate which object will be referred to based on a quantity-type expectation.

However, people's ability to make this kind of predictive reference is not operational in all situations. More specifically, if the context contains an overtly uncooperative speaker or too many over-descriptions, then people will not show anticipatory eye movements. Grodner and Sedivy (2011) conducted an experiment in which participants were told that they were going to execute instructions that were produced by a patient with a neurological disorder. Participants were told that the purpose of the experiment was to determine how well impaired speakers are able to convey information by examining the movements that people made in response to their instructions. The pre-recorded instructions from the "neurological patient" contained several instances in which objects or locations were mislabelled, as well a very high percentage of over-descriptions (i.e. approximately 80% of trials contained an extra modifier). This was referred to as the unreliable speaker condition. A second group of participants was assigned to the reliable speaker condition, which contained no mislabelled objects, and very few over-described utterances (i.e. fewer than 8% of trials contained an extra modifier). The results showed that participants in the reliable speaker condition benefitted from the presence of a contrast, whereas participants in the unreliable speaker condition did not. Therefore, this shows that anticipatory or predictive eye movements based on the expectation for a certain quantity of information are not always generated. This indicates that initial interpretations depend on the perceived cooperativeness of the speaker, as well as an assessment of his or her linguistic abilities.

One question that follows from the Grodner and Sedivy study is how do listener's expectations change so as to affect initial processing, that is, what leads to the attenuation of predictive eye movements? It is possible that the attenuation observed by Grodner and Sedivy was due to the explicit cue about the uncooperativeness of the speaker. To investigate whether predictive eye movements could be eliminated simply by exposure to over-descriptions, Engelhardt (2008) conducted a within subjects experiment that included a block variable. An informative block of trials contained few over-descriptions (i.e. approximately 20% of trials were over-described), and an uninformative block contained many unnecessary modifiers (i.e. approximately 80% of trials were over-described). Half of the participants received the informative block first and the uninformative block second. The other half received the reverse ordering. The prediction was that if participants are sensitive to the informativeness of modifiers across trials, then there should be a reduction in predictive eye movements in the uninformative block. More specifically, when the number of over-descriptions is high, participants should be less likely to interpret a modifier as referring to an object that is member of a contrast set, thus making the effect contextually cancellable.

The results were unexpected. When the informative block of trials was heard first, participants generated predictive or anticipatory eye movements in both blocks. When the uninformative block was heard first, participants did not show predictive eye movements in either block. Therefore, this study shows that the tendency of participants to anticipate reference based on the presence of a prenominal adjective can be reduced by the presence of over-descriptions, but only when the over-descriptions occur early in the experiment. Recall that in the Grodner and Sedivy

(2011) study participants were explicitly told that the instructions were recorded from a patient with a neurological disorder. In contrast, the results from the Engelhardt (2008) study show that it is not the explicit cue to uncooperativeness (or unreliability) that leads to the reduction in anticipatory fixations. Rather it seems that participants adapted to the task situation relatively early in the experiment, and then did not change as the number of over-described instructions changed in the second block of trials. Also, post-experiment interviews did not reveal any tendency for subjects to explicitly pick up on the fact that one block of trials had many more over-descriptions compared to the other. The results from these two experiments showed that over-descriptions **can** eliminate anticipatory eye movements and make participants less efficient with regard to establishing reference.

4.3 Production-Comprehension Asymmetry

The data reviewed in the previous sections leads to an important question: Why would speakers consistently produce over-descriptions if over-descriptions eliminate people's tendency to anticipate a referent? The combination of production and comprehension data seems to indicate that speakers do not adhere to the *Audience Design Hypothesis,* which assumes that speakers will construct their utterances to be cooperative with the communicative needs of their interlocutors (Bell, 1984; Clark & Wilkes-Gibbs, 1986). Grice (1975) also hypothesized that there may be communicative consequences to additional information if listeners assume that there is some purpose to the extra information. However, studies of the effects of over-descriptions on comprehension performance have yielded mixed results.

Several studies have concluded that additional information is beneficial for comprehension (Arts, 2004; Levelt, 1989; Maes, Arts, & Noordman, 2004). In Arts, Maes, Noordman, and Jansen (2011) participants read descriptions, such as *the large square gray button.* In this study, the object was uniquely identifiable with shape information alone (e.g. *the square button*). However, when all three attributes (size, color, and shape) were included, participants were 58 ms faster identifying the object compared to the shape-only instruction. The authors argued, similar to Levelt (1989), that the extra information was useful insofar as it helped create a Gestalt or search template (Malcolm & Henderson, 2009, 2010) for the object, which facilitated search and identification.

Other studies have concluded that additional information is detrimental to comprehension performance (Engelhardt, 2008; Engelhardt et al., 2006; Grodner & Sedivy, 2011). Engelhardt et al. (2006) reported data from an eye tracking study which showed that listeners were slower to execute instructions that contained an unnecessary prepositional phrase modifier. In their study, participants heard, for example, *put the apple in the box* or *put the apple on the towel in the box.* The visual displays consisted of 2 × 2 arrays, and contained, for these examples, an apple on a towel, a pencil, an empty towel, and an empty box. From the onset of the word *box*, which is the disambiguating word, participants were approximately 1 s slower in executing the instruction with the unnecessary modifier (e.g. *on the towel*). The slowdown was primarily due to the fact that the empty towel receives a substantial proportion of fixations because participants get misled into believing that the empty towel is the destination for the apple. In this case, the unnecessary prepositional phrase modifier leads to a garden-path effect or temporary confusion about where the apple should be placed. These comprehension slowdowns were in spite of the fact that an earlier production study showed that participants produced an unnecessary prepositional phrase modifier on one-third of trials when the context contained a single referent. Thus, even if the context contained only a single apple, participants were likely to produce a referring expression, such as *the apple on the towel.*

4.4 Predicting Reference

The Sedivy et al. (1999) results, reviewed above, clearly show that participants can anticipate a referent when hearing a pre-nominal modifier. However, this effect also depends on the visual context being available in advance of the linguistic input. In a recent study, we used an attentional-cuing paradigm to investigate processing impairments associated with over-descriptions (Engelhardt, Demiral, & Ferreira, 2011). In that study, we also manipulated the predictability of the modifier in arrays that contained contrasting objects. This allowed us to tease apart benefits associated with contextual predictability from impairments with over-descriptions. Participants in the study saw two objects side by side on the computer screen (see Figure 1). They had to fixate the cross in the center, and then they heard a modified noun phrase (e.g. *the red circle)*. Participants made a left or right button press depending on which side the indicated object was located on. Notice that with all three types of displays the target object is identifiable at the modifier.

Results showed that reaction times were fastest with the display on the left (821 ms) and slowest with the display on the right (907 ms). The display in the middle had reaction times that were in between the other two (864 ms). With the display on the left, participants can predict that they will hear a color modifier, in this case, either red or blue. With the display in the middle, participants know that they will hear a modifier, but crucially, they cannot predict whether it will be size (e.g. big/small) or color (red/blue). The lack of predictability leads to a significant (43 ms) slow down in reaction time $t(13) = 3.23$, $p < .01$. With the display on the right, a modifier is unnecessary because there are two different shapes, which makes the noun phrase over-described. In this case, participants are likely not expecting a modifier but instead a bare noun phrase (e.g. *the circle*). In this case, the presence of a modifier leads to an additional 43 ms slow down in reaction time.[7] This study provides additional evidence to suggest that over-descriptions are in fact detrimental to establishing reference.

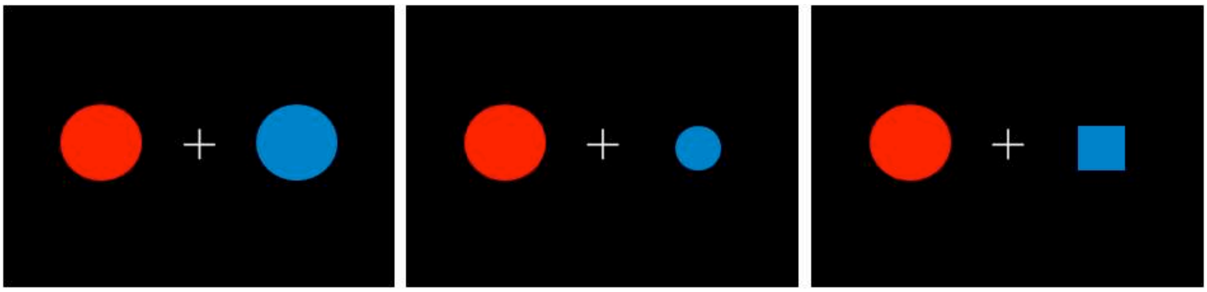

Figure 1. Example displays used in Engelhardt, Demiral, and Ferreira (2011). The object on the left is red and the object on the right is blue for all three arrays.

At this point it is not clear whether over-descriptions are problematic to comprehension in all situations. The majority of studies arguing that over-descriptions are beneficial to comprehension were experiments that presented written descriptions. Arts and colleagues have investigated written descriptions in instructional texts, and also, in experiments in which participants read an object description and then had to identify an object in a subsequently presented array

[7] Size modifiers were also tested and produced similar results.

(Arts, Maes, Noordman, & Jansen, 2011). In this case, the linguistic description and visual context do not co-occur, and thus, this may be the type of situation where the listener can build a more complete mental representation of an object, which then facilitates **visual search.** This is consistent with predictions made by Levelt (1989). However, given that there are at least as many, if not more, studies showing comprehension impairments with over-descriptions, it is still surprising how often speakers include extra modifiers when producing referential expressions.

4.5 Production-Comprehension Asymmetry II

We think that there are at least three potential explanations for speaker's tendency to include extra information. (We should point out that there is little empirical work focused on this issue that would help select from among the explanations.) One is inconsistent with the assumptions of the *Audience Design Hypothesis,* and two are consistent with it. The first explanation is speaker egocentricity. By this explanation, certain object features may be more salient to speakers for any number of reasons. We also know that language production is a resource demanding process, which is why speakers have a tendency to begin their utterances with the most accessible words or concepts first (e.g. Bock, 1987). Therefore, speakers might mention features that are irrelevant to establishing reference and be unable to compute (because of resource limitations) the extent to which these features negatively affect comprehension. The impact of array complexity on speakers' choices has received little attention (cf. V. Ferreira, Slvec, & Rogers, 2005).

The second explanation that could account for speaker's lack of adherence to the Maxim of Quantity is that over-described utterances may lead to more efficient searches. As mentioned in the previous section, speakers may have a tendency to over-describe in cases where the additional information will help a listener create a more detailed representation of an object (a search template), which might facilitate visual search (Levelt, 1989; Nadig & Sedivy, 2002; Sonnenschein, 1984). Here the extra information serves a purpose, because it will help the listener locate an object more quickly, and thus is consistent with the Audience Design Hypothesis.

The final possibility is one that we have been pursuing in recent work (Engelhardt & Ferreira, 2014). We hypothesized that the phonetic properties of unnecessary modifiers might suggest that they are less prominent than modifiers required for referent identification. In a production study, we compared the acoustic properties (i.e. duration, pitch, and intensity) of two types of modifiers. The first were modifiers that distinguished two contrasting objects (e.g. a small triangle and a large triangle), and the second were modifiers that did not distinguish contrasting objects (i.e. were over-descriptions). Results showed that the over-described modifiers were significantly shorter in duration compared to those that were used to distinguish contrasting objects (282ms versus 355ms). Moreover, comprehension data suggests that listeners are sensitive to these length differences. For example, upon hearing a short modifier, participants are quicker to identify an object that is not a member of a contrasting set. This is a second way in which the inclusion of extra modifiers may be consistent with the Audience Design Hypothesis.

In summary, the process of establishing reference based on a modified noun phrase can be predictive in nature. We assume that the mechanism that affords this prediction is similar to the one described in sentence comprehension (see section 3.3). The ability to anticipate a referent has been shown in situations in which the number of over-descriptions is relatively low. Future work will have to investigate the reasons that speakers tend to include extra modifiers and whether over-descriptions are truly inconsistent with the Audience Design Hypothesis. At present, the balance of evidence suggests that over-described referring expressions hinder people's ability to establish reference.

5.0 Conclusions

Across many studies, including both sentence comprehension and reference, it is clear that when linguistic input occurs in the presence of relevant visual context that comprehension processes are predictive in nature. At this point, it is not entirely clear how much prediction is inherent to linguistic processing and how much is due to the combination of linguistic input and visual context together (Altmann & Mirkovic, 2009; cf. Hale, 2001; Lau et al., 2006; Levy, 2008; Rommers, Meyer, Praamstra, & Huettig, 2012; Staub & Clifton, 2006). One issue that strikes us as an important avenue for future research is to manipulate the complexity in the visual displays to test the limits of this predictive mechanism in language comprehension (for an example, see Ferreira, Foucart, & Engelhardt, 2013). In the Introduction, we noted that in most situations it is impossible to separate or remove the effect of visual context on language comprehension because many of the visual world studies have presented the visual context before the auditory input. The concern, of course, is that if the displays are too simple then participants might be able to make reasonable guesses or predictions about what the linguistic input will be. If they can, then the generalizability of the results is limited to situations in which the outcome of parsing is never in doubt.

At this juncture, there is limited data to suggest that people actively predict the content and/or form of utterances in advance of linguistic input (cf. DeLong, Urbach, & Kutas, 2005; Ferreira et al., 2013). What we want to advocate here is that theoretical refinement is needed, and specifically, hypotheses and research questions that consider the limitations of the cognitive processing system. One way of probing the limitations and/or processing thresholds is by increasing the visual and linguistic complexity to the point in which people begin to make errors (Ferreira, Ferraro, & Bailey, 2002; Ferreira, & Patson, 2007; Sorensen & Bailey, 2007). We also believe that such manipulations will lead to greater insights into the nature of how much prediction is based on language comprehension and how much prediction is based on visual representations. Our hypothesis is that as visual and linguistic complexity increase people will have less and less ability to make specific predictions and instead rely on good-enough or more heuristic-type processing strategies (Ferreira, 2003; Gigerenzer, 2008). However, at this point, we do not know where the limitations of the predictive mechanism are or if they even exist.

In summary, eye movements are systematically linked to spoken language comprehension, and reveal the different interpretations that listeners make as a linguistic utterance unfolds. Thus, linguistic input systematically affects how visual attention is deployed to objects in the environment. The overall organization of this chapter was based on the assumption that as words are accessed there is spreading activation to both the features associated with those words and other concepts/objects that are members of the same semantic category. Therefore, the linking between language comprehension and eye movements is based on the activation of conceptual representations from single words and by combinations of words that overlap with a concurrent visual array. Moreover, the linguistic input is mapped onto event representations depicted in the visual scene and based on past event experiences stored in long-term episodic memory. With respect to reference, we showed that people have a tendency to make predictions, similar to the ones observed in sentence comprehension. In all cases, people seem to interpret language in combination with the objects that are concurrently displayed, and future work will have to investigate how much language processing effects are influenced by the content of the visual world and vice versa. Thus, the consensus view integrates many contextually dependent factors between linguistic and visual representations, which lead to anticipatory eye movements in sentence processing and predictive reference assignment.

References

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

Altmann, G. T., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition, 73(3),* 247–264.

Altmann, G. T. M., & Kamide, Y. (2007). The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of Memory and Language*, *57*, 502-518.

Altmann, G. T. M., & Mirkovic, H. (2009). Incrementality and prediction in human sentence processing. *Cognitive Science, 33,* 583-609.

Anderson, J. R. (1983). A Spreading Activation Theory of Memory. *Journal of Verbal Learning and Verbal Behavior, 22*(3), 261-295.

Arts, A. (2004). Overspecification in instructive texts. *Doctoral Dissertation, Tillburg University.*

Arts, A., Maes, A., Noordman, L., & Jansen, C. (2011). Overspecification facilitates object identification. *Journal of Pragmatics, 43,* 361-374.

Barsalou, L.W. (1982). Context-independent and context-dependent information in concepts. *Memory & Cognition 10,* 82-93.

Belke, E. (2006). Visual determinants of preferred adjective order. *Visual Cognition, 14,* 261-294.

Bell, A. (1984). Language style as audience design. *Language in Society, 13,* 145-

204.

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choices in conversation. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 22,* 1482-1493.

Brennan, J., & Pylkkanen, L. (2008). Processing events: Behavioral and neuromagnetic correlates of aspectual coercion. *Brain and Cognition, 106,* 132-143.

Bock, J. K. (1987). An effect of the accessibility of word forms on sentence structures. *Journal of Memory and Language, 26,* 119-137.

Brown, R. (1958). *Words and things.* Macmillan.

Caramazza, A., & Shelton, J. R. (1998). Domain-specific knowledge systems in the brain: the animate–inanimate distinction. *Journal of Cognitive Neuroscience, 10,* 1– 35.

Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition, 22,* 1-39.

Collins, A. M., & Loftus, E. F. (1975). A Spreading-activation theory of semantic processing. *Psychological Review, 82,* 407-428.

Collins, A. M., & Quillian, M. R. (1972). How to make a language user. In E. Tulving, & W. Donaldson (Eds.), *Organization of memory* (pp. 309-351). New York: Academic Press.

Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology, 6*, 84-107.

Cree, G. S., & McRae, K. (2003). Analyzing the factors underlying the structure and computation of the meaning of chipmunk, cherry, chisel, cheese, and cello (and many other such concrete nouns). *Journal of Experimental Psychology: General, 132*, 163-201.

Dahan, D., Magnuson, J. S., & Tanenhaus, M. K. (2001). Time course of frequency effects in spoken-word recognition: Evidence from eye movements. *Cognitive Psychology*, *42*, 317-367.

Dahan, D., & Tanenhaus, M. K. (2004). Continuous mapping from sound to meaning in spoken-language comprehension: Immediate effects of verb-based thematic constraints. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 498-513.

Dahan, D. & Tanenhaus, M. K. (2005). Looking at the rope when looking for the snake: Conceptually mediated eye movements during spoken-word recognition. *Psychonomic Bulletin & Review*, *12*, 453-459.

Dale, R., & Reiter, E. (1995). Computational interpretations of the Gricean maxims in the generation of referring expressions. *Cognitive Science, 19,* 233-263.

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience, 8*(8), 1117-1121.

Deutsch, W., & Pechmann, T. (1982). Social interaction and the development of definite descriptions. *Cognition, 11,* 159-184.

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

Engelhardt, P. E. (2008). Generalized Implicatures: Do uniformative environments eliminate default interpretations? *Doctoral Dissertation. Michigan State University.*

Engelhardt, P. E. (2013). Are speakers ‘fully’ Gricean?: A comment on Davies and Katsos (2013). *Journal of Pragmatics, 14,* 107-113.

Engelhardt, P. E., Bailey, K. G. D., & Ferreira, F. (2006). Do speakers and listeners observe the Gricean Maxim of Quantity? *Journal of Memory and Language, 54,* 554-573.

Engelhardt, P. E., Demiral, S. B., & Ferreira, F. (2011). Over-specified referential expressions impair comprehension: An ERP study. *Brain and Cognition, 77,* 304-314.

Engelhardt, P. E., & Ferreira, F. (2014). Do speakers articulate over-described modifiers differently from modifiers that are required by the context?: Implications for models of reference production. *Language, Cognition, and Neuroscience, 29,* 975-985.

Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive Psychology, 47,* 164-203.

Ferreira, F., Ferraro, V., & Bailey, K. G. D. (2002). Good enough representations in language comprehension. *Current Directions in Psychological Science, 11,* 11-15.

Ferreira, F., Foucart, A., & Engelhardt, P. E. (2013). Language processing in the visual world: Effects of preview, visual complexity, and prediction. *Journal of Memory and Language, 69,* 165-182.

Ferreira, F., & Patson, N. (2007). The good enough approach to language comprehension. *Language and Linguistics Compass, 1,* 71-83.

Ferreira, V. S., Slevc, L. R., & Rogers, E. S. (2005). How do speakers avoid ambiguous linguistic expressions. *Cognition, 96,* 1-22.

Gatt, A., Krahmer, E., van Deemter, K., & van Gompel, R. P. G. (2014). Models and empirical data for the production of referring expressions. *Language, Cognition, and Neuroscience, 29,* 899-911.

Gigerenzer, G. (2008). Why heuristics work. *Perspectives on Psychological Science, 3,* 20-29.

Grice, P. (1975). Logic and Conversation. In: Cole, P., & Morgan, J. (Eds.), *Syntax and semantics: Speech acts* (Vol. III, pp. 41–58).

Grice, P. (1989). *Studies in the ways of words.* Cambridge, MA: Harvard University Press.

Grodner, D., & Sedivy, J. (2011). The effects of speaker-specific information on pragmatic inferences. *The Processing and Acquisition of Reference.* Eds. N. Pearlmutter, and E. Gibson. Cambridge MA: MIT Press.

Hagoort, P., Hald, L. A., Bastiaansen, M. C. M., & Petersson, K. M. (2004). Integration of word meaning and world knowledge in language comprehension. *Science*, *304*, 438-441.

Hale, J. (2001). *A probabilistic Earley parser as a psycholinguistic model.* In Proceedings of NAACL (Vol. 2, pp. 159–166). Pittsburgh, PA.

Henderson, J. M., & Ferreira, F. (Eds.) (2004). *The interface of language, vision, and action: Eye movements and the visual world.* New York: Psychology Press.

Huettig, F., & Altmann, G. T. M. (2004). The online processing of ambiguous and unambiguous words in context: Evidence from head-mounted eye-tracking. In M. Carreiras & C. Clifton (Eds.), *The Online Study of Sentence Comprehension: Eyetracking, ERP and Beyond* (pp. 187-207). New York, NY: Psychology Press.

Huettig, F., & Altmann, G. T. M. (2005). Word meaning and the control of eye fixation: semantic competitor effects and the visual world paradigm. *Cognition*, *96*, B23-B32.

Huettig, F., & Altmann, G. T. M. (2007). Visual-shape competition during language-mediated attention is based on lexical input and not modulated by contextual appropriateness. *Visual Cognition, 15*, 985-1018.

Huettig, F., & Altmann, G. T. M. (2011). Looking at anything that is green when hearing 'frog' – How object surface color and stored object color knowledge influence language- mediated overt attention. *Quarterly Journal of Experimental Psychology, 64,* 122-145.

Huettig, F., & Hartsuiker, R.J. (2008). When you name the pizza you look at the coin and the bread: Eye movements reveal semantic activation during word production. *Memory & Cognition, 36,* 341-360.

Huettig, F., & McQueen, J.M. (2007). The tug of war between phonological, semantic, and shape information in language-mediated visual search. *Journal of Memory and Language, 54*, 460-482.

Huettig, F., Rommers, J., & Meyer, A.S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologia, 137,* 151-171.

Husband, E.M., Kelly, L.A., & Zhu, D.C. (2011). Using complement coercion to understand the neural basis of semantic composition: Evidence from an fMRI study. *Journal of Cognitive Neuroscience, 23,* 3254-3266.

Kamide, Y. (2008). Anticipatory processes in sentence processing. *Language and Linguistics Compass*, *2/4*, 647-670.

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

Kamide, Y., Scheepers, C., & Altmann, G. T. M. (2003). Integration of syntactic and semantic information in predictive processing: Cross-linguistic evidence from German and English. *Journal of Psycholinguistic Research*, 32, 37-55.

Knoeferle, P., & Crocker, M. W. (2006). The coordinated interplay of scene, utterance, and world knowledge: evidence from eye tracking. *Cognitive Science*, *30*, 481-529.

Knoeferle, P. & Crocker, M. W. (2007). The influence of recent scene events on spoken comprehension: evidence from eye-movements. *Journal of Memory and Language*, *57*, 519-543.

Knoeferle, P., Crocker, M.W., Scheepers, C., & Pickering, M.J. (2005). The influence of the immediate visual context on incremental thematic role-assignment: evidence from eye-movements in depicted events. *Cognition, 95,* 95-127.

Koolen, R., Goudbeek, M., & Krahmer, E. (2013). The effect of scene variation on the redundant use of color in definite reference. *Cognitive Science, 31,* 395-411.

Krahmer, E., & van Deemter, K. (2012). Computational generation of referring expressions: A survey. *Computational Linguistics, 38,* 173-218.

Kukona, A., Altmann, G.T.M., & Kamide, Y. (2014). Knowing what, where, and when: Event comprehension in language processing. *Cognition, 133,* 25-31.

Lau, E. F., Stroud, C., Plesch, S., & Phillips, C. (2006). The role of structural prediction in rapid syntactic analysis. *Brain and Language, 98,* 74-88.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation.* Cambridge, MA: MIT.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition, 106,* 1126–1177.

Maes, A., Arts, A., & Noordman, L. (2004). Reference management in instructive discourse. *Discourse Processes, 37,* 117-144.

Magnuson, J. S., Dixon, J. A., Tanenhaus, M. K., & Aslin, R. N. (2007). The dynamics of lexical competition during spoken word recognition. *Cognitive Science, 31*, 1-24.

Mahon, B. Z., & Caramazza, A. (2009). Concepts and categories: a cognitive neuropsychological perspective. *Annual Review of Psychology, 60*, 27-51.

Malcolm, G. L., & Henderson, J. M. (2009). The effects of target template specificity on visual search in real-world scenes: Evidence from eye movements. *Journal of Vision, 9(11):8,* 1-13.

Malcolm, G. L., & Henderson, J. M. (2010). Combining top-down processes to guide eye movements during real-world scene search. *Journal of Vision, 10(2):4,* 1-11.

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken wordrecognition. *Cognition, 25,* 71–102.

Marslen-Wilson, W. D. (1990) Activation, competition, and frequency in lexical access. In G. T. M. Altmann (Ed.), *Cognitive Models of Speech Processing: Psycholinguistics and Computational Perspectives* (pp.148-172). Cambridge, MA: MIT Press.

Marslen-Wilson, W., & Zwitserlood, P. (1989). Accessing spoken words: On the importance of word onset. *Journal of Experimental Psychology: Human Perception and Performance, 15*, 576–585.

Masson, M. E. J. (1995). A distributed memory model of semantic priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 3-23.

Martin, A. (2001). Functional neuroimaging of semantic memory. In R. Cabeza & A. Kingstone (Eds.), *Handbook of Functional NeuroImaging of Cognition* (pp. 153-186): Cambridge: MIT Press.

Martin, A., & Simmons, W. K. (2008). 3.07 - Structural Basis of Semantic Memory. In H. B. Editor-in-Chief: John (Ed.), *Learning and Memory: A Comprehensive Reference* (pp. 113-130). Oxford: Academic Press.

Martin, A., Wiggs, C. L., Ungerleider, L. G., & Haxby, J. V. (1996). Neural correlates of category-specific knowledge. *Nature, 379,* 649-652.

McClelland, J. L., & Rumelhart, D. E. (1985). Distributed Memory and the Representation of General and Specific Information. *Journal of Experimental Psychology: General, 114*(2), 159-188.

McElree, B., Pylkkänen, L., Pickering, M.J., & Traxler, M. (2006). The time course of enriched composition. *Psychonomic Bulletin & Review, 13,* 53-59.

McRae, K. (2004). Semantic memory: Some insights from feature-based connectionist attractor networks. *Psychology of Learning and Motivation: Advances in Research and Theory, 45*, 41-86.

McRae, K, de Sa, V. R., & Seidenberg, M. S. (1997). On the nature and scope of featureal representation of word meaning. *Journal of Experimental Psychology: General, 126,* 99-130.

Moore, C. J., & Price, C. J. (1999). A functional neuroimaging study of the variables that generate category-specific object processing differences. *Brain, 122,* 43–962.

Moores, E., Laiti, L., & Chelazzi, L. (2003). Associative knowledge controls deployment of visual selective attention. *Nature Neuroscience, 6,* 182–189.

Moss, H. E., McCormick, S. F., & Tyler, L. K. (1997). The time course of activation of semantic information during spoken word recognition. *Language and Cognitive Processes*, *12*, 695-731.

Nadig, A. S., & Sedivy, J. C. (2002). Evidence of perspective-taking constraints in children's on-line reference resolution. *Psychological Science, 13,* 329-336.

Olson, D. R. (1970). Language and thought: Aspects of a cognitive theory of semantics. *Psychological Review, 77,* 257-273.

Pechmann, T. (1989). Incremental speech production and referential overspecification. *Linguistics, 27,* 89-110.

Perani, D., Cappa, S. F., Bettinardi, V., Bressi, S., Gorno-Tempini, M., Matarrese, M., & Fazio, F. (1995). Different neural systems for the recognition of animals and man-made tools. *Neuroreport, 6,* 1637-1641.

Reiter, E., & Dale, R. (2000). *Building natural language generation systems.* Cambridge, UK: Cambridge University Press.

Rommers, J., Meyer, A.S., Praamstra, P., & Huettig, F. (2012). The contents of predictions in sentence comprehension: Activation of the shape of objects before they are referred to. *Neuropsychologia, 51,* 437-477.

Scheepers, C., Keller, F., & Lapata, M. (2008). Evidence for serial coercion: A time course analysis using the visual-world paradigm. *Cognitive Psychology, 56,* 1-29.

Sedivy, J. C. (2003). Pragmatic versus form-based accounts of referential contrast: Evidence for effects of informativity expectations. *Journal of Psycholinguistic Research, 32,* 3-23.

Sedivy, J. C. (2006). Evaluating explanations for referential context effects: Evidence for Gricean mechanisms in online language interpretation. In J. C. Trueswell and M. K. Tanenhaus (Eds.), *Approaches to World-Situated language use: Bridging the Language-as-product and Language-as-action traditions (learning, development, and conceptual change).* (pp. 345-364). Cambridge MA: MIT Press.

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

Sonnenschein, S. (1984). The effects of redundant communications on listeners: Why different types may have different effects. *Journal of Psycholinguistic Research, 13,* 147-166.

Sorensen, D. W . & Bailey, K. G. D. (2007). The world is too much: Effects of array size on the link between language comprehension and eye movements. *Visual Cognition*, *15*, 112- 115.

Staub, A., & Clifton, C., Jr. (2006). Syntactic prediction in language comprehension: Evidence from *either…or.* *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*, 425-436.

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

Taylor, K. I., Devereux, B. J., & Tyler, L. K. (2011). Conceptual structure: Towards an integrated neurocognitive account. *Language and Cognitive Processes, 26*(9), 1368-1401.

Taylor, K. I., Moss, H. E., & Tyler, L. K. (2007). The Conceptual Structure Account: A cognitive model of semantic memory and its neural instantiation. In J. Hart & M. Kraut (Eds.), *The Neural Basis of Semantic Memory* (pp. 265-301). Cambridge: Cambridge University Press.

Traxler, M.J., Pickering, M.J., & McElree, B. (2002). Coercion in sentence processing: evidence from eye-movements and self-paced reading. *Journal of Memory and Language, 47,* 530-547.

Tulving, E. (1972). Episodic and Semantic Memory. *Organization of memory.* (pp. 381-403). New York: Academic Press.

Tyler, L. K., & Moss, H. E. (2001). Towards a distributed account of conceptual knowledge. *Trends in Cognitive Science, 5,* 244-252.

van Deemter, K., Gatt, A., van Gompel, R. P. G., & Krahmer, E. (2012). Towards computational psycholinguistics of reference production. *Topics in Cognitive Science, 4,* 166-183.

Warrington, E. K., & McCarthy, R. (1983). Category specific access dysphasia. *Brain, 106,* 859–878.

Warrington, E. K., & Shallice, T. (1984). Category Specific Semantic Impairments. *Brain, 107*, 829-854.

Yee, E., & Sedivy, J. C. (2006). Eye movements to pictures reveal transient semantic activation during spoken word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*, 1-14.